Active Spectrum Sensing with Sub-Nyquist Sampling

Lorenzo Ferrari, and Anna Scaglione

Abstract—At high carriers, enabling low power wireless devices to use opportunistically the available spectrum requires an analog front-end that can sweep different bands quickly, since Nyquist sampling is prohibitively expensive. In this paper we propose a new framework that allows to optimize a sub-Nyquist sampling front-end to combine the benefits of optimum sequential sensing with those of compressive spectrum sensing. The sensing strategy we propose is formulated as an optimization problem whose objective of maximizing a utility decreasing linearly with the number of measurement and increasing monotonically with the components found empty. The optimization designs the optimal measurement selecting the best linear combinations of sub-bands to mix to accrue the maximum utility. The structure of the utility represents the trade-off between exploration, exploitation and risk of making an error that is characteristic of the spectrum sensing problem. We characterize the optimal policy under constraints on the sensing matrix and derive the approximation factor of the greedy approach. Second, we present the analog front-end architecture and map the measurement model into the abstract optimization problem proposed and analyzed in the first part of the paper. Numerical simulations corroborate our claims and show the benefits of combining opportunistic spectrum sensing with sub-Nyquist sampling.

I. INTRODUCTION

Mobile devices are increasingly been used for entertainment (gaming, video streaming). The coexistence of these networks with the “Internet of Things” (IoT) and Machine-to-Machine (M2M) communications means that wireless applications may quickly become starved for bandwidth. Increasing the spectrum to millimeter waves can provide the much needed linear increase in throughput, but it poses the challenge of reversing the trend of bringing high speed sampling closer and closer to the antenna. Also, for IoT, a better match would be decentralized cognitive opportunistic spectrum access rather than scheduling in a Cloud Radio Access Network. But that requires somehow overcoming the bottleneck of Nyquist sampling when sensing the large spectrum available. There are two ways this bottleneck can be eliminated. One is to use a sub-Nyquist sampling front-end which, in principle, can scan the entire spectrum at once, reducing significantly the sensor analog to digital conversion hardware complexity and its energy cost, at the expenses of increased complexity in the reconstruction of the underlying signal. Sub-Nyquist sampling is often referred to as Finite Rate of Innovation (FRI) sampling, e.g. [1]–[3] and the so called X-sampling architecture is one of the popular examples of FRI sampling front-ends [5]. We provide two arguments to reconsider the way in which FRI sampling may benefit the cognitive spectrum access problem. First, in the cognitive sensing problem the direct application of FRI sampling, as proposed in e.g. [6]–[8], poses a serious limitation: it does not work unless the spectrum is sufficiently sparse. This inability to adapt to the ever changing spectrum occupancy rate, makes its real application to cognitive spectrum sensing impractical. Furthermore, the procedure of folding the spectrum creates the phenomenon known as noise folding [9], which creates an SNR deterioration approximately linear in the number of mixed bands, and to overcome this low density measurement matrices have been considered [10]. Second, in the spectrum sensing problem the objective is the detection of the idle channels, not the signal reconstruction. This suggests that the FRI hardware complexity may still exceed what is really necessary, but also that when considering the detection problem, most of the standard results in compressive sensing that bound the \( \ell_2 \) norm of the estimation error do not directly express the detection performances. These two observations motivate our idea of looking at a sequential implementation of an FRI detector whose goal is exclusively carrier sensing over bands, not the signal reconstruction. The other approach consists in selecting opportunistic and in a cognitive fashion a small section of the spectrum at a time, relying on an analog front-end able to swiftly switch between small sub-bands. Several authors proposed to explore optimal sequential sensing methods to address this problem by choosing optimally the band to explore, see e.g. [11]–[13]. More recently, other authors have proposed and studied sensing strategies that would reduce the required number of measurements adaptively using compressive sensing, optimizing the sensing actions based on previous observations [14]–[18]. The common goal in these works is the recovery of the full support of a given vector. Typically, the techniques proposed are shown to be able to cope with lower SNR in the signal reconstruction with low complexity. What these optimization do not capture is the fact that in cognitive spectrum sensing applications it is also desirable to have enough time to exploit the spectrum by making a timely decision. A cross-layer framework to jointly optimize spectrum sensing and scheduling is presented in [19], where however the sensing phase duration is pre-determined and the policy optimizes the average sensing traffic, whereas in our model a dynamic optimization on the number of measurement is discussed based on recovery guarantees. In fact, the receiver collects energy measurements sequentially (sampling at a fraction of the Nyquist rate) under the assump-
tion that the activity of the Primary Users (PUs) in a certain spectrum will persist for several sampling periods (which is a reasonable assumption in this context). The novelty in our architecture stems from the idea of selecting opportunistically the signal used for aliasing the wide-spectrum input, so as to mix selective portions of the spectrum differently each time we perform an energy measurement.

A. Contributions and Outline

We cast the spectrum sensing problem as a sequential sensing procedure where the cognitive receiver (CR) selects a group of sub-bands to sense at each test which do not necessarily need to be contiguous. We consider a scenario where the CR is able to get observations in sequence, each by aliasing a wide-spectrum signal so as to mix selective portions of the spectrum each time differently and then sampling the output of a narrow-band filter. By considering a sequential acquisition of measurements we can use a single non-coherent receiver to further simplify the architecture. We model the trade-off between the benefit in accuracy yielded by more measurements versus the reduction in time available to exploit the knowledge acquired, through a time-dependent utility function. Our main contribution is a new dynamic optimization framework for the design of an adaptive sub-Nyquist spectrum sensing frontend. Our method simultaneously captures the sensing time and the selection of the sub-bands leading to a greedy algorithm for design of the sensing matrix, whose approximation factor relative to the optimum is fully characterized. Note that scanning one sub-band at a time is a possible action of our active spectrum sensing strategy, which means that our method subsumes previous techniques to scan the spectrum optimally without mixing it and is adaptive with respect to the time horizon $K$, the number of resources $N$, the prior probabilities on the states of each resources, and the parameters that characterize the utility function (i.e. reward and penalty for good/bad decisions). In addition, our numerical results clearly illustrate how sparse sensing matrix designs are far superior to dense sensing matrices for two reasons: 1) through belief propagation algorithms, they achieve near-optimum detection performance; 2) they mitigate the noise folding phenomena. We believe this work highlights the importance of being dynamic in the design of the sensing phase and that static approaches limit the applicability of FRI architectures for heterogeneous and non-stationary environments.

We also emphasize in the paper that our model is applicable not only when the utility comes from finding empty entries (e.g. spectrum sensing) but also when one is interested in finding the occupied ones (e.g. in a RADAR application). The paper is organized as follows: Section II is dedicated to the model description, in Section III we study the optimal policy for the Direct Inspection (DI) case (also known as scanning receiver), where there is no mixing of sub-bands, in Section IV we discuss the possibility of mixing introduced different bands, following an approach inspired by Group-Testing (GT) and show that, despite the fact that finding the optimal policy is exponentially complex in the number of resources, it is possible to characterize the approximation factor a greedy procedure can achieve relative to the optimum policy. Numerical results to sustain our claims are presented in Section VII.

B. Notation

We use bold lower-case to represent vectors, bold upper-case for matrices and calligraphic letters to indicate sets. For vectors, we use the same letter upper-case to indicate the $|\cdot|$ norm (i.e. $\Omega = |\omega|^2$) and with the notation $s, A$ we select the entries $i \in A$ of vector $s$. With $|y|^2_A$ we indicate the weighted $\ell_2$ norm $y^T A y$. For any set function $f(A)$ we define the marginal increment for adding element $a$, as $\partial_a f(A) = f(A+a) - f(A)$. The submodularity property for a set function $f(A)$ that will be used in our proofs is

$$f(A) + f(A + b) \geq f(A + a) + f(A).$$

II. Model Description

We consider the situation of an agent that needs to divide $K$ instants of time available between the sensing and the exploitation of a set $N = \{1, 2, \ldots, N\}$ of items/sub-bands. The agent accrues a reward that is a function of the underlying state vector for these items:

$$s \triangleq [s_1, \ldots, s_N] \in \{0, 1\}^N$$

where the entries $s_i \in \{0, 1\}$, as well as the residual time available after sensing. We consider the $s_i$ as indicators of good/bad (0/1) state of a resource; for the cognitive spectrum sensing application 0 would mean the channel is “idle” and 1 would mean the channel is “busy”. The player acquires information about the entries via random observations coming from a known pdf parameterized by an unknown vector. During the times devoted to sensing $k = 1, 2, \ldots, K < K$ the player has the possibility to dynamically and adaptively design each measurement by selecting a subset of entries of $s$ to probe through a sensing vector $b_k = [b_{k1}, b_{k2}, \ldots, b_{kN}]$ which will be non-zero only on the channels that are actively sensed; we assume the measurement provides an observation $y[k]$ drawn from a density $f_{\theta(k)}(y)$ that is a function of both the choice of $b_k$ and the state $s$, i.e. $\theta(k) = \theta(b_k, s)$. We assume the state entries $s_i$ are mutually independent Bernoulli random variables with known prior probabilities given by a vector $\omega = [\omega_1, \omega_2, \ldots, \omega_N]$, where $\omega_i = P(s_i = 0)$.

The decision maker needs to design:

1) the $K \times N$ measurement matrix $B$ whose rows are the measurement vectors $b_k$ for $k = 1, 2, \ldots, K$ for each test and $K$ indicates the sensing (exploration) time to acquire information on the states $s_i$ via the observations $y[k]$.

2) a set of $N$ decision rules $\delta = \{\delta_i \in \{0, 1\} : i = 1, 2, \ldots, N\}$ over the unknown states $s_i$ of the resources at the end of the exploration phase.

The total utility for the player will be proportional to the time left for exploitation $(K-K)$. Let us consider a reward $r_i > 0$ for correctly detecting an empty/busy resource and a penalty
\( \rho_i < 0 \) for failing to detect a busy/empty resource, the utility can be written as

\[
U(s, \mathcal{N}, K, \mathbf{B}, \delta) \triangleq \begin{cases} 
(K - K_i) \sum_{i=1}^N \omega_i r_i (1 - \alpha_i) + (1 - \omega_i) \rho_i \beta_i & \text{case 0} \\
(K - K) \sum_{i=1}^N (1 - \omega_i)(1 - \beta_i) r_i + \omega_i \alpha_i \rho_i & \text{case 1}
\end{cases}
\]

where \( \alpha_i, \beta_i \) denote the type I and type II errors probability respectively, i.e. \( \alpha_i = P(\delta_i = 1 | s_i = 0) \) and \( \beta_i = P(\delta_i = 0 | s_i = 1) \). To differentiate between case 0/1 allows to consider applications where the utility comes from an action on the entries detected as empty/busy: i.e., in a spectrum sensing application, the utility would come from the decision on transmitting over frequency bands found empty, whereas for a RADAR application, it makes more sense to consider the utility comes from taking action on the frequency (spatial directions) found busy. Finding the optimal policy corresponds to solve the following optimization problem

\[
\text{maximize} \quad \mathbb{E}[U(s, \mathcal{N}, K, \mathbf{B}, \delta)]
\]

The theory developed in this paper could be easily modified to fit different observation models and assumptions, but to ease the presentation we will limit ourselves to consider the following form for the function \( \theta \):

\[
\theta[k] = \theta(b_k, s) = b_k (\varphi^T + w^T)
\]

where \( \varphi = [\varphi_1, \varphi_2, \ldots, \varphi_N] \) is a non-negative vector, such that the state variable \( s_i = 1 \) when \( \varphi_i > 0 \) and 0 when \( \varphi_i = 0 \), and \( w = [w_1, w_2, \ldots, w_N] \), in the context of RF spectrum sensing, represents the background zero mean, circularly symmetric, Additive White Gaussian Noise (AWGN) variances on each frequency band. The theory developed in this paper concerns the detection of the non-negative entries of \( \varphi \) and consequently the maximization of the utility accruable from the resources declared to be in the empty/busy state \( \mathcal{N} \). The observation model also assumes that

\[
y[k] \sim f_{\theta[k]}(y) = \frac{1}{\theta[k]} e^{-\frac{y}{\theta[k]}}
\]

where, for convenience, we use the alternative parameterization for the exponential distribution. This choice of distribution corresponds to the case of complex circularly symmetric signal samples in each sub-band with zero mean and variance \( \varphi_i \) embedded in AWGN with variance \( w_i \).

III. DIRECT INSPECTION (DI) CASE

In the Direct Inspection (DI) case, we limit \( b_k \) to have only one non-zero entry \( i \), i.e. \( b_{ki} \neq 0, b_{kj} = 0 \forall j \neq i \). This means that there is an underlying hypothesis testing:

\[
\mathcal{H}_0 : y[k] \sim \text{Exp}(\theta_0[k]) \]
\[
\mathcal{H}_1 : y[k] \sim \text{Exp}(\theta[k])
\]

with \( \theta_0[k] = b_{ki} w_i \) and \( \theta[k] = b_{ki} (\varphi_i + w_i) > \theta_0[k] \) and the quantity the test should maximize is

\[
\begin{cases}
\omega_i (1 - \alpha_i) r_i + (1 - \omega_i) \beta_i \rho_i & \text{case 0} \\
(1 - \omega_i)(1 - \beta_i) r_i + \omega_i \alpha_i \rho_i & \text{case 1}
\end{cases}
\]

in \([5]\), which is equivalent to minimize the Bayesian risk

\[
\begin{cases}
\omega_i \alpha_i r_i + (1 - \omega_i) \beta_i \rho_i & \text{case 0} \\
\omega_i \alpha_i \rho_i + (1 - \omega_i) \beta_i r_i & \text{case 1}
\end{cases}
\]

therefore the optimal detector is the one that chooses \( \mathcal{H}_1 \) if

\[
\frac{f_{\theta_0}[y[k]]}{f_{\theta_1}[y[k]]} \underbrace{\gamma_i}_{\mathcal{H}_0} \triangleq \frac{\frac{r_i}{(1 - \omega_i)\beta_i}}{\frac{r_i}{(1 - \omega_i)\alpha_i}} \quad \text{case 0}
\]

In a typical scenario we could know \( w_i \) (i.e. the background noise variance over a certain bandwidth) but have only a statistical characterization of \( \varphi_i \) (i.e. the average received power from an existing transmission over the i-th frequency band) so one could use a Bayesian approach to integrate out the unknown \( \varphi_i \) and write the test as

\[
\frac{f_{\theta_0}[y[k]]}{f_{\theta_1}[y[k]]} \underbrace{\gamma_i}_{\mathcal{H}_0} \quad \text{case 1}
\]

or, in absence of prior knowledge, use a GLRT and replace the MLE for \( \varphi > 0 \) under the true hypothesis, i.e.

\[
\frac{\sum_{\varphi_i > 0} f_{\theta_0}[y[k]]}{\sum_{\varphi_i \leq 0} f_{\theta_1}[y[k]]} \underbrace{\gamma_i}_{\mathcal{H}_0}
\]

Assumption 1: \( \forall i \in \mathcal{N}, \omega_i < \frac{\rho_i}{\rho_i + r_i} \) (case 0) / \( \omega_i > \frac{\rho_i}{\rho_i + r_i} \) (case 1).

From Assumption 1 one can derive the optimal decision rule for the resources that have not been sensed, by simply checking that for \( \delta_i = 1 \) (case 0) / \( \delta_i = 0 \) (case 1) one would have \( U_j(i, \omega_i) = 0 \), whereas \( \delta_i = 0 / \delta_i = 1 \) would give a negative expected utility. This means that no resource can be declared empty / busy a priori, i.e. no utility can be accrued from resources that have not been sensed.

For the remainder of this work, we will consider the Generalized Likelihood Ration Test (GLRT) approach. It is clear that the optimality of \( \mathcal{H}_1 \) for the threshold \( \gamma_i \) completely characterizes the set of decision rules \( \delta \) for the sensed resources, while Assumption 1 gives us the decision rule for the non-sensed resources. This implies that for the DI case, the optimization in \([4]\) can be expressed solely in terms of \( B \).

Notice however that if \( \omega_i \leq \frac{\rho_i}{\rho_i + r_i} \), then from \([9]\), we have \( \gamma_i \leq 1 \), and for value of \( \gamma_i < 1 \) even when the resource is sensed, the test in \([11]\) would always choose the alternative hypothesis. To overcome this, we constrain the alternative hypothesis \( \varphi_i \geq \varphi_{\min} > 0 \quad \forall i \in \mathcal{N} \), assuming that if a transmission exists over that frequency band, there is a minimum value \( \varphi_{\min} \) strictly greater than 0, that is received on average, and rewrite our test as

\[
\max_{\varphi_i \geq \varphi_{\min}} \frac{f_{\theta_0}(\varphi_i + w_i)(y[k])}{f_{\theta_1}(w_i)(y[k])} \underbrace{\gamma_i}_{\mathcal{H}_0}
\]

1The threshold in \([9]\) is the optimal threshold that minimizes \([8]\) for the case of binary hypothesis testing, where \( \varphi_i \) is known. Notice that no theoretical guarantees are given for the GLRT case for the single sample, but it is of common practice to replace the MLE estimate for the unknown parameter and then reduce to the binary case, using the same threshold. A Local Most Powerful test exists for \( \theta \to \theta_0 \) but the GLRT is preferred since we want to consider high SNR range.
We then derive the error probabilities $\alpha_i$ and $\beta_i$ for the test in (12) and add the superscript $\text{DI}$ since this test is the one used in the DI strategy.

$$
\alpha_i^{\text{DI}} = \min \left\{ \left( \frac{\rho_i(1 - \omega_i)}{\rho_i(1 + \frac{\omega_i}{\beta_i})} \right)^{1 + \frac{\omega_i}{\beta_i}}, 1 \right\}
$$

$$
\beta_i^{\text{DI}} = 1 - \left( \alpha_i^{\text{DI}} \right)^{\frac{1}{\beta_i^{\text{max}}}}
$$

where as expected the false alarm probability is independent from the alternative hypothesis, whereas the detection improves with the true average transmitted power $\varphi_i$. What we can guarantee, since $\varphi_i \geq \varphi_{\text{min}}$ is that

$$
\beta_i^{\text{DI}} \leq 1 - \left( \frac{\rho_i(1 - \omega_i)}{\rho_i(1 + \frac{\omega_i}{\beta_i})} \right)^{\frac{\omega_i}{\beta_i^{\text{max}}}} = \beta_i^{\text{max}}
$$

Remark 1: The performances of the GLRT for the DI case do not depend on $b_k$, therefore, for the DI case no further optimization is needed over the sensing matrix $B$, other than selecting the non-zero entries.

The impossibility of knowing the true $\beta_i$ of each test motivates us to optimize the minimum guaranteed achievable utility, using the bound in (15). This consideration together with Assumption 1 leads to rewrite the optimization problem in (4) for the DI case as

$$
\max_{A \subseteq N} U^{\text{DI}}(A)
$$

where

$$
U^{\text{DI}}(A) \triangleq (K - |A|) \sum_{i \in A} u_i^{\text{DI}}
$$

$$
u_i^{\text{DI}} \triangleq \omega_i r_i (1 - \alpha_i^{\text{DI}}) - (1 - \omega_i) \rho_i \beta_i^{\text{max}}
$$

We then introduce the following Lemma

**Lemma 1:** $U^{\text{DI}}(A)$ is a normalized, non-monotone, non-negative sub-modular function of $A$.

**Proof** See Appendix A.

Lemma 1 implies that there are diminishing returns in augmenting sets by adding a certain action to bigger and bigger sets. The maximization of a non-monotonic sub-modular function is generally NP-hard, but the case of interest is not as difficult. In fact, it is clear that by sorting the resources $i$ so that

$$
u_i^{\text{DI}} \geq \nu_2^{\text{DI}} \geq \ldots \geq \nu_N^{\text{DI}}
$$

then the set of size $i$, $A_i = \{1, \ldots, i\}$ will be such that for any set $X$ of size $|X| = i$

$$
\sum_{k=1}^i u_k^{\text{DI}} \geq \sum_{k \in X} u_k^{\text{DI}}
$$

Therefore, what remains is to find the best set size $i$ such that

$$
U^{\text{DI}}(A) \leq U^{\text{DI}}(A_i) \leq \max_i \left( (K - i) \sum_{k=1}^i u_k^{\text{DI}} \right)
$$

The maximum in (20) is attained for

$$
i^* = \inf_i \{ i : \partial_i U^{\text{DI}}(A_i) < 0 \}
$$

where $\partial_i U^{\text{DI}}(A_i) = (K - i) u_i^{\text{DI}} - \sum_{k=i+1}^{K} u_k^{\text{DI}}$. In fact, given the function is sub-modular as soon as this condition is attained it is maintained for $i + 2$, $i + 3$ etc. given that the marginal returns continue to decrease. This maximization is greedy and stops when the marginal reward becomes negative.

IV. GROUP TESTING INSPIRED STRATEGY

We now allow the test to mix different sub-bands, i.e. the vector $b_k$ to have more than one non-zero entry. The main idea of this section is to develop a dynamic simple strategy that can be characterized in closed form and gives us a sufficient condition to claim an MT strategy would outperform the DI alternative. From the sensing matrix $B$, let us define the sets $A_k = \{ i \in N : b_{ki} \neq 0 \}$ and $B_i = \{ 1 \leq k \leq K : b_{ki} \neq 0 \}$, and let us use the convention that whenever $b_k$ is the argument of a function, then the set $A_k$ is also the argument of that function and that when $B$ is the argument of a function, then all the sets $A_k$ for $0 \leq k \leq K - 1$ and all the sets $B_i$ for $i \in N$ are arguments of that function as well. As outlined in the Introduction, aliasing of the spectrum creates with an associated noise folding phenomenon. Its impact is particular severe in a non coherent scheme as ours, where it is appropriate to consider the signal samples as complex circularly symmetric gaussian with zero mean given that the samples are collected sequentially and not in parallel (see Section for a detailed discussion on the signal model). To mitigate the noise folding effects it is necessary to use low density measurement matrices. A common approach for recovery with low density measurement matrices is to use belief propagation via message passing whose most well known application is the decoding of LDPC error correction codes. In designing LDPC matrices it is preferable to avoid having short cycles, which result to be problematic for the convergence of the message passing algorithm: it has in fact shown numerically that, even if no theoretical guarantee is given for belief propagation in loopy networks, a very good accuracy is obtained in absence of short cycles in the graph. Nevertheless, being interested in the detection problem and not necessarily the recovery of the signal, rather than a design with asymptotical guarantees or bounds on the $\ell_2$-norm, it would be preferable to have a tool that allows us to quantify the objective in (4) for a given matrix and this problem. When looking at detection performances, the sensing problem falls more naturally in the group testing framework rather than compressive sensing, and in the context of group testing little is known in presence of measurement errors that depend on the group size, which is the scenario this work considers, as the remainder of the paper will detail. Asymptotic results on the targeting rate for measurement-dependent noise, using an information-theoretic approach, are given in [20], where the noise is modeled as independent additive Bernoulli with bias dependent on the

\[\text{in our model, an uninformative prior can be assigned to the } \varphi_i \text{'s to run the belief propagation message-passing algorithm on the obtained measurements}\]
test size, giving therefore the same false-alarm and missed-detection probabilities, relative to the single test. An additional noise, called \textit{dilation effect} was considered in \cite{21}, where each resource could independently flip from 1 to 0 before the grouped test, and information-theoretic bounds were provided. In our model the false alarm and missed-detection probabilities are dependent on the optimization of the threshold for the GLRT, therefore the noise is not independently added (nor an independent dilution can be considered) and the strategy derived depends on the finite horizon for $K$, i.e. our results are not asymptotic. The same considerations apply to similar information-theoretic approaches in \cite{22-24}.

Let us start by considering a matrix $\mathbf{B}$ without length-4 cycles, i.e., two different measurements do not mix more than one sub-band in common. For each test we define a binary \textit{group testing-like} hypothesis testing where

\[
\begin{align*}
\mathcal{H}_0 : & \forall i \in A_k \quad s_i = 0 \\
\Rightarrow & \theta_0[k] = b_k \mathbf{w}^T \\
\mathcal{H}_1 : & \exists i \in A_k \text{ s.t. } s_i = 1 \\
& \Rightarrow \theta[k] \geq (\min b_{ki}) \varphi_{\min} + b_k \mathbf{w}^T = \theta_{\min}[k]
\end{align*}
\]  

(22)

The test can then be written

\[
\max_{\theta(k) \geq \theta_{\min}[k]} \frac{f_{\theta(k)}(y[k])}{f_{\theta_0(k)}(y[k])} \overset{\mathcal{H}_1}{\geq} \alpha_{\min}(\gamma_k)
\]

(23)

for which we can derive

\[
\alpha(b_k, \gamma_k) = \left(1 - \frac{1}{\gamma_k (1 + \frac{\theta_{\min}}{\theta_1})} \right)^{1 + \frac{\theta_{\min}}{\theta_1}}
\]

(24)

\[
\beta(b_k, \gamma_k) = 1 - (\alpha(b_k, \gamma_k))^{1 + \frac{\theta_{\min}}{\theta_1}}
\]

(25)

The decision declares that resource $i$ is busy ($\mathcal{H}_1$ is true) if the majority of the tests where resource $i$ is involved is positive, else it accepts the null hypothesis $\mathcal{H}_0$ for resource $i$. This would give us:

\[
\pi_0(i, \mathbf{b}, \gamma) \triangleq \left(1 - \prod_{j \in A_k \setminus i} \omega_j \right) (1 - \beta_i(b, \gamma; 0)) + \alpha(b, \gamma) \prod_{j \in A_k \setminus i} \omega_j
\]

(26)

\[
\pi_1(i, \mathbf{b}, \gamma) \triangleq 1 - \beta_i(b, \gamma; 1)
\]

(27)

where the functions $\pi_0(i, \mathbf{b}, \gamma)$ and $\pi_1(i, \mathbf{b}, \gamma)$ are only defined when $b_i \neq 0$. Notice that the error probabilities $\alpha$, $\beta$ refer to each \textit{group hypothesis testing} defined in (22). The notation for $\beta_i(\mathbf{b}, \gamma; s_i)$ indicates the probability of having a missed-detection conditioned on the state $s_i$ of one of the resources. It then follows that

\[
\alpha_i^{MT}(\mathbf{B}, \gamma) \triangleq 1 - F_{PBD} \left( \left\lfloor \frac{|B_i|}{2} \right\rfloor - 1; |B_i|, \{ \pi_0(i, \mathbf{b}_k, \gamma_k) : k \in B_i \} \right)
\]

(28)

\[
\beta_i^{MT}(\mathbf{B}, \gamma) \triangleq F_{PBD} \left( \left\lfloor \frac{|B_i|}{2} \right\rfloor - 1; |B_i|, \{ \pi_1(i, \mathbf{b}_k, \gamma_k) : k \in B_i \} \right)
\]

(29)

where $F_{PBD}(k; n, p)$ indicates the CDF of a Poisson Binomial Distribution parameterized by $p \in [0,1]^n$. One can then replace (28)-(29) in (3) to then solve the optimization in (4), where the equivalence between the decision rules $\delta$ and the selection of the thresholds $\gamma$ is essentially the same as for the DI case. Notice that the condition of no length-4 cycles for $\mathbf{B}$ allows to write (28)-(29), i.e. to consider each of the $B_i$ tests independent, conditioned on the state of the resource $i$. The optimization remains extremely complex due to the complexity of the decision space for $\mathbf{B}$ and the sum of an exponentially growing number of terms for the probabilities defined in (28)-(29). Nevertheless, it gives a method to evaluate the objective of our optimization for any sensing matrix $\mathbf{B}$, where the optimization over $\gamma$ can be numerically solved, since (28)-(29) are monotonic functions of the probabilities $\pi_0, \pi_1$ defined in (26)-(27), which are monotonic in the $\gamma_k$’s, and therefore we know that a unique solution for $\gamma$ exists.

Next, we introduce additional constraints to (4), in particular on the structure of $\mathbf{B}$, in order to evaluate whether an MT strategy could be superior to the DI approach.

Note that an ML or a MAP estimator, for a rank-deficient sensing matrix, do not provide optimality guarantees in terms of minimum error probability or minimum Bayesian risk. Nevertheless, for the same sensing matrix, we expect the MAP estimator to outperform the binary \textit{group-testing} hypothesis in \cite{22} by simply adding more degrees of freedom to the decision $\delta$ in the $K$-th dimensional space of the observations. Therefore the evaluation of the objective in (4) via (28)-(29) provides a benchmark for the utility obtainable with a more refined detection method.

\subsection{A. The pairwise tests case}

We start by considering matrices $\mathbf{B}$ that have the following property: each resource is sensed only one time, either directly inspected or mixed with another resource, and no test mixes more than 2 resources, i.e. $|\phi_k| \leq 2$; $|B_i| \leq 1$ \forall $k = 1, \ldots, K$, \quad $i = 1, \ldots, N$. Let us discuss the test that mixes entries $i$ and $j$. According to the strategy derived at the beginning of the section one can use (26)-(27) to write out the per-time instant utility obtainable after the decision. First, from (28), without prior knowledge over $\varphi_i, \varphi_j$ other than the threshold $\varphi_{\min}$, the best choice to minimize $\alpha$ is to set $b_i = b_j$ (we refer to this false alarm probability as $\alpha_{ij}$). After that, a missed detection in (22) can occur for 3 different states of the resources $i, j$ but we upper-bound these by always considering $\theta = \theta_{\min}$ (we refer to this missed detection probability as $\beta_{ij, \max}$). We then obtain

\[
u_{ij}^{MT} \triangleq \omega_i \omega_j (r_i + r_j) (1 - \alpha_{ij}^{MT}) + [(\omega_i (1 - \omega_j) (r_i + r_j) + \omega_j (1 - \omega_i) (r_i + r_j)] \beta_{ij, \max}^{MT}
\]

(30)

where the threshold for this test $\gamma_{ij}$ has been set to maximize (30), i.e.

\[
\gamma_{ij} = \frac{\omega_i \omega_j (r_i + r_j)}{(1 - \omega_i) (|\rho_i| - \omega_j r_j) + (1 - \omega_j) (|\rho_j| - \omega_i r_i)}
\]

(31)

Let us then consider a graph where each resource is a vertex and the edge weight $u_{ij}$ between two vertices $ij$ is the utility
(per time instant) $u_{ii}^{MT}$ just defined (the weight of the loops $u_{ii}$ are given by $u_{ii}^{DI}$ in (18)). We can then translate our problem into a particular instance of a max-cut problem: picking a subset of the edges and form a subgraph, where each edge represents a test, to maximize the objective in (34). Formally, we can write

$$\max_{E} U^{MT}(E)$$

subject to $\deg_E(i) \leq 1 \ \forall i \in \mathcal{V}$

where

$$U^{MT}(E) \triangleq (K - |E|) \left( \sum_{i,j \in E} u_{ij} \right)$$

and $\deg_E(i)$ is the nodal degree of node $i$ induced by the undirected graph $G = (\mathcal{V}, \mathcal{E})$. Notice that $u_{ij} = u_{ij}^{MT}$ when $i \neq j$ and $u_{ii}^{MT} = u_{ii}^{DI}$ when $i = j$. It is possible to map the constraint on the nodal degree in the objective of (32) by adding a penalty for the violation of such constraint, which guarantees the optimal solution will be equivalent to (32), i.e. no set of edges that does not respect the constraint can improve the objective, and any feasible set of edges would have the same objective in the two problems, i.e.: 

$$\max_{E} U^{MT}(E) - M \sum_{i \in \mathcal{V}} \Upsilon(\deg_E(i))$$

where

$$\Upsilon(n) \triangleq \begin{cases} 0 & \text{for } n \leq 1 \\ n - 1 & \text{for } n \geq 2 \end{cases}$$

and $M$ is a positive constant.

**Lemma 2:** For $M > 0$ the objective in (34) is a non-monotone sub-modular function of $E$ and it is possible to find $M^* > 0$ such that for any $M > M^*$ the two optimizations (32) and (34) are equivalent.

**Proof** See Appendix [B].

We know discuss the extension of our approach for $\phi > 2$.

**B. Extension to $\phi > 2$**

The approach in the previous section can be extended to tests that mix more than 2 channels. However, instead of just edges or self loops to indicate the tests, we could have cycles of length up to $\phi$. The nodal degree in (34) will then be interpreted as the number of cycles a node is in and the set of edges will be replaced with the set of cycles, extending the validity of Lemma 2 to this case. We then replace the set $\mathcal{E}$ of edges with the set $\mathcal{C}$ of possible cycles, and use $c$ to indicate the generic cycle (which could be a self-loop, an edge or a cycle with length 3 or greater). In light of the constraint $|B_i| \leq 1$ we will have that no node can be in two cycles. In Fig. 1 we show two possible set of cycles of length up to 4. On the right we have a set of tests that respect our constraint: there is a test that only considers one resource and three tests that combine 2, 3, and 4 resources respectively, but no resource is considered in two different tests. On the left, instead, a resource is considered in two tests: one where is combined with other 3 resources and one where is inspected directly, and such configuration is therefore not acceptable.

**Algorithm 1: Greedy Maximization of $U^{MT}(C)$**

1. **Initialize**: $C = \emptyset$.
2. **While** $\exists \ c \in \mathcal{C}^c$ such that $\partial_c U^{MT}(C) > 0$
3. \textbf{Find} $e^* = \arg \max_{c \in \mathcal{C}^c} \partial_c U^{MT}(C)$
4. $C \leftarrow C \cup e^*$
5. **End**

In the greedy procedure in Algorithm 1 there is a constant number of operations per query, which indicates the overall complexity of the algorithm is dominated by the sorting of the cycles utilities. The set $\mathcal{C}^c$ indicates the set of cycles that are not adjacent (share a node) with any of the cycle in $C$. Since in the worst case, sorting $n$ values require $O(n^2)$ operations, the complexity will be given by the total number of possible tests $O \left( \left( \sum_{c \in \mathcal{C}} (N_c)^2 \right)^2 \right) = O \left( N^{2\phi} \right)$, i.e. polynomial in $N$ and exponential in $\phi$.

**C. The factor approximation of the greedy algorithm**

Having proven the sub-modularity of (34) in Lemma 2 it is natural to resort to a greedy procedure, however it is important to highlight that the objective in (34) does not respect the non-negativity property and to the best of our knowledge, there is no known procedure in the literature on approaching the maximization of a general sub-modular non monotone function, if the minimum value is not known, and therefore no constant approximation factor guarantee can be given. Nevertheless, due to the particular structure of our problem it is possible to find a factor approximation for the output of the greedy algorithm.

**Lemma 3:** Algorithm 1 guarantees a $\alpha$-constant factor approximation of the optimal solution for (34), where $\alpha = \frac{1}{\min \{ \phi, \frac{2}{K} \} K - \min \{ \phi, \frac{2}{K} \} + 1}$. 

**Proof** See Appendix [C].

Note that

$$\partial_c U^{MT}(C) = - \sum_{u \in C} u_c + (K - |C| - 1) u_c$$

so, as long as the number of tests $|C|$ added in the greedy maximization is less than the time horizon $K$ we have

$$\arg \max_{C \in \mathcal{C}^c} \partial_c U^{MT}(C) = \arg \max_{c \in \mathcal{C}^c} u_c$$

which in the greedy procedure, edges are added in decreasing order of utility, respecting the constraint on the nodal degree in light of the equivalence between (32) and (34), shown in

![Fig. 1: Example of two sets of tests.](attachment:image.png)
Appendix [B] Notice, also, that from [16] it is easy to find that the optimal $|C|$ will never be greater than $\left\lfloor \frac{K-1}{2} \right\rfloor$.

V. APPROXIMATE ML ESTIMATE FOR MIXED TESTS

In the previous sections we have provided methods that find a low density measurement matrix. As will be apparent in our numerical results the noise folding phenomenon justifies the use of sparse sensing matrices and these kind of matrices are ideal when one wants to use belief propagation to the decision problem. However, for the sake of comparison here we propose a possible alternative approach, which can be applied to any measurement matrix $B$. Let us in fact consider $K$ measurements have been collected that involve the mixing of a set $A \subseteq \mathcal{N}$ of resources. One could ignore the prior $\omega_k$ and derive the ML estimate for $\gamma$. First, let us write the log-likelihood

$$
\log (f(y|\varphi_A)) = -\sum_{k=1}^{K} \log \theta[k] + \frac{y[k]}{\theta[k]}
$$

$$
\theta[k] \approx \sum_{k=1}^{K} 1 + \log y[k] + \frac{1}{2} \left( \frac{y[k] - \theta[k]}{y[k]} \right)^2
$$

(38)

where the linearization corresponds to the Taylor expansion of the likelihood for the observations around their mean (recall (5)-(6)). One could aim at solving the following weighted $\ell_1$ minimization in a LASSO fashion

$$
\varphi_A = \arg \min_{\varphi_A} ||\Gamma_A \varphi_A^T||_1 + \frac{1}{2} ||y - B(\varphi_A^T + w_A^T)||^2_{\mathbb{C}^{1\times 1}}
$$

(39)

with $\Gamma = \{\gamma_i \text{ from (9)} : i \in \mathcal{N}\}$ and $C = \text{diag}(y)$ being the covariance of the observations. The first penalty term in the objective enhances sparsity, favoring the entries with lower threshold $\gamma_i > 0$ whereas the second term in the objective is the objective in the ML estimate in (38). Note that compared to the non-sequential sampling models (i.e. those using a filterbank) the application of the LASSO is an approximation since we do not have an observation that is a noisy linear combination of a sparse input, but rather the samples have a p.d.f that depends on those linear combinations.

VI. APPLICATION TO SEQUENTIAL SPECTRUM SENSING

In the context of spectrum sensing for cognitive radio, since data transmission includes large amounts of control overhead in addition to the data payload, it is natural to assume that the activity of the Primary Users (PUs) in a certain spectrum will persist for several sampling periods (see Fig.2). However, assuming this interval lasts $T = KT_s$, the sensing mechanisms should be providing the fastest decision it can.

A. Analog Front-End Sampling

The aliasing sequence, folds the spectrum present in specific sub-bands onto the center frequency of the receiver during what we can refer to as a sub-Nyquist carrier sensing phase. The samples are spaced by intervals of duration $T_s = 1/R_s$ which is a factor $1/N$ smaller than the total spectrum and, as the diagram in in Fig. 3 shows, rather than having a filter bank architecture as in [14], our approach is to do sequential non-coherent tests, according to the strategy, i.e. the sensing matrix $B$ designed via our greedy algorithm, to maximize the utility accruable from utilizing the sub-bands (channels) detected as free. We assume that the complex envelope of the analog signal we are exploring is a multicomponent signal, whose components are a frequency band width equal to $W = NR_s$ and, hence, during the interval $0 \leq t < T$ the received signal is:

$$
y(t) = x(t) + w(t)
$$

(40)

$$
x(t) = \sum_{i=1}^{K} s[i|m] x_i(t) e^{-j2\pi R_s (i-1)t}
$$

(41)

where we already introduced the state variables $s[i|m]$ previously, $w(t) \sim N(0, N_0 \delta(\tau))$ is Additive White Gaussian Noise, and $x_i(t)$ are the transmitter signals modeled as band-limited random processes with bandwidth $R_s$, i.e. it is equal with probability one to:

$$
x_i(t) = \sum_{k=1}^{K} x_i[k] \text{sinc}(\pi(R_s t - k + 1)).
$$

(42)

The sequential receiver we propose first modulates the received signal at each antenna over the period $(k-1)T_s \leq t < kT_s$ with:

$$
\beta_k(t) = \sum_{i=1}^{N} \sqrt{b_{ki}} e^{j(2\pi R_s (i-1)t + \phi_i)},
$$

(43)

where $b_{ki}$ are the coefficients of the sensing matrix $B$ previously introduced and the phase $\phi_i$ accounts for the delay in...
generating the tone at the $i$-th frequency plus the oscillator phase, i.e. our receiver can be implemented by combining different oscillators that do not require to be synchronized or phase-locked (a benefit of our incoherent sensing model). Then, after convolving the modulated signal with an ideal low-pass filter with impulse response $sinc(\pi R_s t)$, it samples the output $c(t)$ at times $k T_s$, $k = 1, \ldots, K$. This operation is equivalent\(^\text{1}\) to an orthogonal projection, as shown below:

$$
c[k] = [y(t)\beta_k(t)] \ast R_s sinc(\pi R_s t)|_{t=k T_s} = \sum_{i=1}^{N} \sqrt{b_{ki}} e^{j\phi_i} Y_{ki}
$$

(44)

where $Y_{ki}$ represent the orthogonal projections over the period $(k-1)T_s \leq t < kT_s$ of $y(t)$ over the following signals:

$$
Y_{ki} = y(t), R_s e^{j2\pi R_s (i-1) t} sinc(\pi (R_s t - k + 1)) > (45)
$$

Considering that the signals

$$
\left\{e^{j2\pi R_s (i-1) t} sinc(\pi (R_s t - k + 1))\right\}_{i,k \in \mathbb{Z}}
$$

form a orthogonal basis, and that (40) is equivalent to:

$$
x(t) = \sum_{k=1}^{K} \sum_{i=1}^{N} s_i x_i[k] e^{j2\pi R_s (i-1) t} sinc(\pi (R_s t - k + 1))
$$

(46)

$$
Y_{ki} = s_i x_i[k] + w[k]
$$

(47)

where $w[k] \sim \mathcal{CN}(0, N_0)$. If we model $x_i[k]$ also as i.i.d. $x_i[k] \sim \mathcal{CN}(0, \varphi_i)$ we get that for a given state $s$:

$$
Y_{ki} \sim \mathcal{CN}(0, \varphi_i + N_0).
$$

(48)

where $\varphi$ (introduced in the presentation of our model) is a vector collecting the average, unknown a priori, power received from the existing communications. The receiver samples for $k = 1, \ldots, K$ are:

$$
c[k] = \sum_{i=1}^{N} \sqrt{b_{ki}} e^{j\phi_i} (s_i x_i[k] + w[k])
$$

(49)

and therefore (assuming $\phi_i$’s are independent and uniformly distributed in $[0, 2\pi]$) they are also conditionally zero mean Gaussian random variables:

$$
c[k] \sim \mathcal{CN}(0, \theta[k] \triangleq b_k (\varphi^T + w^T))
$$

(50)

It follows that the information for the detection of the PU communications is in the variance and by considering as observations

$$
y[k] \triangleq |c[k]|^2
$$

(51)

then one has $y[k] \sim \mathcal{E}xp(|\theta[k]|)$ in accordance with the model presented in Section III.

Remark 2: Note that while modulated signals are discrete and non-Gaussian, here it is reasonable to assume that its distribution is well approximated by a Gaussian p.d.f since the receiver is not synchronized with the active source and the signal received, while remaining in its original band, is most likely subject to linear distortion due to a multi-path channel. It would be more appropriate potentially to include a certain correlation among the samples $x_i[t]$ and it is just for simplicity that we do not consider it, given the generalization is straightforward and does not impact the derivation of the opportunistic strategy.

VII. SIMULATION RESULTS

In this Section we showcase the ability of our approach to dynamically switch between a Direct Inspection (scanning receiver) and a Group Testing approach, based on the expected occupancy (the vector of priors $\omega$), the time available $K$, the minimum SNR threshold $SNR_{\text{min}} = \frac{\sum_{i} b_i}{w}$ and the number of resources $N$. In the context of spectrum sensing (case 0), the parameters $v_i$ and $\rho_i$ can be mapped into a maximization of the overall weighted network throughput (see [25]): the reward $r_i$ can be proportional to the achievable rate over the channel $i$ in the absence of PU communications, i.e. $r_i \propto \log(1+SNR_i, S)$ (where the suffix $S$ indicates the secondary communication), whereas the penalty $\rho_i$ can be made proportional to the loss in rate caused to the primary communication due to the interference added by the secondary. Notice also that from (3)-(4), one can normalize the $r_i$’s and $\rho_i$’s over the same constant without altering the optimization, therefore the bandwidth can be ignored in the definition of $r_i$ and $\rho_i$.

Note that for the cognitive radio application the concept of exploitation of the resource is tied to the definition of utility function chosen in [3] which is expressed in bits/s/Hz. The longer the time available to transmit the larger is number of bits that can be transmitted over that band. For the other case, i.e. when the reward comes from detecting correctly which resources that are busy (for example a RADAR application) it is not immediately clear why the utility would be proportional to the number of remaining time instants. To interpret this, we model the action upon declaration of $s_i = 1$ as a Bernoulli trial which accrues a reward $r_i$ if such action is successful (i.e. the target is actually hit) and this happens with a certain probability $p_i$ for each attempt. The number of attempts $T_i$ necessary to hit the target will then be geometrically distributed. One can find then that the expected reward is equal to

$$
v_i P(T_i \leq (K - K)) = r_i \sum_{k=1}^{K-K} p_i (1-p_i)^{k-1} = r_i (1-(1-p_i)^{K-K}) \approx (K-K) r_i p_i
$$

for small $p_i$, which would motivate having an expected utility which increases linearly with time. The $\rho_i$ associated with this case would model an intervention cost, which main purpose is to limit the false alarm rate. It is important to highlight, however, the time dependency in the objective would prevent our formulation to return a standard constant false alarm rate CFAR detection method. Nevertheless, our model can apply to electronic warfare (tentatives of create jamming), wake-up radio and other problems where the action (and the associated utility) is on the channels that are declared busy.

Notice that in light of the symmetry in the definition of the threshold $\gamma_i$, one can switch the $r$’s and $\rho$’s to go from case 0 (cognitive radio) to case 1 (RADAR) and find the same

---

\(\text{1}\) If the periodic signals where not truncated in time the relationship would be exact, in practice there will be some approximation error due to the windowing of the signal over the prescribed interval $[k-1]T_s, kT_s]$. The effect of this can be mitigated by using raised cosine filtering and a non-rectangular window to reduce the effect of side lobes.
trends, even for the combined tests. However, we highlight the difference in the two scenarios in the first simulation we present. For this experiment we set $K = 30, N = 60$ and $r_i = r, \rho_i = \rho$ and $\omega_i = \omega, SNR_i = SNR_{\text{min}}(10dB) \forall i \in \mathcal{N}$ we have that for $\omega$ equal to $\frac{\rho}{\rho+r}$ or $\frac{r}{\rho+r}$ for case 0 and case 1 respectively, i.e. the threshold value given in Assumption [1] which guarantees no resource can give positive utility if not tested. As we can see, in both scenarios the utility increases with the ratio $\frac{\omega}{r}$, since the prior increases favorably with respect to the utility function. However, for the spectrum sensing application the Group Testing (GT) approach gives higher accuracy. For this experiment we looked at case GT approach outperforms whereas when the horizon increases SNR for procedure in Algorithm 1. In principle, the optimal value maximum number of resources per test allowed in our greedy maximization is potentially worse for higher values of $\rho$ introduces additional degrees of freedom. However, as proved in our Lemma [3] the approximation factor of the greedy maximization is potentially worse for higher values of $\phi$ and this can be seen in the numerical results that follow. We also indicate with “Group Testing” the utility obtained with our GT approach whereas the “MAP Estimator” is the estimator that knows the true values $\varphi_i$, uses the same matrix $\mathbf{B}$ of the GT approach, but then decides on each resource based on the posterior for $\omega_i$ using belief-propagation. In Fig. 5 we plot the utility (normalized over $K^2$) over the ratio $\frac{\omega}{r}$ for two different horizons, i.e. $K = 10$ and $K = 30$ and $SNR_{\text{min}} = 10dB$. We can see that only for $\frac{K}{N} \approx 0.75$ the GT approach outperforms whereas when the horizon increases almost no benefit is given by mixing resources since, roughly speaking, there is enough time to test them independently and have higher accuracy. For this experiment we looked at case 0 and set $\omega_i \sim \left(0.7, \frac{\rho_i}{\rho_i+r_i}\right)$, where $r_i = \log(1 + SNR_i, S)$ and $\rho_i = 5r_i$ with $SNR_i, S_{1dB} \sim \mathcal{U}([10,20])$. The SNR for the test, i.e. $\frac{\omega}{w}$ is generated uniformly between 10 and 20 dB but the only information used in our algorithm is the minimum value, i.e. in this case $10dB$. In the regime we show, the DI is approximately constant since it is easy to show $U^{DI,OPT} \leq \frac{K}{2}u_{\text{max}}$ and therefore, for a fixed $K$ there is basically no benefit in increasing $N$ over $K$ except for having $\frac{K}{N}$ higher rewards due to the random generation of the parameters of the utility function. We then looked at how the utility behaves versus the SNR of each test. In this case the SNR was drawn uniformly between $SNR_{\text{min},dB}$ and $SNR_{\text{min},dB} + 10$, and once again only the value of $SNR_{\text{min}} = \frac{\omega}{w}$ was used in the algorithm, which is shown in the abscissa of the figures. Matching our intuition, we can see how the GT approach outperforms the DI only when $SNR_{\text{min}}$ is high enough and also that the gain in utility is larger for $K = 10$ than for $K = 30$. In fact, for this experiment the number of resources has been set constant to $N = 20$ and as previously highlighted, increasing $K$ for fixed $N$ diminishes the advantages of combining resources in a test. In this case we also plotted the utility obtainable with the ML estimate via Compressive Sensing described in Section [V] with a dense matrix that has the same aspect ratio of the one found via GT approach (i.e. that scans the same set of resources for the test) or a relatively dense matrix.

![Graph](image-url)
The absence of any optimization in the choice of which and how many resources to test produces a utility which for low SNR is lower than the DI approach proposed, and only for high enough SNR can outperform the DI approach, while still giving a utility lower than our GT strategy with $\phi = 2$. This highlights the benefit of having an active sub-Nyquist receiver compared to a static offline selection of the parameters.

VIII. CONCLUSIONS

In this work, we proposed a new framework to optimize the performances of an opportunistic spectrum access strategy with Sub-Nyquist sampling, and described the connection between such strategy and the design of the front-end sampling architecture. For the problem proposed we characterized the actor approximation of the greedy strategy and showed, via numerical results, how our dynamic design for the sensing matrix can outperform the DI approach, while still giving a utility lower than our GT strategy with $\phi = 2$. This highlights the benefit of having an active sub-Nyquist receiver compared to a static offline selection of the parameters.

APPENDIX

A. Proof of Lemma [2]

To prove the submodularity of $U^{DI}(A)$ we show that to prove the property in (1) is equivalent to prove

$$U_a(\omega_a) + U_b(\omega_b) \geq 0$$

(52)

which is true by assumption on the function $U_i(\omega_i)$: since $U_i(\omega_i) = 0$ for $\alpha_i = 1$, $\beta_i = 0$, we have $U_i(\omega_i) \geq 0$, $\forall i \in N$ for the optimized $\alpha_i^*, \beta_i^*$.

B. Proof of Lemma [2]

The function is the sum of two terms, to prove the first one is sub-modular one can follow the same steps in Appendix A. For the second term, it is enough to show that, for any $i$, $-\gamma(\deg(i))$ is a sub-modular function of $\gamma$. The function is clearly sub-modular since it a concave function of the nodal degree, and from this we can conclude the second term is a positive sum of sub-modular functions, hence sub-modular. To prove the equivalence of the two optimizations in (32)-(34), we first note that for any $\gamma$ that satisfies the constraints in (32), the second term of the objective in (34) is equal to 0 and the two objectives are equal. It follows that we simply need to verify that no set of edges that violates the constraint on the nodal degree would be the optimal solution for (34). To show this, we note that any infeasible set of edges (according to (32)) can be made feasible by removing some edges. For $M$ large enough, i.e. $M > K \max_{ij} u_{ij}$, it is relatively straightforward to verify that such removal of edges would improve the objective, preventing an infeasible solution for (32) to be optimal for (34), and this concludes the proof.

C. Proof of Lemma [2]

We want to prove

$$U^{MT}(C^G) \geq \alpha U^{MT}(C^{OPT})$$

(53)

where $\alpha = \frac{1}{\min_{\gamma}(\phi_{\min}, \frac{K-1}{\gamma})}$ and $\phi_{\min} \leq \phi$ is the largest test size returned by the greedy algorithm. We also rewrite

$$U^{MT}(C^G) = (K - |C^G|)U_{CG}$$

$$U^{MT}(C^{OPT}) = (K - |C^{OPT}|)U_{C^{OPT}}$$
To prove the claim we look at the graph obtained by the union of the cycles in the optimal and the greedy solution. Since in each of the solution, no node can be in two cycles it follows that in the obtained graph, no node can be in more than two cycles. Let us start by assuming there is a cycle \( c \) with associated utility \( u_c \) in the optimal solution that does not share any node with the greedy solution. This means

\[
\frac{U_{\text{OPT}} - u_c}{K - |C^{\text{OPT}}|} \leq u_c \leq \frac{U_G}{K - |C^G|} - 1
\]

\[
\Rightarrow U_{\text{OPT}} \leq (K - |C^{\text{OPT}}| + 1)u_c \leq U_G \frac{K - |C^{\text{OPT}}|}{K - |C^G|}
\]

where \( K \) follows from the fact that adding \( c \) to \( C^{\text{OPT}} \setminus \{c\} \) improves the objective but would not improve the objective for the greedy solution. From \( K \) we could then conclude \( |C^G| > |C^{\text{OPT}}| \), since for \( |C^G| \leq |C^{\text{OPT}}| \) we would find from \( K \) that \( U_{\text{OPT}}(c^G) \geq U_{\text{OPT}}(C^{\text{OPT}}) \). This means we can replace a cycle in \( C^G \) with this isolated cycle, to form a set of cycle \( C^G \) whose objective is lower than \( C^G \) by greedy search, and that in light of \( K \) we can iterate this process by always picking the cycle to be replaced in such a way that all the cycles in the optimal solution have at least one node with the set of cycles in \( C^G \). Now we have that all the cycles in the optimal solution have at least one node with a cycle in \( C^G \). If instead one has that no cycle \( c \) in the optimal solution is isolated and that there are isolated cycles in the greedy solution, then the set \( C^G \) is formed by removing these cycles from \( C^G \), lowering the objective (by submodularity and greedy search) and one would again obtain that all the cycles in the optimal solution have at least one node with the set of cycles in \( C^G \). We now iteratively remove cycles from \( C^G \) and \( C^{\text{OPT}} \), while bounding the loss in performance and therefore obtain the factor approximation we want to prove. We can remove cycles from \( C^G \) in decreasing order of utility and since we know that for each cycle \( c' \) of length \( \phi \) in \( C^G \) there are at most \( \phi \) different cycles in the optimal solution that share a node with \( c' \), by greedy search we have that \( \phi \cdot u_{c'} \) is greater than the utility given by the cycles in the optimal solution that are adjacent to cycle \( c' \). Let us then define \( C^G \) as the minimal subset of \( C^G \) that can cause the removal of all the cycles in the optimal solution when iterating the procedure just described, i.e. the set containing the first \( |C^G| \) in decreasing order of utility contained in \( C^G \). Again by sub-modularity and greedy search one can easily find that the objective for \( C^G \) is lower than \( C^G \), since if the objective could not be improved by removing a cycle from \( C^G \), then it also cannot improve the objective for \( C^G \) which has utility strictly greater than \( C^G \). At this point we can prove \( K \) for \( C^G \) and this will prove it for \( C^G \). At this point we use \( d \) to define the number of cycles in \( C^{\text{OPT}} \) that can be removed by removing the cycle \( c \) in \( C^G \) and \( d \triangleq \max d_c \). By then iterating our procedure described above we end up having

\[
(K - |C^G|)U_{C^G} = \frac{1}{d} \frac{K - |C^G|}{K - |C^{\text{OPT}}|} (K - |C^{\text{OPT}}|)d \cdot U_{C^G}
\]

and since \( (K - |C^G|)U_{C^G} \geq (K - |C^G|)U_{C^G} \) this concludes the proof. We have used the fact that \( d \leq \phi \) and that the function \( d(K - d) \) has its maximum in \( d = \frac{K}{2} \). Fig. 7 shows an example of the iterative procedure to obtain the bound just derived.

REFERENCES

[1] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” IEEE transactions on Signal Processing, vol. 50, no. 6, pp. 1417–1428, 2002.
[2] C. S. Seelamantula and M. Unser, “A generalized sampling method for finite-rate-of-innovation-signal reconstruction,” IEEE Signal Processing Letters, vol. 15, pp. 813–816, 2008.
[3] V. Y. Tan and V. K. Goyal, “Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm,” IEEE Transactions on Signal Processing, vol. 56, no. 10, pp. 5135–5146, 2008.
[4] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, “Performance bounds and design criteria for estimating finite rate of innovation signals,” IEEE Transactions on Information Theory, vol. 58, no. 8, pp. 4993–5015, 2012.
[5] M. Mishali and Y. C. Eldar, “Wideband spectrum sensing at sub-nyquist rates [applications corner],” IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 102–135, 2011.
[6] X. Li, A. Rueetschi, A. Scaglione, and Y. C. Eldar, “Compressive link acquisition in multiuser communications,” IEEE Transactions on Signal Processing, vol. 61, no. 12, pp. 3229–3245, 2013.
[7] Z. Tian and G. B. Giannakis, “Compressed sensing for wideband cognitive radios,” in Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, vol. 4. IEEE, 2007, pp. IV–1357.
[8] P. Shukla and P. L. Dragotti, “Sampling schemes for multidimensional signals with finite rate of innovation,” IEEE Transactions on Signal Processing, vol. 55, no. 7, pp. 3670–3686, 2007.

[9] E. Arias-Castro and Y. C. Eldar, “Noise folding in compressed sensing,” IEEE Signal Processing Letters, vol. 18, no. 8, pp. 478–481, 2011.

[10] D. Baron, S. Sarvotham, and R. G. Baraniuk, “Bayesian compressive sensing via belief propagation,” IEEE Transactions on Signal Processing, vol. 58, no. 1, pp. 269–280, 2010.

[11] Q. Zhao, B. Krishnamachari, and K. Liu, “On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance,” IEEE Transactions on Wireless Communications, vol. 7, no. 12, 2008.

[12] J. Unnikrishnan and V. V. Veeravalli, “Algorithms for dynamic spectrum access with learning for cognitive radio,” IEEE Transactions on Signal Processing, vol. 58, no. 2, pp. 750–760, 2010.

[13] K. Wang and L. Chen, “On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach,” IEEE Transactions on Signal Processing, vol. 60, no. 1, pp. 300–309, 2012.

[14] S. Bagheri and A. Scaglione, “The restless multi-armed bandit formulation of the cognitive compressive sensing problem,” Signal Processing, IEEE Transactions on, vol. 63, no. 5, pp. 1183–1198, March 2015.

[15] J. Hao, F. Tosato, and R. J. Piechocki, “Sequential compressive sensing in wireless sensor networks,” in Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th. IEEE, 2012, pp. 1–5.

[16] D. M. Malioutov, S. R. Sanghavi, and A. S. Willsky, “Sequential compressed sensing,” IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 435–444, 2010.

[17] J. Zhao, X. Wang, and Q. Liu, “Cooperative sequential compressed spectrum sensing over wide spectrum band,” in Sensing, Communication, and Networking (SECON), 2015 12th Annual IEEE International Conference on. IEEE, 2015, pp. 1–9.

[18] G. Braun, S. Pokutta, and Y. Xie, “Info-greedy sequential adaptive compressed sensing,” IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 4, pp. 601–611, 2015.

[19] N. Michelusi and U. Mitra, “Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics,” IEEE Transactions on Cognitive Communications and Networking, vol. 1, no. 1, pp. 128–145, 2015.

[20] Y. Kaspi, O. Shayevitz, and T. Javidi, “Searching for multiple targets with measurement dependent noise,” in Information Theory (ISIT), 2015 IEEE International Symposium on. IEEE, 2015, pp. 969–973.

[21] G. K. Atia and V. Saligrama, “Boolean compressed sensing and noisy group testing,” IEEE Transactions on Information Theory, vol. 58, no. 3, pp. 1880–1901, 2012.

[22] J. Scarlett and V. Cevher, “Converse bounds for noisy group testing with arbitrary measurement matrices,” in Information Theory (ISIT), 2016 IEEE International Symposium on. IEEE, 2016, pp. 2868–2872.

[23] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, “Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms,” in Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on. IEEE, 2011, pp. 1832–1839.

[24] A. Sharma and C. R. Murthy, “On finding a subset of non-defective items from a large population using group tests: Recovery algorithms and bounds,” in Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015, pp. 4150–4154.

[25] L. Ferrari, Q. Zhao, and A. Scaglione, “Utility maximizing sequential sensing over a finite horizon,” 2017, accepted for publication in Signal Processing, IEEE Transactions on.