

Channel Estimation for Opportunistic Spectrum Sensing: Uniform and Random Sensing

Quanquan Liang^{*†}, Mingyan Liu[†], Dongfeng Yuan^{*}

[†]Department of Electrical Engineering and Computer Science

University of Michigan, Ann Arbor, MI 48109-2122

Email: liangq@umich.edu, Mingyanliu@eecs.umich.edu

^{*}School of Information Science and Engineering

Shandong University, Shandong, China 250100

Abstract—The knowledge of channel statistics, as a result of random fading, interference, and primary user activities, can be very helpful for a secondary user in making sound opportunistic spectrum access decisions in a cognitive radio network. It is therefore desirable to be able to efficiently and accurately estimate channel statistics, even for resource constrained secondary users like wireless sensors. In this paper we focus on the traditional ML (maximum likelihood) estimator. However, rather than using equal or uniform sampling/sensing intervals as is typically done, we introduce a random sampling/sensing based ML estimation strategy. The randomization of the sampling intervals allows us to catch channel variations on a finer (time) granularity; the associated likelihood function is also more sensitive to channel variations. Consequently, this scheme significantly reduces the average sampling rate compared to uniform sampling. Analysis and simulation both show that random sampling significantly outperforms uniform sampling at low sampling rate. We further propose a randomized uniform sampling scheme which achieves a better tradeoff between good performance of random sampling and the low complexity of uniform sampling.

I. INTRODUCTION

Recent advances in software defined radio and cognitive radio [1] have given wireless devices greater ability and opportunity to dynamically access spectrum, thereby potentially significantly improving spectrum efficiency and user performance [2], [3]. To be able to fully utilize spectrum availability (either as a secondary user seeking opportunities of idle periods in the presence of primary users, or as one of many peer users in a multi-user system seeking channels with the best condition), a key enabling ingredient in dynamic spectrum access is high quality channel sensing that allows the user to obtain accurate real-time information on the condition of wireless channels.

Spectrum sensing is often studied in two contexts: at the physical layer and at the MAC layer. Physical layer spectrum sensing typically focuses on the *detection* of instantaneous primary user signals. Several detection methods, such as matched filter detection, energy detection and feature detection, have been proposed for cognitive radios [4]. MAC layer spectrum sensing [5], [6] is more of a resource allocation issue, where we are concerned with the *scheduling* problem of when to sense the channel and the *estimation* problem of extracting statistical properties of the random variation in the channel, assuming physical layer spectrum sensing provides sufficiently

accurate results on instantaneous channel availability. Such channel statistics can be very helpful in making good channel access decisions, and most studies on opportunistic spectrum access assume such knowledge.

In this paper we focus on the scheduling of channel sensing and study the effect different scheduling algorithms have on the accuracy of the resulting estimate we obtain on channel parameters. In particular, we are interested in the *sparse sensing/sampling* regime where we can use only a limited number of measurements over a given period of time. The goal is to decide how these limited number of measurements should be scheduled so as to minimize the estimation error within the maximum likelihood (ML) estimator framework. Throughout the paper the terms *sensing* and *sampling* will be used interchangeably.

There has been relatively little work on MAC layer channel estimation within the context of cognitive radios. Below we briefly review those most relevant to the present paper. Kim and Shin [5] introduced a ML estimator for renewal channels using a uniform sampling/sensing scheme where samples of the channel are taken at regular time intervals. A more accurate, but also much more computationally costly Bayesian estimator was introduced in [7], again based on uniform sensing. [8] analyzed the relationship between estimation accuracy, number of samples taken and the channel state transition probabilities by using the sampling and estimation framework of [5] and focusing on Markovian channels. [9] proposed a Hidden Markov Model (HMM) based channel status predictor using reinforcement learning techniques. This predictor predicts next channel state based on past information obtained through uniformly sampling the channel. [10] presented a channel estimation technique based on wavelet transform followed by filtering. This method relies on dense sampling of the channel.

In most of the above cited work, the focus is on estimation given (sufficiently dense) uniform sampling of the channel, i.e., with equal time periods between successive samples. This scheme will be referred to as *uniform sensing* in the remainder of this paper. By contrast, sampling schemes where time intervals between successive samples are drawn from a certain probability distribution will be referred to as *random sensing* throughout the paper. We observe that due to constraints on

time, energy, memory and other resources, a user may wish to perform channel sensing at much lower frequencies while still hoping for good estimates. It is this sparse sampling scenario that we will focus on in this study, and the goal is to judiciously schedule these limited number of sampling opportunities. As we shall see the key to achieving this is the use of random sensing.

Our main contributions are summarized as follows.

- Under the same channel statistics and the same average sampling interval (or frequency), the selection of random sampling times affects the estimation accuracy through the higher-order central moments of the random sampling intervals.
- In the special case of exponentially distributed on/off channels, we show that uniform sensing is the *worst* one can do under a sparse sampling condition, and any deviation from it improves the estimation accuracy. In particular, in this case using exponentially distributed sampling intervals results in much better estimation performance compared to using sampling intervals drawn from uniform or normal distributions.
- We propose a *randomized uniform sensing* scheme that utilizes the above observation and attains a better tradeoff between estimation accuracy and its computational cost.

The remainder of this paper is organized as follows: Section II presents the channel models and Section III gives the detail of the ML estimator and its functional form under a random sequence of sampling times. Then in Section IV we present how the sampling scheme affects the estimation performance, a detailed comparative analysis between uniform sensing and random sensing for exponentially distributed channel model is also given in this section. In Section V we propose a randomized uniform sensing scheme. Validating numerical results are shown in these two sections, respectively. Section VI concludes the paper.

II. THE CHANNEL MODEL

In this paper we will limit our attention to MAC layer spectrum sensing as mentioned in the introduction. Within this context, the channel state as perceived by a secondary user is represented by a binary sequence. This is a model commonly used in a large volume of literature, from channel estimation (e.g., [5], [8]) to opportunistic spectrum access (e.g., [6]) to spectrum measurement (e.g., [11]). Specifically, let $Z(t)$ denote the state of the channel at time t , such that

$$\begin{cases} Z(t) = 1 & \text{if the channel is sensed busy at time } t, \\ Z(t) = 0 & \text{otherwise.} \end{cases}$$

The advantage of such a model is its simplicity and tractability in many instances. The weakness lies in the fact that the actual energy present or detected in the channel is hardly binary. The raw channel measurement data will have to go through a binary hypothesis test (e.g., via thresholding) to be reduced to the above form, a process that comes with probabilities of error. Consequently, the channel is sensed to

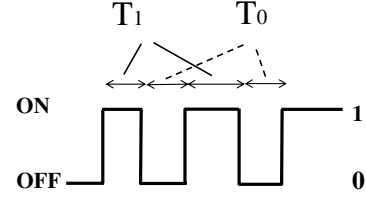


Fig. 1. Channel model: alternating renewal process with on and off states

be in either state with a detection probability and a false alarm probability.

In this paper our focus is on extracting and *estimating* essential statistics given a sequence of measured channel states (0s and 1s) rather than the (binary) *detection* of channel state (deciding between 0 and 1 given the energy reading). For this purpose, we will assume that the channel state measurements are error-free. If we have side information on what the detection and false alarm probabilities are, then the estimation results may be adjusted accordingly to utilize such knowledge.

The channel state process $Z(t)$ is assumed to be a continuous-time alternating renewal process, alternating between on/busy (state “1”) and off/idle (state “0”), an illustration is given in Fig. 1. Typically, it is assumed that a secondary user can utilize the channel only when it is sensed to be in the off states (i.e., when the channel is idle or the primary user is absent). When the channel state transitions to the on state, the secondary user is required to vacate the channel so as not to interfere with the primary user (also referred to as the spectrum underlay paradigm, see e.g., [12]).

This random process is completely defined by two probability density functions $f_1(t)$ and $f_0(t)$, $t > 0$, i.e., the probability distribution of the sojourn times of the on periods (denoted by the random variable T_1) and the off periods (denoted by the random variable T_0), respectively. The channel utilization u is defined as

$$u = \frac{E[T_1]}{E[T_1] + E[T_0]}, \quad (1)$$

which is also the average fraction of time the channel is occupied or busy. By the definition of a renewal process, T_1 and T_0 are independent and all on (off) periods are independently and identically distributed (i.i.d). It’s worth pointing out that the widely used Gilbert-Elliot model (a two-state Markov chain) is a special case of the alternating renewal process where the on (off) periods are exponentially (in the case of continuous time) or geometrically (in the case of discrete time) distributed.

III. MAXIMUM LIKELIHOOD (ML) BASED CHANNEL ESTIMATION

We proceed to describe the maximum likelihood (ML) estimator [13] we will use to estimate channel parameters from a sequence of channel state observations.

Recall that the channel state is assumed to follow an alternating renewal process. Such a process is completely characterized by the set of conditional probabilities $P_{ij}(\Delta t)$,

$i, j \in \{0, 1\}$, $\Delta t \geq 0$, defined as the probability that given i was observed Δt time units ago, j is now observed. This quantity is also commonly known as the semi-Markov kernel of an alternating renewal process [17]. Assuming the process is in equilibrium, standard results from renewal theory [17] suggest the following Laplace transforms of the above transition probabilities:

$$\begin{aligned} P_{00}^*(s) &= \frac{1}{s} - \frac{\{1 - f_1^*(s)\} \{1 - f_0^*(s)\}}{E[T_0]s^2 \{1 - f_1^*(s)f_0^*(s)\}}, \\ P_{01}^*(s) &= \frac{\{1 - f_1^*(s)\} \{1 - f_0^*(s)\}}{E[T_0]s^2 \{1 - f_1^*(s)f_0^*(s)\}}, \\ P_{10}^*(s) &= \frac{\{1 - f_1^*(s)\} \{1 - f_0^*(s)\}}{E[T_1]s^2 \{1 - f_1^*(s)f_0^*(s)\}}, \\ P_{11}^*(s) &= \frac{1}{s} - \frac{\{1 - f_1^*(s)\} \{1 - f_0^*(s)\}}{E[T_1]s^2 \{1 - f_1^*(s)f_0^*(s)\}}. \end{aligned} \quad (2)$$

where $f_1^*(s)$ and $f_0^*(s)$ are the Laplace transforms of $f_1(t)$ and $f_0(t)$, respectively. We see from this set of equations that the channel statistics is completely defined by the probability density functions $f_1(t)$ and $f_0(t)$.

The above set of equations are very useful in recovering the time-domain expressions of the semi-Markov kernel (often times this is the only viable method). For example, in the special case where the channel has exponentially distributed on/off periods, we have

$$\begin{cases} f_1(t) = \theta_1 e^{-\theta_1 t} \\ f_0(t) = \theta_0 e^{-\theta_0 t} \end{cases}. \quad (3)$$

Their corresponding Laplace transforms and expectations are

$$\begin{cases} f_1^*(s) = \theta_1 / (s + \theta_1) \\ f_0^*(s) = \theta_0 / (s + \theta_0) \end{cases}, \quad \begin{cases} E[T_1] = 1/\theta_1 \\ E[T_0] = 1/\theta_0 \end{cases}.$$

Substituting the above expressions into Eqn (2) followed by an inverse Laplace transform we get the state transition probability as follows:

$$P_{ij}(\Delta t) = u^j (1-u)^{1-j} + (-1)^{j+i} u^{1-i} (1-u)^i e^{-(\theta_0 + \theta_1)\Delta t}, \quad (4)$$

where $u = \frac{E[T_1]}{E[T_1] + E[T_0]}$ as defined earlier.

In this paper we consider the following estimation problem. Assume that the on/off periods are given by certain known distribution functions $f_0(t)$ and $f_1(t)$ but with unknown parameters. Suppose we obtain m samples $\{z_1, z_2, \dots, z_m\}$, taken at sampling times $\{t_1, t_2, \dots, t_m\}$, respectively. We wish to use these samples to estimate the unknown parameters.

First note that the channel utilization factor u can be estimated through the sample mean of the m measurements as follows

$$\hat{u} = \frac{1}{m} \sum_{i=1}^m z_i, \quad (5)$$

which is an unbiased estimator for the true quantity u .

Let $\bar{\theta}$ be the unknown parameters of the on/off distributions: $\bar{\theta} = \{\bar{\theta}_1, \bar{\theta}_0\}$. Note that in general $\bar{\theta}_1$ and $\bar{\theta}_0$ are vectors

themselves. Then the likelihood function is given by

$$\begin{aligned} L(\bar{\theta}) &= Pr\{\bar{Z}; \bar{\theta}\} \\ &= Pr\{Z_{t_m} = z_m, Z_{t_{m-1}} = z_{m-1}, \\ &\quad Z_{t_{m-2}} = z_{m-2}, \dots, Z_{t_1} = z_1; \bar{\theta}\}. \end{aligned} \quad (6)$$

The idea of ML estimation is to find the value of $\bar{\theta}$ that maximizes the log likelihood function $\ln L(\bar{\theta})$. That is, the estimate $\hat{\bar{\theta}}$ is such that $\frac{\partial \ln L(\bar{\theta})}{\partial \bar{\theta}} \Big|_{\hat{\bar{\theta}}} = 0$. This method has been used extensively in the literature [14]–[16]. For a fixed set of data and underlying probability model, the ML estimator selects the parameter value that makes the data “most likely” among all possible choices. Under certain (fairly weak) regularity conditions the ML estimator is asymptotically optimal [18].

The question we would like to investigate is what impact does the selection of the sampling time sequence $\{t_1, t_2, \dots, t_m\}$ have on the performance of this estimator, given a limited number of samples m . Specifically, we question whether random sampling is a better way of sensing the channel than uniform sampling where the measurement samples are taken at regular time intervals. As we mentioned in the introduction, most existing literature focuses on the latter method.

For the remainder of our analysis we will limit our attention to the case where the channel on/off durations are given by exponential distributions. This is for both mathematical tractability and simplicity of presentation. We will explore more general cases in our numerical experiments.

Since the exponential distribution is defined by a single parameter, we have now $\bar{\theta} = \{\theta_1, \theta_0\}$, where θ_1 and θ_0 are the two unknown scalar parameters of the on and off exponential distributions, respectively. Using the memoryless property, the likelihood function becomes

$$\begin{aligned} L(\bar{\theta}) &= Pr\{\bar{Z}; \bar{\theta}\} \\ &= Pr\{Z_{t_1} = z_1; \bar{\theta}\} \cdot \prod_{i=2}^m Pr\{Z_{t_i} = z_i | Z_{t_{i-1}} = z_{i-1}; \bar{\theta}\} \\ &= Pr\{Z_{t_1} = z_1; \bar{\theta}\} \cdot \prod_{i=2}^m P_{z_{i-1}z_i}(\Delta t_i; \bar{\theta}). \end{aligned} \quad (7)$$

where $\Delta t_i = t_i - t_{i-1}$. The first quantity on the right is taken to be

$$Pr\{z_{t_1} = z_1; \bar{\theta}\} = u^{z_1} (1-u)^{1-z_1}. \quad (8)$$

That is, the unconditioned probability of finding the channel in a particular state (LHS) is taken to be the stationary distribution given on the RHS. This choice is justified by assuming that the channel is in equilibrium.

The second quantity $P_{z_{i-1}z_i}(\Delta t_i; \bar{\theta})$ is given in (4). Combining these two quantities, we have

$$\begin{aligned} L(\theta_0, \theta_1) = L(\bar{\theta}) &= u^{z_1} (1-u)^{1-z_1} \prod_{i=2}^m \left(u^{z_i} (1-u)^{1-z_i} \right. \\ &\quad \left. + (-1)^{z_i + z_{i-1}} u^{1-z_{i-1}} (1-u)^{z_{i-1}} e^{-(\theta_0 + \theta_1)\Delta t_i} \right). \end{aligned} \quad (9)$$

The estimates for the parameters are found by solving

$$\begin{cases} \frac{\partial \ln L(\theta_0, \theta_1)}{\partial \theta_0} = 0 \\ \frac{\partial \ln L(\theta_0, \theta_1)}{\partial \theta_1} = 0 \end{cases} \quad (10)$$

Conceptually to get the estimates for both θ_0 and θ_1 one needs to solve the above two equations simultaneously. This however proves to be computationally complex and analytically intractable. Instead, we adopt the following estimation procedure. We first estimate u using (5), and take $\theta_1 = \frac{(1-u)\theta_0}{u}$. The likelihood function (9) can then be written as

$$\begin{aligned} L(\theta_0) = & u^{z_1} (1-u)^{1-z_1} \prod_{i=2}^m (u^{z_i} (1-u)^{1-z_i} \\ & + (-1)^{z_i+z_{i-1}} u^{1-z_{i-1}} (1-u)^{z_{i-1}} e^{-\theta_0 \Delta t_i / u}). \end{aligned} \quad (11)$$

The estimation of θ_0 is then derived by solving the equation $\frac{\partial \ln L(\theta_0)}{\partial \theta_0} = 0$.

In the remainder of this paper, we will use this procedure by treating u as a known constant and solely focus on the estimation of θ_0 , with the understanding that u can be unbiasedly estimated, and once we have the estimate for θ_0 we have the estimate for θ_1 . It has to be noted that this procedure is in general not equivalent to solving (10) simultaneously, i.e., the resulting estimates may have different distributions. However, we have found this to be much more amenable to analysis and more computationally feasible (this is also the method used in [5] but without such an explicit justification).

IV. RANDOM SENSING VS. UNIFORM SENSING

The goal of this study is to judiciously schedule a very limited number of sampling opportunities so that the estimation accuracy is least affected. In this section we show that when the number of samples are small (a sparse sampling scenario), then random sensing significantly outperforms uniform sensing. In what follows we first make a qualitative comparison using a few intuitive arguments. We then perform a more precise analysis of the two through the use of Fisher information, in the case of exponential on/off distributions. In particular, we will show that using this measure, uniform sensing is the *worst* type of scheduling in terms of the estimation accuracy. We further compare the performance of a few random scheduling schemes. Numerical results are also provided.

A. An intuitive comparison

Uniform sensing, where samples are taken at constant time intervals, is a natural, easy-to-implement, and easy-to-analyze scheme. Specifically, with the on/off durations being exponential the likelihood function has a particularly simple form; there is also a closed-form solution to the maximization of the log likelihood function, see e.g., [5]. On the other hand, there are a number of drawbacks to uniform sensing compared to random sensing, where samples are taken at random time intervals (generated according to certain probability distribution), given the same average sampling rate.

One of the first things to note is that since there is no variation across sampling intervals under uniform sensing, the uniform interval needs to be upper-bounded in order to catch potential channel state changes that occur over small intervals¹. By contrast, under random sensing, even if the average sampling interval is large, there can be significant probability for sufficiently small sampling intervals to exist in any realization of the sampling time sequence $\{t_1, t_2, \dots, t_m\}$.

To see why this variability is important, consider the transition probabilities $P_{ij}(\Delta t)$, $i, j \in \{0, 1\}$. As shown in the previous section, these probabilities completely define the likelihood function. They will approach the stationary probabilities as Δt increases. For instance, we have $P_{01}(\Delta t) \rightarrow \frac{E[T_1]}{E[T_1]+E[T_0]} = u$ as $\Delta t \rightarrow \infty$, and so on. This stationary quantity represents the average fraction of time the channel is busy, which contains little direct information on the average length of a busy period, the parameter we are trying to estimate. Depending on the mixing time of the underlying renewal process, this convergence can occur rather quickly. What this means is that if sampling is sparsely done, in which case Δt is a relatively large constant Δt_o under uniform sensing, then these transition probabilities will become constant-like (i.e., approaching the stationary value), which in turn causes the likelihood function to look like a constant. This makes it difficult for the ML estimator to produce accurate estimates [13].

Loosely speaking, the transition probabilities being constant-like means that samples are of a similar quality in the sense noted above, each providing little additional information. In particular, there is zero probability that we will collect two samples less than Δt_o apart. By contrast, under a random sensing scheme Δt is a variable, meaning with positive probability there will be sampling intervals both *smaller* and *larger* than Δt_o . This variability in the interval realizations results in measurements of $P_{01}(\Delta t)$ that contain more information about the parameter to be estimated (we will shortly make this notion more precise).

Next we show that using random sensing the resulting likelihood function is more sensitive to the estimated parameter, a helpful feature in its maximization. Recall that the estimate of the parameter $\hat{\theta}_0$ is derived by solving $\frac{\partial \ln L(\theta_0)}{\partial \theta_0} |_{\theta_0=\hat{\theta}_0} = 0$. In Fig. 2 we plot an example of the likelihood function $L(\theta_0)$ under the uniform sensing (US) and random sensing (RS) schemes, respectively. In this example the true parameter value of θ_0 and θ_1 are 0.05 and 0.1, respectively. We use 50s as the uniform sensing interval over a total sensing period of 5000s. The random sensing times are generated by randomly distributing 100 points over the same period using a uniform distribution, resulting in an average sensing interval of 50s. The measurement sequence $\{z_t\}$ in both cases are then used in (11) to obtain the respective likelihood function as shown in Fig. 2. The true parameter value of θ_0 is at the center of the horizontal axis, this is also the ML estimate $\hat{\theta}_0$. We see that while both functions attain their maximum value at the point

¹One such upper bound was proposed in [5].

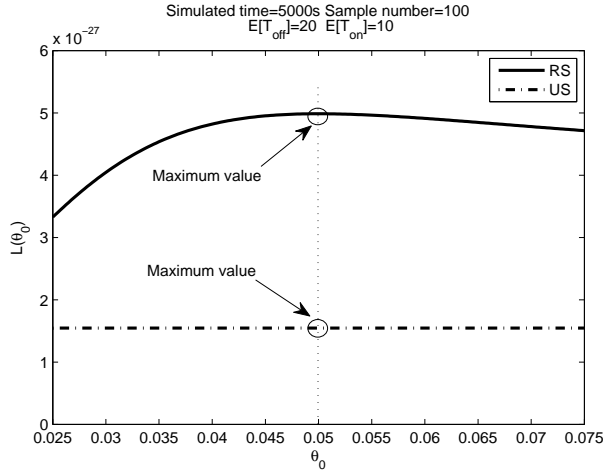


Fig. 2. Comparison of sensitivity to channel parameter

$\theta_0 = 0.05$, they behave very differently around this point. The one under random sensing is clearly more sensitive to perturbations in the estimated value, and thus it is easier to obtain the correct value of θ_0 given the data using this function.

B. An analysis using Fisher information

We now analyze this notion of information content more formally via a measure known as the *Fisher information* [21]. It is defined as follows for a given log likelihood function:

$$I(\bar{\theta}) = -E\left[\frac{\partial^2 \ln L(\bar{\theta})}{\partial \bar{\theta}^2}\right]. \quad (12)$$

The Fisher information is a measure of the amount of information an observable random variable conveys about an unknown parameter. This measure of information is particularly useful when comparing two observation methods of random processes (see e.g., [22]). For instance, the precision to which we can estimate $\bar{\theta}$ is fundamentally limited by the Fisher information of the likelihood function.

Due to the product form of the likelihood function given in Eqn (7), the Fisher information can be written as the summation of functions $g(\Delta t_i; \bar{\theta})$, $i = 2, 3, \dots, m$, $I(\bar{\theta}) = \sum_{i=2}^m g(\Delta t_i; \bar{\theta})$, where

$$g(\Delta t_i; \bar{\theta}) = -E\left[\frac{\partial^2 \ln P_{z_{i-1}z_i}(\Delta t_i; \bar{\theta})}{\partial \bar{\theta}^2}\right]. \quad (13)$$

The function $g()$ will be referred to as the *Fisher function* in the remainder of our discussion. Note that $g()$ is a function of both Δt_i and $\bar{\theta}$. However, in the remainder of this section we will suppress $\bar{\theta}$ from the argument in the Fisher function $g()$ and write it simply as $g(\Delta t)$. This is because our analysis focuses on how this function behaves as we select different Δt (the sampling interval) while holding $\bar{\theta}$ constant. Note that the first term in Eqn (7) does not appear in the above expression. This is because this first term is only a function of u (see Eqn (8)), which is separately estimated using Eqn (5) and not

viewed as a function of $\bar{\theta}$. Therefore the term disappears after the differentiation.

Similarly, using the likelihood function given in Eqn (11), the Fisher information can be written as:

$$\begin{aligned} I(\theta_0) &= -E\left[\frac{\partial^2 \ln L(\theta_0)}{\partial \theta_0^2}\right] \\ &= -E\left[\sum_{i=2}^m \frac{\partial^2 \ln[\alpha_i + \beta_i e^{-\theta_0 \Delta t_i / u}]}{\partial \theta_0^2}\right] \\ &= \sum_{i=2}^m \frac{\Delta t_i^2}{u^2} E\left[\frac{-\alpha_i \beta_i e^{-\theta_0 \Delta t_i / u}}{(\alpha_i + \beta_i e^{-\theta_0 \Delta t_i / u})^2}\right], \end{aligned} \quad (14)$$

where $\alpha_i = u^{z_i}(1-u)^{1-z_i}$ and $\beta_i = (-1)^{z_i+z_{i-1}}u^{1-z_{i-1}}(1-u)^{z_{i-1}}$. Again each term within the summation will be referred to as the Fisher function, and the Fisher information can be written as $I(\theta_0) = \sum_{i=2}^m g(\Delta t_i)$.

The expectation on the right-hand side of the above expression can be calculated by considering all four possibilities for the pair (z_{i-1}, z_i) , i.e., (0,0), (0,1), (1,0), and (1,1). Using Eqn (4), we obtain the transition probability of each case to be $(1-u)P_{00}(\Delta t)$, $(1-u)P_{01}(\Delta t)$, $uP_{10}(\Delta t)$ and $uP_{11}(\Delta t)$, respectively. We can therefore calculate the Fisher function as follows:

$$\begin{aligned} g(\Delta t) &= \frac{\Delta t^2}{u^2} e^{-\theta_0 \Delta t / u} \left[\frac{u^2(1-u)}{u - u e^{-\theta_0 \Delta t / u}} \right. \\ &+ \frac{u(1-u)^2}{(1-u) - (1-u)e^{-\theta_0 \Delta t / u}} \\ &\left. - \frac{u(1-u)^2}{(1-u) + u e^{-\theta_0 \Delta t / u}} - \frac{u^2(1-u)}{u + (1-u)e^{-\theta_0 \Delta t / u}} \right]. \end{aligned} \quad (15)$$

Below we show that under a certain *sparsity* condition on the sampling rate, the Fisher function is convex, and that the Fisher information is minimized when uniform sampling is used. We begin by introducing this sparsity condition.

Condition 1: Let $\alpha = \max\{2 + \sqrt{2}, \ln(\frac{1-u}{u}), \ln(\frac{u}{1-u})\}$. This condition requires that $\Delta t > \alpha u / \theta_0$.

Taking Δt to be the time between two consecutive sampling points, the above condition states that these two points cannot be too close together with respect to the average off duration ($1/\theta_0$) and the channel utilization u .

Lemma 1: The Fisher function $g(\Delta t)$ given in Eqn (15) is convex under Condition 1 (i.e., for $\Delta t > \alpha u / \theta_0$).

Proof: For simplicity in presentation, we first write $g(\Delta t) = h_o(\Delta t)h(\Delta t)$, where

$$\begin{aligned} h_o(\Delta t) &= \frac{\Delta t^2}{u^2} e^{-\theta_0 \Delta t / u}, \\ h(\Delta t) &= h_1(\Delta t) + h_2(\Delta t) + h_3(\Delta t), \end{aligned}$$

where

$$\begin{aligned} h_1(\Delta t) &= \frac{2u(1-u)}{1 - e^{-\theta_0 \Delta t / u}}, \\ h_2(\Delta t) &= -\frac{u(1-u)^2}{(1-u) + u e^{-\theta_0 \Delta t / u}}, \\ h_3(\Delta t) &= -\frac{u^2(1-u)}{u + (1-u)e^{-\theta_0 \Delta t / u}}. \end{aligned}$$

We proceed to show that each of the above functions is convex under Condition 1.

We first show that $h_o(\Delta t)$ is convex for $\Delta t > (2 + \sqrt{2})u/\theta_0$. Under this condition and noting $0 < u < 1$ and $\theta_0 > 0$ we have

$$\begin{aligned} h'_o(\Delta t) &= \frac{\Delta t}{u^2} e^{-\theta_0 \Delta t/u} \left(2 - \frac{\theta_0 \Delta t}{u}\right) < 0, \\ h''_o(\Delta t) &= \frac{e^{-\theta_0 \Delta t/u}}{u^2} \left[\left(\frac{\theta_0 \Delta t}{u} - 2\right)^2 - 2\right] > 0. \end{aligned}$$

Therefore for $\frac{\theta_0 \Delta t}{u} > 2 + \sqrt{2}$, $h_o(\Delta t)$ is convex.

That $h_1(\Delta t)$ is convex is straightforward. Since $0 < u < 1$ and $\theta_0 > 0$, we have:

$$\begin{aligned} h'_1(\Delta t) &= \frac{-2(1-u)\theta_0 e^{-\theta_0 \Delta t/u}}{(1 - e^{-\theta_0 \Delta t/u})^2} < 0, \\ h''_1(\Delta t) &= \frac{2(1-u)\theta_0^2 e^{-\theta_0 \Delta t/u} (1 + e^{-\theta_0 \Delta t/u})}{u(1 - e^{-\theta_0 \Delta t/u})^3} > 0. \end{aligned}$$

Next we show that $h_2(\Delta t)$ is convex for $\Delta t > \frac{u}{\theta_0} \ln\left(\frac{u}{1-u}\right)$. This condition is equivalent to $u e^{-\theta_0 \Delta t/u} < 1 - u$. Under this condition and again noting $0 < u < 1$ and $\theta_0 > 0$, we have

$$\begin{aligned} h'_2(\Delta t) &= \frac{-u(1-u)^2 \theta_0 e^{-\theta_0 \Delta t/u}}{[(1-u) + u e^{-\theta_0 \Delta t/u}]^2} < 0, \\ h''_2(\Delta t) &= \frac{(1-u)^2 \theta_0^2 e^{-\theta_0 \Delta t/u} [(1-u) - u e^{-\theta_0 \Delta t/u}]}{[(1-u) + u e^{-\theta_0 \Delta t/u}]^3} > 0. \end{aligned}$$

Similar, $h_3(\Delta t)$ is convex under the condition $\Delta t > \frac{u}{\theta_0} \ln\left(\frac{1-u}{u}\right)$, since

$$\begin{aligned} h'_3(\Delta t) &= \frac{-u(1-u)^2 \theta_0 e^{-\theta_0 \Delta t/u}}{[u + (1-u)e^{-\theta_0 \Delta t/u}]^2} < 0, \\ h''_3(\Delta t) &= \frac{(1-u)^2 \theta_0^2 e^{-\theta_0 \Delta t/u} [u - (1-u)e^{-\theta_0 \Delta t/u}]}{[u + (1-u)e^{-\theta_0 \Delta t/u}]^3} > 0. \end{aligned}$$

Therefore under the condition $\Delta t > \alpha u/\theta_0$, h_1 , h_2 and h_3 are all monotonically decreasing convex functions. It follows that $h = h_1 + h_2 + h_3$ is also monotonically decreasing and convex. Furthermore, for any $\Delta t > 0$, $h_o(\Delta t) > 0$, and $h(\Delta t) > h(+\infty) = 0$. We can now show that g is convex under this condition:

$$\begin{aligned} g''(\Delta t) &= (h_o(\Delta t)h(\Delta t))'' \\ &= h''_o(\Delta t)h(\Delta t) + 2h'_o(\Delta t)h'(\Delta t) + h_o(\Delta t)h''(\Delta t) \\ &> 0, \end{aligned} \tag{16}$$

where the inequality holds because every term on the right hand side is positive under the condition $\Delta t > \alpha u/\theta_0$ as summarized above. ■

Lemma 2: For any $n \in \mathbb{N}, n \geq 1, T \in \mathbb{R}, T > (n+1)\alpha u/\theta_0$, and $\alpha u/\theta_0 < \Delta t < T - n\alpha u/\theta_0$, the function $G(\Delta t) = ng\left(\frac{T-\Delta t}{n}\right) + g(\Delta t)$ has a minimum of $(n+1)g\left(\frac{T}{n+1}\right)$ attained at $\Delta t = \frac{T}{n+1}$.

Proof: Setting the first derivative of G to zero and solving for Δt results in solving the equation $g'(\Delta t) = g'\left(\frac{T-\Delta t}{n}\right)$. Since the arguments on both side satisfy Condition 1 by the assumption of the lemma, g is convex according to Lemma 1 and g' is a strictly monotonic function. Therefore there exists a unique solution within the range of $(\alpha u/\theta_0, T - n\alpha u/\theta_0)$ to this equation at $\Delta t = \frac{T}{n+1}$.

Next we calculate the second derivative of G at this point. Since $G''(\Delta t) = g''(\Delta t) + \frac{1}{n}g''\left(\frac{T-\Delta t}{n}\right)$, we have $G''\left(\frac{T}{n+1}\right) = (1 + \frac{1}{n})g''\left(\frac{T}{n+1}\right)$. Since $T > (n+1)\alpha u/\theta_0$, g is convex at this stationary point by Lemma 1. Hence G is convex at this point and it is thus a global minimum within the range $(\alpha u/\theta_0, T - n\alpha u/\theta_0)$; the minimum value is $(n+1)g\left(\frac{T}{n+1}\right)$, completing the proof. ■

Theorem 1: Consider a period of time $[0, T]$, in which we wish to schedule $m > 3$ sampling points, including one at time 0 and one at time T . Denote the sequence of time spacings between these samples as $\underline{\Delta t} = [\Delta t_2, \Delta t_3, \dots, \Delta t_m]$, where $\sum_{i=2}^m \Delta t_i = T$. For a given sequence of $\underline{\Delta t}$, define the Fisher information $I(\theta_0)$ as in Eqn (14) and rewrite it as $I(\theta_0; \underline{\Delta t})$ to emphasize its dependence on $\underline{\Delta t}$. Assuming $T > (m-1)\alpha u/\theta_0$, then we have

$$\min_{\underline{\Delta t} \in \mathcal{A}_m} I(\theta_0; \underline{\Delta t}) = (m-1)g\left(\frac{T}{m-1}\right),$$

where $\mathcal{A}_m = \{\Delta t_i : \sum_{i=2}^m \Delta t_i = T, \Delta t_i > \alpha u/\theta_0, i = 2, \dots, m\}$, and with the minimum achieved at $\Delta t_i = \frac{T}{m-1}, i = 2, \dots, m$.

This theorem says that among all sampling strategies with sampling intervals satisfy the sparsity Condition 1, the worst strategy as measured by the Fisher information is the uniform sampling strategy. In other words, any deviation from this strategy improves the Fisher information.

Proof: We prove this by induction on m .

Induction basis: For $m = 3$,

$$I(\theta_0; \underline{\Delta t}) = g(\Delta t_2) + g(\Delta t_3).$$

Using Lemma 1 in the special case of $n = 1$ the result follows.

Induction step: Suppose the result holds for $3, 4, \dots, m$, we want to show it also holds for $m+1$ for $T > m\alpha u/\theta_0$. Note that in this case $\underline{\Delta t} \in \mathcal{A}_{m+1}$ implies that $\alpha u/\theta_0 < \Delta t_{m+1} < T - (m-1)\alpha u/\theta_0$. We thus have

$$\begin{aligned} &\min_{\underline{\Delta t} \in \mathcal{A}_{m+1}} \{I(\theta_0; \underline{\Delta t})\} \\ &= \min_{\underline{\Delta t} \in \mathcal{A}_{m+1}} \left\{ \sum_{i=2}^m g(\Delta t_i) + g(\Delta t_{m+1}) \right\} \\ &= \min_{\Delta t_{m+1} \in \mathcal{A}_{m+1}} \left\{ \min_{\sum \Delta t_i = T - \Delta t_{m+1}} \left\{ \sum_{i=2}^m g(\Delta t_i) \right\} + g(\Delta t_{m+1}) \right\} \\ &= \min_{\Delta t_{m+1} \in \mathcal{A}_{m+1}} \left\{ (m-1)g\left(\frac{T - \Delta t_{m+1}}{m-1}\right) + g(\Delta t_{m+1}) \right\} \\ &= mg\left(\frac{T}{m}\right), \end{aligned}$$

where the third equality is due to the induction hypothesis and the first term on the RHS is obtained at $\Delta t_i = \frac{T - \Delta t_{m+1}}{m-1}$,

$i = 2, \dots, m$. The last equality invokes Lemma 2 in the special case of $n = m - 1$, and is obtained at $\Delta t_{m+1} = \frac{T}{m}$. Combining these we conclude that the minimum value of Fisher information is $mg(\frac{T}{m})$, when $\Delta t_i = \frac{T}{m}, i = 2, \dots, m + 1$. Thus the case $m + 1$ also holds, completing the proof. \blacksquare

Theorem 1 states that given the total sensing period T and the total number of samples m , provided that the sampling is done sparsely (with sufficiently large sampling intervals), the Fisher information attains its minimum when all sampling intervals have the same value, i.e when using a uniform sensing schedule. In this sense uniform sensing is the *worst* possible sensing scheme in that any deviation from it, while keeping the same average sampling interval $T/(m - 1)$, can only increase the Fisher information. This increase in Fisher information become more significant when sampling gets sparser, when $T/(m - 1)$ increases.

Note that for the above result to hold we need the condition $\Delta t_i > \alpha u/\theta_0$. This is because the analysis involves comparing specific sampling sequences, i.e., the average performance (over channel on/off realizations) of an exact sampling sequence. When a random algorithm is used, the sampling sequences are generated randomly, with some sequences performing better than others. The average performance is now also averaged over all possible sampling sequences. Our numerical results indicate that in this average sense a random sensing algorithm performs better than uniform sensing as long as the *average* sampling interval satisfies the sparsity condition (i.e., $T/(m - 1)$ is sufficiently large), rather than requiring the same for every sampling interval for every sampling sequence.

C. A comparison between different random sampling schemes

Having established that random sampling schemes result in higher value of the Fisher information than uniform sensing, we next compare different random sampling schemes. Assuming that the sampling intervals Δt_i 's are generated independently according to some pdf $f(\Delta t)$ we now examine the expectation of the Fisher function, averaged over randomly generated sampling sequences, calculated as follows.

$$\begin{aligned} E[g(\Delta t)] &= \int_0^\infty g(\Delta t)f(\Delta t)d\Delta t & (17) \\ &= \int_0^\infty [g(\mu_o) + g'(\mu_o)(\Delta t - \mu_o) \\ &\quad + \dots + \frac{g^{(n)}(\mu_o)(\Delta t - \mu_o)^n}{n!} + \dots]f(\Delta t)d\Delta t \\ &= g(\mu_o) + g'(\mu_o)\mu_1 + \dots + \frac{g^{(n)}(\mu_o)\mu_n}{n!} + \dots \end{aligned}$$

where the Taylor expansion is around the expected sampling interval $\mu_o = E[\Delta t]$, or $T/(m - 1)$ for given window T and m number of samples taken, and $\mu_n = \int_0^\infty (\Delta t - \mu_o)^n f(\Delta t)d\Delta t$ is the n th order central moment of Δt .

In order to have a fair comparison (and to ensure the same sparsity condition), we will fix T and m , and thus fix the average sampling interval μ_o under different sampling schemes. Also note that the value $g^{(n)}(\mu_o)$ is completely determined by the channel statistics and not the sampling

sequence. Consequently the expected value of the Fisher function is affected by the selection of a sampling scheme only through the higher order central moments resulting from the distribution $f(\cdot)$. Note that the expectation of the Fisher function under uniform sampling with constant sampling interval μ_o is simply $g(\mu_o)$ (i.e., only the first term on the right hand side remains). Therefore any random scheme would improve upon this if it results in a positive sum over the higher order terms. While the above equation does not immediately lead to an optimal selection of a random scheme, it is possible to seek one from a family of distribution functions through optimization over common parameters.

In Table I we have listed the higher order central moments of normal, uniform and exponential distributions². It can be easily concluded that among these three choices the Fisher function has the largest expectation under the exponential distribution. Recall that all quantitative analysis in this section is based on the assumption that the channel has exponentially distributed on/off durations. Since the expectation of the Fisher function is a quadrature of g and pdf f as show in (17), a random sampling scheme with a pdf matching the shape of g will achieve a larger value. This may explain why in this case exponentially distributed sampling intervals perform the best.

As an example, the Fisher function of exponentially distributed on/off channel model with $E[T_1] = 1$ and $[T_0] = 2$ is shown in Fig. 3, and the Fisher information under different sampling distributions are given in Fig. 4. The sampling times under different distributions are generated as follows. We fix a sensing window T and an average number of samples m ³. We place the first and the last sampling times at time 0 and T , respectively. We then sequentially generate $\Delta t_2, \Delta t_2, \dots$ according to a given pdf $f(\cdot)$ with parameters normalized such that it has a mean (sampling interval) of $T/(m - 1)$. For each Δt_i we generate we place a sampling point at time $\sum_{k=2}^i \Delta t_k$. This process stops when this quantity exceeds T . Note that under this procedure the last sampling interval will not be exactly according to $f(\cdot)$ since we have placed a sampling point at time T . However, this approximate seems unavoidable. Alternatively we can allow T to be different from one trial to another while maintaining the same average. As long as T is sufficiently large this procedure does not affect the essence or the fairness of the comparison. (Note that Fig. 4 could also be obtained using numerical integration rather than simulation.)

D. Simulation results

In this subsection we numerically compare the performance of uniform sensing and random sensing. Our simulation is done in Matlab and uses a discrete time model. We assume a time unit of 1s; the on/off periods are rounded to the closest

²Here uniform distribution refers to the sampling interval being randomly selected according to a uniform distribution, not to be confused with uniform sensing where sampling intervals are a constant. For normal distribution the probability distribution function is cut off at zero and then renormalized.

³The reason m is only an average and not an exact requirement is because we cannot guarantee to have exactly m samples within a window of T if we generate sampling intervals randomly according to a given pdf. By allowing m to be an average we can simply require the pdf to have a mean of $T/(m - 1)$.

TABLE I
HIGHER CENTRAL MOMENTS

	Normal	Uniform	Exponential
n is even	$\frac{n! \sigma^n}{(\frac{n}{2})! 2^{\frac{n}{2}}}$	$\frac{\mu^n}{n+1}$	$\mu^n \sum_{k=0}^n \frac{(-1)^k n!}{k!}$
n is odd	0	0	

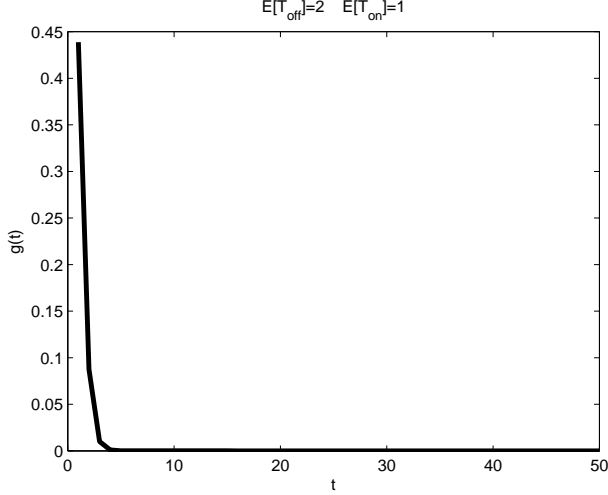


Fig. 3. Fisher function of exponentially distributed channel model

integers⁴. The sampling intervals under uniform sensing are $\lceil T/(m-1) \rceil$ where T is the total length of the sensing period and m the number of samples allowed. Under the random sensing scheme, we fix the average sampling interval

⁴We note that the choice of the time unit is rather arbitrary and inconsequential for our purpose.

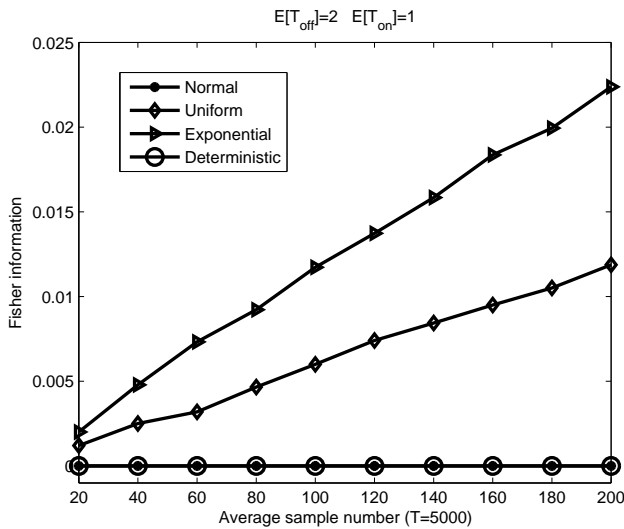


Fig. 4. Comparison of Fisher information: Normal vs. Uniform vs. Exponential

to be $\lceil T/(m-1) \rceil$ and sequentially generate random sampling intervals as described in the previous subsection.

Fig. 5 compares the performance for an exponentially distributed channel model under different values of m over a duration of $T = 4096$ s. The maximum sample number is 4096; this is because the on/off periods are rounded to integers, so there is no point in sampling faster than once per 1 unit of time. Sampling points are randomly placed using a uniform distribution. For both parameters we see that random sensing outperforms uniform sensing, and significantly so when m is small. This validates our earlier analysis.

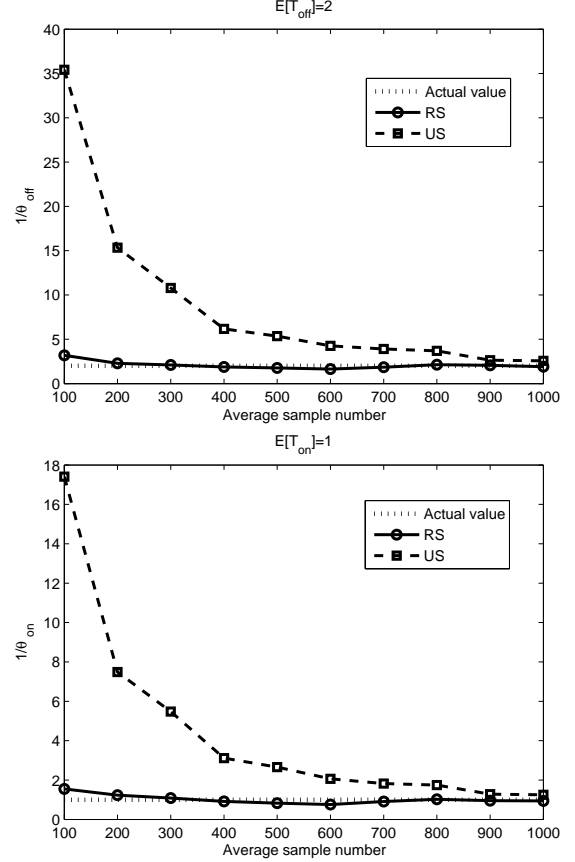


Fig. 5. Performance comparison: random sensing vs. uniform sensing

Fig. 6 shows the performance comparison between random sensing and uniform sensing for a channel with Gamma distributed on/off periods. The probability density functions of such a channel model can be expressed as

$$\begin{cases} f_1(t) = t^{k_1-1} \frac{e^{-t/\lambda_1}}{\lambda_1^{k_1} \Gamma(k_1)} \\ f_0(t) = t^{k_0-1} \frac{e^{-t/\lambda_0}}{\lambda_0^{k_0} \Gamma(k_0)} \end{cases} \quad (18)$$

They are each parameterized by a shape parameter k and a scale parameter λ , both of which are positive. In this case, the Laplace transforms of $f_0(t)$ and $f_1(t)$ are $(1 + \lambda_0 s)^{-k_0}$ and $(1 + \lambda_1 s)^{-k_1}$, respectively, and the expectation of the on/off periods are $E[T_1] = k_1 \lambda_1$ and $E[T_0] = k_0 \lambda_0$. In the simulation both k_1 and k_0 are set to 2, with a simulated

time of 5000s. For random sensing the sampling intervals are randomly generated following an exponential distribution. We see that random sensing again outperforms uniform sensing using such a channel model.

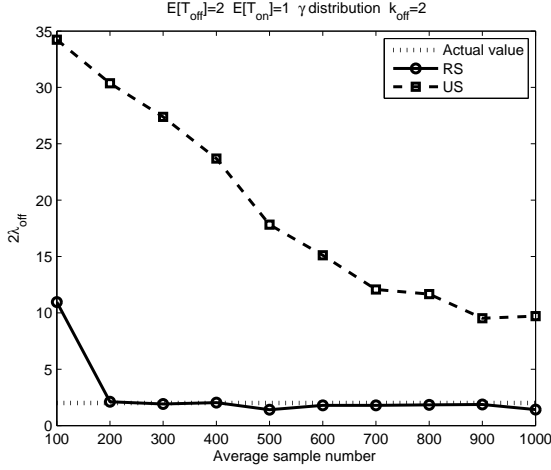


Fig. 6. Performance comparison of gamma distributed channel model

In Fig. 7, we compare the performance between different random sampling schemes. Keeping the same average, the sampling intervals are generated following normal, uniform, and exponential distributions, respectively. We see that exponential random sampling outperforms the other two. This is consistent with our earlier analysis.

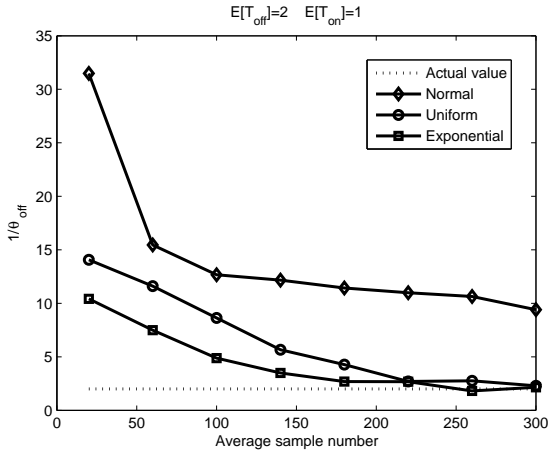


Fig. 7. Performance comparison of random sampling: Normal vs. Uniform vs. Exponential

V. RANDOMIZED UNIFORM SENSING

We see from the preceding analysis that making the sampling intervals non-uniform can improve the performance of the ML estimator. On the other hand, a non-uniform scheme, and in particular a purely random sequence of sampling intervals, will result in a more complex form of the likelihood function and hence computationally more costly.

This motivates us to propose a simple variation to uniform sensing, a hybrid between uniform sensing and random sensing. This is referred to as the *randomized uniform sensing*. As we have pointed out, the main reason why uniform sensing does not perform well is its lack of variability in the sampling intervals. Thus the main idea behind the randomized uniform sensing scheme is to adopt two different sampling interval values, a long one and a short one, denoted by Δt_l and Δt_s , respectively. Assume we sample with the short interval m_s times. Then given the sensing period T and a required number of samples m , we can sense the channel m_l times with long interval Δt_l , where $m_l = m - m_s$ and $\Delta t_l = \frac{T - (m_s - 1)\Delta t_s}{m - m_s}$, assuming we place a sample point at times 0 and T , respectively.

To further simplify the scheme, we can group together long and short intervals. In our numerical experiments presented next we simulated the following specific instance. The algorithm starts by sensing the channel with long intervals Δt_l . Then at a randomly selected time (which could be time 0) or upon completing all m_l samples, whichever occurs first, it switches to sensing with short intervals and collects m_s samples. At this point the algorithm switches back to sensing with long intervals and completes the collection of whatever remains of the m_l samples. Essentially the algorithm seeks to place a group of m_s sampling times, each spaced Δt_s apart, at a random location within the sensing period.

Fig. 8 shows the comparison among random sensing, uniform sensing and randomized uniform sensing. The simulated time is $T = 5000$ s, the short interval is set to be one third of the average sampling interval, the number of short interval samples is set to be $\lfloor m/2 \rfloor$, while maintaining the same interval average of $\lfloor T/m \rfloor$ by adjusting the long intervals. We see clearly that this randomization scheme is very effective yet simple. Since the channel is sampled at alternate but deterministic intervals, the estimation computation is far simpler than for purely random sensing.

VI. CONCLUSION

In this paper, we studied three sensing schemes for the channel estimation problem in cognitive radio networks. Analysis and simulations shows that random sensing with randomized sampling intervals outperforms uniform sensing for exponentially distributed on/off channel model at low sampling rate. A randomized uniform sensing scheme was proposed to obtain better tradeoff between estimation accuracy and its computational cost.

ACKNOWLEDGMENT

This work was supported by NSF grants ANI-0238035 and CCF-0910765, NSF of China grant 60832008, and Key Project of Shandong Provincial Scientific Foundation grant Z2008G01.

REFERENCES

- [1] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE journal on selected areas in communications*, vol. 23, pp: 201-220, Feb. 2005.

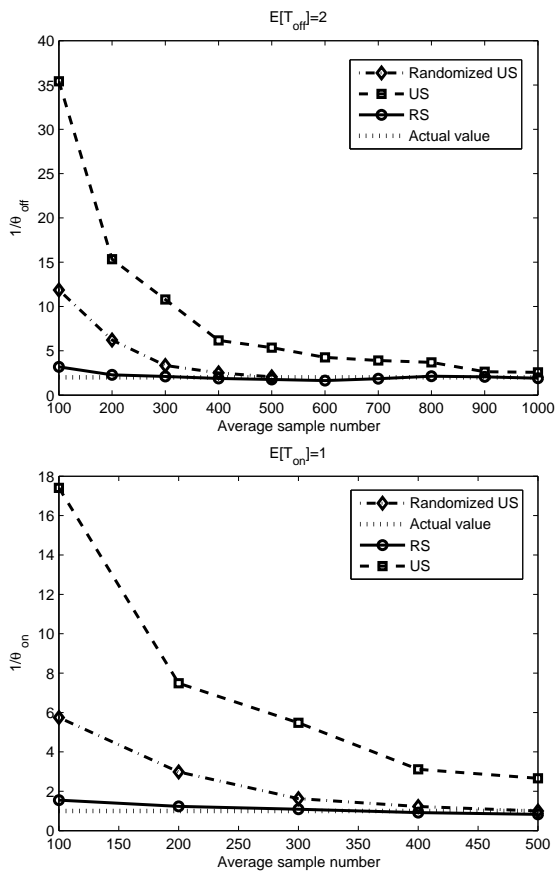


Fig. 8. Performance comparison: RS vs. US vs. randomized US

- [2] K. Challapali, C. Cordeiro, and D. Birru, "Evolution of Spectrum-Agile Cognitive Radios: First Wireless Internet Standard and Beyond," *Proceedings of ACM International Wireless Internet Conference*, August 2006.
- [3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation dynamic spectrum access cognitive radio wireless networks: A survey," *Computer Networks Journal (Elsevier)*, pp: 201-220, Sept. 2006.
- [4] D. Cabric, S. M. Mishra, R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2004.
- [5] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol.7, no.5, pp: 533-545, May 2008.
- [6] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp: 589-599, Apr. 2007.
- [7] H. Kim and K. G. Shin, "Fast Discovery of Spectrum Opportunities in Cognitive Radio Networks," *Proceedings of the 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (IEEE DySPAN)*, Oct. 2008.
- [8] X. Long, X. Gan, Y. Xu, J. Liu, M. Tao, "An Estimation Algorithm of Channel State Transition Probabilities for Cognitive Radio Systems," *Proceedings of Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, 15-17 May 2008
- [9] C. H. Park, S. W. Kim, S. M. Lim, M. S. Song, "HMM Based Channel Status Predictor for Cognitive Radio," *Proceedings of Asia-Pacific Microwave Conference*, 11-14 Dec. 2007.
- [10] A. A. Fuqaha, B. Khan, A. Rayes, M. Guizani, O. Awwad, G. Ben Brahim, "Opportunistic Channel Selection Strategy for Better QoS in Cooperative Networks with Cognitive Radio Capabilities," *IEEE Journal on Selected Areas in Communications*, Vol. 26, No. 1, pp: 156-167, Jan. 2008.
- [11] D. Chen, S. Yin, Q. Zhang, M. Liu and S. Li, "Mining Spectrum Usage Data: a Large-scale Spectrum Measurement Study," *ACM MobiCom*, September 2009, Beijing, China.
- [12] P. J. Kolodzy, "Cognitive radio fundamentals," *Proceedings of SDR Forum*, Singapore, Apr. 2005.
- [13] R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Mathematical Foundations of Theoretical Statistics* vol. 222, pp: 309-368, 1922.
- [14] J. Lee, J. Choi, H. Lou, "Joint Maximum Likelihood Estimation of Channel and Preamble Sequence for WiMAX Systems," *IEEE Transactions on Wireless Communications*, Vol. 7, No. 11, pp: 4294-4303, Nov. 2008.
- [15] J. Wang, A. Dogandzic, A. Nehorai, "Maximum Likelihood Estimation of Compound-Gaussian Clutter and Target Parameters," *IEEE Transactions on Signal Processing*, vol. 54, no.10, pp: 3884-3898 Oct. 2006.
- [16] U. Orguner, M. Demirekler, "Maximum Likelihood Estimation of Transition Probabilities of Jump Markov Linear Systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, Part 2, pp: 5093-5108 Oct. 2008.
- [17] D. R. Cox, *Renewal Theory*, Butler and Tanner, 1967.
- [18] A.W. van der Vaart, *Asymptotic Statistics* (Cambridge Series in Statistical and Probabilistic Mathematics) (1998)
- [19] H.A. Cirpan, M.K. Tsatsanis, "Maximum likelihood blind channel estimation in the presence of Doppler shifts," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp: 1559-1569, Jun. 1999.
- [20] M. Abuthinien, S. Chen, L. Hanzo, "Semi-blind Joint Maximum Likelihood Channel Estimation and Data Detection for MIMO Systems," *IEEE Signal Processing Letters*, vol. 15, pp: 202-205, 2008.
- [21] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh, 1935.
- [22] J.A. Legg, D.A. Gray, "Performance Bounds for Polynomial Phase Parameter," *IEEE Transactions on Signal Processing*, vol. 48, no.2, pp: 331-337 Feb. 2000.
- [23] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp: 4036-4048, 2006.
- [24] E. Candés and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp: 5406-5425, 2006.
- [25] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp: 489-509, 2006.
- [26] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham, and R.G. Baraniuk, "Distributed compressed sensing," 2005, Preprint.
- [27] E. Cands and T. Tao, "Decoding by linear programming," *IEEE Transactions Information Theory*, vol. 51, no. 12, pp: 4203-4215, Dec. 2005.
- [28] Z. Tian and G. Giannakis, "Compressed Sensing for Wideband Cognitive Radios," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. 4, pp: 1357-1360, Honolulu, Apr. 2007.
- [29] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer and M. Zorzi, "On the Interplay Between Routing and Signal Representation for Compressive Sensing in Wireless Sensor Networks," *Information Theory and Applications Workshop (ITA 2009)*, San Diego, CA.
- [30] C. Luo, F. Wu, C. W. Chen and J. Sun, "Compressive Data Gathering for Large-Scale Wireless Sensor Networks," *ACM MobiCom*, September 2009, Beijing, China.
- [31] J. A. Tropp, A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp: 4655-4666, 2007.