Channel Estimation for Opportunistic Spectrum Access: Uniform and Random Sensing

Quanquan Liang, Mingyan Liu, *Senior Member, IEEE*, Dongfeng Yuan, *Senior Member, IEEE*

Abstract—The knowledge of channel statistics can be very helpful in making sound opportunistic spectrum access decisions. It is therefore desirable to be able to efficiently and accurately estimate channel statistics. In this paper we study the problem of optimally placing sensing/sampling times over a time window so as to get the best estimate on the parameters of an on-off renewal channel. We are particularly interested in a sparse sensing regime with a small number of samples relative to the time window size. Using Fisher information as a measure, we analytically derive the best and worst sensing sequences under a sparsity condition. We also present a way to derive the best/worst sequences without this condition using a dynamic programming approach. In both cases the worst turns out to be the uniform sensing sequence, where sensing times are evenly spaced within the window. Interestingly the best sequence is also uniform but with a much smaller sensing interval that requires a priori knowledge of the channel parameters. With these results we argue that without a priori knowledge, a robust sensing strategy should be a randomized strategy. We then compare different random schemes using a family of distributions generated by the circular β ensemble, and propose an adaptive sensing scheme to effectively track time-varying channel parameters. We further discuss the applicability of compressive sensing for this problem.

Index Terms—Spectrum sensing, channel estimation, Fisher information, random sensing, sparse sensing, uniform sensing.

1 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]. A key enabling ingredient is thus 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. Examples like 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.

- Q. Liang and M. Liu are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, quanquan406@gmail.com, mingyan@umich.edu. Q. Liang is also affiliated with the School of Information Science and Engineering, Shangdong University, Shangdong, China. D. Yuan is with the School of Information Science and Engineering, Shangdong University.
- An earlier version of this paper appeared at the Information Theory and Application Workshop (ITA), UC San Diego, CA, February 2010. This work is partially supported by NSF grant CCF 0910765 and NSF of China grant 60832008, 61071122.

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 a *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.

MAC layer channel estimation has been studied in recent years. Below we review those most relevant to the present paper. In [5] an ML estimator was introduced for renewal channels using a uniform sampling scheme. A more accurate but 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 framework of [5]. [9] proposed a Hidden Markov Model (HMM) based channel status predictor using reinforcement learning techniques. [10] presented a channel estimation technique based on wavelet transform followed by filtering.

In most of the above cited work, the focus is on the estimation problem 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. 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. This could be relevant for instance in cases where a user wants to track the channel condition in between active data communication, or where a user needs to track a large number of different channels. 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 samples.

Our main contributions are summarized as follows.

- We demonstrate that when sampling is done sparsely, *random sensing* significantly outperforms uniform sensing.
- In the special case of exponentially distributed on/off durations, we obtain the best and worst possible sampling schemes measured by the Fisher information. We show that uniform sensing is the *worst*; any deviation from it improves the estimation accuracy. Interestingly, the best sensing sequence is also uniform, but with a much smaller sampling interval that requires the knowledge of the underlying channel parameter, thus unimplementable.
- We present an *adaptive random sensing* scheme that can very effectively track time-varying channel parameters, and is shown to outperform its counterpart using uniform sensing.

The remainder of this paper is organized as follows: Section 2 presents the channel models and Section 3 gives the detail of the ML estimator. In Section 4 we present how the sampling scheme affects the estimation performance. In Section 5 we use the circular β ensemble to examine different random sampling schemes. Section 6 presents an adaptive random sensing scheme, and Section 7 concludes the paper.

2 THE CHANNEL MODEL

The channel state perceived by a secondary user is modeled as a binary random variable. This is 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 $Z(t) = 1$ if the channel is sensed busy at time t, and $Z(t) = 0$ otherwise.

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 throughout our analysis. Random detection errors are examined in Section 5.5.

The channel state process $Z(t)$ is assumed to be a continuous-time alternating renewal process. Typically, it is assumed that a secondary user can utilize the channel only when it is sensed to be in the off states. 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.

This random process is completely defined by two probability density functions $f_1(t)$ and $f_0(t)$, $t > 0$, i.e., that 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

$$u = \frac{E[T_1]}{E[T_1] + E[T_0]}, \tag{1}$$

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.

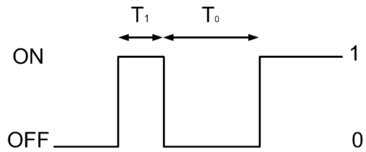


Fig. 1. Channel model: alternating renewal process with on and off states

3 MAXIMUM LIKELIHOOD ESTIMATION

We proceed to describe the maximum likelihood (ML) estimator [12] we will use to estimate channel parameters from a sequence of channel state observations.

An alternative renewal 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 [13]. Assuming the process is in equilibrium, standard results from renewal theory [13] suggest the following Laplace transforms of the above transition probabilities:

$$\begin{aligned}
P^*_{00}(s) &= \frac{1}{s} - \frac{\{1 - f^*_1(s)\}\{1 - f^*_0(s)\}}{E[T_0]s^2\{1 - f^*_1(s)f^*_0(s)\}}\ , \\
P^*_{01}(s) &= \frac{\{1 - f^*_1(s)\}\{1 - f^*_0(s)\}}{E[T_0]s^2\{1 - f^*_1(s)f^*_0(s)\}}\ , \\
P^*_{10}(s) &= \frac{\{1 - f^*_1(s)\}\{1 - f^*_0(s)\}}{E[T_1]s^2\{1 - f^*_1(s)f^*_0(s)\}}\ , \\
P^*_{11}(s) &= \frac{1}{s} - \frac{\{1 - f^*_1(s)\}\{1 - f^*_0(s)\}}{E[T_1]s^2\{1 - f^*_1(s)f^*_0(s)\}}\ ,
\end{aligned} \tag{2}$$

where $f^*_1(s)$ and $f^*_0(s)$ are the Laplace transforms of $f_1(t)$ and $f_0(t)$, respectively. We see that these are 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

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

Substituting the above expressions into (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.

The relevant estimation problem is stated as follows. 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 $[z_1, z_2, \cdots, z_m]$, taken at sampling times $[t_1, t_2, \cdots, t_m]$, 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

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

Let $\underline{\theta}$ be the unknown parameters of the on/off distributions: $\underline{\theta} = [\underline{\theta_1}, \underline{\theta_0}]$. Note that in general $\underline{\theta}_1$ and $\underline{\theta}_0$ are vectors themselves. Then the likelihood function is given by

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

The idea of ML estimation is to find the value of $\underline{\theta}$ that maximizes the log likelihood function $\ln L(\underline{\theta})$. This method has been used extensively in the literature [14]–[18]. 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 [19].

The question we wish to investigate is what impact the selection of the sampling time sequence $[t_1, t_2, \cdots, t_m]$ has on the performance of this estimator, given a limited number of samples m. 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. Other distributions are explored in our numerical experiments. It's worth noting that the exponential distribution assumption is adopted widely in channel models [5]–[7]; spectrum measurement studies have also found it to be a good approximation of real channel vacancy durations [11], [20].

Since the exponential distribution is defined by a single parameter, we have now $\underline{\theta} = [\theta_1, \theta_0]$, 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 is given by (7) where $\Delta t_i = t_i - t_{i-1}$. The first quantity on the right is taken to be

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

the steady state probability of finding the channel in a particular state. This is justified by assuming that the channel is in equilibrium.

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

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

The estimates for the parameters are found by solving

$$(\hat{\theta}_1, \hat{\theta}_0) = \arg\max_{\theta_1,\theta_0} \; \ln L(\theta_0, \theta_1) . \quad (10)$$

The above joint optimization proves to be computationally complex and analytically intractable. Instead, we adopt the following sub-optimal estimation procedure, also used in [5]. We first estimate u using Eqn (5), and take $\theta_1 = \frac{(1-u)\theta_0}{u}$. Due to the exponential assumption, it can be shown that this estimate of u is unbiased regardless of the sequence $[t_1, \cdots, t_m]$ as long as it is determined offline. The likelihood function (9) can then be re-written as

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

The estimation of θ_0 is then derived by solving $\max_{\theta_0} \ln L(\theta_0)$

In our analysis, 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 is separately and 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). However, this is a reasonable approach, computationally feasible, and much more amenable to analysis.

4 Best and worst sensing sequences

The goal of this study is to judiciously schedule a very limited number of sampling times so that the estimation accuracy is least affected. We first argue intuitively why the commonly used uniform sampling

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

does not perform well when the number of samples allowed is limited. We then present a precise analysis through the use of Fisher information in the case of exponential on/off distributions. We show that using this measure, under a certain sparsity condition uniform sensing is the *worst* schedule in terms of estimation accuracy. We also derive an upper bound on the Fisher information as well as the sampling sequence achieving this upper bound. These provide us with useful benchmarks to assess any arbitrary sampling sequence.

4.1 An intuitive explanation

We begin this section by presenting a comparison between sensing periodically (i.e., uniform sensing) and sensing randomly when the average sensing frequency is low. Uniform sensing is an easy-to-implement, and easy-to-analyze scheme. 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]. Figure 2 shows a comparison between these two sensing schemes. In this comparison the sampling times of the random sensing scheme are randomly placed using a uniform distribution[1] within a window of 5000 time units. The on/off periods are exponentially distributed with parameters $E[T_0] = 2$, $E[T_1] = 1$ time units, respectively. The figure shows the estimated value of $E[T_0]$ as a function of the number of samples taken within the window of 5000. We see that random sensing outperforms uniform sensing, and significantly so when m is small.

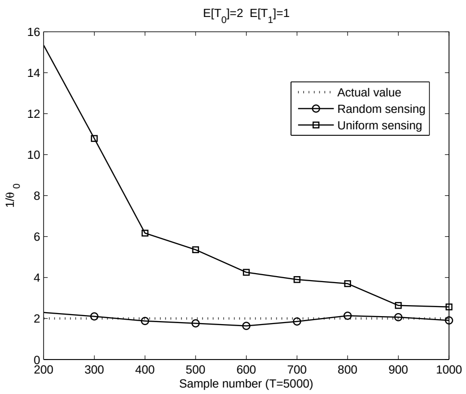


Fig. 2. Estimation accuracy on θ_0: uniform sensing vs. random sensing

1. Here uniform distribution refers to the sampling times being randomly placed within the window following a uniform distribution, not to be confused with uniform sensing where sampling intervals are a constant.

To understand the reasons behind this performance difference, we first note that since there is no variation across sampling intervals under uniform sensing, the uniform interval in general needs to be upper-bounded in order to catch potential channel state changes that occur over small intervals. This bound cannot be guaranteed under sparse sensing. If sensing is done randomly, then even if the average sampling interval is large, there can be significant probability for sufficiently small sampling intervals to exist in any realization.

Furthermore, consider the transition probabilities $P_{ij}(\Delta t)$, $i, j \in \{0, 1\}$, which completely define the likelihood function. They approach the stationary probabilities as Δt increases: $P_{01}(\Delta t) \rightarrow \frac{E[T_1]}{E[T_1]+E[T_0]} = u$ as $\Delta t \rightarrow \infty$. This stationary quantity represents the average fraction of time the channel is busy, but 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 Markov chain, this convergence can occur rather quickly. Therefore, under sparse uniform sensing, these transition probabilities will become constant-like. This in turn causes the likelihood function to be constant-like, making it difficult for the ML estimator to produce accurate estimates [12]. It should be noted however that for our estimation problem high variability can also be introduced in a deterministic sampling sequence. In this sense a sequence does not have to be randomly generated; as long as it contains sufficient variability, estimation accuracy can be improved.

4.2 Fisher information and preliminaries

We now analyze this notion of information content more formally via a measure known as the *Fisher information* [21]. For the likelihood function given in Eqn (11), the Fisher information is defined as:

$$I(\theta_0) \quad = \quad -E[\frac{\partial^2 \ln L(\theta_0)}{\partial \theta_0^2}] \ . \qquad (12)$$

This 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]). The precision to which we can estimate θ_0 is fundamentally limited by the Fisher information of the likelihood function. Note that in (12) we have suppressed u from the argument as u is estimated separately and taken as a constant in our subsequent analysis.

Due to the product form of the likelihood function, we have

$$\begin{aligned} I(\theta_0) &= -E\Big[\sum_{i=2}^{m} \frac{\partial^2 \ln[\alpha_i + \beta_i e^{-\theta_0 \Delta t_i/u}]}{\partial \theta_0^2}\Big] \\ &= \sum_{i=2}^{m} \frac{\Delta t_i^2}{u^2} E\Big[\frac{-\alpha_i \beta_i e^{-\theta_0 \Delta t_i/u}}{(\alpha_i + \beta_i e^{-\theta_0 \Delta t_i/u})^2}\Big] , \quad (13) \end{aligned}$$

where $\alpha_i = u^{z_i}(1-u)^{1-z_i}$ and $\beta_i = (-1)^{z_i+z_{i-1}} u^{1-z_{i-1}}(1-u)^{z_{i-1}}$. Define:

$$g(\theta_0; \Delta t_i) = \frac{\Delta t_i^2}{u^2} E\Big[\frac{-\alpha_i \beta_i e^{-\theta_0 \Delta t_i/u}}{(\alpha_i + \beta_i e^{-\theta_0 \Delta t_i/u})^2}\Big] , \quad (14)$$

so that the Fisher information can be simply written as $I(\theta_0) = \sum_{i=2}^{m} g(\theta_0; \Delta t_i)$. The function $g()$ will be referred to as the *Fisher function* in our discussion. Note that $g()$ is a function of both Δt_i and θ_0. However, we will suppress θ_0 from the argument 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 θ_0 constant. Note that the first term in Eqn (11) 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 θ_0. Therefore the term disappears after the differentiation.

The expectation on the RHS of (13) 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 in Eqn (15). Below we show that under a certain *sparsity* condition on the sampling rate, the Fisher function is strictly convex and the Fisher information is minimized under uniform sampling.

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

Lemma 1: The Fisher function $g(\Delta t)$ given in Eqn (15) is strictly convex under Condition 1.

The proof of this lemma can be found in the Appendix. Using this lemma we next derive tight lower and upper bounds of the Fisher information.

4.3 A tight lower bound on the Fisher information

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 $G(\Delta t) = ng(\frac{T-\Delta t}{n}) + g(\Delta t)$ has a minimum of $(n+1)g(\frac{T}{n+1})$ 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 $g^{'}(\Delta t) = g^{'}(\frac{T-\Delta t}{n})$. Since the arguments on both side satisfy Condition 1, by the assumption of the lemma, g is strictly 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 $G^{''}(\Delta t) = g^{''}(\Delta t) + \frac{1}{n} g^{''}(\frac{T-\Delta t}{n})$, we have $G^{''}(\frac{T}{n+1}) = (1+\frac{1}{n}) g^{''}(\frac{T}{n+1})$. 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 $(n+1)g(\frac{T}{n+1})$, completing the proof. □

Theorem 1: Consider a period of time $[0, T]$, in which we wish to schedule $m \geq 3$ sampling points, including one at time 0 and one at time T. Denote the sequence of time spacings between these samples as $\underline{\Delta t} = [\Delta t_2, \Delta t_3, \cdots, \Delta t_m]$, where $\sum_{i=2}^{m} \Delta t_i = T$. For a given sequence $\underline{\Delta t}$, define the Fisher information $I(\theta_0)$ as in Eqn (13) 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

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

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

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 $3, 4, \ldots m$, 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 $\alpha u/\theta_0 < \Delta t_{m+1} < T - (m-1)\alpha u/\theta_0$, which will be denoted as $\Delta t_{m+1} \in \mathcal{A}_{m+1}$ below for convenience. We thus have

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

where the third equality is due to the induction hypothesis and the first term on the RHS is obtained

$$g(\Delta t) = \frac{\Delta t^2}{u^2} e^{-\theta_0 \Delta t/u} [\frac{u^2(1-u)}{u - ue^{-\theta_0 \Delta t/u}} + \frac{u(1-u)^2}{(1-u) - (1-u)e^{-\theta_0 \Delta t/u}} - \frac{u(1-u)^2}{(1-u) + ue^{-\theta_0 \Delta t/u}} - \frac{u^2(1-u)}{u + (1-u)e^{-\theta_0 \Delta t/u}}] \ . \tag{15}$$

at $\Delta t_i = \frac{T - \Delta t_{m+1}}{m-1}$, $i = 2, \ldots, m$. 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 $\Delta t_i = \frac{T}{m}, i = 2, \ldots, m+1$. Thus the case $m+1$ also holds, completing the proof. □

Theorem 1 states that given the total sensing period T and the total number of samples m, provided that the sampling is done sparsely (per Condition 1), the Fisher information attains its minimum when all sampling intervals are the same, i.e a uniform sensing schedule. In this sense uniform sensing is the *worst* possible sensing scheme; any deviation from it, while keeping the same average sampling interval $T/(m-1)$, can only increase the Fisher information. As we have seen in Figure 2, this increase in Fisher information becomes more significant when sampling gets sparser, i.e., when m decreases.

4.4 A tight upper bound on the Fisher information

The derivation of the upper bound follows very similar steps as those for the lower bound.

Lemma 3: For any $T \in \mathbb{R}, T > 2\alpha u/\theta_0$, and $\alpha u/\theta_0 < \Delta t < T - \alpha u/\theta_0$, the function $F(\Delta t) = g(T - \Delta t) + g(\Delta t)$ has a maximum of $g(\alpha u/\theta_0) + g(T - \alpha u/\theta_0)$ attained at $\Delta t = \alpha u/\theta_0$ or $\Delta t = T - \alpha u/\theta_0$.

Proof: We first prove that F is convex under the stated conditions. We have $F^{'}(\Delta t) = g^{'}(\Delta t) - g^{'}(T - \Delta t)$. Since g is strictly convex under the stated conditions, by Lemma 1 $g^{'}$ is monotonic increasing. Thus $F^{'}$ is also monotonic increasing, hence F is convex. It follows that the maximum of $F(\Delta t)$ is attained at one and/or the other extreme point of Δt. In either case we have

$$F(\alpha u/\theta_0) = F(T - \alpha u/\theta_0) = g(\alpha u/\theta_0) + g(T - \alpha u/\theta_0).$$

□

Theorem 2: Consider a period of time $[0, T]$, in which we wish to schedule $m \geq 3$ sampling points, including one at time 0 and one at time T. Denote the sequence of time spacings between these samples as $\underline{\Delta t} = [\Delta t_2, \Delta t_3, \cdots, \Delta t_m]$, where $\sum_{i=2}^{m} \Delta t_i = T$. Assuming $T > (m-1)\alpha u/\theta_0$, then we have

$$\begin{aligned} \max_{\underline{\Delta t} \in \mathcal{A}_m} I(\theta_0; \underline{\Delta t}) &= (m-2)g(\alpha u/\theta_0) \\ &\quad + g(T - (m-2)\alpha u/\theta_0), \end{aligned}$$

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

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 3 the result immediately follows.

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

$$\begin{aligned} &\max_{\underline{\Delta t} \in \mathcal{A}_{m+1}} \{I(\theta_0; \underline{\Delta t})\} \\ = &\max_{\underline{\Delta t} \in \mathcal{A}_{m+1}} \left\{ \sum_{i=2}^{m} g(\Delta t_i) + g(\Delta t_{m+1}) \right\} \\ = &\max_{\Delta t_{m+1} \in \mathcal{A}_{m+1}} \left\{ \max_{\sum \Delta t_i = T - \Delta t_{m+1}} \left\{ \sum_{i=2}^{m} g(\Delta t_i) \right\} \right. \\ &\left. \qquad + g(\Delta t_{m+1}) \right\} \\ = &\max_{\Delta t_{m+1} \in \mathcal{A}_{m+1}} \left\{ (m-2)g(\alpha u/\theta_0) \right. \\ &\left. \quad + g(T - \Delta t_{m+1} - (m-2)\alpha u/\theta_0) + g(\Delta t_{m+1}) \right\} \\ = &\ (m-1)g(\alpha u/\theta_0) + g(T - (m-1)\alpha u/\theta_0) \ , \end{aligned}$$

where the third equality is due to the induction hypothesis and the first term on the RHS is obtained at $\Delta t_i = \alpha u/\theta_0$, $i = 2, \ldots, m-1$ and $\Delta t_m = T - \Delta t_{m+1} - (m-2)\alpha u/\theta_0$. The last equality invokes Lemma 3, and is obtained at $\Delta t_{m+1} = T - (m-1)\alpha u/\theta_0$ or $\Delta t_{m+1} = \alpha u/\theta_0$. Thus the case $m+1$ also holds, completing the proof. □

We see from this theorem that under the sparsity condition, the best sensing sequence is to sample at the smallest interval that the condition would allow, till we use all the $m-2$ samples we have the freedom of placing. This produces a *pseudo uniform* sequence of sampling times; it forms a uniform sequence except for the last sampling interval. It can be shown that if we remove the constraint of having a window of T, but rather seek to optimally place m points subject to the sparsity condition, then the optimal sequence would be exactly uniform with the interval $\Delta t_i = \alpha u/\theta_0$. However, it should be emphasized that since θ_0 is the very thing we are trying to estimate, it would be unreasonable to suggest that this optimal interval is known a priori. Thus this optimal sequence is not implementable. It nevertheless sheds light on the nature of the sequence that maximizes the Fisher information.

4.5 Best and worst sampling schemes without the sparsity condition

We next show how to obtain the best and worst sensing sequences in a more general setting, without the requirement of Condition 1, via the use of dynamic programming. While this result is more general compared to those derived under the sparsity condition, structurally they are not as easy to identify and are thus given in a numerical form. We also note that these sequences are not practically implementable as they assume a priori knowledge of the parameters to be estimated. They are derived only to serve as benchmarks.

Denote by π a sampling policy given by the time sequence $[t_1, t_1, \cdots, t_m]$. Then the optimal sampling policy is given by

$$\pi^* = \arg\max_{\pi \in \Pi} I(\theta_0) \ , \qquad (16)$$

where the set of admissible policies $\Pi = \{t_i : t_1 = 0, t_m = T, 0 < t_2 < \cdots < t_{m-1} < T\}$.

The maximum $I(\theta_0)$ can be recursively solved through the set of dynamic programming equations given below:

$$\begin{aligned} V(1,t) &= g(T-t), \ \ \forall\ 0 \le t < T; \\ V(k,t) &= \max_{t<x<T} [g(x-t) + V(k-1,x)], \\ & \quad \forall\ 0 \le t < T,\ k = 2, 3, \cdots, m-1 \ , \end{aligned} \qquad (17)$$

and

$$\max I(\theta_0) = \max_{0<t<T} [g(t) + V(m-1,t)] \ . \qquad (18)$$

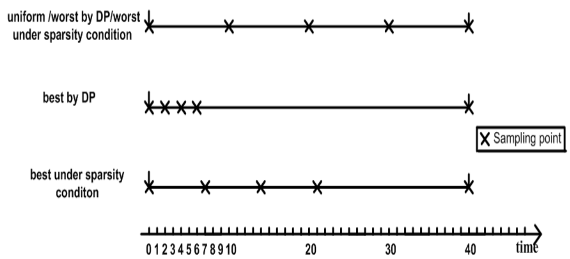
Here the value function $V(k,t)$ denotes the maximum achievable Fisher information given we last sampled at time t, with k points remaining to be placed between $(t, T]$.

Note that since t is continuous, the pair (k,t) has an uncountable state space. In computing the DP equation (17) we discretize t and T into small steps and require that both be integer multiples of this small quantity. The resulting DP has a finite state space and is solved backwards in time [23].

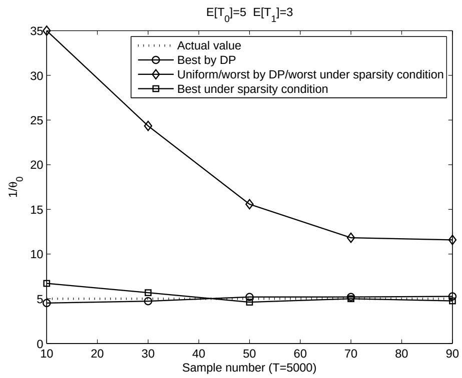
It is straightforward to see the exact same procedure can be used to find the sampling sequence that *minimizes* the Fisher information, thus giving the worst sampling sequence. It turns out that the worst sampling sequence in this case coincides with the worst sequence derived under the sparsity condition, i.e., it is also the uniform sequence.

4.6 A comparison

We now compare the different sensing sequences we obtained in this section using an example. They are illustrated in Figure 3(a). In this example the channel parameters are $E[T_0] = 5$ and $E[T_1] = 3$ time units, respectively. The time window is set to be 40 time units, and the channel can only be sensed 5 times. Shown in the figure are the uniform sensing sequence, the best/worst sensing sequences derived under the sparsity condition, and the best/worst sequences derived using dynamic programming. As mentioned earlier, the worst obtained via dynamic programming coincides with the uniform sampling sequence. The worst under the sparsity condition also coincides with the uniform sequence, a fact proven in Theorem 1, as the sparsity condition holds in this case. In Figure 3(b), we compared the performance of these sampling strategies, by setting the time window to 5000 time units. The estimated value under each strategy is shown as a function of the number of samples taken. The true value is also shown for comparison. These are used as benchmarks in the next section in evaluating random sensing schemes.



(a) Illustration of different sampling sequences



(b) Performance comparison

Fig. 3. Comparison of different sampling sequence

As we can see from Figure 3(a), the best sensing sequence produced by dynamic programming without the sparsity condition also appears to be *pseudo uniform*, as in the case with the best sequence under the sparsity condition[2]. The difference is that the former uses a smaller interval value that violates the sparsity condition. As mentioned earlier, if we were to remove the requirement that one sample be placed at time T, then the optimal sequence of m would appear to be uniform (again, this conclusion is drawn empirically in the case of no sparsity requirement, and

2. Note however that this conclusion is drawn empirically from a large amount of numerical experiment in the case of not requiring sparsity. By contrast, under the sparsity condition the conclusion is drawn analytically in Theorem 2.

precisely and analytically in the case of sparsity), with the optimal interval being the value that maximizes (15). Interestingly, the worst sequence is also uniform with or without the sparsity condition.

What this result suggests is that in the ideal case if we have a priori knowledge of the channel parameters, to maximize the Fisher information the best thing to do is indeed to sense uniformly. The difficulty of course is that without this knowledge we have no way of deciding what the optimal interval should be, and uniform sensing would be a bad decision as it could turn out to be the worst with an unfortunate choice of the sampling interval.

In such cases, the robust thing to do is simply to sense randomly, so that with some probability we will have sampling intervals close to the actual optimum. This is investigated in the next section.

5 Random sensing

Under a random sensing scheme, the sampling intervals Δt_i are generated according to some distribution $f(\Delta t)$ (this may be done independently or jointly). Below we first analyze how the resulting Fisher information is affected, and then use a family of distributions generated by the circular β ensemble to examine the performance of different distributions.

5.1 Effect on the Fisher information

We begin by examining the expectation of the Fisher function, averaged over randomly generated sampling intervals, calculated as (19), 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 nth order central moment of Δt.

In order to have a fair comparison we will assume T and m are fixed, thus fixing 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 of the distribution $f()$. 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.

Before we proceed with this in the next subsection, we compare the normal, uniform and exponential random sampling schemes using the above analysis. In Table 1 we list the higher order central moments of normal, uniform and exponential distributions [3]. It can be easily concluded that among these three choices the Fisher function has the largest expectation under the exponential distribution.

TABLE 1
Higher central moments

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

We further compare their performance in Fig. 4 as we increase the number of samples m over a window of $T = 5000$ time units. Our simulation is done in Matlab and uses a discrete time model; all time quantities are in the same time units. All values presented here are the averages over 100 independent runs. The maximum number of samples is 5000; this is because the on/off periods are integers, so there is no reason to sample faster than once per unit of time. The sampling intervals under the uniform sensing are $\lfloor T/(m-1) \rfloor$. The sampling times under random schemes are generated as follows. We fix the window T and take m to be the average number of samples[4]. We place the first and the last sampling times at time 0 and T, respectively. We then sequentially generate $\Delta t_2, \Delta t_3, \cdots$ according to the given pdf $f()$ 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 $f()$ since we have placed a sampling point at time T. However, this approximation 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 accuracy or the fairness of the comparison. For each value of m, the result shown on the figure is the average of 100 randomly generated sensing schedules. We see that exponential random sampling outperforms the other two; this is consistent with our earlier analysis on the Fisher information.

5.2 Circular β ensemble

We now use the circular β ensemble [24] to study a family of distributions. The advantage of using this

3. For normal distribution the probability distribution function is cut off at zero and then renormalized.

4. The 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 $T/(m-1)$.

$$E[g(\Delta t)] = \int_0^\infty g(\Delta t)f(\Delta t)d\Delta t = \int_0^\infty [g(\mu_o) + g'(\mu_o)(\Delta t - \mu_o) + \cdots + \frac{g^{(n)}(\mu_o)(\Delta t - \mu_o)^n}{n!} + \cdots]f(\Delta t)d\Delta t$$
$$= g(\mu_o) + g'(\mu_o)\mu_1 + \cdots + \frac{g^{(n)}(\mu_o)\mu_n}{n!} + \cdots \quad (19)$$

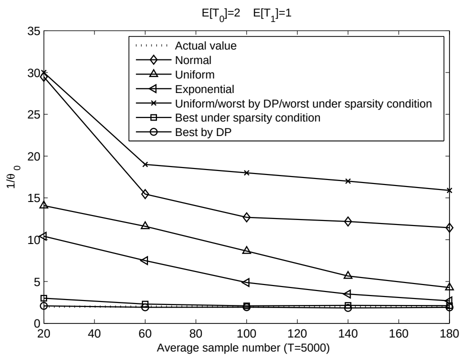


Fig. 4. Performance comparison of random sensing

ensemble is that with a single tunable parameter we can approximate a wide range of different distributions while keeping the same average sampling rate.

The circular β ensemble may be viewed as given by n eigenvalues, denoted as $\lambda_j = e^{i\phi_j}$, $j = 1, \cdots, n$. These eigenvalues have a joint probability density function proportional to the following:

$$\prod_{1\le k<l\le n} |e^{i\phi_k} - e^{i\phi_l}|^\beta, \quad -\pi < \phi_j \le \pi, \quad j,k,l = 1,\cdots,n,$$

where $\beta > 0$ is a model parameter. In the special cases $\beta = 1, 2$ and 4, this ensemble describes the joint probability density of the eigenvalues of random orthogonal, unitary and sympletic matrices, respectively [24].

We use the set of eigenvalues generated from the above joint pdf to determine the placement of sample points in the interval $[0, T]$ in the following manner. In [25] a procedure is introduced to generate a set of values ϕ_j, $j = 1, 2, \cdots, n$ that follow the above joint pdf. Setting $n = m$, these n eigenvalues are then placed along a unit circle (each at the position given by ϕ_j), which are subsequently mapped onto the line segment $[0, 1]$. Scaling this segment to $[0, T]$ gives us the m sampling times. The intervals between these points now follow a certain joint distribution indexed by β. Below we will refer to this method of generating sample points/intervals as using the circular β ensemble.

In Fig. 5 we give the pdfs of intervals generated by the circular ensemble with different β. For each value of β, we obtain 200 random variables in [0, 1], then scale them to be in [0, 5000]. The successive intervals between neighboring points are collected with the their pdf shown in the figure. We see that as β approaches 0^+ the pdf becomes exponential-like and as β approaches $+\infty$, the pdf becomes deterministic; these are well known facts about circular ensembles.

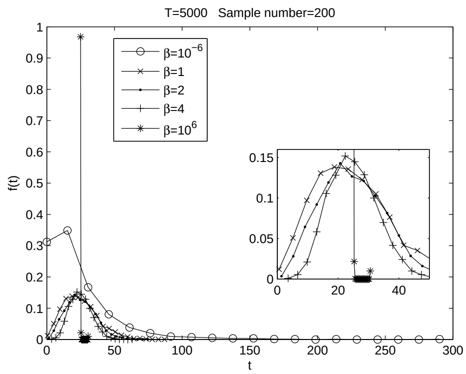


Fig. 5. Probability distribution function of intervals generated by the circular β ensemble

5.3 A comparison between different random sensing schemes

In Fig. 6 we show the Fisher information with sampling intervals generated by the circular β ensemble. The corresponding estimation performance comparison is given in Fig. 7. The performance of the best and worst sequences with and without the sparsity condition are also shown for comparison. Note that when $\beta = 10^6$, the sampling sequence coincide with the worst obtained via dynamic programming, the worst under sparsity condition and uniform sensing, therefore their performances are the same. We see again that exponentially generated sampling intervals performs the best. This may be due to the fact that the on/off durations are also exponentially distributed, thereby creating a good "match" between the fisher function $g()$ and the pdf $f()$ that results in a larger value of the expected Fisher function value (see Eqn. (19)).

5.4 Discussions on other channel models

We now examine a channel model with on/off durations following the gamma distribution, given by

$$\begin{cases} f_1(t) = t^{k_1-1}\frac{e^{-t/\lambda_1}}{\lambda_1^{k_1}\Gamma(k_1)} \\ f_0(t) = t^{k_0-1}\frac{e^{-t/\lambda_0}}{\lambda_0^{k_0}\Gamma(k_0)} \end{cases}. \quad (20)$$

They are each parameterized by a shape parameter k and a scale parameter λ, both positive. In this

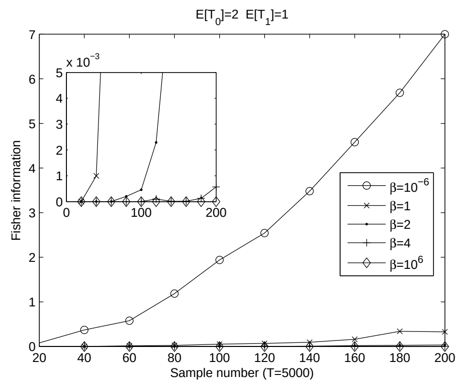


Fig. 6. Fisher information with intervals generated by the circular β ensemble for exponential channel

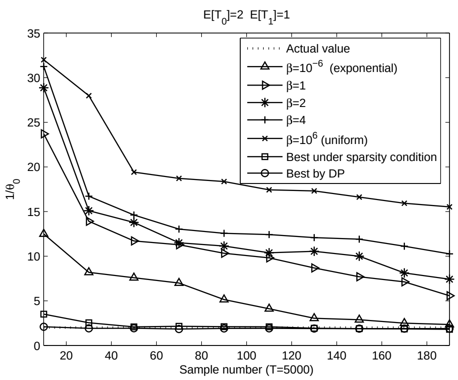


Fig. 7. Performance of random sampling using circular β ensemble for an exponential channel

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 following simulation both k_1 and k_0 are set to 2, with a simulated time of 5000 time units. The sampling intervals are randomly generated by the circular β ensemble. We see in Fig. 8 that random sensing again outperforms uniform sensing using such a channel model.

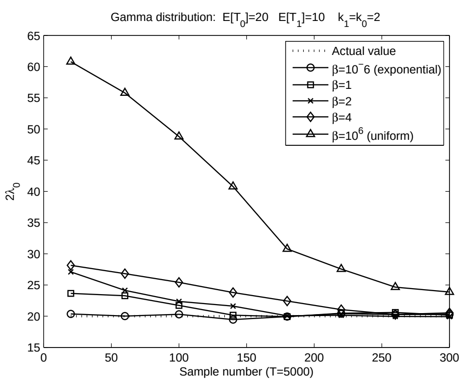


Fig. 8. Performance of random sampling using circular β ensemble for a gamma channel

Obtaining similar result for other channel distributions becomes computationally prohibitive. There are two reasons. Firstly, for most distributions the Laplace transform is complex, resulting in the complexity in obtaining the corresponding time domain expressions. Secondly, with the exception of the exponential distribution, without the memoryless property the likelihood function also becomes intractable.

We next examine a 3-state Markov channel model recently proposed in [26] as an alternative to the commonly used 2-state Markov model (the so-called Gilbert-Elliot model, which is also the exponential on/off model analyzed in this paper). This 3-state Markov chain produces higher variance of the on/off durations and is shown to better match spectrum measurement data. Under this model, the state "0" denotes an off state, while states "1" and "2" are both on states. Each state i has a probability $1-p_i$ of remaining in the same state, $i = 0, 1, 2$. From state "0" the chain visits states "1" and "2" with transition probabilities αp_0 and $(1-\alpha)p_0$, respectively, while from states "1" and "2" the transition probabilities to state "0" are p_1 and p_2, respectively.

The results of applying our sensing algorithms to traces generated by this 3-state Markovian model are shown in Fig. 9, by using $\alpha = 0.7, p_0 = 0.49, p_1 = 0.91, p_2 = 0.06$ (taken from [26]). In performing estimation we treated the channel as if it were a memoryless on/off channel which is not the case here due to the two distinct on states. We see that random sensing again outperforms uniform sensing.

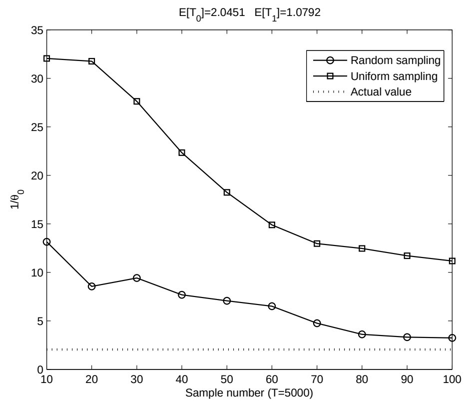


Fig. 9. Performance under a 3-state Markov model

5.5 Effect of channel state detection error

So far we have assumed that the channel state process $Z(t)$ is perfectly detected. Fig. 10 shows the estimation performance in the presence of detection error. We inserted random detection errors into the 0-1 sequence with varying probabilities, and performed estimation using the revised sequence. The actual mean value refers to the sequence before error insertion. We see that random sampling is very robust to random detection errors and consistently outperforms the uniform sampling method. Our explanation is that since sampling is sparsely done, when detection error probability is not very high (the highest shown is 20%) the chance of taking a sample that happens to be an error is even lower. Note that when using

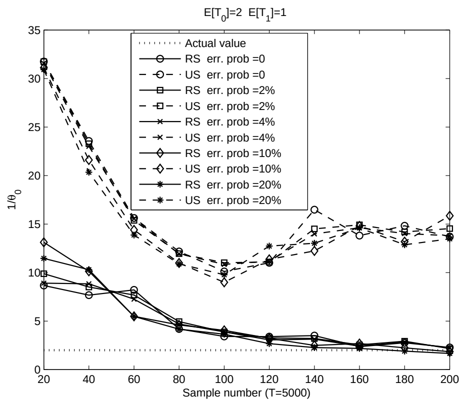


Fig. 10. Estimation accuracy on θ_0 with imperfect channel detection: uniform sensing vs. random sensing

uniform sampling the estimation performance first improves with the increasing sampling rate, but then (around 100 samples) it starts to deteriorate. This is because as the sampling rate increases the number of errors in the sampled sequence also increases. Similar phenomenon also exists for random sampling, but occurs at a much higher sampling rate beyond the sparse sampling regime. It is interesting to note that under uniform sampling the estimate starts to deviate before it could converge to the actual value.

6 Adaptive Random Sensing for Parameter Tracking

Using insights we have obtained on uniform sensing and random sensing, we now present a method of estimating and tracking a time-varying parameter. The overall sensing duration T is divided into windows of lengths T_w. In each window samples are taken and an estimate produced at the end of that window. This estimate is then used to determine the optimal number of samples to be taken in the next window. This will be referred to as the *adaptive random sensing* scheme; the adaptive nature lies in adjusting the number of samples taken in each window based on past estimates.

Specifically, at the end of the i-th window of T_w, we obtain the ML estimate $\hat{\theta}_0^{(i)}$ and $\hat{u}^{(i)}$ based on samples collected during that window. Now *assuming* that we will use uniform sensing in the $(i+1)$th window with a sampling interval Δt_p, and assuming that $\hat{\theta}_0^{(i)}$ and $\hat{u}^{(i)}$ are the true parameter values in the $(i+1)$th window, we can obtain the *expectation* of the next estimate, denoted as $\tilde{\theta}_0^{(i+1)}$, as a function of $(T_w, \Delta t_p, \hat{u}^{(i)}, \hat{\theta}_0^{(i)})$. The optimal sampling interval $\Delta t_p^{(i+1)}$ for the $(i+1)$th window is then calculated as follows:

$$\Delta t_p^{(i+1)} = \arg\min_{\Delta t_p} \left| |\tilde{\theta}_0^{(i+1)} - \hat{\theta}_0^{(i)}| - \varepsilon \right| , \qquad (21)$$

where ε is an error factor introduced to lower bound the minimizing interval $\Delta t_p^{(i+1)}$. Without this factor the interval will end up being very small, i.e., requiring a large number of samples for the next window. The intuition behind the above formula is that assuming the channel parameters are relatively slow varying in time, the estimate from the previous window $\hat{\theta}_0^{(i)}$ may be viewed as true. So for the next window we would like to find the sampling interval that allows us to get as close as possible to this value subject to an error.

Note that the above calculation relies on the availability of $\tilde{\theta}_0^{(i+1)}$, a quantity obtained assuming uniform sampling will be used in the next window. In the actual execution of the algorithm, we simply use this to obtain $\Delta t_p^{(i+1)}$ as shown above. This gives us the desired number of samples to be taken in the next window: $M^{(i+1)} = \lceil T_w / \Delta t_p^{(i+1)} \rceil$. Following this, random sensing is used to generate $M^{(i+1)}$ random sampling times within the next window. An estimate is then made and this process repeats.

It remains to show how $\tilde{\theta}_0^{(i+1)}$ is obtained. As mentioned earlier, when the on/off periods are exponentially distributed there is a simple closed-form solution to the ML estimator. This was calculated in [5] and we will use that result directly below. Specifically, with $M = \lceil T_w / \Delta t_p \rceil$ samples uniformly taken, the estimate of channel utilization u is given by $\hat{u} = \frac{1}{M} \sum_{i=1}^{M} z_i$. The estimate of θ_0 is given by

$$\hat{\theta}_0 = -\frac{u}{\Delta t_p} \ln[\frac{-B + \sqrt{B^2 - 4AC}}{2A}], \qquad (22)$$

where

$$\begin{cases} A = (u - u^2)(M-1) \\ B = -2A + (M-1) - (1-u)n_0 - un_3 \\ C = A - un_0 - (1-u)n_3 \end{cases} . \qquad (23)$$

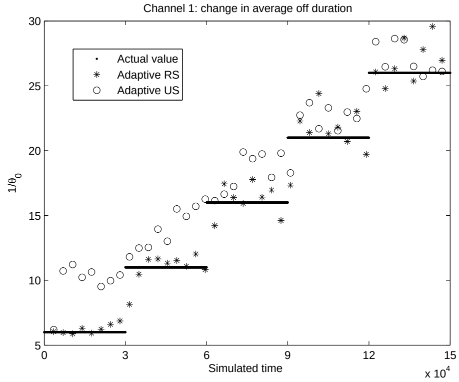
Here $n_0/n_1/n_2/n_3$ denotes the number of $(0 \to 0)/(0 \to 1)/(1 \to 0)/(1 \to 1)$ transitions out of the total $(M-1)$ transitions. Their respective expectations are given by

$$\begin{array}{l} E[n_0] = M(1-u)P_{00}(\Delta t_p; \theta_0), \\ E[n_2] = MuP_{10}(\Delta t_p; \theta_0), \\ E[n_1] = M(1-u)P_{01}(\Delta t_p; \theta_0), \\ E[n_3] = MuP_{11}(\Delta t_p; \theta_0). \end{array} \qquad (24)$$

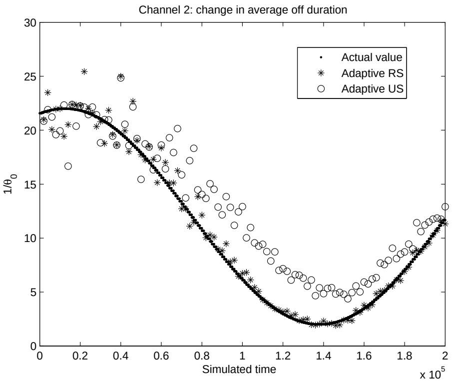
Taking these quantities into (23) and (22), we obtain the expectation of $\hat{\theta}_0$, $\tilde{\theta}_0$, which is a function of $(T_w, \Delta t_p, u, \theta_0)$. Replacing u with $\hat{u}^{(i)}$, θ_0 with $\hat{\theta}_0^{(i)}$, and $\tilde{\theta}_0$ with $\tilde{\theta}_0^{(i+1)}$ we obtain the desired result.

Figure 11 shows the tracking performance of the adaptive random sensing algorithm, where within each moving window the sampling times are randomly place following a uniform distribution. In the simulation the size of the time window is set to be 3500 time units and the error factor ε is set at 1. In Figure 11(a) the channel parameter $E[T_0]$ varies as a step function: starting from 6 time units, it is increased by 5 every 30000 time units, while $E[T_1]$ is set to $E[T_0]/2$. In Figure 11(b) the channel parameter changes more smoothly as shown. The dashed line represents the actual channel parameter. The adaptive uniform sensing scheme follows the exact same procedure as the adaptive random sensing scheme , with the only difference that in the i-th window uniform sensing is used, instead of random sensing, with a constant sampling interval of $\Delta t_p^{(i)}$ as calculated in (21). We see that the estimation under adaptive random sensing (RS) can closely track the time-varying channel, and clearly outperforms adaptive uniform sensing (US) at short on/off periods.

The number of samples taken in each window (or estimation cycle), for the two channel models shown in Figure 11, following this adaptive scheme is given in Figures 12(a) and 12(b), respectively. We see that as the on/off periods vary, the sampling rate is automatically adjusted as an outcome of the tracking. For instance, the decrease in the sample number shown in Figure 12(a) corresponds to the consistent increase in the average off durations, while the large peak in Figure 12(b) corresponds to the valley of small average off durations shown in Figure 11(b).



(a)



(b)

Fig. 11. Estimation performance of time-varying channel

7 Conclusion and Future Work

In this paper we studied sensing schemes for a channel estimation problem under a sparsity condition. Using Fisher information as a performance measure, we derived the best and worst sensing sequences both with and without the sparsity condition. An adaptive random sensing scheme was also proposed to effectively track time-varying channel parameters. Recent advances in compressive sensing theory [27], [28], [29] allow us to represent sparse signals with significantly fewer samples than required by the Nyquist sampling theorem. It is thus tempting to examine whether this technique brings any advantage for our channel estimation problem. The idea is to randomly sample the channel state, use compressive sensing techniques to reconstruct the entire sequence of channel state evolution, and then use the ML estimator to determine the channel parameter. So far we have not found this method to work well. The main obstacle lies in finding a good basis matrix that can both sparsify the channel signal vector and at the same time be sufficiently incoherent with the measurement matrix. A similar difficulty was noted in [30] in trying to use compressive sensing for a data gathering problem. This remains an interesting direction of future work.

Acknowledgment

The authors wish to thank Prof. Raj Rao Nadakuditi for pointing us to the use of circular β ensemble and for helpful discussions, and the anonymous reviewers for constructive comments.

References

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

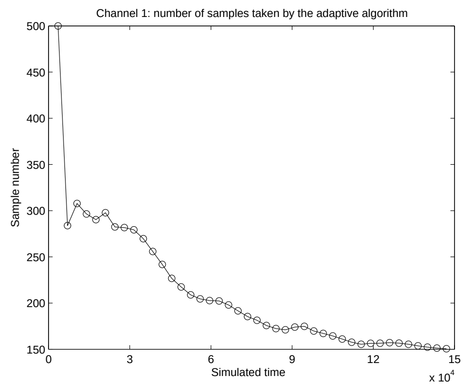
[2] K. Challapali, C. Cordeiro, and D. Birru, "Evolution of Spectrum-Agile Cognitive Radios: First Wireless Internet Standard and Beyond," *Proceedings of ACM International Wireless Internet Conference*, August 2006.

[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.

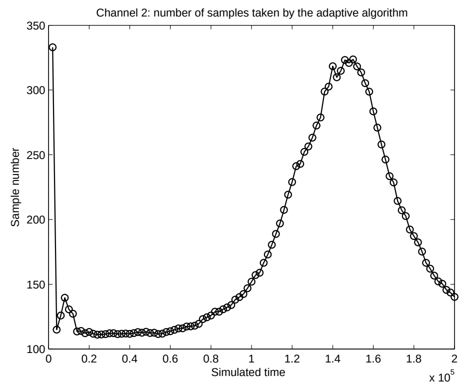
[4] D. Cabric, S. M. Mishra, R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2004.

[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.



(a)



(b)

Fig. 12. Corresponding number of samples in each estimation window

[7] H. Kim and K. G. Shin, "Fast Discovery of Spectrum Opportunities in Cognitive Radio Networks," *Proceedings of the 3rd IEEE Symposia 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. *IEEE Transactions on Mobile Computing*, vol. 8, No. 12, Dec. 2009.

[12] R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Mathematical Foundations of Theoretical Statistics* vol. 222, pp: 309-368, 1922.

[13] D. R. Cox, *Renewal Theory*, Butler and Tanner,1967.

[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] 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.

[18] 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.

[19] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge Series in Statistical and Probabilistic Mathematics) (1998)

[20] M. Wellens, J. Riihijrvi, and P. Mhnen, "Modelling Primary System Activity in Dynamic Spectrum Access Networks by Aggregated ON/OFF-Processes," *Proceedings of the 4th IEEE Workshop on Networking Technologies for SDR Networks*, Jun. 2009.

[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] R. Bellman, S. Dreyfus, *Princeton University Press*, 2010.

[24] M. L. Mehta, *Random matrices*. Elsevier/Academic Press, 2004.

[25] A. Edelman, B.D. Sutton, "The Beta-Jacobi Matrix Model, the CS Decomposition, and Generalized Singular Value Problems," *Foundations of Computational Mathematics*, vol. 8 , no. 2, pp: 259-285, May 2008.

[26] Y. Li, S.-P. Sheng, R. Saigal, M. Liu, D. Chen and Q. Zhang, "A stochastic differential equation model for spectrum utilization," *Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2011, Princeton, NJ.

[27] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp: 4036-4048, 2006.

[28] 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.

[29] 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.

[30] 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.

Appendix A
Proof of Lemma 1

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

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

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

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

1 and $\theta_0 > 0$ we have

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

Therefore for $\frac{\theta_0 \Delta t}{u} > 2 + \sqrt{2}$, $h_o(\Delta t)$ is strictly convex. That $h_1(\Delta t)$ is strictly 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 strictly convex for $\Delta t > \frac{u}{\theta_0} \ln(\frac{u}{1-u})$. This condition is equivalent to $ue^{-\theta_0 \Delta t/u} < 1 - u$. Under this condition and again noting $0 < u < 1$ and $\theta_0 > 0$, we have

$$\begin{aligned} h_2^{'}(\Delta t) &= \frac{-u(1-u)^2 \theta_0 e^{-\theta_0 \Delta t/u}}{[(1-u) + ue^{-\theta_0 \Delta t/u}]^2} < 0, \\ h_2^{''}(\Delta t) &= \frac{(1-u)^2 \theta_0^2 e^{-\theta_0 \Delta t/u}[(1-u) - ue^{-\theta_0 \Delta t/u}]}{[(1-u) + ue^{-\theta_0 \Delta t/u}]^3} \\ &> 0. \end{aligned}$$

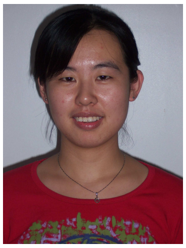
Similarly, $h_3(\Delta t)$ is strictly convex under the condition $\Delta t > \frac{u}{\theta_0} \ln(\frac{1-u}{u})$, since

$$\begin{aligned} h_3^{'}(\Delta t) &= \frac{-u(1-u)^2 \theta e^{-\theta_0 \Delta t/u}}{[u + (1-u)e^{-\theta_0 \Delta t/u}]^2} < 0, \\ h_3^{''}(\Delta t) &= \frac{(1-u)^2 \theta^2 e^{-\theta_0 \Delta t/u}[u - (1-u)e^{-\theta_0 \Delta t/u}]}{[u + (1-u)e^{-\theta_0 \Delta t/u}]^3} \\ &> 0. \end{aligned}$$

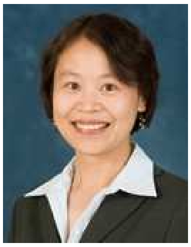
Therefore under the condition $\Delta t > \alpha u/\theta_0$, h_1, h_2 and h_3 are all monotonically decreasing convex functions. It follows that $h = h_1 + h_2 + h_3$ is also monotonically decreasing and convex. Furthermore, for any $\Delta t > 0$, $h_o(\Delta t) > 0$, and $h(\Delta t) > h(+\infty)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

$$\begin{aligned} g^{''}(\Delta t) &= (h_o(\Delta t) h(\Delta t))^{''} \\ &= h_o^{''}(\Delta t) h(\Delta t) + 2h_o^{'}(\Delta t) h^{'}(\Delta t) \\ &\quad + h_o(\Delta t) h^{''}(\Delta t) > 0 \ , \qquad (25) \end{aligned}$$

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. □



Quanquan Liang received her B.S. degree and Ph.D. in communication and information systems in the School of Information Science and Engineering, Shandong University, China, in 2004 and 2010. From 2008 to 2010, she was with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, as a visiting student. Her research interests focus on wireless sensor networks and cognitive radio networks.



Mingyan Liu Mingyan Liu received her Ph.D. Degree in electrical engineering from the University of Maryland, College Park, in 2000. She joined the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, in September 2000, where she is currently an Associate Professor. Her research interests are in optimal resource allocation, performance modeling and analysis, and energy efficient design of wireless, mobile ad hoc, and sensor networks.

Dongfeng Yuan received his M. S. degree from Department of Electrical Engineering, Shandong University, China, 1988, and got his Ph. D degree from Department of Electrical Engineering, Tsinghua University, China in 2000. Currently he is a full professor and dean in School of Information Science and Engineering, Shandong University, China. His research interests are in the area of wireless communications, with a focus on cross-layer design and resource allocations.