Multi-Objective Clustering Optimization for Multi-Channel Cooperative Spectrum Sensing in Heterogeneous Green CRNs

Abdulkadir Celik, Ahmed E. Kamal
Dept. of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50011

***Abstract*—In this study, we address *energy efficient* (EE) *cooperative spectrum sensing* (CSS) policies for large scale heterogeneous *cognitive radio networks* (CRNs) which consist of multiple *primary channels* (PCs) and large number of *secondary users* (SUs) with heterogeneous sensing and reporting channel qualities. We approach this issue from macro and micro perspectives; *Macro perspective* groups SUs into clusters with the objectives: 1) Total energy consumption minimization, 2) Total throughput maximization, and 3) Inter-cluster energy and throughput fairness. We adopt and demonstrate how to solve these using the Non-dominated Sorting Genetic Algorithm-II (NSGA-II). *The micro perspective*, on the other hand, operates as a sub-procedure on cluster formations decided by the macro perspective. For the micro perspectives, we first propose a procedure to select the cluster head (CH) which yields: 1) The best CH which gives the minimum total multi-hop error rate, and 2) the optimal routing paths from SUs to the CH. Exploiting *Poisson-Binomial* distribution, a novel and generalized K-out-of-N voting rule is developed for heterogeneous CRNs to allow SUs to have different local detection performances. Then, a convex optimization framework is established to minimize the intra-cluster energy cost by jointly obtaining the optimal sensing durations and thresholds of *feature detectors* for the proposed voting rule. Likewise, instead of a common fixed sample size test, we developed a weighted sample size test for quantized soft decision fusion to obtain a more EE regime under heterogeneity. We have shown that the combination of proposed CH selection and cooperation schemes gives a superior performance in terms of energy efficiency and robustness against reporting error wall.**

I. Introduction

A. Background and Motivation

The motivation behind the cognitive radio (CR) technology is rooted in the deficiency of the current rigid spectrum allocation policy to meet the high *quality of service* (QoS) demands of today's wireless communication networks. With traditional spectrum allocation policies, radio spectrum is allocated for long-running time periods and exploited merely by licensees, such that spectrum is usually underutilized [1]. CRs have been introduced to detect and utilize unused spectrum bands in an opportunistic and non-intrusive manner, such that, primary (licensed) users (PUs) are protected against performance degradation caused by CRs which are also referred to as secondary users (SUs).

Since a substantial proportion of mobile and wireless network devices have limited energy resources, energy efficiency of CRNs led itself into being an inevitable design consideration. Moreover, considering the fact that 30% of the energy expenditure of mobile devices is caused by wireless networking and computing hardware [2], *energy efficient* (EE) *cognitive radio networks* (CRNs) play a vital role to provide portable devices with more spectrum for less energy consumption. Since approximately 2% of the worldwide CO_2 emissions is caused by the communications and information technologies [3], optimizing energy consumption leads to not only a more affordable network with reduced cost, but also an environmentally friendly network. Therefore, EE policies are now at the forefront of CRN research due to environmental and operational energy costs.

Nevertheless, detection performance of individual SUs is severely affected by channel impairments, therefore, *cooperative spectrum sensing* (CSS) has been exploited to take advantage of the spatial diversity of SUs to detect the primary channel (PC) activity with higher confidence [4]. In a cluster based CSS, members of an SU cluster report local results via an erroneous common control channel (CCC) to a *cluster head* (CH) which fuses local reports to obtain a global decision and feeds it back to the cluster members. In this regard, EE-CSS is a favored sensing method due to its performance in trading off the domains of QoS, PU interference and network complexity [5]. However, considering certain objectives and constraints, formation of clusters and achievement of EE-CSS is a non trivial task in a large scale heterogeneous network which is defined as follows: *A large scale heterogeneous CRN consists of large number of PCs and SUs with heterogeneous signal-to-noise-ratios (SNRs), i.e., sensing qualities, on different primary channels (PCs). SUs also experience heterogeneous CCC errors, i.e., reporting qualities, among themselves.*

B. Related Work

Some of the recent research effort addressing the EE-CSS can be exemplified as follows: By use of amplify-and-forward relaying, Huang et al. consider an EE-CSS scheme with the goal of minimizing the total energy consumption subject to detection performance constraints [6]. In [7], Deng et al. divide the sensing nodes into a number of nondisjoint feasible subsets which satisfy the detection requirements. Then, they extend the network lifetime by successively activating each subset while keeping others in a sleep mode. Reference [8] considers a single SU and multiple primary channels (PCs) and optimizes the average energy cost of spectrum sensing, channel switching, and data transmission subject to some sensing reliability, delay and throughput constraints. In [9], the authors provide a theorem along with the closed-form expression for determining the optimal number of CRs to obtain an EE-CSS. Authors of [10] propose a convex optimization framework to minimize the energy consumption by deriving a closed form expression for SU priority and detection thresholds. From a different perspective, [11] achieves EE-CSS by reducing the total number of reports exchanged between the SUs and the CH, which is an efficient approach where the reporting energy is dominant.

Since data fusion requires extensive amount of bandwidth for the CCC, quantized SDF is considered to be a practical

yet an efficient method in [12] and [13]. Employing auto-correlation based feature detector, authors in [12] study the effect of CCC errors and clearly show the existence of an error wall at which successful detection is not possible. On the other hand, [13] uses an energy detector with a Lloyd-Max (LM) quantizer and shows that it gives a better performance than the Maximum-Output entropy quantizer in [12]. However, studies in [13] and [12] do not consider the heterogeneity, energy efficiency, clustering, multihop reporting, and the CH selection.

The study in [14] is first to address the energy efficiency in cluster based CSS. They first propose a voting scheme based on SUs' own confidence then develop a cluster-collect-forward scheme to save energy spent on vote-collection and information exchange. In [15], to reduce reporting time and bandwidth requirement, a dynamic CH selection scheme is proposed based on the sensing qualities of SUs. Kozal et al. propose a multi-hop reporting scheme to reduce the reporting power consumption. [16].

C. Main Contributions and Novelty

We consider clusters as groups of SUs dedicated to sense a single PC, therefore, there is one to one correspondence between a PC and the cluster of SUs sensing this PC. We then approach the fair energy and throughput optimization of cluster formations from *macro* and *micro* perspectives. In this respect, the macro perspective primarily tackles the SU and cluster association issue with the following goals: 1) Minimization of total energy consumption of clusters, 2) Maximization of total throughput of clusters, and 3) Inter-cluster energy cost and throughput fairness. Accordingly, we formulate a multi-objective clustering optimization (MOCO) as a mixed-integer non-linear programming (MINLP) problem. Then, we adopt and demonstrate how to use the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) to heuristically solve the MOCO.

On the other hand, the micro perspective is a sub-routine of the macro perspective to evaluate the performance of a candidate solution which is simply a single instance of all possible cluster formations. That is, given a cluster formation, the following tasks are performed at the micro level: 1) Selection of CHs along with the optimal reporting paths with minimum reporting error between cluster members and CHs, 2) Determination of optimal sensing parameters (ie. sensing durations and detection thresholds) of each SU to minimize the energy consumption of each cluster subject to PU protection and spectrum utilization constraints. Our contributions in micro perspective are detailed as follows:

Instead of using common single-hop reporting links between SUs and CHs, employing multi-hop path diversity might result in a superior reporting performance in terms of robustness, delay and communication range. Hence, we propose a procedure to select the CH which yields: 1) The best CH which gives the minimum total multi-hop error rate, and 2) the optimal routing paths from SUs to the CH using *Dijkstra's algorithm*. Obtained results show that the multihop diversity has a superior robust reporting and a potential to alleviate the bit error probability (BEP) wall phenomenon in [12].

Exploitation of commonly studied Binomial based K-out of-N rule would not yield an EE-CSS in a heterogeneous environment since it treats each SU equivalently by enforcing them to have identical local detector performance. To illustrate, enforcing heterogeneous SUs to sense with identical detection performances will make SUs with relatively low SNRs to sense longer. In this case, while the total energy cost increases, the available time left for secondary transmission decreases since the CH will wait for the slowest SU to announce the global decision. Therefore, a novel hard decision fusion (HDF) based CSS scheme is proposed to tolerate SUs to report with various local detection accuracy according to their sensing and reporting quality. We further develop a convex framework which jointly optimizes detection threshold and sensing duration of each SU to minimize the total energy consumption of the cluster. Numerical results clearly demonstrate that the taking the heterogeneity into account yield a lower total energy consumption and higher throughput.

For the quantized soft decision fusion (SDF), commonly studied fixed sample size test (FSST) also has significant disadvantages since it requires heterogeneous SUs to have identical sensing durations. After deriving the distributions of the test statistic over the optimal multihop reporting path, we propose a weighted sample size test (WSST) to obtain a more EE regime by assigning sensing duration of SUs proportional to their SNRs. Results revealed that the achieved performance of proposed scheme is superior to FSST.

The rest of this paper is organized as follows: Section-II introduces the system model. Then, Section-III describes the best CH selection procedure. Section-IV gives the details of HDF-based and SDF-based CSS. Section-V develops MOCO and explains its solution with NSGA-II. Finally, simulation results and analysis are presented in Section-VI and Section-VII concludes the paper with a few remarks.

II. System Model

We consider a cluster based heterogeneous CSS with N SUs and M PCs and assume $N >> M$. Each cluster is responsible for sensing only one channel at a time. Time is divided into fixed-length slots, T, in each of which a PC is either in the busy or idle state for the whole slot. SUs can join at most one cluster during a time slot. Assuming that PCs and CCC experience slow and block fading, we adopt the simplified path loss model which inherits necessary design parameters from an empirical model for the path loss [17]. Albeit its simplicity and popularity, *energy detector* (ED) operates on a perfect noise variance knowledge which results in a poor detection performance under estimation errors. Also, noting that ED is subject to an SNR wall under which accurate detection is impossible, exploiting known features of primary signals can provide SUs with an improved detection performance and robustness [18].

Accordingly, we prefer to use an auto-correlation based *feature detector* (FD) which detects PUs using the second-order statistics of primary signals.

For instance, orthogonal frequency-division multiplexing (OFDM) inserts a cyclic prefix (CP) to each data block.

Denoting the number of symbols in a CP and data block as T_c and T_d, respectively, the autocorrelation coefficient of SU_n on PC_m is given as [19]

$$\rho_n^m = \frac{T_c}{T_c + T_d} \frac{\gamma_n^m}{1 + \gamma_n^m} \tag{1}$$

where $\gamma_n^m = {}^{\sigma_{n,m}^2}/_{\sigma^2}$ is the SNR of SU_n on PC_m, $\sigma_{n,m}^2$ is the received primary signal power by SU_n on PC_m, and σ^2 is the noise power. Please note that T_c and T_d can either be estimated by SUs and recorded for different bands or be obtained from technical standardizations. As a case study, we consider commonly exploited OFDM based primary signals in Section VI and obtain T_c and T_d values from IEEE 802.11 standard [20]. The fundamental objective of detector design is to choose the test statistic $\mathcal{K}_n^m$ and determine the detection threshold ε_n^m and number of samples S_n^m which is proportional to sensing duration and the primary channel bandwidth, W_m. In this respect, we employ log-likelihood ratio (LLR) of the received complex PU signal vector, $\mathbf{r} \in \mathbb{C}^{S_n^m}$, as our test statistic which is given as

$$\Lambda_n^m = \log\left(\frac{f\left(\mathbf{r}|\mathcal{H}_1\right)}{f\left(\mathbf{r}|\mathcal{H}_0\right)}\right) \tag{2}$$

where $f\left(\mathbf{r}|\mathcal{H}_1\right)$ and $f\left(\mathbf{r}|\mathcal{H}_0\right)$ are the conditional probability distribution functions (pdf) of $\mathbf{r}$. Accordingly, LLR test can be expressed as

$$\mathcal{K}_n^m = \Lambda_n^m \underset{\mathcal{H}_m^0}{\overset{\mathcal{H}_m^1}{\gtrless}} \varepsilon_n^m \tag{3}$$

Accordingly, conditional distributions of Λ_n^m are given by

$$\Lambda_n^m|\mathcal{H}_m^0 \sim \mathcal{CN}\left(0, \frac{2S_n^m(\rho_n^m)^2}{[1-(\rho_n^m)^2]^2}\right) \tag{4}$$

$$\Lambda_n^m|\mathcal{H}_m^1 \sim \mathcal{CN}\left(\frac{2S_n^m(\rho_n^m)^2}{1-(\rho_n^m)^2}, 2S_n^m(\rho_n^m)^2\right) \tag{5}$$

where $\mathcal{CN}(\cdot)$ denotes circularly symmetric complex Gaussian distribution [19]. Based on (4) and (5), the probabilities of false alarm and detection of SU_n on PC_m are given by

$$P_{n,m}^f\left(S_n^m, \varepsilon_n^m\right) = \frac{1}{2}\operatorname{erfc}\left(\sqrt{S_n^m}\ \eta_n^m\right) \tag{6}$$

$$P_{n,m}^d\left(S_n^m, \varepsilon_n^m, \rho_n^m\right) = \frac{1}{2}\operatorname{erfc}\left(\sqrt{S_n^m}\ \frac{\eta_n^m - \rho_n^m}{1-(\rho_n^m)^2}\right) \tag{7}$$

where $\eta_n^m = \frac{1-(\rho_n^m)^2}{2S_n^m\rho_n^m}\left(\varepsilon_n^m + S_n^m \log\left(1-(\rho_n^m)^2\right)\right) + \rho_n^m$ [19].

Although the autocorrelation function is spread due to the time dispersion in a multipath fading channel, averaging the second-order statistics over multiple OFDM symbols can alleviate the impact of multipath fading to achieve a detection performance close to the that of AWGN channel [18]. Hence, $P_{n,m}^f$ and $P_{n,m}^d$ under the multipath fading can be calculated from (6) and (7) by substituting γ_n^m with its average $\bar{\gamma}_n^m = \gamma_n^m \sum_i \mathbb{E}\left[|h_{n,m}^i|^2\right]$ where $h_{n,m}^i$ is the i^{th} channel tap coefficient. Without loss of generality, we will address these two scenarios together by assuming $\sum_i \mathbb{E}\left[|h_{n,m}^i|^2\right] = 1$.

III. Cluster Head Selection

Assuming cluster members are given by the macro perspective, we will first consider the micro perspective and omit cluster indices m until Section V for the sake of tractability and without loss of generality. Even though single-hop reporting links between SUs and CHs are extensively studied, this may not always result in a reliable and energy efficient cooperation, especially when SUs with limited maximum transmission power are spread over a wide area. Alternatively, exploiting a multi-hop diversity does not only alleviate the communication range limitation and poor reporting channel quality but also gives a chance to save energy [21] and exploit an algorithm which finds the best CH and optimal multi-hop paths with minimum error rate from SUs to CHs. However, it is worth noting that the multi-hop reporting requires extra time overhead which is considered in the formulation of available throughput in Section V.

Table of Notations	
Notation	**Description**
M	Number of clusters/PCs with indexing $1 \leq m \leq M$
N	Number of SUs with indexing $1 \leq n \leq N$
T	Timeslot duration
T_c/T_d	Cyclic prefix/data block duration of primary signals
W_m	Bandwidth of PC_m
S_m^n	Number of Samples of SU_n on PC_m
ε_n^m	Local detection threshold of SU_n on PC_m
γ_n^m	SNR of SU_n on PC_m
ρ_n^m	Auto-correlation coefficient of SU_n on PC_m
$\Lambda_n^m/\hat{\Lambda}_n^m/\tilde{\Lambda}_n^m$	Observed/quantized/received LLR of SU_n^m on PC_m
$\mathcal{K}_n^m$	Local test statistic of SU_n on PC_m
$\mathcal{H}_m^0/\mathcal{H}_m^1$	Binary hypotheses for idle/busy state of PC_m
$P_{n,m}^f/P_{n,m}^d$	Local false alarm/detection prob. of SU_n on PC_m
$\tilde{P}_{n,m}^f/\tilde{P}_{n,m}^d$	Received false alarm/detection prob. of SU_n by CH_m
$\mathcal{C}$	Cluster set with cardinality, i.e., $\lvert\mathcal{C}\rvert = C$
$\mathcal{K}_{opt}/\mathcal{K}_{hd}/\mathcal{K}_{sd}$	Global test statistic for ideal/HDF/SDF schemes
$\kappa_{opt}/\kappa_{hd}/\kappa_{sd}$	Global thresholds for ideal/HDF/SDF CSS schemes
$E_m^{opt}/E_m^{hd}/E_m^{sd}$	Energy cost of cluster m for ideal/HDF/SDF based schemes.
$E_s/P_s/\tau_s$	Energy/power/duration per sample
$E_x/P_x/\tau_x$	Energy/power/duration per reported bit
Q_f/Q_d	Global detection/false alarm probabilities
Q_{th}^f/Q_{th}^d	Global detection/false alarm thresholds
L	Number of quantization levels, i.e., $L = 2^b$
ℓ_n^i	i^{th} quantization level of SU n, $1 \leq i \leq L$
$n \rightsquigarrow k$	An arbitrary routing path from SU_n to SU_k with H_n^k hops.
$n \rightarrow k$	Optimal routing path from SU_n to SU_k
$\boldsymbol{T_h}$	$L \times L$ Channel transition matrix of hop h, $1 \leq h \leq H_n^k$.
c_n^h	Transmitted codeword of length $b = \log_2 L$
$\boldsymbol{T}_{n\rightsquigarrow k}$	End-to-end channel transition matrix from SU_n to SU_k
$t_{i,j}^{n\rightsquigarrow k}$	(i,j) entry of $\boldsymbol{T}_{n\rightsquigarrow k}$, ie. $t_{i,j}^{n\rightsquigarrow k} = \mathcal{P}(\tilde{\Lambda}_n = \ell_n^i \lvert \hat{\Lambda}_n = \ell_n^j)$

Table I: Table of Notations

A. Cluster Topology and Path Characterization

The graph representing a cluster, $\mathcal{G}(\mathcal{C}, \mathcal{E}, \Omega)$, is defined by the set of SU nodes $\mathcal{C}$ with cardinality C, the set of edges $\mathcal{E}$ with cardinality E, and the set of edge weights Ω. We first denote the neighboring set of $SU_n \in \mathcal{C}$ as $\Phi_n = \{k|\, \gamma_n^k \geq \gamma_{th}, \forall k \in \mathcal{C}\}$ where γ_n^k and γ_{th} are the received signal SNR by SU_k from SU_n and the SNR threshold for the communication range, respectively. Please note that Φ_n and γ_n^k can be estimated via pilot and hello signaling mechanisms. Then, $\mathcal{E}_n = \{e_n^k \,|\, n, k \in \mathcal{C},\ n \neq k,\ k \in \Phi_n\}$ represents the direct edge/hop from SU_n to SU_k.

We assume link symmetry between SU pairs; however, the proposed scheme can easily be generalized to asymmetric links since the shortest path algorithm used in this paper is applicable to both directed and undirected graphs, and the undirected graph is used to capture link asymmetry. In this section, we will generalize our previously proposed single-bit multi-hop reporting and CH selection procedure [22] to support multi-bit reporting (quantized SDF) methods.

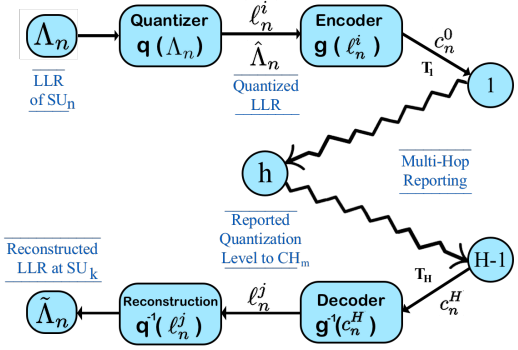

Figure 1: Multihop Reporting from SU_n to SU_k

Assuming SUs employ $L-level$ LM quantizers, Fig. 1 illustrates the multi-bit reporting between SU_n and SU_k over an arbitrary multi-hop path, $n \rightsquigarrow k$, which consists of H_n^k independent hops. SU_n first quantize the observed local LLR Λ_n such that $\hat{\Lambda}_n = q(\Lambda_n) = \ell_n^j$, $1 \leq j \leq L$. Then, the encoder maps ℓ_n^j into a $b = \log_2(L)$-bit binary codeword, $c_n^0 = g(\ell_n^j)$. SUs operate on these codewords using a decode and forward (DF) protocol in a bit-by-bit basis over the path from SU_n to the SU_k. Each hop is modeled as a binary symmetric channel (BSC) and is characterized by its $L \times L$ channel transition matrix, $\mathbf{T_h}, 1 \leq h \leq H_n^k$. The entry in the j^{th} row and i^{th} column of $\mathbf{T_h}$, $t_{i,j}^h$, is the probability of detecting c_n^i given that c_n^j is transmitted from the previous node, which is given by

$$t_{i,j}^h = \mathcal{P}_h(c_n^i|c_n^j) = (1-\epsilon_h)^{b-\delta(i,j)}\epsilon_h^{\delta(i,j)} \tag{8}$$

where $\delta(i,j)$ is the Hamming distance between c_n^i and c_n^j, $1 \leq i,j \leq L$ and ϵ_h is the crossover probability of the h^{th} hop that could be measured via plot signals.

Exploiting $\mathbf{T_h}$ matrices, the end-to-end multihop transition matrix can be calculated as

$$\mathbf{T}_{n \rightsquigarrow k} = \prod_{h=1}^{H_n^k} \mathbf{T_h} \tag{9}$$

which follows from the independent hop assumption and Markov property. Consequently, the i,j entry of $\mathbf{T}_{n \rightsquigarrow k}$ provides us the probability of receiving ℓ_n^i given that the SU_n transmitted ℓ_n^j,

$$t_{i,j}^{n \rightsquigarrow k} = \mathcal{P}(\tilde{\Lambda}_n = \ell_n^i | \hat{\Lambda}_n = \ell_n^j) \tag{10}$$

B. The Best Cluster Head Selection

In an erroneous CCC, SUs have to perform more accurate local detection to achieve a target performance. However, CCC imperfection can not be mitigated by spending more energy after the BEP wall. Therefore, following a routing path with minimum reporting error is crucial to mitigate this wall and to reduce the additional energy cost induced from CCC imperfections. Accordingly, we weight the direct edges from any $SU_a \in \mathcal{C}$ to any $SU_b \in \mathcal{C}$ with the symbol success probability (SSP) w_a^b which is any main diagonal element of $\boldsymbol{T_h}$. Depending upon this edge weighting[1], SSP of $n \rightsquigarrow k$ is given by $\Omega(n \rightsquigarrow k) = \prod_{a,b \in n \rightsquigarrow c} w_a^b$. In fact, the optimal path which maximizes the $\Omega(n \rightsquigarrow k)$ is the one which minimizes the negative sum of logarithms of $\Omega(n \rightsquigarrow k)$ as follows

$$n \rightarrow k = \underset{n \rightsquigarrow k}{\operatorname{argmin}} \left(- \sum_{a,b \in n \rightsquigarrow c} \log(w_a^b) \right) \tag{11}$$

By transforming the multiplication into a summation, *Dijkstra's algorithm* can be employed to calculate the reporting route with minimum path cost. Defining Ω_m^k as the total reporting error probability induced from selecting SU_k to be the CH, $\Omega_m^k = -\sum_{n \in \mathcal{C}} \log\left(\Omega\left(n \rightarrow k\right)\right)$, the SU with the minimum total reporting error is chosen to be the best CH as

$$CH = \underset{k \in \mathcal{C}}{\operatorname{argmin}}\ \Omega_m^k \tag{12}$$

IV. Cooperative Spectrum Sensing

A. Ideal CSS

We first introduce the ideal case since the performance evaluation of the HDF and SDF schemes will be based on the energy difference between the proposed and ideal cases. In the ideal case, CH knows the observed LLRs exactly and exploits the global summary test statistic as follows

$$\mathcal{K}_{opt} = \sum_{n=1}^{C} \Lambda_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa_{opt} \tag{13}$$

Using (4) and (5), conditional distributions of the global test statistic are given as

$$\mathcal{K}_{opt}|\mathcal{H}_0 \sim \mathcal{CN}\left(0,\ 2\sum_{n=1}^{C} \frac{S_n \rho_n^2}{[1-\rho_n^2]^2}\right) \tag{14}$$

$$\mathcal{K}_{opt}|\mathcal{H}_1 \sim \mathcal{CN}\left(2\sum_{n=1}^{C} \frac{S_n \rho_n^2}{1-\rho_n^2},\ 2\sum_{n=1}^{C} S_n \rho_n^2\right) \tag{15}$$

which reduces to the fixed sample size test (FSST) in [19] if we set $S_n = S$, $\forall n$. Following from (14) and (15), global false alarm and detection probabilities, $Q_f = \mathcal{P}\left(\mathcal{K}_{opt} > \kappa_{opt} \mid \mathcal{H}_0\right)$ and $Q_d = \mathcal{P}\left(\mathcal{K}_{opt} > \kappa_{opt} \mid \mathcal{H}_1\right)$, can be derived similar to (6) and (7). Accordingly, we refer to the ideal total energy consumption of the cluster as

$$E_{opt} = \min_{\substack{S_n, \kappa_{opt} \\ n \in \mathcal{C}}} \left\{ \sum_n S_n \mid Q_{th}^d \leq Q_d,\ Q_f \leq Q_{th}^f \right\} \tag{16}$$

which can be equivalently written as a convex problem due to the linearity of the objective function and the log-concavity of

[1]We ignore the case in which a symbol may be corrupted in an even number of hops, hence resulting in correct reception of the symbol, since the probability of this occurrence is very small

the constraints using similar steps followed in Appendix A.

B. Single-bit HDF-Based CSS

Suppose that SU_k is determined as the CH by (12), each $SU_n \in \mathcal{C}$ reports final binary decision $u_n \in \{0,1\}$ to the CH. If we denote the received decision bit by the CH as $\tilde{u}_n \in \{0,1\}$, local false alarm and detection probabilities received by the CH are given by

$$\tilde{P}_n^d = t_{1,0}^{n\to k}\left(1-P_n^d\right) + t_{1,1}^{n\to k}P_n^d \tag{17}$$

$$\tilde{P}_n^f = t_{1,1}^{n\to k}P_n^f + t_{1,0}^{n\to k}\left(1-P_n^f\right) \tag{18}$$

where P_n^f and P_n^d are given in (6) and (7), respectively. We define the test statistic for the global decision employed in CH as

$$\mathcal{K}_{hd} = \sum_{n\in\mathcal{C}} \tilde{u}_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa_{hd} \tag{19}$$

which is known as the K-out-of-N rule where CH decides on $\mathcal{H}_1$ if at least κ_{hd} of SUs report 1. Under the homogeneity assumption, i.e., $\tilde{P}_n^d = \tilde{P}^d,\ \tilde{P}_n^f = \tilde{P}^f, \forall n$, $\mathcal{K}_{hd}$ is a *Binomial* random variable. This assumption yields an optimal result if SUs have identical SNRs and the same reporting errors, which is hardly the case. However, we propose a general HDF scheme which employs the heterogeneity of SUs to achieve a more EE-CSS framework. In this case, $\mathcal{K}_{hd}$ has a *Poisson-Binomial* distribution and we have

$$Q_d = \sum_{i=\kappa_{hd}}^{C} \sum_{A\in F_i} \prod_{j\in A} \tilde{P}_j^d \prod_{k\in A^c}\left(1-\tilde{P}_k^d\right) \tag{20}$$

$$Q_f = \sum_{i=\kappa_{hd}}^{C} \sum_{A\in F_i} \prod_{j\in A} \tilde{P}_j^f \prod_{k\in A^c}\left(1-\tilde{P}_k^f\right) \tag{21}$$

where F_i is the set of all subsets of i integers that can be selected from $\{1,2,3,\ldots,C\}$ where C is cardinality of $\mathcal{C}$ [23]. Since F_i has $\binom{C}{i}$ elements, using an efficient method to calculate Q_f and Q_d is very important, especially when C is very large. For this purpose, probability mass function (pmf) and cumulative distribution function of *Poisson-Binomial* random variables can be expeditiously calculated in order of $\mathcal{O}(C\log_2 C)$ from polynomial coefficients of the probability generating function of $\mathcal{K}_{hd}$ [24].

HD-GP ($\boldsymbol{S}, \boldsymbol{\mathcal{E}}$) :

1: $\min\limits_{\boldsymbol{S},\boldsymbol{\mathcal{E}}} E_{hd} = E_s\mathbf{1}^T\boldsymbol{S} + E_x\mathbf{1}^T\boldsymbol{H}$

2: s.t. $Q_{th}^d \le Q_d$

3: $Q_f \le Q_{th}^f$

Regardless of the type of detector employed in SUs, detection performance is a function of the observation duration and sensing channel characteristics. Although obtaining more samples from the primary signal yields more accurate results, it is extravagant with the power consumed in Analog to Digital Conversion (ADC) and fast Fourier transformation (FFT) which are known to be the two major energy demanding components of the receiver [25]. We assume the energy spent for sensing and processing per sample is identical for each SU, i.e., $E_n^s = E_s = P_s\tau_s$ where P_s and τ_s denote the sensing power and sample duration. We also assume that SUs transmit with identical reporting energy per bit, $E_n^x = E_x = P_x\tau_x$ where P_x and τ_x denote the reporting power and reporting duration. To reduce the notational complexity, the variables and parameters is represented in a vectorized form, for instance, $\boldsymbol{S} = [S_1,\ldots,S_C]^T$. Then, the total energy cost of the cluster is given by

$$E_{hd} = E_s\mathbf{1}^T\boldsymbol{S} + E_x\mathbf{1}^T\boldsymbol{H} \tag{22}$$

where $\mathbf{1}$ and $\boldsymbol{H}$ represent vector of ones and number of hops from SUs to the CH.

Accordingly, Hard Decision based General Problem (HD-GP) formulate the problem of minimizing the total energy consumption induced from sensing and reporting activities of cluster members for a given voting rule κ_{hd} and the number of hops $\boldsymbol{H}$ calculated from Section III. In HD-GP, the objective in Line 1 is a linear function of $\boldsymbol{S}$ which needs to be minimized with respect to $\boldsymbol{S}$ and $\boldsymbol{\mathcal{E}}$. Lines 2 and 3 are PU protection and spectrum utilization constraint, respectively. HD-GP is not in a convex optimization problem form since the error functions in (6) and (7) are not jointly convex in $(S_n,\ \varepsilon_n)$. To alleviate this issue, we transform HD-GP into a *bilevel optimization* with convex upper level problem (ULP) and lower level problem (LLP) based on the convexity (concavity) of $\mathrm{erfc}(\cdot) \le 0.5$ ($\mathrm{erfc}(\cdot) \ge 0.5$) and the following parameterized nature of HD-GP. For a fixed $\bar{\varepsilon}_n$ and increasing $(S_n,\ E_{hd},\ P_n^d,\ \tilde{P}_n^d,\ Q_d)$ increase, $\forall n \in \mathcal{C}$. Hence, optimum E_{hd} is attained at the point $Q_d = Q_{th}^d$. For a fixed $\bar{S}_n$ and decreasing ε_n, on the other hand, $(P_n^f,\ P_n^d,\ \tilde{P}_n^f,\ \tilde{P}_n^d,\ Q_f,\ Q_d)$ increase. Noting that E_m^{hd} is not a function of ε_n, increase of Q_f is bounded by Q_{th}^f, thus, the minimum feasible ε_n is attained at $Q_f = Q_{th}^f$. In this fashion, we obtain the $\boldsymbol{\mathcal{E}}$ by parameterizing $\boldsymbol{S}$ in the LLP and achieve the minimum E_{hd} by parameterizing $\boldsymbol{\mathcal{E}}$ in the ULP.

HD-ULP ($\boldsymbol{S} \mid \bar{\boldsymbol{\mathcal{E}}}$) :

1: $\min\limits_{\boldsymbol{S}}$ $E_{hd} = E_s\mathbf{1}^T\boldsymbol{S} + E_x\mathbf{1}^T\boldsymbol{H}$

2: s.t. $-\log(Q_d) \le -\log(Q_{th}^d)$

3: $-\boldsymbol{P_d} \preceq -0.5$

4: $^{\mathbf{3}\boldsymbol{A}\bar{\boldsymbol{\mathcal{E}}}}/_{\boldsymbol{B}} \preceq \boldsymbol{S} \preceq {}^{\mathbf{3}\boldsymbol{A}\bar{\boldsymbol{\mathcal{E}}}}/_{\boldsymbol{B}-\boldsymbol{\rho_m}}$

HD-LLP ($\boldsymbol{\mathcal{E}} \mid \bar{\boldsymbol{S}}$) :

1: $\min\limits_{\boldsymbol{\mathcal{E}}}$ $\left|\log(Q_{th}^f) - \log(Q_f)\right|$

2: s.t. $\boldsymbol{P_f} \preceq 0.5$

where $A = \frac{1-\rho_n^2}{2\rho_n}$, $B = A\log(1-\rho_n^2) + \rho_n$, and $\preceq$ denotes element wise non-equality. In HD-ULP, the total energy cost is minimized for a given $\bar{\boldsymbol{\mathcal{E}}}$ which is obtained by exploiting the optimal solution set of HD-LLP for a given $\bar{\boldsymbol{S}}$. The constraint in Line 2 is the standard convex form of the first constraint in HD-GP. Constraints in Line 3 and 4 are introduced to preserve the convexity of the problem. In HD-LLP, the minimum feasible $\boldsymbol{\mathcal{E}}$ is obtained for given $\bar{\boldsymbol{S}}$ at the point of $Q_f = Q_{th}^f$. We refer interested readers to Appendix A for a more detailed convexity analysis.

C. Multi-bit SDF-Based CSS

As a remedy for the communication overhead which stems from reporting the raw LLR values directly to the CH, quantization is an attractive tool to achieve more accurate sensing yet a reasonable CCC overhead. An LM quantizer divides the support of the probability distribution functions (pdf) of Λ_n into non-overlapping intervals and determines the corresponding quantization levels of each intervals. Therefore, quantizer converts the pdf into a pmf $p_{\hat{\Lambda}_n}(\hat{\Lambda}_n = \ell_n^j),\ 1 \le j \le L$, which is calculated by integrating the pdf over corresponding intervals of ℓ_n^j.

However, the pmf of the received quantization levels, $p_{\tilde{\Lambda}_n}(\ell_n^i),\ 1 \le i \le L$, does not follow the the same distribution any more because of the imperfect reporting. That is, a quantization level received/reconstructed at the CH may result from any transmitted level at the source. Taking all possibilities into account, the $p_{\tilde{\Lambda}_n}(\ell_n^i)$ can be written in terms of the end-to-end transition matrix entries as follows

$$p_{\tilde{\Lambda}_n}(\ell_n^i) = \sum_{j=1}^{L} \mathcal{P}(\hat{\Lambda}_n = \ell_n^j)\ t_{i,j}^{n \rightsquigarrow k} \tag{23}$$

After collecting all reports, the CH sums $\tilde{\Lambda}_n$s and compare the global test statistic as follows

$$\mathcal{K}_{sd} = \sum_{n=1}^{C} \tilde{\Lambda}_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa_{sd} \tag{24}$$

We note that the global test statistic is the summation of C_m independent discrete random variable, hence, its pmf can be derived using convolution sum of marginal pmfs given in (23) [26]. Thus, the support values of global pmf is the Cartesian product of reporting SUs' quantization levels, i.e., $\mathcal{L} = \mathcal{L}_1 \times \cdots \times \mathcal{L}_C$.

Accordingly, the Soft Decision based General Problem (SD-GP) formulate the minimization problem of the total energy consumption of the cluster for a given number of quantization levels, L. Exploiting the FSST which assumes $S_n = S,\ \forall n$, SD-GP can be formulated as

SD-GP $(S, \kappa_{sd}|L)$ **:**

$$\begin{aligned} 1:\ &\min_{S,\kappa_{sd}} && E_{sd} = CSE_s + E_x b \mathbf{1}^T \boldsymbol{H} \\ 2:\ &\text{s.t.} && Q_{th}^d \le Q_d \\ 3:\ & && Q_f \le Q_{th}^f \end{aligned}$$

where Q_f and Q_d are monotonically decreasing with κ_{sd} while Q_d is monotonically increasing with S for a given κ_{sd}, $\bar{\kappa}_{sd}$. Then, by parameterizing the detection threshold as $\bar{\kappa}_{sd} = \min_{x \in \mathcal{L}} \{\mathcal{P}(\mathcal{K}_{sd} > x \mid \mathcal{H}_0) \le Q_{th}^f\}$, the best E_{sd} is attained at $S = \min\{S \mid Q_d \ge Q_{th}^d\}$. Due to the linearity of the objective and monotonically increasing Q_d, SD-GP can simply be solved by *Golden Section Method*. Under the venue of heterogeneous cluster, we propose a weighted sample size test (WSST) as follows

$$S_n = CS \frac{\rho_n}{\sum_{n \in \mathcal{C}} \rho_n} \tag{25}$$

In other words, instead of sharing the total CS number of samples equally as in FSST, SUs share it proportional to their sensing quality, i.e., ρ_n. Even if it may not be the optimal solution, numerical results will show a significant performance enhancement.

V. Multi-Objective Clustering Optimization

For a given sensing period, if there exists N SUs available to help with sensing and there exists M potential PCs to sense, a clustering of the SUs is required with the following objectives: 1) Total energy cost minimization of the clusters, 2) Total throughput maximization of clusters, and 3) Inter-cluster energy cost and throughput fairness. Accordingly, we formulate these objectives as follows:

$$F_1 = \sum_m E_m, \qquad F_2 = \sum_m T_m = \sum_m \max_{n \in \mathcal{C}_m} (S_m^n \tau_s + \tau_x b H_m^n)$$

$$F_3 = \max_m\ E_m - \min_m\ E_m, \qquad F_4 = \max_m\ T_m - \min_m\ T_m$$

where E_m represents the energy cost of the cluster m and can be substituted by E_m^{opt}, E_m^{hd} and E_m^{sd} according to preferred CSS schemes. T_m is the sensing+reporting duration of the slowest SU in cluster m, which determines the available time left for the secondary transmission. Therefore, minimizing T_m is equivalent to maximizing the transmission time. F_3 and F_4 close the gap between the maximum and minimum of E_m and T_m to achieve fairness in energy cost and throughput gain among the clusters, respectively. Defining $\mathcal{I}_m(n)$ as the indicator for the membership of SU_n in cluster m, we formulate MOCO as in Algorithm 1 where lines 2 and 3 are global PU protection and spectrum utilization constraints. Since, an SU can sense at most one channel during a sensing period, $\sum_{m=1}^{M} \mathcal{I}_n(m)$ in Line 2. Moreover, Line 3 makes sure that each PC is sensed by at least one SU. The constraint in Line 6 on the sensing time is especially beneficial to take SUs with unnecessarily low sensing quality out of consideration.

Algorithm 1 : MOCO

$$\begin{aligned} 1:\ &\text{Min} && \mathbf{F} = [F_1, F_2, F_3, F_4] \\ 2:\ &\text{s.t.} && Q_{th}^d \le Q_m^d, \forall m \\ 3:\ & && Q_m^f \le Q_{th}^f, \forall m \\ 4:\ & && \textstyle\sum_{m=1}^{M} \mathcal{I}_n(m) \le 1, \forall n \\ 5:\ & && \textstyle\sum_{n=1}^{N} \mathcal{I}_n(m) \le \bar{C}, \forall m \end{aligned}$$

Algorithm 1 is an NP-hard multi-objective MINLP problem which requires impractical time complexity even for moderate size of problems with a single objective. Due to the conflicting objectives, there may exists a set of *nondominated solutions* by which none of the objective functions can be improved without degrading some of the other objective values. Therefore, employing meta-heuristic methods to obtain a sufficient solution within a reasonable time frame is preferable in practice. Multi-objective genetic algorithms (MOGA), which are generic population based meta-heuristic approaches inspired by biological evolution, were shown to be performing well for many problems if it is adapted and applied carefully.

Hereupon, we will use the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [27] which is $\mathcal{O}\left(\mathbb{O}\mathbb{P}^2\right)$ where $\mathbb{O}$

and $\mathbb{P}$ are number of objectives and populations, respectively. Encoding is the first step of the implementation, which can be described as determination of the form in which a candidate solution (chromosome) translated into a genetic string. Based on the encoding scheme, NSGA-II first creates a random parent population P_0 for the initial generation and the genetic operators such as *crossover*, *mutation*, and *selection* are used to create an offspring population Q_0. Crossover is a way of introducing variations into the population of designs by mixing two different populations. Mutation operation, on the other hand, is introduced for maintaining the genetic diversity from one generation to the next. Selection involves choosing the chromosomes from current generation to be employed in the next generation based on the fitness values of the solutions which are calculated using the methods explained in Sections III and IV. For other generations $1 \leq g \leq \mathbb{G}$, a combined population $R_g = P_g \cup Q_g$ of size $2\mathbb{P}$ is formed by merging P_g and Q_g and the population R_g is sorted according to non-domination. Then, the best $\mathbb{P}$ members of the current generation are selected to be the parents of the next generation, P_{g+1}, using the crowded-distance operator. After that, selection, crossover and mutation operators are executed on P_{g+1} to create the offspring population of the $(g+1)^{th}$ generation, Q_{g+1}.

Since SUs may not always have good sensing quality on all PCs, the potential candidacy of SUs for clusters is one of the practical issues which affects the convergence of the Algorithm 1. Therefore, for a predetermined cluster size upper bound $\bar{C}$, we can sort SUs with respect to γ_n^m, $\forall n$ and select best $\bar{C}$ SUs as candidates. $\bar{C}$ can be determined using the 'OR' rule under homogeneity assumption as

$$\bar{C} = \max_{\substack{0.5 \leq P_d \leq 1 \\ 0 \leq P_f \leq 0.5}} \left(\left\lceil \frac{\log(1-Q_{th}^f)}{\log(1-P_f)} \right\rceil, \left\lfloor \frac{\log\left(1-Q_{th}^d\right)}{\log\left(1-P_d\right)} \right\rfloor \right)$$

Since 'OR' rule under the traditional HDF scheme is shown to be the one of the less EE case in Section VI, other schemes already satisfy the Q_{th}^f and Q_{th}^d with $\bar{C}$ for less energy expenditure. Denoting $\bar{N}$ as the number of SUs which are candidates for at least one cluster and $\mathcal{M}_n$ as the set of PCs/clusters for which SU_n is a candidate, chromosome design is shown in Table II.

PCs	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_8$	$\cdots$	$\mathcal{M}_n$	$\cdots$	$\mathcal{M}_{\bar{N}-1}$	$\mathcal{M}_{\bar{N}}$
SUs	1	3	8	$\cdots$	n	$\cdots$	$\bar{N}-1$	$\bar{N}$

Table II: A random chromosome representation for solution **s**

Using the coding scheme given in Table II, the constraint in Line 2 which requires an SU can be assigned at most one PU is already satisfied. For the constraint in Line 3, chromosomes are checked at the end of every genetic operation and genes violating these constraints are replaced with a proper value randomly. Constraints in Line 4 and 5 are handled directly by the method proposed in NSGA-II.

VI. Results and Analysis

Since orthogonal frequency-division multiplexing (OFDM) is a key technique with a broad range of employment in contemporary wireless systems, we assume that primary network communicates using OFDM technology. Due to the practical and theoratical considerations, we will use a symbol and the cyclic prefix length of 32 and 8, respectively. Thereupon, a single OFDM block will be enough to invoke the central limit theorem to keep results consistent with the theory. All simulation results are obtained and plotted using Matlab running on a PC with 2.7 GHz Intel Core i5 processor. We note that simplified path loss model with free space path loss exponent is used throughout the simulations.

Parameter	Value	Parameter	Value	Parameter	Value
P_s	1 $W/sample$ [25]	W_m	1 MHz	σ^2	1
τ_s	1 $\mu s/sample$	T	2 s	$\mathbb{G}$	10
P_x	0.7 W/bit [25]	T_c	8 [20]	$\mathbb{P}$	50
τ_x	0.1 ms/bit	T_d	32 [20]	M	16
Q_{th}^d	0.99	Q_{th}^f	0.01	N	250

Table III: Default parameter values used for simulations.

A. Cluster Head Selection

We assume that SUs use the CSMA/CA protocol for reporting and controlling purposes. Since we assume that DF operates on a bit by bit basis, crossover probabilities of each hop is calculated using BPSK modulation under Rayleigh fading. To compare the proposed multi-hop based proposed scheme with the common single-hop reporting, we consider two different geographical areas: $25\ m \times 25\ m$ and $50\ m \times 50\ m$ with three different cluster sizes, $C = (5, 10, 20)$. Averaging over 5000 different network scenarios, cluster's average number of edges, average number of hops, the average maximum number of hops and average multi-hop gain are shown in Table IV, where the multi-hop gain is defined as the ratio of single-hop total error rate to multi-hop total error rate. All values in Table IV increase with cluster size and area as a result of the increasing number of nodes and the distance among themselves. In particular, the effect of multi-hop diversity on reporting accuracy becomes more significant as the cluster size and the coverage area increase in size. In (12), sorting and computing Ω_m^ks take $\mathcal{O}\left(C \log C\right)$ and $\mathcal{O}\left(C\left(E + C \log C\right)\right)$, respectively. The average number of edges given in Table IV shows that $\mathcal{O}(E) \approx \mathcal{O}(C)$ in practical network scenarios, therefore, overall complexity of the CH selection procedure can be given as $\mathcal{O}\left(C^2 + C^2 \log C\right) \approx \mathcal{O}\left(C^2\right)$ since $\bar{C} = 10$ for $Q_d^{th} = 0.99$ and $Q_f^{th} = 0.01$. Furthermore, partial path calculation property of the Dijsktra's algorithm eliminates the recalculation of all routes in case of an SU inclusion/exclusion.

Distance	# Edges	# Hops	Max. # Hops	Multi-hop Gain
$25\ m \times 25\ m$	(5.1, 7.3, 9.1)	(1.1, 1.8, 3.2)	(1.7, 3.8, 6.5)	(1.9, 4.5, 5.6)
$50\ m \times 50\ m$	(0.9, 2.2, 4.3)	(1.2, 1.9, 3.3)	(1.9, 3.9, 6.6)	(2.2, 4.8, 10.5)

Table IV: Averaged multi-hop results for different areas and sizes of clusters. The 3 numbers between parentheses correspond to cluster sizes of 5, 10 and 20, respectively.

B. CSS with Hard Decision Fusion

To analyze the inherited features of the proposed HDF-based CSS, we first consider an example heterogeneous cluster which consists of 5 SUs with SNRs $[0, -5, -15, -20, -25]$ dB. We are interested in a comparison of the proposed K-out of-N rule and traditional K-out of-N rule which treats each SU equivalently and employs the *Binomial* distribution by

enforcing them to report with identical local detection and false alarm probabilities. Using traditional and proposed HDF with majority voting rule and global detection and false alarm probability targets of $Q_{th}^{d} = 0.99$ and $Q_{th}^{f} = 0.01$, respectively, obtained results are demonstrated in Fig. 2 where the FD and the proposed HDF scheme always give better energy and throughput efficiency with respect to the ED and traditional HDF scheme.

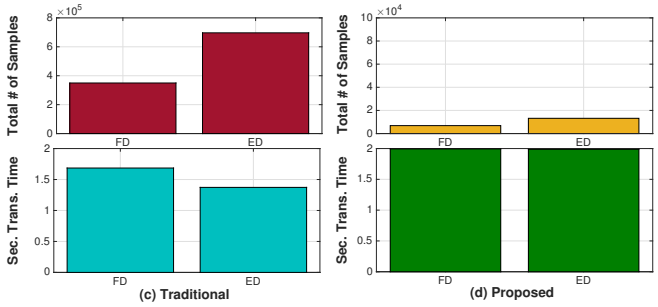

Figure 2: A simple heterogeneous cluster instance to compare EDs and FDs under traditional and proposed HDF-based CSS.

Local Prob.	**0 dB**	**-5 dB**	**-15 dB**	**-20 dB**	**-25 dB**
Trad. P_n^d	0.8944	0.8944	0.8944	0.8944	0.8944
Prop. P_n^d	0.9999	0.9993	0.9547	0.5622	0.5057
Trad. P_n^f	0.1056	0.1056	0.1056	0.1056	0.1056
Prop. P_n^f	0.0001	0.0007	0.0391	0.5000	0.5000
# Samples $\times 10^5$	**0 dB**	**-5 dB**	**-15 dB**	**-20 dB**	**-25 dB**
ED Trad.	0.0003	0.0008	0.0645	0.6313	6.2702
ED Prop.	0.0010	0.0055	0.1229	0.0023	0.0003
FD Trad.	0.0003	0.0005	0.0332	0.3188	3.1450
FD Prop.	0.0008	0.0033	0.0633	0.0013	0.0003

Table V: P_n^d, P_n^f and number of samples values for Fig. 2.

We note that the FD values are obtained using the false alarm and detection probabilities given in (6) and (7), respectively. Similarly, the ED values are obtained based on the false alarm and detection probabilities given in [28, (5), (10)]. While the superiority of FD is a result of the exploitation of known features of OFDM signals, the advantage of the proposed HDF scheme is resulted from enforcing SUs with relatively low and high SNRs to sense higher and lower local confidence, respectively, as can be seen from Table V. We note that the energy and throughput loss of traditional HDF is mainly because of enforcing the slowest SU with the lowest SNR value to achieve $P_d^n = 0.8944$ and $P_f^n = 0.1056$, which results in more energy loss and less time left for secondary transmission.

Next, we analyze the effect of the voting rule and the imperfection of the reporting environment on the *energy loss* which is defined as any extra sensing and reporting energy cost beyond that of the ideal case. For a cluster size of $N = 5$, Fig. 3 depicts the energy loss induced from the non-ideality of the HDF-based CSS for different voting rules $K = 1, \ldots, 5$ with respect to average BEP of the single hop reporting case. We note that, SUs experience different channel sensing quality of the target PC, so that, SNRs and autocorrelation coefficients of SUs are not identical. In this respect, there are two sets of voting rule results in Fig. 3: one for traditional and other for the proposed K-out of-N rule. The effect of the reporting error on the energy loss for both cases can be observed from Fig. 3. For $K = 3$, for example, energy loss of traditional approach

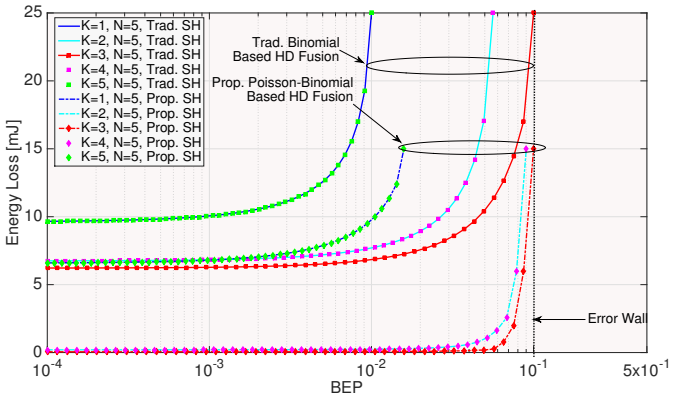

Figure 3: Comparison of energy loss caused from traditional and proposed HDFschemes using single-hop reporting.

is not significantly affected by BEP until 10^{-3}. After that, the energy loss notably increases with the BEP and goes to infinity around 0.1, which is known as the *BEP wall* and its existence has been shown first in [29] for SNR loss under the traditional K-out of-N fusion rule.

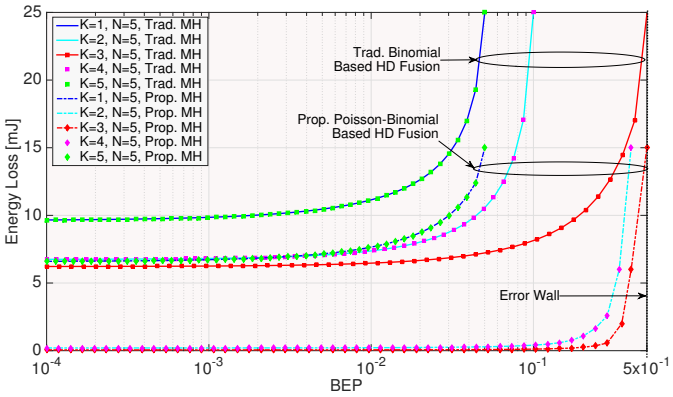

Figure 4: Comparison of energy loss caused from traditional and proposed HDF schemes using proposed CH selection method.

We note that the majority voting rule gives the best performance in terms of the robustness against the energy loss and BEP wall compared to other voting rules. On the orher hand, the proposed method, which takes the sensing and reporting quality heterogeneity of the SUs into account, outperforms the traditional one in terms of the energy loss and the error wall. Therefore, the proposed method with the majority voting rule gives the best performance as can be seen from Fig. 3. Likewise, Fig. 4 shows the performance of the traditional and proposed methods using multi-hop reporting and the best CH selection algorithm given in Section III. It is obvious that the proposed CH selection method increases the robustness of the energy loss against the BEP wall effect from 0.1 to 0.5.

C. CSS with Soft Decision Fusion

For a cluster size of $N = 5$, Fig. 5 depicts the energy loss induced from the non-ideality of the quantized SDF with respect to average BEP of the single hop reporting case. We again note that SUs experience different channel sensing quality of the target PC. To put an evident comparison between hard and soft decision fusion, we also plot the best case of the

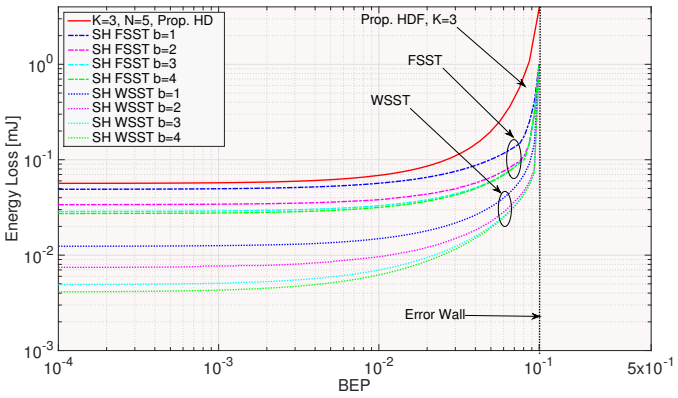

Figure 5: Comparison of energy loss caused from FSST and WSST schemes for different number of bits using single-hop reporting.

HDF, which is the majority voting rule of the proposed method in Fig. 3. For the quantized SDF, we have two different sets of results: one for the FSST approach, and another for the WSST where sensing duration of SUs are weighted based on their sensing quality metrics. We first point out that single bit LM quantizer's FSST energy loss performance is superior to the best case HDF. On the other hand, single bit WSST gives a better energy efficiency than the $4-bit$ FSST scheme. Similar to Fig. 4, the benefit of the proposed multihop reporting based CH selection is apparent in terms of the BEP wall in Fig. 6 where the BEP wall is located at 0.5.

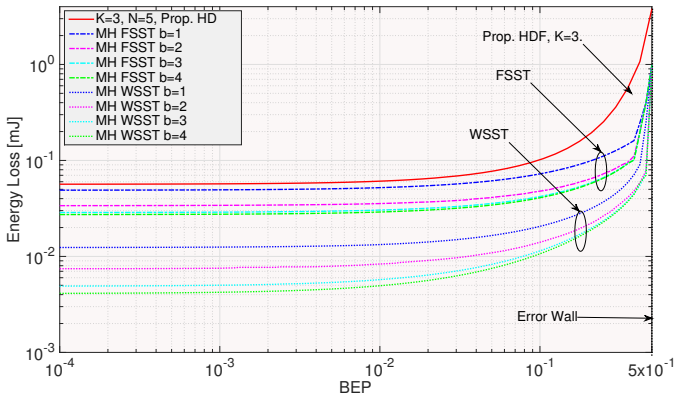

Figure 6: Comparison of energy loss caused from different FSST and WSST schemes using proposed CH selection method.

D. Multi-Objective Clustering Optimization

For the population and generation sizes given in Table III, Fig. 7 shows the results of MOCO with 16 PCs and 250 SUs spread over $1\ km \times 1\ km$ area. In Fig. 7, F_1 represents the total consumed energy, F_2 demonstrates the total maximum sensing duration, F_3 shows the difference between the maximum and minimum energy cost of clusters, and F_4 depicts the difference between the maximum and minimum sensing+reporting duration of clusters. The total energy consumed for multihop reporting is shown in red colors in F_1. We note that the F_3 and F_4 values are primarily determined by the clusters with the maximum energy consumption and the slowest sensing SU with the longest sensing duration, respectively.

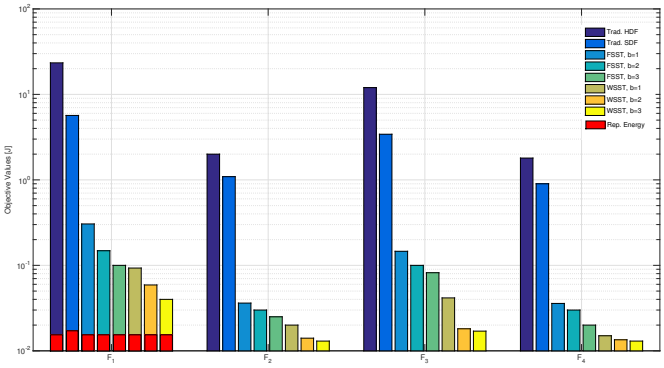

Figure 7: MOCO results for $\mathbf{F}$ and H_1 for different CSS schemes.

While the HDF-based micro perspective converges in an average of $I = 50$ iterations, SDF-based micro perspective converges in an average of $I = 20$ iterations. Neglecting the number of objectives which is a constant, NSGA-II complexity is $\mathcal{O}\left(\mathbb{P}^2\right)$. However, for each population member we are required to determine the CH and sensing parameters which is $\mathcal{O}\left(\mathbb{P}M\left(\bar{C}^2+I\right)\right)$. Thus, the worst case complexity of the MOCO is $\mathcal{O}\left(\mathbb{P}^2+M\mathbb{P}\left(\bar{C}^2+I\right)\right)$.

However, it is worth noting that candidacy approach proposed in Section V greatly reduces the implementation time since the best SUs are set as the initial members for each cluster. Thus, NSGA-II mainly determines the membership of SUs which are potential candidates for more than one cluster. Although repeating the MOCO in a fast fading environment is not practical, the micro perspective can easily be repeated based on sensing and reporting channel changes without recalculating the SU↔cluster association using the MOCO. As explained in Section VI-A, we also do not require repeating the entire CH selection procedure in case of SU inclusion or exclusion.

VII. Conclusions

In this paper, we have focused on energy and throughput efficient clustered CSS from the micro and macro perspectives. By use of the multihop diversity, we have developed a procedure to find the best CH and optimal routing paths from SUs to CH, which is shown to be better in terms of the robustness to the reporting channel imperfection and energy cost. Results has clearly shown that the proposed multihop reporting and CH selection procedure mitigate the BEP wall phenomenon. From the micro perspective, we have revealed that consideration of the sensing and reporting heterogeneity under both HDF and SDF based CSS has a significant impact on the intra-cluster energy consumption and achievable throughput. We have shown that the proposed novel HDF scheme tends to enforce SUs with relatively high SNRs to have a perfect local detection performance while the low SNR SUs' local detection is released. Subject to global detection and false alarm probability constraints, this nature has been shown to be more energy efficient than the traditional one. On the other hand, the proposed WSST method, which decides on SUs' sample sizes proportional to their SNRs, has been shown to have a much more energy efficient performance than the well

known FSST for SDF schemes. For the macro perspective, on the other hand, we have formulated the MOCO problem and have solved it using NSGA-II to obtain network wide fair energy and throughput efficient partitioning of SUs.

APPENDIX A
CONVEXITY ANALYSIS OF HARD DECISION CASE

We will handle the convexity analysis of HD-LLP and HD-ULP in two subsections: A) We prove the parameterized concavity / convexity of $P^d_{n,m}$/$P^f_{n,m}$. B) We show that Q^d_m/Q^f_m is a log-concave function of $\tilde{P}^d_{n,m}$/$\tilde{P}^f_{n,m}$, so that line 2 in HD-LLP and lines 2-3 in HD-ULP are both convex constraints. Throughout the appendix, we omit cluster indices m for the sake of tractability and without loss of generality. To distinguish from each other, we denote the inside expressions of $\mathrm{erfc}(\cdot)$ in (6-7) as follows

$$f(S_n,\varepsilon_n)=\frac{A\varepsilon_n}{\sqrt{S_n}}+B\sqrt{S_n}$$
$$g(S_n,\varepsilon_n,\rho_n)=\frac{1}{1-\rho_n^2}\left(\frac{A\varepsilon_n}{\sqrt{S_n}}+(B-\rho_n)\sqrt{S_n}\right)$$

where $A=\frac{1-\rho_n^2}{2\rho_n}$ and $B=A\log(1-\rho_n^2)+\rho_n$. Since convex composition rules will be extensively exploited in this section, we will remind the readers of these rules as follows [30].

Remark 1: *1) $a(\boldsymbol{x})=b(c(\boldsymbol{x}))$ is convex if b is convex and non-increasing in each argument, and c_i is concave in $\boldsymbol{x}$.*
2) $a(\boldsymbol{x})=b(c(\boldsymbol{x}))$ is concave if b is concave and non-increasing in each argument, and c_i is convex in $\boldsymbol{x}$.

A. Concavity / convexity of P^d_n / P^f_n

Lemma 1: *$P^f_n(f)$ and $P^d_n(g)$ are neither convex nor concave functions of (S_n,ε_n), neither are Q_d and Q_f. Hence, HD-GP is not a convex optimization problem.*

Proof: Although $\mathrm{erfc}(\cdot)$ is a non-increasing convex and concave function for cases $\mathrm{erfc}(\cdot)\le 0.5$ and $\mathrm{erfc}(\cdot)\ge 0.5$, respectively, $\mathrm{erfc}(f)$/$\mathrm{erfc}(g)$, f/g must be concave/convex in (S_n,ε_n). However, we show that f and g are neither convex nor concave in (S_n,ε_n), since the Hessian matrix of f in (26) and that of g in (27) are neither positive nor negative semi-definite as follows

$$\bigtriangledown^2 f=S_n^{-3/2}\begin{bmatrix}\frac{3A\varepsilon}{4S_n}-\frac{B}{4} & -\frac{A}{2}\\ -\frac{A}{2} & 0\end{bmatrix} \tag{26}$$

$$\bigtriangledown^2 g=\frac{S_n^{-3/2}}{1-\rho^2}\begin{bmatrix}\frac{3A\varepsilon}{4S_n}-\frac{B-\rho}{4} & -\frac{A}{2}\\ -\frac{A}{2} & 0\end{bmatrix} \tag{27}$$

Therefore, $P^f_n(f)$ and $P^d_n(g)$ are neither convex nor a concave function of (S_n,ε_n). Thus, this result directly affects the convexity / concavity of Q_f, Q_d, and HD-GP. ∎

Lemma 2: *This Lemma provides the basis for convexity of parameterized approach used in bilevel optimization.*
(a) For a parameterized S, $\bar{S}_n$, f and g are both linear functions of ε_n. Thus, $P^f_n(f)\le 0.5$ / $P^d_n(g)\ge 0.5$ is a convex / concave function of ε_n.
(b) For a parameterized ε, $\bar{\varepsilon}_n$, f is a concave function of S_n if $S_n\ge 3A\varepsilon_n/B$, and a convex function of S if $S\le {}^{3A\varepsilon_n}/_{B-\rho_n}$. Thus, $P^f_n(f)\le 0.5$ / $P^d_n(g)\ge 0.5$ is a convex / concave function of S_n for $\frac{3A\varepsilon_n}{B-\rho_n}\ge S_n\ge 3A\varepsilon_n/B$.

Proof:

(a) Because of $\frac{\partial^2 f}{\partial\varepsilon_n^2}=\frac{\partial^2 g}{\partial\varepsilon_n^2}=0$, f and g are both convex and concave functions of ε_n, i.e., linear. Since $\mathrm{erfc}(\cdot)\le 0.5$ is a non-increasing convex function of f and f is a concave function of ε_n, their composition $P^f_n(f)\le 0.5$ is also convex in ε_n. Similarly, since $\mathrm{erfc}(\cdot)\ge 0.5$ is a non-increasing concave function of g and g is a convex function of ε_n, their composition $P^d_n(g)\ge 0.5$ is also concave in ε_n.

(b) By using the same composition rules in (a), we need $\frac{\partial^2 f}{\partial S_n^2}\le 0$ to assure the concavity of f in S_n. Noting that the sensing duration cannot be negative (i.e., $S_n\ge 0$), this condition reduces to $S_n\ge 3A\bar{\varepsilon}_n/B$. Likewise, we need $\frac{\partial^2 g}{\partial S_n^2}\ge 0$ to assure the convexity of g. Due to $S_n\ge 0$, this condition reduces to $\frac{3A\bar{\varepsilon}_n}{B-\rho_n}\ge S_n$.

As a result of (a) and (b), line 2 in HD-LLP and lines 3-4 in HD-ULP are convex constraints. ∎

Lemma 3: *Results of Lemma 2 also hold for $\tilde{P}^f_n$ and $\tilde{P}^d_n$.*

Proof: In (17-18), $\tilde{P}^d_n$ and $\tilde{P}^f_n$ are non-negative weighted sum of P^d_n and P^f_n, respectively. Since non-negative weighted summation preserves convexity, Lemma 3 is an immediate result of Lemma 2. ∎

B. Log-concavity of Q_d and Q_f

Sum of independent *Bernoulli* variables forms a log-concave random variable since log-concavity is closed under convolution [31]. Consequently, $\mathcal{K}_{hd}$ given in (19) is log-concave under both homogeneity and heterogeneity. Since log-concavity is preserved by integration [30], cumulative distribution function (*cdf*), $\mathcal{P}(\mathcal{K}_{hd}\le\kappa_{hd})$, and survivability function (*sf*), $\mathcal{P}(\mathcal{K}_{hd}\ge\kappa_{hd})$, of a log-concave random variable are also log-concave functions. However, this is true with respect to κ_{hd} which is a parameter in our case. Therefore, we need to analyze the log-concavity of these functions with respect to $\tilde{P}^d_n$ and $\tilde{P}^f_n$ to establish connection to concavity in (S_n,ε_n).

Lemma 4: *Pmf of a Binomial random variable is a log-concave function of $\tilde{P}^d$ / $\tilde{P}^f$, so are its cdf and sf.*

Proof: Without loss of generality, we show the proof for $\tilde{P}^d$, which can be repeated for $\tilde{P}^f$. The logarithm of *Binomial pmf* and its second derivative are given as

$$\log\left(P\left(\mathcal{K}_{hd}=i\right)\right)=\log\left[\binom{C}{i}\right]+i\log(\tilde{P}^d)+(C-i)\log(1-\tilde{P}^d)$$

$$\frac{\partial^2\log\left(P(\mathcal{K}_{hd}=i)\right)}{\partial(\tilde{P}^d)^2}=-\frac{i}{(\tilde{P}^d)^2}-\frac{C-i}{(1-\tilde{P}^d)^2},\ 1\le i\le C$$

where the second derivative is always non-positive since $C\ge i$, hence, $\log(P(\mathcal{K}_{hd}=i))$ is a log-concave function of $\tilde{P}^d$. Since log-concavity is preserved by integration as indicated above, *cdf* and *sf* of $\mathcal{K}_{hd}$ are also log-concave functions. ∎

Lemma 5: *Pmf of a Poisson-Binomial random variable is a log-concave function of $\tilde{\boldsymbol{P}}^{\boldsymbol{d}}_{\boldsymbol{n}}$ / $\tilde{\boldsymbol{P}}^{\boldsymbol{f}}_{\boldsymbol{n}}$, so are its cdf and sf.*

Proof: Fernandez et. al. show that the distribution of *Poisson-Binomial* random variable is given by the following probability generating function [24]

$$A_0+A_1z+A_2z^2+\cdots+A_Cz^C=\alpha(z-r_1)(z-r_2)\cdots(z-r_C)$$

where $\alpha=\prod_{n=1}^{C}\tilde{P}_n^d$, $r_n=-(1-\tilde{P}_n^d)/\tilde{P}_n^d$, and polynomial coefficients on the left hand side represent the *pmf* in terms of the real roots on the right hand side, i.e., $A_i=\mathcal{P}(\mathcal{K}_{hd}=i)$. First, we note that the coefficients of a polynomial with real negative roots are log-concave functions of the roots [31], which is the case here since $r_n<0,\ 1\leq n\leq C$. Second, r_n is an increasing (non-decrasing) and strictly concave function of $\tilde{P}_n^d$ since $\partial^2 r_n/\partial(\tilde{P}_n^d)^2=-2(\tilde{P}_n^d)^{-3}$ is always negative and r_n increases as $\tilde{P}_n^d$ increases. Contingent upon Lemma 3 and Remark 1, r_n is a concave function of $\tilde{P}_n^d$. Third, α is obviously a linear function of $\tilde{P}_n^d$. Combining these three steps proves the log concavity of the *Poisson-Binomial pmf* with respect to $\tilde{P}_n^d$. Finally, log-concavity of *Poisson-Binomial cdf* and *sf* again follow from the integration property of log-concave functions. ■

Lemma 6: *Binomial and Poisson-Binomial cdf / sf is an increasing / decreasing function of $\tilde{P}_n^d$ and $\tilde{P}_n^f$.*

Proof: In above polynomial, coefficient A_{C-k} is the sum of products of k roots such that

$$A_{C-k}=(-1)^k\sum_{1\leq j_1,j_2,\ldots,j_k\leq C} r_{j_1}r_{j_2}\ldots r_{j_k}$$

where negative sign of roots are eliminated for both odd and even numbers of k, thus, coefficients are decreasing functions of $\tilde{P}_n^d$, so are *pmf*s. Based on this, *cdf* is a decreasing function of $\tilde{P}_n^d$, hence, tail *cdf* is an increasing function of $\tilde{P}_n^d$. Proof for Binomial distribution is a special case of Poisson-Binomial. ■

Consequently, based on Lemmas 2 and 3, constraints $-\boldsymbol{P_d}\preceq-0.5$ and $\boldsymbol{P_f}\preceq 0.5$ are convex. Furthermore, using Lemmas 2-5, $-\log(Q_d)\leq-\log(Q_{th}^d)$ and $|\log(Q_{th}^f)-\log(Q_f)|$ are also convex.

REFERENCES

[1] F. C. Commission *et al.*, "Spectrum policy task force report, fcc 02-155," 2002.

[2] G. Gur and F. Alagoz, "Green wireless communications via cognitive dimension: an overview," *Network, IEEE*, vol. 25, no. 2, pp. 50–56, 2011.

[3] M. Webb *et al.*, "Smart 2020: Enabling the low carbon economy in the information age," *The Climate Group. London*, vol. 1, no. 1, pp. 1–1, 2008.

[4] I. F. Akyildiz *et al.*, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.

[5] S. Eryigit, G. Gur, S. Bayhan, and T. Tugcu, "Energy efficiency is a subtle concept: fundamental trade-offs for cognitive radio networks," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 30–36, 2014.

[6] S. Huang, H. Chen, Y. Zhang, and F. Zhao, "Energy-efficient cooperative spectrum sensing with amplify-and-forward relaying," *IEEE Communications Letters*, vol. 16, no. 4, pp. 450–453, 2012.

[7] R. Deng, J. Chen, C. Yuen, P. Cheng, and Y. Sun, "Energy-efficient cooperative spectrum sensing by optimal scheduling in sensor-aided cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 716–725, 2012.

[8] S. Wang, Y. Wang, J. P. Coon, and A. Doufexi, "Energy-efficient spectrum sensing and access for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 906–912, 2012.

[9] D. Huang, G. Kang, B. Wang, and H. Tian, "Energy-efficient spectrum sensing strategy in cognitive radio networks," *IEEE Communications Letters*, vol. 17, no. 5, pp. 928–931, 2013.

[10] A. E. et. al., "Sensor selection and optimal energy detection threshold for efficient cooperative spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1565–1577, 2015.

[11] C. Mousavifar, S.A. ; Leung, "Energy efficient collaborative spectrum sensing based on trust management in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 1927–1939, 2015.

[12] S. Chaudhari *et al.*, "Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?" *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 18–28, 2012.

[13] N. Nguyen-Thanh and I. Koo, "Log-likelihood ratio optimal quantizer for cooperative spectrum sensing in cognitive radio," *IEEE Communications Letters*, vol. 15, no. 3, pp. 317–319, 2011.

[14] C.-h. Lee and W. Wolf, "Energy efficient techniques for cooperative spectrum sensing in cognitive radios," in *IEEE CCNC*. IEEE, 2008, pp. 968–972.

[15] N. Nguyen-Thanh and I. Koo, "A cluster-based selective cooperative spectrum sensing scheme in cognitive radio," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–9, 2013.

[16] A. S. Kozal, M. Merabti, and F. Bouhafs, "Spectrum sensing-energy tradeoff in multi-hop cluster based cooperative cognitive radio networks," in *proc. IEEE INFOCOM*. IEEE, 2014, pp. 765–770.

[17] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[18] E. Axell *et al.*, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[19] S. Chaudhari *et al.*, "Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2690–2700, 2009.

[20] IEEE Standards Association *et al.*, "802.11-2012-ieee standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications," 2012.

[21] V. Rodoplu and T. H. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1333–1344, 1999.

[22] A. Celik and A. E. Kamal, "Multi-objective clustering optimization for multi-channel cooperative sensing in crns," in *IEEE GLOBECOM*. IEEE, 2014, pp. 3441–3446.

[23] Y. H. Wang, "On the number of successes in independent trials," *Statistica Sinica*, vol. 3, no. 2, pp. 295–312, 1993.

[24] M. Fernandez and S. Williams, "Closed-form expression for the poisson-binomial probability density function," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 2, pp. 803–817, 2010.

[25] C. Jiang *et al.*, "Energy-efficient non-cooperative cognitive radio networks: micro, meso, and macro views," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 14–20, 2014.

[26] G. Casella and R. L. Berger, *Statistical inference*. Duxbury Pacific Grove, CA, 2002, vol. 2.

[27] K. Deb *et al.*, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[28] Y.-C. Liang *et al.*, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, 2008.

[29] S. Chaudhari *et al.*, "Bep walls for cooperative sensing in cognitive radios using k-out-of-n fusion rules," *Signal Processing*, vol. 93, no. 7, pp. 1900–1908, 2013.

[30] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[31] O. Johnson and C. Goldschmidt, "Preservation of log-concavity on summation," *ESAIM: Probability and Statistics*, vol. 10, pp. 206–215, 2006.