Physical Layer Data Fusion via Distributed Co-Phasing with General Signal Constellations

Manesh A., Chandra R. Murthy, *Senior Member, IEEE*, and Ramesh Annavajjala, *Senior Member, IEEE*

***Abstract*—This paper studies the design and analysis of a pilot-assisted physical layer data fusion technique known as Distributed Co-Phasing (DCP). In this two-phase scheme, sensors first estimate the channel to the fusion center (FC) using pilot symbols sent by the latter; and then they simultaneously transmit their common data symbols by pre-rotating them by the estimated channel phase, thereby achieving physical layer data fusion. First, by analyzing the symmetric mutual information of the system, it is shown that the use of higher order constellations can significantly improve the throughput performance of DCP in comparison with binary signaling considered heretofore. Using a higher order constellation in the DCP setting requires the estimation of the composite DCP channel at the FC for data decoding. To this end, two blind algorithms are proposed: 1) power method, and 2) modified K-means algorithm. The latter algorithm is shown to be computationally efficient and converges significantly faster than the conventional K-means algorithm. Analytical expressions for the probability of error are derived, and it is found that even at moderate to low SNRs, the proposed modified K-means algorithm achieves a probability of error comparable to that achievable with perfect channel estimate at the FC, while requiring no pilot symbols to be transmitted from the sensor nodes. Also analyzed is the problem of signal corruption due to imperfect DCP, and constellation shaping to minimize the probability of signal corruption is proposed and analyzed. The analysis is validated and the promising performance of DCP for energy-efficient physical layer data fusion is illustrated using Monte Carlo simulations.**

***Index Terms*—Distributed co-phasing, K-means algorithm, sensor networks, data fusion, mutual information, Nakagami-m fading.**

I. Introduction

Wireless sensor networks (WSN) are spatially distributed sensors which are deployed to accomplish some specialized task. To this end, they typically pass their information (e.g., data and/or channel state) among themselves, or cooperatively transmit some common information to a *fusion center* (FC) [2]. In the latter context, Distributed Co-Phasing (DCP) has recently been proposed as an energy-efficient physical layer technique for data fusion [3], [4]. In DCP, the phase of the transmitted signal from the sensor nodes are adjusted to ensure that the signals combine coherently at the FC. This leads to both array gains as well as diversity benefits, resulting in a lower probability of error, lower required transmit power, increased communication range, robustness to node failures, etc. In pilot-assisted DCP, a pilot signal is first sent from the FC to all the sensor nodes, using which the nodes estimate the wireless channel. In a Time Division Duplex (TDD) system with a quasi-static and reciprocal channel, the nodes can exploit the channel state information (CSI) obtained from the pilot signal to cooperatively transmit their common information to the FC. An interesting feature of this scheme is that with uncoded BPSK modulation, CSI is not required at the FC for data detection. Hence, there is no need to transmit any pilot symbols from the power-starved sensor nodes prior to data transmission. We note that transmission of pilots from the sensors to the FC is inefficient, as it must necessarily occur in a round-robin fashion. Also, since the transmit power of the sensors is typically low, multiple pilot symbols may be necessary for accurate channel estimation at the FC. On the other hand, the FC could be connected to the mains, and have much better power resources at its disposal to transmit high-power pilots, thereby facilitating accurate channel estimation with only a small number of pilot symbols. Hence, in this work, we consider the pilot-assisted DCP scheme described above.

DCP has been investigated by various researchers in the context of wireless sensor networks. Coherent communications from multiple antennas to a distant stationary antenna was perhaps first studied in [5], for AWGN channels. The authors proposed a master-slave architecture to achieve clock synchronization across distributed antennas. The feasibility of DCP in fading channels, and practical issues such as carrier and symbol synchronization under a master-slave architecture were further studied in [3]. Carrier phase synchronization between two transmitting nodes was practically demonstrated in [6], where the authors proposed a consensus-based algorithm for achieving global frequency synchronization. Other studies on distributed phase synchronization across a wireless sensor network include [7]–[13]. We also note that, while channel estimation has to be performed at least once per channel coherence duration, synchronization needs to be repeated only when sensors fall out of sync. The latter could be much longer than the channel coherence duration. Consequently, in our work, we assume that such phase synchronization schemes are employed by the sensors and FC to achieve the cophased transmission; and we do not consider the cost of synchronization in our analysis. Information sharing among nodes in a wireless sensor network was studied in [14] and [15], where distributed transmit beamforming (DTB) was employed as the underlying

Manesh A. is with the Naval Physical and Oceanographic Laboratory, Defense Research and Development Organization, Kochi, 682 012, India. He was with the Dept. of Electrical Communication Engineering (ECE), Indian Institute of Science (IISc), Bangalore 560 012, India, during the course of this work. Email: a.manesh@gmail.com. C. R. Murthy is with the Dept. of ECE, IISc Bangalore, 560 012, India. Email: cmurthy@ece.iisc.ernet.in. R. Annavajjala is with the Charles Stark Draper Laboratory, Cambridge, MA, USA. Email: ramesh.annavajjala@gmail.com.

This work was presented in part in [1].

communication scheme. The distribution of the beam pattern with DTB using uniform and Gaussian distributed sensor nodes was studied in [16] and [17]. In [18], the authors studied the use of DTB for communication between a multiple antenna base station and a single antenna mobile receiver. The impact of imperfect channel estimation on the BER performance of transmit beamforming was studied in [19]. A comparative study of DCP with other co-phasing transmission schemes such as maximum ratio transmission (MRT) and truncated channel inversion (TCI), in terms of performance metrics such as the SNR and BER with BPSK modulation was presented in [20]. However, the DCP schemes considered in [20], by design, were limited to using BPSK (or M-PSK) modulation as the signal constellation. This is because knowledge of CSI at the FC, which is required for data detection, is not directly available, as no pilots are sent from the sensors to the FC. Another drawback of the DCP scheme studied in [20] is the issue of channel corruption. Due to the channel estimation errors at the nodes, the co-phasing is imperfect, and the phase angle of the effective DCP channel could exceed half the rotational symmetry of the constellation (e.g., $\frac{\pi}{2}$ with BPSK signaling). When this happens, it results in catastrophic detection errors at the FC. In this paper, we overcome both these drawbacks.

In this work, we start by analyzing the mutual information performance of the DCP schemes with higher order constellations, and show that higher order constellations far outperform BPSK at moderate to high SNRs, even in the absence of CSI at the FC. This motivates one to explore methods within the framework of DCP that can support non-constant modulus constellations. This, however, requires CSI at the FC in order to set the decision boundaries for coherent data detection. Channel knowledge at the CSI can be acquired either by sending a pilot symbol from all sensors to the FC (reverse training) or by the FC estimating the CSI from the received data itself, i.e., blind channel estimation. In a typical WSN, the nodes are battery-power driven, and have limited resources for computation and communication. It is important, therefore, for the design to account for both the time and power overhead incurred in the training process. In particular, transmitting high-power pilots or multiple pilot symbols on the reverse link for channel estimation at the FC are both undesirable, as they could leave very low power or time resources, respectively, for data transmission. The FC, on the other hand, is typically connected to the power mains and has significantly better resources at its disposal. As will be shown in this paper, our proposed computationally efficient algorithm for *blind* channel estimation at the FC works well even at low data SNR, thus obviating the need for spending valuable resources of the sensor nodes on reverse channel training. For example, our simulation results show that with $N = 10$ sensors, and with 16 QAM modulation, just 20 data symbols are sufficient to achieve a performance very close to that of a genie-aided receiver with perfect CSI at the FC, even at an SNR as low as -10 dB. This provides a strong motivation for adopting blind channel estimation methods as proposed in this paper.

Our main contributions are as follows:

- We analyze the mutual information achievable by the DCP scheme. The analysis serves to underline the benefits of higher order constellations over the binary transmission, as the SNR increases.
- We propose two blind methods to estimate the effective channel at the FC.
- We analyze the symbol error rate performance at the FC. Our analysis is based on a Nakagami-m approximation of the composite channel formed after imperfect DCP from the sensors based on the estimated channel at the individual sensor nodes. The analysis in [20] was based on an improved Gaussian approximation [21], which is valid only at very low data SNRs, and only for M-PSK modulation. In contrast, the analysis in this paper is very accurate for a wide range of data and pilot SNRs, as will be shown through simulations.
- We explore the use of asymmetric signal constellations as a remedy to the problem of channel corruption, which is a major cause of error at low pilot SNRs. The asymmetric constellations also require channel estimation at the FC, for which we repurpose our proposed blind methods. We also extend our probability of error analysis to asymmetric signal constellations, and show that asymmetric signaling can outperform symmetric constellations with the same cardinality at low pilot SNRs.
- We corroborate our analytical expressions with Monte Carlo simulations. The results (see Sec. VII) also provide useful and interesting insights on the relative effects of the operating parameters such as the data SNR, pilot SNR, and number of sensors, on the performance of DCP.

Our conclusion is that DCP, along with the proposed modified K-means based blind channel estimation method, is a promising candidate for reliable and energy-efficient physical layer data fusion from a given set of sensor nodes. Furthermore, the analytical expressions derived in this work can be of independent interest, for example, in studying the effect of channel estimation errors on the performance of equal gain combining receivers.

The rest of the paper is organized as follows. In section II, we introduce our system model. In section III, we derive the mutual information between the sensors' inputs and the FC output, to show that there is a performance improvement in using a higher order constellation compared to BPSK. Section IV proposes two blind channel estimation methods for using DCP with higher constellations. Section V analyzes the SER performance of the proposed schemes, and Section VI compares the performance of symmetric and asymmetric signal constellations in the DCP setup. Section VII presents the Monte Carlo simulation results, and finally, Section VIII concludes the paper. Some of the detailed mathematical derivations are relegated to the Appendix.

II. System Model

We consider N sensors that share some common information, for example, a packet of data, that they wish to cooperatively transmit to an FC. Such a common information can be acquired by the sensors by observing a correlated random field and then exchanging information to arrive at consensus. In this work, we concentrate on the sensors-to-FC communication link, its design and performance analysis. The baseband equivalent complex-valued channel from the k^{th} sensor to the FC is denoted by $g_k =$

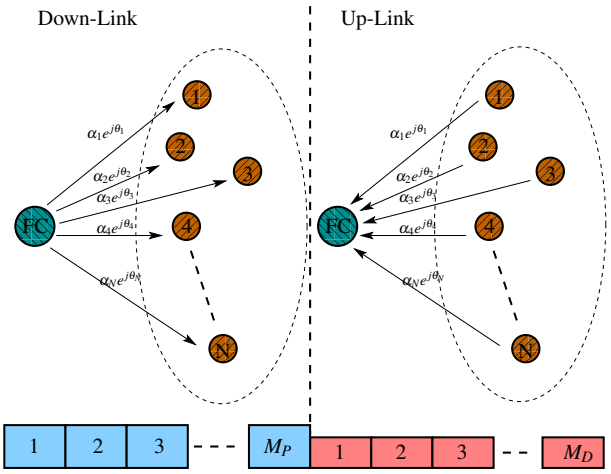

Fig. 1. Two phases of DCP in a wireless sensor network.

$\alpha_k e^{j\theta_k}$, where the channel magnitudes α_k are independent Rayleigh distributed and θ_k are i.i.d. uniform random variables (r.vs), $1 \leq k \leq N$. The channel is assumed to be quasi-static and block fading, i.e., it remains constant for the coherence time of duration M_T symbols, and changes in an i.i.d. fashion from block to block.

We consider a TDD system, and a two stage communication protocol between the sensors and FC, as illustrated in Fig. 1. In the first stage, the FC transmits M_P known pilot symbols to the sensors, where $1 \leq M_P \leq M_T$. The sensor nodes use the pilot symbols to estimate their respective channels to the FC, by reciprocity[1] [20]. In the second stage, the sensors simultaneously transmit their common data to the FC, for a duration of $M_D = M_T - M_P$ symbols. The sensors attempt to coherently combine their signals at the FC by pre-rotating the data by their respective estimated channel phase angles.

The received signal at the k^{th} sensor during the first stage of pilot transmission from the FC to the sensors is

$$r_k[n] = g_k\sqrt{E_P} + \eta_k[n], \quad n = 1, \ldots, M_P, \tag{1}$$

where M_P is the number of pilot symbols, E_P is the pilot power, $\eta_k[n]$ is a complex Gaussian r.v with zero mean and variance per dimension $N_0/2$ and g_k is the complex channel gain between the k^{th} sensor and the FC.

From this, the ML estimate of the channel phase can be obtained at the k^{th} sensor node as [20]

$$\widehat{\theta}_k = \tan^{-1}\left(\frac{\Im\{\frac{1}{M_P}\sum_{n=1}^{M_P} r_k[n]\}}{\Re\{\frac{1}{M_P}\sum_{n=1}^{M_P} r_k[n]\}}\right) \tag{2}$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real and imaginary parts of the argument, respectively.

Note that sensor k uses its estimate $\widehat{\theta}_k$ to compensate for the phase offset introduced by the channel to the FC. With ideal phase estimation, i.e., when $\widehat{\theta}_k = \theta_k$, the signals from the sensors combine coherently at the FC, leading to an improved probability of error performance due to the diversity advantage of the combining scheme. The received signal at the FC during the second phase of DCP data transmission is

$$r[n] = \sum_{k=1}^{N} x_k[n]e^{-j\widehat{\theta}_k}g_k + v[n], \quad n = M_P+1, \ldots, M_T, \tag{3}$$

where $x_k[n]$ is the data symbol from sensor k at time n and $v[n]$ is a complex Gaussian r.v with zero mean and variance per dimension $N_0/2$. Since all sensors have the same symbol to transmit to the FC, i.e., $x_k[n] = x[n]$, (3) can be written as

$$r[n] \;\; = \;\; x[n]H_{\text{DCP}} + v[n], \quad n = M_P + 1, \ldots, M_T. \tag{4}$$

In the above, H_{DCP} is the effective DCP channel, defined as

$$H_{\text{DCP}} \triangleq \sum_{k=1}^{N} \alpha_k e^{j\theta_{e,k}} = |H_{\text{DCP}}|e^{j\phi_H} \tag{5}$$

where $|H_{DCP}|$ and ϕ_H are the magnitude and phase of the effective channel, respectively, and $\theta_{e,k} \triangleq \widehat{\theta}_k - \theta_k$ is the phase estimation error at the k^{th} sensor node.

Note that, when the channel is not estimated at the FC, the ML detection of a constant modulus constellation such as BPSK involves simply comparing the real part of the received signal to zero [20]. In the next section, using a Mutual Information (MI) based analysis, we show that one can achieve a larger throughput if a higher order constellation is used instead of BPSK, at moderate-to-high SNRs. This motivates one to explore methods within the framework of DCP that can support higher order constellations.

III. Mutual Information Analysis of the DCP Scheme

In this section, we compute the MI between the sensors and the FC, and use it to show that a larger constellation size improves the achievable data rate as the SNR increases. For simplicity, we consider equi-probable constellation symbols for which the MI is also referred to as the symmetric MI. The MI of the system model in (4) can be evaluated with and without the channel knowledge at the FC, as follows [22]:

- With CSI at the FC:

$$I(x; r|H_{\text{DCP}}) = \mathbb{E}_{H_{\text{DCP}}}\left\{\sum_{i=1}^{M}\frac{1}{M}\int_{r\in\mathbb{C}} p(r|H_{\text{DCP}}, x_i) \log\frac{p(r|H_{\text{DCP}}, x_i)}{p(r|H_{\text{DCP}})}dr\right\} \tag{6}$$

  where $I(x; r|H_{\text{DCP}})$ is used to denote the MI with perfect CSI at the FC, x_i is the i^{th} element of the transmitted constellation of cardinality M, $\mathbb{C}$ represents complex field and $\mathbb{E}_{H_{\text{DCP}}}$ computes the expectation with respect to the distribution of the effective DCP channel H_{DCP}. Note that $p(r|H_{\text{DCP}}, x_i)$ is simply a circularly symmetric complex Gaussian distribution with mean $H_{\text{DCP}}x_i$ and variance $N_0/2$ per dimension. Also, due to equi-probable constellation symbols, $p(r|H_{\text{DCP}}) = \sum_{m=1}^{M}\frac{1}{M}p(r|H_{\text{DCP}}, x_m)$.
- Without CSI at FC:

$$I(x; r) = \sum_{i=1}^{M}\frac{1}{M}\int_{r\in\mathbb{C}} p(r|x_i)\log\frac{p(r|x_i)}{p(r)}dr. \tag{7}$$

[1]Note that channel reciprocity requires well calibrated radio-frequency chains, which is assumed here.

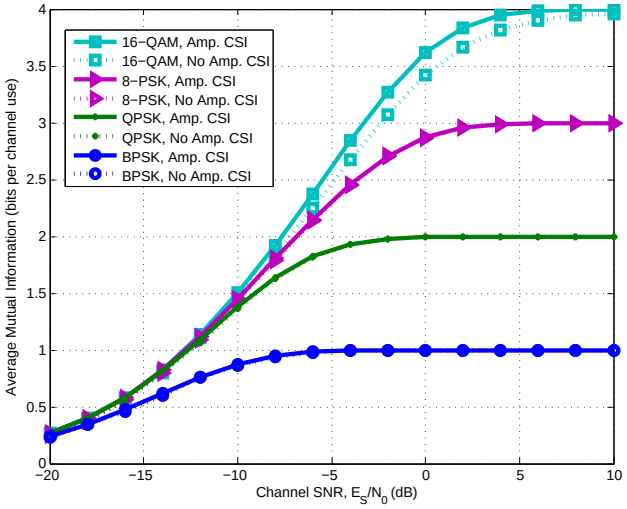

Fig. 2. Comparison of mutual information for ideal DCP with no CSI and perfect CSI for different constellations, with $N = 5$ sensors.

where $I(x;r)$ is used to denote the MI without CSI at the FC, $p(r|x_i) \triangleq \mathbb{E}_{H_{\text{DCP}}}\{p(r|x_i, H_{\text{DCP}})\}$ and $p(r) \triangleq \mathbb{E}_{H_{\text{DCP}}}\{p(r|H_{\text{DCP}})\} = \sum_{i=1}^{M} \frac{1}{M} p(r|x_i)$.

The above MI expressions in (6) and (7) are hard to evaluate in closed form, hence we compute them using Monte Carlo simulations.

In Fig. 2, we plot the MI as a function of the transmit energy per sensor (in dB, relative to the noise power spectral density), for various signal constellations. We see that, as the SNR increases, higher order constellations perform better than BPSK. Also, there is only a marginal loss in MI due to the absence of receiver CSI. This is because the effective channel phase ϕ_H has zero mean, and, at reasonable pilot SNRs, its distribution concentrates around the mean. As the number of nodes increases, there is a channel hardening effect, so that the instantaneous channel is often close to the average channel. Due to this, the absence of CSI at the FC results in only a small loss in MI. This suggests that one could possibly achieve a performance close to the perfect CSI case with the help of a blind demodulator at the FC.

IV. DCP for Non-Constant Modulus Constellations

As seen in the previous section, higher-order constellations can improve the spectral efficiency of a DCP system. However, demodulation of a non-constant modulus constellation such as 16-QAM requires knowledge of the effective DCP channel, H_{DCP}, at the FC, in order to determine the detection thresholds, or, in the coded case, the bit log-likelihood ratios. Assuming all the data symbols to be equally likely, the Maximum Likelihood (ML) estimate of the CSI can be obtained at the FC using the received data symbols as

$$\widehat{H}_{\text{DCP,ML}} = \arg\max_{H\in\mathbb{C}} \frac{1}{(M\pi\sigma^2)^{M_D}} \sum_{\mathbf{x}\in\mathcal{X}^{M_D}} e^{-\sum_{k=1}^{M_D} \frac{|r[k]-Hx_k|^2}{\sigma^2}} \quad (8)$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_{M_D}]$ is the transmitted message in the M_D data symbols, the set $\mathcal{X}$ is the signal constellation, and M is the constellation size. As obtaining a closed-form expression for the ML estimate is hard, we propose two computationally simple methods for channel estimation at the FC: 1) A power-based method, and 2) a modified K-means algorithm. We elaborate upon these methods in the following subsections.

A. The Power Method

Note that, since the channel does not change for the duration of M_D symbols, it follows from (4) that

$$\mathbb{E}\{|r|^2\} = |H_{\text{DCP}}|^2 E_S + N_0, \quad (9)$$

where E_S is the average power in the signal transmitted by the individual sensors. Using the weak law of large numbers to compute $\mathbb{E}\{|r|^2\}$ in (9), we propose to use the following channel estimate:

$$|\widehat{H}_{\text{DCP}}| = \left[\frac{\frac{1}{M_D}\sum_{n=1}^{M_D} |r[n]|^2 - N_0}{E_S}\right]^{\frac{1}{2}}. \quad (10)$$

It can be seen that the above estimate satisfies $\mathbb{E}\{|\widehat{H}_{\text{DCP}}|^2\} = |H_{\text{DCP}}|^2$. Notice that, when the pilot power from the FC is relatively high, and correspondingly, when the phase error in the DCP is low, $|\widehat{H}_{\text{DCP}}|$ closely approximates H_{DCP}, and hence is sufficient for data detection at the FC.

B. The Modified K-means Algorithm

Here, we use the K-means clustering algorithm [23] to perform joint channel estimation and data detection at the FC. The joint ML estimate of the effective DCP channel and the transmitted symbols can be written as

$$\left(\widehat{H}_{\text{DCP,ML}}, \widehat{\mathbf{x}}_{\text{ML}}\right) = \arg\max_{H\in\mathbb{C}, \mathbf{x}\in\mathcal{X}^{M_D}} \frac{1}{(\pi\sigma^2)^{M_D}} e^{-\sum_{k=1}^{M_D} \frac{|r[k]-Hx_k|^2}{\sigma^2}}. \quad (11)$$

The K-means algorithm is an iterative method that involves two steps: (a) a *nearest neighbor step*, which partitions a given set of vectors into K groups such that their squared distances to K centroids are minimized, and (b) a *centroid update* step, which updates the centroids based on the given partition. In our context, the received signal constellation constitute the centroids, and the received data signal points constitute the data to be decoded. Hence, the K-means clustering algorithm can be used to perform the function of the joint ML estimator.

In the following, we present a modified K-means algorithm that achieves more accurate and faster convergence than the original K-means algorithm. The motivation for the modified algorithm stems from the fact that the K-means algorithm is designed to operate on arbitrarily located centroids. However, in the case of DCP, centroids of the received data at the FC are a scaled version of the transmitted constellation, and the constellation itself is known to the FC. We can use this property to improve the centroid update step in the original K-means algorithm. Instead of updating the centroids individually, we seek to find an optimum scaling factor β that minimizes the mean squared error between the received data and a scaled version of the transmitted constellation, as follows:

$$\widehat{\beta} = \arg\min_{\beta} J(\beta) \quad (12)$$

TABLE I
THE MODIFIED K-MEANS ALGORITHM

Input: received data $r[n]$
Outputs: channel estimate, demodulated data.
Initialize $J(\beta_{\text{old}}) = \infty$
Initialize β_{new} using power method; set $J(\beta_{\text{new}}) = 10^6$
while $J(\beta_{\text{old}}) - J(\beta_{\text{new}}) > \epsilon$
Nearest neighbor-based classification of data points:
$R(i) = \{r[j] : \lvert r[j] - \beta_{\text{new}}\ s_i\rvert^2 \leq \lvert r[j] - \beta_{\text{new}}\ s_l\rvert^2, \forall\ l \neq i\}$
Set $M_i = \lvert R(i)\rvert$
Set $\beta_{\text{old}} = \beta_{\text{new}}$ and $J(\beta_{\text{old}}) = J(\beta_{\text{new}})$
Centroid step:
$\beta_{\text{new}} = \frac{\sum_{k=1}^{M} s_k^* \sum_{l=1}^{M_k} r_{lk}}{\sum_{k=1}^{M} M_k \lvert s_k\rvert^2}$
Compute $J(\beta_{\text{new}}) = \sum_{k=1}^{M} \sum_{l=1}^{M_k} \lvert\beta_{\text{new}} s_k - r_{lk}\rvert^2$
end while

where $J(\beta) \triangleq \sum_{k=1}^{M} \sum_{l=1}^{M_k} |\beta s_k - r_{lk}|^2$ is the squared error metric, s_k is the k^{th} point in the signal constellation, M_k is the number of points in the k^{th} group, and r_{lk} is the l^{th} point (received signal) in the k^{th} group.

Since $J(\beta)$ is a convex function of β, taking the derivative of $J(\beta)$ with respect to β, setting to zero and solving, we get the optimum value of β for a given partition of received points into the M constellation points as

$$\widehat{\beta} = \frac{\sum_{k=1}^{M} s_k^* \sum_{l=1}^{M_k} r_{lk}}{\sum_{k=1}^{M} M_k |s_k|^2}. \tag{13}$$

A reasonable initialization for β is $\widehat{\beta} = \widehat{H}_{\text{DCP}}$, where $\widehat{H}_{\text{DCP}}$ is the estimate of the channel obtained using the power method in the previous subsection. The modified K-means algorithm is summarized in Table I.

The convergence of the modified K-means algorithm to a local minimum follows directly from the convergence properties of the original K-means algorithm [24]. As a consequence of exploiting the signal constellation structure, the modified K-means algorithm converges faster, and estimates the effective DCP channel more accurately, compared to the conventional K-means algorithm. This confirmed by the simulation results presented in Sec. VII. We note that a two step, non-iterative algorithm is presented in [25] for detection of On-Off Keying (OOK) signaling constellation in burst-mode optical receivers. The algorithm consists of an *initial partition* and a *partition adjustment* step. However, the algorithm presented above is applicable to general signal constellations; and its iterative nature results in much better performance.

Note that, one could also perform joint channel estimation and soft data detection at the FC by using an EM framework [26], [27]. This approach would result in soft detection of the data symbols rather than hard detection in the K-means algorithm, leading to better coded BER performance [28]. However, we do not pursue this approach in this paper due to lack of space, and also because our purpose is simply to show that reliable data detection is possible at the FC even without using explicit pilot signals from the sensors to the FC for channel estimation.

V. PERFORMANCE ANALYSIS

In this section, we derive the probability of symbol error for data detection at the FC based on the estimated channel using the modified K-means algorithm proposed in Section IV.

First, note that, at low pilot SNR, the phase angle of the effective channel, ϕ_H, could be large due to the channel estimation errors at the individual nodes. If ϕ_H is more than half the angle of rotational symmetry of the signal constellation, it leads to catastrophic detection errors due to the phase ambiguity at the FC. This problem exists for all blind detection algorithms operating on rotationally symmetric constellations, and we refer to it as *channel corruption*. When channel corruption occurs, since all the symbols are detected incorrectly, the probability of symbol error is $(M-1)/M$ which can be upper bounded by 1. We note that channel corruption occurs due to the channel estimation errors at the individual nodes, while the performance of the modified K-means based channel estimation depends on the signal power from the sensors to the FC. Hence, at reasonable data SNRs, even under channel corruption, the K-means based channel estimate converges, in mean-squared sense, to the true magnitude of the effective DCP channel $|H_{\text{DCP}}|$: it still finds the optimum scale factor β, whose magnitude is independent of ϕ_H. Hence, in the analysis to follow, we assume that the K-means channel estimate converges to $\widehat{H}_{\text{DCP}}$ such that $|\widehat{H}_{\text{DCP}}| = |H_{\text{DCP}}|$. This simplifies the analysis while retaining its accuracy, as will be shown via simulation results in Sec. VII. Further, we will see that, at moderate-to-high pilot SNR the probability of error of a system using the K-means based channel estimate is almost the same as that of a system having perfect CSI at the FC. This implies that $\widehat{H}_{\text{DCP}}$ in fact converges, in mean-square sense, to H_{DCP} at moderate-to-high pilot SNR as the number of data symbols increases.

In light of the above, there are two possible events that lead to a symbol error: either channel corruption occurs, or the FC obtains a perfect estimate of the composite DCP channel from the K-means algorithm. In the former case, the probability of symbol error is approximated to be unity, while in the latter case, it is the same as the probability of error of a genie-aided system with perfect CSI at the FC. Hence, the probability of symbol error can be well approximated as

$$P_e \approx P_{cc} + (1 - P_{cc}) P_{e,CSIR}, \tag{14}$$

where P_{cc} is the probability of channel corruption and $P_{e,CSIR}$ is the probability of error with perfect knowledge of H_{DCP} at the FC. We derive expressions for P_{cc} and $P_{e,CSIR}$ for M-ary and PAM and square QAM constellations below. Extensions to other constellations is straightforward.

A. Probability of Symbol Error for M-PAM Signaling

1) Derivation of $P_{e,CSIR,PAM}$: Probability of error for a general M-PAM constellation with perfect CSIR is [29]

$$P_{e,CSIR,PAM} = \frac{2(M-1)}{M} \mathbb{E}_H \left\{ Q\left(R\right) \right\}, \tag{15}$$

where $R = \sqrt{\frac{6E_S}{(M^2-1)N_0}} |H_{\text{DCP}}|$, and $Q(x) = 1/\sqrt{2\pi} \int_x^{\infty} e^{-u^2/2} du$ is the complimentary cumulative distribution function of the standard Gaussian random variable.

The expectation above is over the distribution of $|H_{\text{DCP}}|$. As this distribution is not available in closed-form, it is hard to

compute an exact solution for the probability of error. However, note that, with ideal DCP, i.e., when the phase estimate at the sensors is error-free, $H_{\text{DCP}} = \sum_{k=1}^{N} \alpha_k$ can be well-approximated as a Nakagami-m distributed r.v. [30]. Therefore, even in the presence of phase estimation errors, we propose to approximate the r.v. R with a Nakagami-m r.v. with its parameters chosen using the moment matching method. For simplicity, and to facilitate tractable performance analysis, the estimated Nakagami parameter is rounded to the nearest integer (denoted by $\lceil \cdot \rfloor$). With this, the Nakagami fading parameters m_R and $\bar{\gamma}_R$ can be written in terms of the moments of R as [31]

$$m_R = \left\lceil \frac{(\mathbb{E}\{R^2\})^2}{\text{var}\{R^2\}} \right\rfloor = \left\lceil \frac{\left(\mathbb{E}\left\{|H_{\text{DCP}}|^2\right\}\right)^2}{\text{var}\left\{|H_{\text{DCP}}|^2\right\}} \right\rfloor, \quad (16)$$

$$\bar{\gamma}_R = \mathbb{E}\{R^2\} = \frac{6E_S}{(M^2-1)N_0} \mathbb{E}\left\{|H_{\text{DCP}}|^2\right\}. \quad (17)$$

Computing the above requires the mean and variance of $|H_{\text{DCP}}|^2$, which is given by Lemma 5.1 below for i.i.d. channels. The interested reader is referred to [32] for the result on independent but non-identical fading.

Lemma 5.1: Let H_{DCP} be defined as in (5). When the channels from the sensors to the FC are i.i.d. Rayleigh distributed, the mean and variance of $|H_{\text{DCP}}|^2$ are given by

$$\mathbb{E}\{|H_{\text{DCP}}|^2\} = \frac{N\Omega\left[4 + (4 + (N-1)\pi)\,\gamma_p\Omega\right]}{4(1+\gamma_p\Omega)}, \quad (18)$$

where $\Omega = \mathbb{E}\left\{\alpha_k^2\right\}$ and $\gamma_p = \frac{E_P M_P}{N_0}$, and

$$\begin{aligned}\text{var}\{|H_{\text{DCP}}|^2\} &= \mathbb{E}\{H_{\text{R,DCP}}^4\} + \mathbb{E}\{H_{\text{I,DCP}}^4\} \\ &\quad + 2\mathbb{E}\{H_{\text{R,DCP}}^2 H_{\text{I,DCP}}^2\} - \left(\mathbb{E}\{|H_{\text{DCP}}|^2\}\right)^2, \quad (19)\end{aligned}$$

where $H_{R,\text{DCP}} \triangleq \Re\{H_{\text{DCP}}\}$ and $H_{I,\text{DCP}} \triangleq \Im\{H_{\text{DCP}}\}$ are the real and imaginary parts of H_{DCP}, respectively. The terms in (19) are given by

$$\begin{aligned}\mathbb{E}\{H_{\text{R,DCP}}^4\} &= N\mathbb{E}\left\{\alpha_k^4 \cos^4\theta_{ek}\right\} \\ &\quad + N(N-1)(N-2)(N-3)\left[\mathbb{E}\left\{\alpha_k \cos\theta_{ek}\right\}\right]^4 \\ &\quad + 3N(N-1)\left[\mathbb{E}\left\{\alpha_k^2 \cos^2\theta_{ek}\right\}\right]^2 \\ &\quad + 6N(N-1)(N-2)\mathbb{E}\left\{\alpha_k^2 \cos^2\theta_{ek}\right\}\left[\mathbb{E}\left\{\alpha_k \cos\theta_{ek}\right\}\right]^2 \\ &\quad + 4N(N-1)\mathbb{E}\left\{\alpha_k^3 \cos\theta_{ek}^3\right\}\mathbb{E}\left\{\alpha_k \cos\theta_{ek}\right\}. \quad (20)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\{H_{\text{I,DCP}}^4\} &= N\mathbb{E}\left\{\alpha_k^4 \sin^4\theta_{ek}\right\} \\ &\quad + 3N(N-1)\left[\mathbb{E}\left\{\alpha_k^2 \sin^2\theta_{ek}\right\}\right]^2. \quad (21)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\{H_{\text{R,DCP}}^2 H_{\text{I,DCP}}^2\} &= N\mathbb{E}\left\{\alpha_k^4 \cos^2\theta_{ek}\sin^2\theta_{ek}\right\} \\ &\quad + N(N-1)\mathbb{E}\left\{\alpha_k^2 \cos^2\theta_{ek}\right\}\mathbb{E}\left\{\alpha_k^2 \sin^2\theta_{ek}\right\} \\ &\quad + 2N(N-1)\mathbb{E}\left\{\alpha_k^3 \sin^2\theta_{ek}\cos\theta_{ek}\right\}\mathbb{E}\left\{\alpha_k \cos\theta_{ek}\right\} \\ &\quad + N(N-1)(N-2)\left[\mathbb{E}\left\{\alpha_k \cos\theta_{ek}\right\}\right]^2 \mathbb{E}\left\{\alpha_k^2 \sin^2\theta_{ek}\right\}. \quad (22)\end{aligned}$$

In Appendix A, we provide closed-form expressions for the various expectations in (20)-(22), and the proof of the above Lemma.

Under the Nakagami-m approximation for the distribution of $|H_{\text{DCP}}|$, $P_{e,CSIR,PAM}$ can now be obtained in closed-form using straightforward integration as [33]

$$\begin{aligned}P_{e,CSIR,PAM} &= \left(\frac{M-1}{M}\right)\frac{\phi(\bar{\gamma}_R, m_R)}{\sqrt{\pi}}\frac{\Gamma\left(m_R+\frac{1}{2}\right)}{\Gamma\left(m_R+1\right)} \\ &\quad {}_2F_1\left(1, m_R+\frac{1}{2}; m_R+1; \frac{2m_R}{2m_R+\bar{\gamma}_R}\right). \quad (23)\end{aligned}$$

where $\phi(\bar{\gamma}_R, m_R) = \left(1+\frac{\bar{\gamma}_R}{2m_R}\right)^{-m_R-\frac{1}{2}}$, $\Gamma(n)$ is the gamma function, and ${}_2F_1\left(a, b; c; x\right)$ is the Gauss hypergeometric function [34].

2) Derivation of $P_{cc,PAM}$: For the M-PAM constellation, from (5), channel corruption occurs whenever $|\phi_H| > \frac{\pi}{2}$, since the constellation is rotationally symmetric with period π. This is same as the event $H_{R,\text{DCP}} < 0$. That is,

$$P_{cc,PAM} = P\{H_{R,\text{DCP}} < 0\}. \quad (24)$$

Again, since the exact pdf of $H_{R,\text{DCP}}$ is hard to compute in closed-form, we use the central limit theorem to approximate it as a real Gaussian r.v., $\mathcal{N}\left(\mu_R, \sigma_R^2\right)$, where μ_R and σ_R are respectively the mean and variance of the r.v. $H_{R,\text{DCP}}$, and are worked out in [20] as

$$\mu_R \triangleq \mathbb{E}\left\{H_{R,\text{DCP}}\right\} = N\sqrt{\frac{2E_S}{N_0}}\sqrt{\frac{\pi\Omega}{4}\frac{\gamma_p\Omega}{(1+\gamma_p\Omega)}} \quad (25)$$

$$\sigma_R^2 \triangleq \text{var}\left\{H_{R,\text{DCP}}\right\} = \frac{NE_S\Omega\left(2+(4-\pi)\gamma_p\Omega\right)}{2N_0(1+\gamma_p\Omega)}. \quad (26)$$

Under this approximation, $P_{cc,PAM}$ can be simplified to

$$P_{cc,PAM} \approx Q\left(\frac{\mu_R}{\sigma_R}\right). \quad (27)$$

The probability of error for M-PAM signaling can now computed by substituting these expressions in (14), to get

$$\begin{aligned}P_{e,PAM} &\approx Q\left(\frac{\mu_R}{\sigma_R}\right) + \left(1 - Q\left(\frac{\mu_R}{\sigma_R}\right)\right)\left(\frac{M-1}{M}\right) \\ &\quad \times \frac{\phi(\bar{\gamma}_R, m_R)}{\sqrt{\pi}}\frac{\Gamma\left(m_R+\frac{1}{2}\right)}{\Gamma\left(m_R+1\right)} \\ &\quad \times {}_2F_1\left(1, m_R+\frac{1}{2}; m_R+1; \frac{2m_R}{2m_R+\bar{\gamma}_R}\right). \quad (28)\end{aligned}$$

It will be shown via simulations in the next section that the above expression is accurate for a wide range of values for the data and pilot SNR, and the number of sensors.

B. Probability of Error for Square M-QAM Signaling

As with the PAM constellation, we compute the probability of error with square M-QAM signaling by decomposing it as the sum of the two terms; the first term, denoted $P_{e,CSIR,QAM}$, corresponding to the probability of error when the K-means channel estimate is accurate, and the second term, denoted $P_{cc,QAM}$, corresponding to the probability of channel corruption. The derivation of these two terms is detailed below.

1) Derivation of $P_{e,CSIR,QAM}$: The probability of error with CSI H_{DCP} at the FC for the square M-QAM constellation can be computed from Section V-A, by decomposing the M-ary QAM constellation into two orthogonal $\sqrt{M}$-PAM constellations, as:

$$\begin{aligned} P_{e,CSIR,QAM}(H_{\text{DCP}}) &= 2P_{e,CSIR,PAM}(H_{\text{DCP}}) \\ &\quad -P^2_{e,CSIR,PAM}(H_{\text{DCP}}), \end{aligned} \tag{29}$$

where

$$\begin{aligned} P_{e,CSIR,PAM}(H_{\text{DCP}}) &= \frac{2(\sqrt{M}-1)}{\sqrt{M}} \\ &\quad \times Q\left(\sqrt{\frac{3E_S}{(M-1)N_0}}|H_{\text{DCP}}|\right). \end{aligned} \tag{30}$$

The average error probability is obtained by taking the expectation of (30) over the distribution of the r.v. H_{DCP}. Following the analysis in Section V-A, the first term in (30) can be evaluated in closed-form by using in (23) $\sqrt{M}$ in place of M, and $\widehat{\bar{\gamma}}_R$ in place of $\bar{\gamma}_R$, where

$$\widehat{\bar{\gamma}}_R = \frac{3E_S}{(M-1)N_0}\mathbb{E}\left\{|H_{\text{DCP}}|^2\right\}. \tag{31}$$

However, the expectation of the second term in (29) can be worked out by using the alternate expression for $Q^2(x)$ which is given by [35]

$$Q^2(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{4}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right)d\theta,\ x \geq 0. \tag{32}$$

Upon using (32), the expectation of the second term in (29) using the Nakagami approximation for H_{DCP} simplifies to

$$\begin{aligned} &\mathbb{E}\left[P^2_{e,CSIR,PAM}(H_{\text{DCP}})\right] = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2 \\ &\times \int_0^{\frac{\pi}{4}} \mathbb{E}\left[e^{-\frac{3E_S}{2N_0(M-1)\sin^2\theta}|H_{\text{DCP}}|^2}\right]d\theta \\ &= \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2 \int_0^{\frac{\pi}{4}} \left(1+\frac{\widehat{\bar{\gamma}}_R}{2m_R\sin^2\theta}\right)^{m_R} d\theta \\ &= \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2 \Bigg\{\frac{1}{4}-\frac{\alpha}{\pi}\Big[\left(\frac{\pi}{2}-\arctan\alpha\right) \\ &\times \sum_{k=0}^{m_R-1}\binom{2k}{k}\frac{1}{4^k(1+c)^k} - \sin\arctan\alpha \\ &\times \sum_{k=1}^{m_R-1}\sum_{i=1}^{k}\frac{T_{i,k}}{(1+c)^k}\cos^{2(k-i)+1}\arctan\alpha\Big]\Bigg\}, \end{aligned} \tag{33}$$

where the final simplification in (33) is due to [35] and $c = \frac{\widehat{\bar{\gamma}}_R}{2m_R}$, $\alpha = \sqrt{\frac{c}{1+c}}$, and $T_{i,k} = \frac{\binom{2k}{k}}{\binom{2(k-i)}{k-i}4^i[2(k-i)+1]}$. Finally, upon using (33) along with the aforementioned changes to (23), we arrive at a closed-form expression for the average symbol error probability of M-QAM constellation with Nakagami approximation to the effective DCP channel H_{DCP}.

2) Derivation of $P_{cc,QAM}$: For the square M-QAM constellation, channel corruption occurs when $|\phi_H| > \frac{\pi}{4}$, since its rotational symmetry is $\frac{\pi}{2}$. Thus, $P_{cc,QAM}$ can be written as

$$P_{cc,QAM} = P\left\{|\phi_H| > \frac{\pi}{4}\right\}. \tag{34}$$

Computation of the above probability requires the joint pdf of $(H_{R,\text{DCP}}, H_{I,\text{DCP}})$, which is hard to evaluate in closed-form. Hence, we use the central limit theorem to approximate it as a bivariate Gaussian random variable. The mean vector $\mathbf{m} = [\mathbb{E}[H_{R,\text{DCP}}], \mathbb{E}[H_{I,\text{DCP}}]]^T$, from (25) and [20], is given by $\mathbf{m} = [\mu_R, 0]^T$. Also, since the conditional pdf of $\theta_{e,k}$ given α_k is an odd function of $\theta_{e,k}$, we have

$$\begin{aligned} \mathbb{E}[H_{R,\text{DCP}}H_{I,\text{DCP}}] &= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[\alpha_i\cos\theta_{e,i}\alpha_j\sin\theta_{e,j}] \\ &= \frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\left[\alpha_i^2\sin 2\theta_{e,i}\right] \\ &\quad + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\mathbb{E}[\alpha_i\cos\theta_{e,i}]\,\mathbb{E}[\alpha_j\sin\theta_{e,j}] \\ &= 0. \end{aligned} \tag{35}$$

With the above, the covariance matrix of $(H_{R,\text{DCP}}, H_{I,\text{DCP}})$ is

$$\mathbf{C} = \begin{bmatrix} \sigma_R^2 & 0 \\ 0 & \sigma_I^2 \end{bmatrix}, \tag{36}$$

where σ_R^2 is the variance of $H_{R,\text{DCP}}$ which is given by (26) and σ_I^2 is the variance of $H_{I,\text{DCP}}$, which is given by

$$\sigma_I^2 = \frac{N\Omega}{2(1+\gamma_p\Omega)}. \tag{37}$$

Using the transformation of r.vs, $X \triangleq \frac{1}{\sqrt{2}}(H_{R,\text{DCP}} - H_{I,\text{DCP}})$ and $Y \triangleq \frac{1}{\sqrt{2}}(H_{R,\text{DCP}} + H_{I,\text{DCP}})$, $P_{cc,QAM}$ can be expressed in terms of the joint cumulative distribution function $F_{XY}(x,y)$ of the bivariate normal r.v. $(X,Y)^T$ [30] as

$$P_{cc,QAM} \approx 2Q\left(\frac{\mu_R}{\sigma_R}\right) - F_{XY}(0,0), \tag{38}$$

with the mean and covariance of $(X,Y)^T$ given by

$$\mu_{XY} = \left[\frac{\mu_R}{\sqrt{2}}, \frac{\mu_R}{\sqrt{2}}\right]^T \tag{39}$$

$$\Sigma_{XY} = \begin{bmatrix} \frac{\sigma_R^2+\sigma_I^2}{2} & \frac{\sigma_R^2-\sigma_I^2}{2} \\ \frac{\sigma_R^2-\sigma_I^2}{2} & \frac{\sigma_R^2+\sigma_I^2}{2} \end{bmatrix}. \tag{40}$$

Finally, the probability of error with M-QAM signaling can be computed by substituting these expressions in (14) to get

$$\begin{aligned} P_{e,QAM} &\approx 2Q\left(\frac{\mu_R}{\sigma_R}\right) - F_{XY}(0,0) \\ &\quad + \left(1-2Q\left(\frac{\mu_R}{\sigma_R}\right) + F_{XY}(0,0)\right) \\ &\quad \times \mathbb{E}[P_{e,CSIR,QAM}(H_{\text{DCP}})], \end{aligned} \tag{41}$$

where $\mathbb{E}[P_{e,CSIR,QAM}(H_{\text{DCP}})]$ is as derived earlier in this section.

VI. Robust Signaling Schemes

At low pilot SNR, the phase estimation at the sensor nodes is unreliable, which could result in large offsets in the phase of the effective channel to the fusion center. As mentioned earlier, whenever the phase of the effective channel exceeds half the rotational symmetry of the constellation (e.g., $\frac{\pi}{2}$ for BPSK), it results in catastrophic demodulation errors at the FC. One possible solution to the problem of channel corruption is to use an asymmetric constellation, such as energy shift keying or OOK. Such constellations are naturally immune to channel corruption, as there is no phase ambiguity in blindly estimating the channel due to the signal asymmetry. However, asymmetric constellations come at the cost of a lower separation between the constellation points compared to symmetric constellations, as they are nonzero mean signals. This makes them more vulnerable to noise and signal fading. To get insight into the effect of the constellation on the performance, we compare the performance of symmetric BPSK $\{-1,+1\}$ and OOK $\{0,\sqrt{2}\}$ constellations. We use the analysis in the previous section to determine which constellation offers the best performance at a given pilot and data SNR. The probability of error for the two signaling schemes can be written as

$$\begin{aligned} P_{e,BPSK} &= P_{cc,PAM} + (1 - P_{cc,PAM}) P_{e,CSIR,BPSK} \\ P_{e,OOK} &= P_{e,CSIR,OOK}. \end{aligned} \quad (42)$$

In the above, $P_{e,CSIR,BPSK}$ and $P_{e,CSIR,OOK}$ are both given by $P_{e,CSIR,PAM}$ in (23) with $M=2$, and with $\bar{\gamma} = \bar{\gamma}_{BPSK}$ and $\bar{\gamma} = \bar{\gamma}_{OOK}$ for BPSK and OOK, respectively, where

$$\begin{aligned} \bar{\gamma}_{BPSK} &= \frac{E_S}{N_0} \frac{N\Omega\left[4 + (4 + (N-1)\pi)\gamma_p\Omega\right]}{4(1+\gamma_p\Omega)} \\ \bar{\gamma}_{OOK} &= \frac{E_S}{2N_0} \frac{N\Omega\left[4 + (4 + (N-1)\pi)\gamma_p\Omega\right]}{4(1+\gamma_p\Omega)}, \end{aligned} \quad (43)$$

with $\Omega = \mathbb{E}\left\{\alpha_k^2\right\}$ and $\gamma_p \triangleq \frac{E_P M_P}{N_0}$.

Finally, the expression for probability of error in both cases can be written as

$$\begin{aligned} P_{e,BPSK} &= Q\left(\frac{\mu_R}{\sigma_R}\right) + \left(1 - Q\left(\frac{\mu_R}{\sigma_R}\right)\right) \\ &\frac{\phi(\bar{\gamma}_{BPSK}, m_R)}{2\sqrt{\pi}} \frac{\Gamma\left(m_R + \frac{1}{2}\right)}{\Gamma\left(m_R + 1\right)} \\ &{}_2F_1\left(1, m_R + \frac{1}{2}; m_R + 1; \frac{2m_R}{2m_R + \bar{\gamma}_{BPSK}}\right), \end{aligned} \quad (44)$$

and

$$\begin{aligned} P_{e,OOK} &= \frac{\phi(\bar{\gamma}_{OOK}, m_R)}{2\sqrt{\pi}} \frac{\Gamma\left(m_R + \frac{1}{2}\right)}{\Gamma\left(m_R + 1\right)} \\ &{}_2F_1\left(1, m_R + \frac{1}{2}; m_R + 1; \frac{2m_R}{2m_R + \bar{\gamma}_{OOK}}\right), \end{aligned} \quad (45)$$

where μ_R, σ_R and m_R are given in (25), (26), and (16), respectively.

The above expressions can be used to determine which of the two signal constellations performs the best, for a given data/pilot SNR and the number of sensor nodes. The simulation results in the next section show the accuracy of the above expressions and illustrate the interesting trade-off that occurs as the data and pilot SNRs are varied.

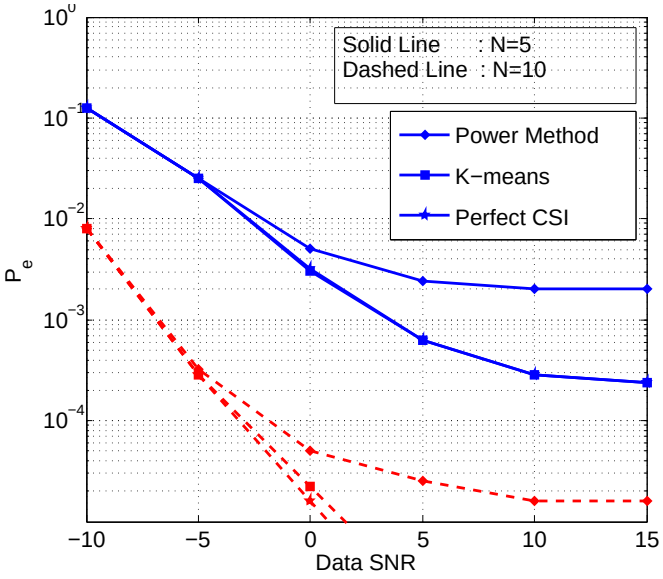

Fig. 3. Probability of symbol error vs. the data SNR for the 4-PAM constellation; pilot SNR = 5dB.

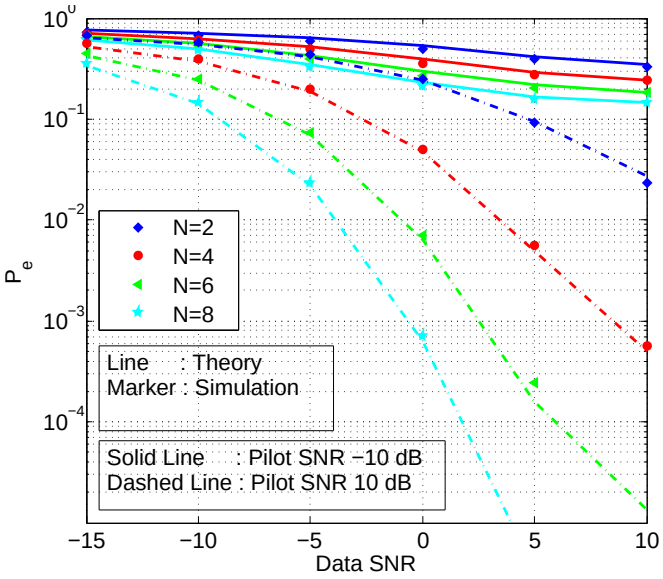

Fig. 4. Probability of symbol error vs. the data SNR for the 4-PAM constellation, using K-means algorithm for channel estimation at the FC.

VII. Simulation Results

In this section, we present Monte Carlo simulation results to validate the theoretical expressions derived in the previous sections and illustrate the performance of the proposed blind channel estimation and data detection schemes.

Simulations were carried out for different values of pilot and data SNR at the FC and with the number of sensors varying from 2 to 8. A single training symbol is sent from the FC to the sensors, followed by M_D data symbols from the sensors to the FC using DCP based on the estimated channel phase angles. The probability of symbol error performance was evaluated by averaging the performance over $1,000$ channel instantiations.

Figure 3 compares the SER performance of the proposed algorithms against the perfect CSI case, where H_{DCP} is known

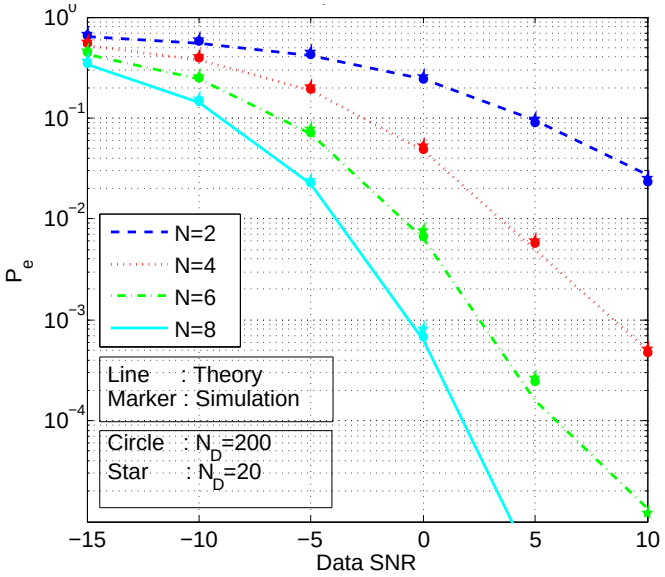

Fig. 5. Probability of symbol error vs. the data SNR, 4-PAM, using K-means algorithm for channel estimation at the FC, with $M_D = 20$ and 200. Pilot SNR $= 10$ dB for channel phase estimation at the nodes.

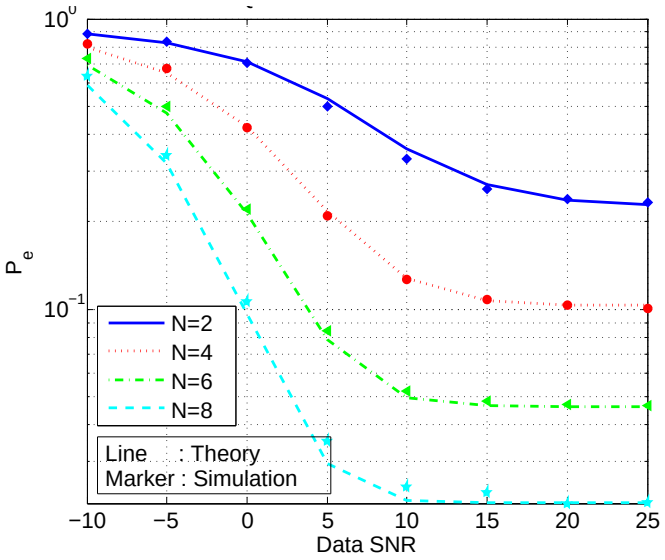

Fig. 6. Probability of symbol error vs. the data SNR, 16-QAM constellation, using K-means for channel estimation at the FC; pilot SNR $= 0$ dB for channel phase estimation at the nodes.

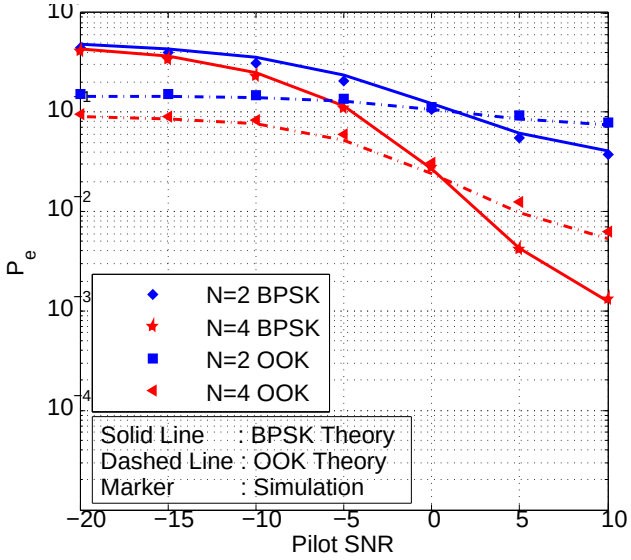

Fig. 7. Performance of symmetric BPSK and OOK with varying pilot SNR and number of sensors, data SNR $= 0$ dB.

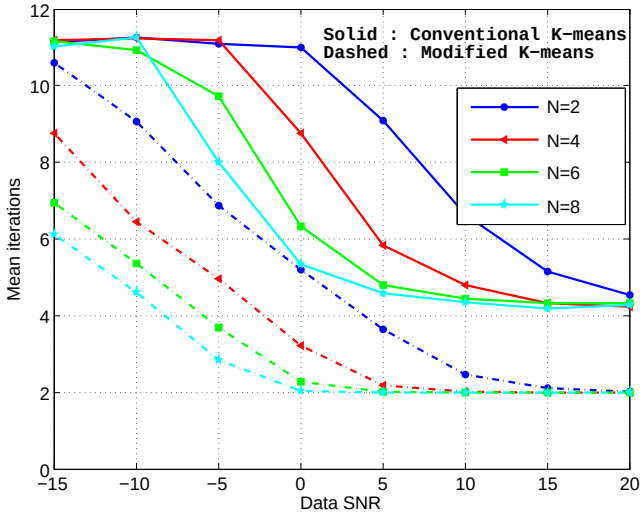

Fig. 8. Comparison of mean number of iterations for the modified K-means algorithm and the conventional K-means algorithm, $M_D = 200$.

perfectly at the FC, for $N = 5$ and 10 sensors, and with a pilot SNR of 5 dB. The modified K-means algorithm outperforms the power method and performs very close to the perfect CSI case. This justifies the assumption in Sec. V, that the K-means algorithm can correctly estimate the channel with a small number of data symbols.

Figure 4 shows the SER plot for different pilot SNRs and number of sensors, and with the 4-PAM constellation. The plot shows that the theoretical expressions closely match with the simulation results for a wide range of pilot SNR, data SNR and number of sensors.

The number of data transmissions M_D is important in determining the performance of the K-means algorithm. The more the number of data points, the better the channel estimation accuracy. Figure 5 compares the performance of the K-means algorithm for different values of $M_D = 20$ and 200 as a function of the data SNR. We note that the performance is nearly the same, irrespective of the value of M_D. This substantiates our approach of excludingM_D from the probability of error analysis.

Figure 6 shows the performance of the K-means algorithm based approach with 16-QAM constellation. The plot shows that the theoretical expressions closely match with the simulation results over a wide range of N and data SNR. The flooring of the curves is due to the effect of channel corruption at the FC, which dominates the performance at high data SNR. Hence, the plot also verifies the accuracy of our analysis of the probability of channel corruption.

Figure 7 shows a comparison of BPSK and OOK in terms of

the probability of bit error as a function of the pilot SNR, with the data SNR fixed at 0 dB. As expected from the analysis in the previous section, at low pilot SNR, where channel corruption is the major cause of error, OOK has better performance. At higher pilot SNR, where noise and fading are the dominant factors, it is better to use symmetric BPSK as it has a higher separation between the constellation points compared to OOK. We also note the accuracy of the expressions derived for the probability of error with OOK in the previous section.

In Fig. 8, we compare the performance of our modified K-means algorithm to the conventional K-means algorithm, in terms of the mean number of iterations required for convergence. The proposed modified K-means algorithm outperforms the conventional algorithm for all ranges of SNR and number of sensors N. Typically, a reduction in number of iterations of the order of 50% is observed for the modified K-means algorithm.

VIII. Conclusion

In this work, we showed that using a higher order constellation for transmission improves the performance of a DCP-based transmission scheme for data fusion from multiple sensor nodes. We proposed two methods, a power based method and a modified K-means algorithm, for blindly estimating the composite channel at the FC. We derived closed-form expressions for the probability of symbol error performance of the proposed DCP and modified K-means based data detection scheme based on a Nakagami-m approximation. We presented results from an extensive simulation study to validate the theoretical results. We showed, through simulations, that the performance of the K-means algorithm based approach is almost the same as that of a system having perfect CSI at the fusion center. Hence, transmission of a training signal from the sensors to the FC is not necessary. We also considered the problem of error flooring due to channel corruption at low pilot SNR inherent to symmetric constellations, and investigated the use of asymmetric constellations such as OOK as a solution. Future work could consider designing the signal constellation for maximizing the throughput from the sensors to the FC, and the design and analysis of power control schemes, in conjunction with DCP-based transmission.

Appendix

A. Moments of $|H_{DCP}|$ with I.I.D. Fading

In this section, we derive the second and fourth moments of $|H_{\text{DCP}}|$ with i.i.d. fading. We use the following notation:

$$\gamma_k = \alpha_k^2; \quad \Omega = \mathbb{E}\{\alpha_k^2\}; \quad \gamma_p = \frac{E_P M_P}{N_0} \text{ and } s = \frac{\gamma_p}{2}. \quad (46)$$

1) Derivation of the Second Moment of $|H_{DCP}|$: We have

$$\begin{aligned}\mathbb{E}\{|H_{\text{DCP}}|^2\} &= \mathbb{E}\left[\sum_{k=1}^{N}\alpha_k\cos\theta_{ek}\right]^2 + \mathbb{E}\left[\sum_{k=1}^{N}\alpha_k\sin\theta_{ek}\right]^2 \\ &= N\mathbb{E}\left[\alpha_k^2\cos^2\theta_{ek}\right] + N(N-1)\mathbb{E}^2\left[\alpha_k\cos\theta_{ek}\right] \\ &\quad + N\mathbb{E}\left[\alpha_k^2\sin^2\theta_{ek}\right] + N(N-1)\mathbb{E}^2\left[\alpha_k\sin\theta_{ek}\right]. \quad (47)\end{aligned}$$

The first two terms in (47) are worked out in [20] and the third term can be easily derived from the first term. The fourth term in (47) simply reduces to zero, since $f_{\theta_{ek}|\alpha_k}(\theta_{ek}|\alpha_k)$ [36] is an even function of θ_{ek}. Therefore, after simplification, we get

$$\mathbb{E}\{|H_{\text{DCP}}|^2\} = \frac{N\Omega\left[4 + (4 + (N-1)\pi)\,\gamma_p\Omega\right]}{4(1+\gamma_p\Omega)}. \quad (48)$$

2) Derivation of The Variance of $|H_{DCP}|^2$: The variance of $|H_{DCP}|^2$ can be obtained as follows:

$$\begin{aligned}\text{var}\{|H_{\text{DCP}}|^2\} &= \mathbb{E}\{H_{\text{R,DCP}}^4\} + \mathbb{E}\{H_{\text{I,DCP}}^4\} \\ &\quad + 2\mathbb{E}\{H_{\text{R,DCP}}^2 H_{\text{I,DCP}}^2\} - \left(\mathbb{E}\{|H_{\text{DCP}}|^2\}\right)^2 \quad (49)\end{aligned}$$

where $H_{\text{R,DCP}}$ and $H_{\text{I,DCP}}$ are the real and imaginary parts of H_{DCP}, respectively. We find the expectations of the first three terms of the above equation as follows. Using a binomial expansion and grouping similar terms, we obtain:

$$\begin{aligned}\mathbb{E}\{H_{\text{R,DCP}}^4\} &= \mathbb{E}\left\{\left[\sum_{k=1}^{N}\alpha_k\cos\theta_{ek}\right]^4\right\} \\ &= N\mathbb{E}\left\{\alpha_k^4\cos^4\theta_{ek}\right\} + 3N(N-1)\left[\mathbb{E}\left\{\alpha_k^2\cos^2\theta_{ek}\right\}\right]^2 \\ &\quad + N(N-1)(N-2)(N-3)\left[\mathbb{E}\left\{\alpha_k\cos\theta_{ek}\right\}\right]^4 \\ &\quad + 6N(N-1)(N-2)\mathbb{E}\left\{\alpha_k^2\cos^2\theta_{ek}\right\}\left[\mathbb{E}\left\{\alpha_k\cos\theta_{ek}\right\}\right]^2 \\ &\quad + 4N(N-1)\mathbb{E}\left\{\alpha_k^3\cos\theta_{ek}^3\right\}\mathbb{E}\left\{\alpha_k\cos\theta_{ek}\right\}. \quad (50)\end{aligned}$$

In a similar manner, using the even symmetry of $f_{\theta_{ek}|\alpha_k}(\theta_{ek}|\alpha_k)$,

$$\begin{aligned}\mathbb{E}\{H_{\text{I,DCP}}^4\} &= \mathbb{E}\left\{\left[\sum_{k=1}^{N}\alpha_k\sin\theta_{ek}\right]^4\right\} = N\mathbb{E}\left\{\alpha_k^4\sin^4\theta_{ek}\right\} \\ &\quad + 3N(N-1)\left[\mathbb{E}\left\{\alpha_k^2\sin^2\theta_{ek}\right\}\right]^2. \quad (51)\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\{H_{\text{R,DCP}}^2 H_{\text{I,DCP}}^2\} \\ &\quad = \mathbb{E}\left\{\left[\sum_{k=1}^{N}\alpha_k\cos\theta_{ek}\right]^2\left[\sum_{k=1}^{N}\alpha_k\sin\theta_{ek}\right]^2\right\} \\ &\quad = N\mathbb{E}\left\{\alpha_k^4\cos^2\theta_{ek}\sin^2\theta_{ek}\right\} \\ &\quad + N(N-1)\mathbb{E}\left\{\alpha_k^2\cos^2\theta_{ek}\right\}\mathbb{E}\left\{\alpha_k^2\sin^2\theta_{ek}\right\} \\ &\quad + 2N(N-1)\mathbb{E}\left\{\alpha_k^3\sin^2\theta_{ek}\cos\theta_{ek}\right\}\mathbb{E}\left\{\alpha_k\cos\theta_{ek}\right\} \\ &\quad + N(N-1)(N-2)\left[\mathbb{E}\left\{\alpha_k\cos\theta_{ek}\right\}\right]^2\mathbb{E}\left\{\alpha_k^2\sin^2\theta_{ek}\right\}. \quad (52)\end{aligned}$$

We now derive analytical expressions for all the terms involved in (50), (51) and (52).

1) Derivation of $\mathbb{E}\{\alpha_k^4\cos^4\theta_{ek}\}$: Using a trigonometric identity from [37]

$$\begin{aligned}\mathbb{E}\{\alpha_k^4\cos^4\theta_{ek}\} &= \frac{1}{8}\left[\mathbb{E}\left\{\alpha_k^4\ \cos 4\theta_{ek}\right\}\right. \\ &\quad \left. + 4\mathbb{E}\left\{\alpha_k^4\ \cos 2\theta_{ek}\right\} + 3\mathbb{E}\{\alpha_k^4\}\right]. \quad (53)\end{aligned}$$

We derive the conditional expectations as

$$\begin{aligned}\mathbb{E}\{\cos 4\theta_{ek}|\alpha_k\} &= \mathbb{E}\{\cos(|4\theta_{ek}|)|\alpha_k\} \\ &= \int_{\theta=0}^{\pi} \cos 4\theta_{ek}\ f(|\theta_{ek}||\alpha_k) d\theta_{ek} \\ &= \int_{\theta=0}^{\pi} \cos 4\theta \left[\frac{e^{-\gamma_p \alpha_k^2}}{\pi} + \frac{\gamma_p \alpha_k^2 \sin 2\theta}{\pi} \right. \\ &\quad \left. \int_{x=0}^{\pi-\theta} \frac{e^{-\frac{\gamma_p \alpha_k^2 \sin^2\theta}{\sin^2 x}}}{\sin^2 x} dx \right] d\theta \\ &= \frac{\gamma_p \alpha_k^2}{\pi} \int_{x=0}^{\pi} \frac{1}{\sin^2 x} \int_{\theta=0}^{\pi-x} \cos 4\theta \sin 2\theta \\ &\quad e^{-\frac{\gamma_p \alpha_k^2 \sin^2\theta}{\sin^2 x}} d\theta dx \quad (54)\end{aligned}$$

where $f(|\theta_{ek}||\alpha_k)$ is from [36]. Substituting $\sin^2\theta = t$ and simplifying

$$\begin{aligned}\mathbb{E}\{\cos 4\theta_{ek}|\alpha_k\} &= \frac{\gamma_p \alpha_k^2}{\pi} \int_{x=0}^{\pi} \frac{1}{\sin^2 x} \\ &\quad \int_{t=0}^{\sin^2 x} \left(8t^2 - 8t + 1\right) e^{-\frac{\gamma_p \alpha_k^2 t}{\sin^2 x}} dt dx. \quad (55)\end{aligned}$$

Integrating by parts

$$\begin{aligned}\mathbb{E}\{\cos 4\theta_{ek}|\alpha_k\} &= 1 + \frac{6}{\gamma_p^2 \alpha_k^4} - \frac{4}{\gamma_p \alpha_k^2} \\ &\quad - \frac{6}{\gamma_p^2 \alpha_k^4} e^{-\gamma_p \alpha_k^2} - \frac{2}{\gamma_p \alpha_k^2} e^{-\gamma_p \alpha_k^2}. \quad (56)\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}\{\alpha_k^4 \cos 4\theta_{ek}\} &= \mathbb{E}\left\{ \frac{6}{\gamma_p^2} - \frac{4\alpha_k^2}{\gamma_p} + \alpha_k^4 - \frac{6}{\gamma_p^2} e^{-\gamma_p \alpha_k^2} \right. \\ &\quad \left. - \frac{2\alpha_k^2}{\gamma_p} e^{-\gamma_p \alpha_k^2} \right\}. \quad (57)\end{aligned}$$

The fourth and fifth terms in the above expectation can be computed for a Rayleigh fading channel as follows. First, note that since α_k is Rayleigh distributed, α_k^2 is exponentially distributed with mean Ω. Then, the Laplace transform of $\gamma_k = \alpha_k^2$ and its derivative are given by

$$\begin{aligned}\mathcal{L}_{\gamma_k}(u) &= \mathbb{E}\{e^{-u\gamma_k}\} = \frac{1}{1+u\Omega} \quad (58) \\ \frac{\partial}{\partial u}\mathcal{L}_{\gamma_k}(u) &= -\mathbb{E}\{\gamma_k e^{-u\gamma_k}\} = -\frac{\Omega}{(1+u\Omega)^2}. \quad (59)\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{E}\{\alpha_k^4 \cos 4\theta_{ek}\} &= 2\Omega^2 + \frac{6}{\gamma_p^2} - \frac{4\Omega}{\gamma_p} - \frac{6}{\gamma_p^2} \frac{1}{(1+\gamma_p\Omega)} \\ &\quad - \frac{2}{\gamma_p} \frac{\Omega}{(1+\gamma_p\Omega)^2}. \quad (60)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\{\alpha_k^4 \cos 2\theta_{ek}\} &= \mathbb{E}\left\{\alpha_k^4 \mathbb{E}\{\cos 2\theta_{ek}|\alpha_k\}\right\} \\ &= \mathbb{E}\left\{\gamma_k^2 \left[1 - \frac{1 - e^{-\gamma_p\gamma_k}}{\gamma_p\gamma_k}\right]\right\} \\ &= 2\Omega^2 - \frac{\Omega}{\gamma_p} + \frac{\Omega}{\gamma_p(1+\gamma_p\Omega)^2}, \quad (61)\end{aligned}$$

where the second equality follows from [20]. From (53),

$$\begin{aligned}\mathbb{E}\{\alpha_k^4 \cos^4\theta_{ek}\} &= \frac{1}{8}\left[\mathbb{E}\{\alpha_k^4 \cos 4\theta_{ek}\}\right. \\ &\quad \left. + 4\mathbb{E}\{\alpha_k^4 \cos 2\theta_{ek}\} + 3\mathbb{E}\left\{\alpha_k^4\right\}\right] \\ &= 2\Omega^2 - \frac{\Omega^2}{4} \frac{(5+4\gamma_p\Omega)}{(1+\gamma_p\Omega)^2}. \quad (62)\end{aligned}$$

2) Derivation of $\mathbb{E}\{\alpha_k^4 \sin^4\theta_{ek}\}$: Using a trigonometric identity from [37], we have

$$\begin{aligned}\mathbb{E}\{\alpha_k^4 \sin^4\theta_{ek}\} &= \frac{1}{8}\left[\mathbb{E}\{\alpha_k^4 \cos 4\theta_{ek}\}\right. \\ &\quad \left. - 4\mathbb{E}\{\alpha_k^4 \cos 2\theta_{ek}\} + 3\mathbb{E}\{\alpha_k^4\}\right] \\ &= \frac{3}{4}\left[\frac{1}{\gamma_p^2} - \frac{1}{\gamma_p^2(1+\gamma_p\Omega)} - \frac{\Omega}{\gamma_p(1+\gamma_p\Omega)^2}\right] \\ &= \frac{3}{4}\frac{\Omega^2}{(1+\gamma_p\Omega)^2} \quad (63)\end{aligned}$$

where the second term in the above is obtained from [20].

3) Derivation of $\mathbb{E}\{\alpha_k^3 \cos^3\theta_{ek}\}$ We write

$$\mathbb{E}\{\alpha_k^3 \cos^3\theta_{ek}\} = \mathbb{E}\{\alpha_k^3\ \mathbb{E}\{\cos^3\theta_{ek}|\alpha_k\}\}. \quad (64)$$

Using the expression for $f_{\theta_{ek}|\alpha_k}(\theta_{ek}|\alpha_k)$ from [36], the conditional expectation can be written as

$$\begin{aligned}\mathbb{E}\{\cos^3\theta_{ek}|\alpha_k\} &= \int_0^{2\pi} \cos^3\theta_{ek} f_{\theta_{ek}|\alpha_k}(\theta_{ek}) d\theta_{ek} \\ &= \int_0^{2\pi} \cos^3\theta_{ek} \left[\frac{e^{-\gamma_p\alpha_k^2}}{2\pi} + \sqrt{\frac{\gamma_p\alpha_k^2}{\pi}} \cos\theta_{ek} \right. \\ &\quad \left. e^{-\gamma_p\alpha_k^2 \sin^2\theta_{ek}} \phi\left(\sqrt{2\gamma_p\alpha_k^2}\cos\theta_{ek}\right)\right] d\theta_{ek}\end{aligned}$$

$$\begin{aligned}&= \sqrt{\frac{\gamma_p\alpha_k^2}{\pi}} \int_0^{2\pi} \cos^4\theta_{ek} e^{-\gamma_p\alpha_k^2 \sin^2\theta_{ek}} \\ &\quad \phi\left(\sqrt{2\gamma_p\alpha_k^2}\cos\theta_{ek}\right) d\theta_{ek} \\ &= \sqrt{\frac{\gamma_p\alpha_k^2}{\pi}} e^{-\frac{\gamma_p\alpha_k^2}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^4\theta_{ek} e^{\frac{\gamma_p\alpha_k^2 \cos 2\theta_{ek}}{2}} d\theta_{ek}\end{aligned}$$

$$\begin{aligned}&= \sqrt{\frac{\gamma_p\alpha_k^2}{\pi}} e^{-\frac{\gamma_p\alpha_k^2}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{3}{8} + \frac{\cos 4\theta_{ek}}{8}\right. \\ &\quad \left. + \frac{\cos 2\theta_{ek}}{2}\right) e^{\frac{\gamma_p\alpha_k^2 \cos 2\theta_{ek}}{2}} d\theta_{ek} \\ &= \sqrt{\frac{\gamma_p\alpha_k^2}{\pi}} e^{-\frac{\gamma_p\alpha_k^2}{2}} \left[\frac{3\pi}{8} I_0\left(\frac{\gamma_p\alpha_k^2}{2}\right)\right. \\ &\quad \left. + \frac{\pi}{2} I_1\left(\frac{\gamma_p\alpha_k^2}{2}\right) + \frac{\pi}{8} I_2\left(\frac{\gamma_p\alpha_k^2}{2}\right)\right], \quad (65)\end{aligned}$$

where $\phi(x)$ is the cumulative distribution function of a standard Gaussian random variable, $I_n(x) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{x\cos\theta}\cos(n\theta)d\theta$ is the n^{th} order modified Bessel

function of the first kind [37]. By a change of variables, this can be written as

$$I_0(z) = \frac{1}{\pi}\int_{-1}^{1}\frac{e^{-tz}}{\sqrt{1-t^2}}dt \quad (66)$$

$$I_1(z) = \frac{z}{\pi}\int_{-1}^{1}e^{-tz}\sqrt{1-t^2}dt \quad (67)$$

$$\text{and} \quad I_2(z) = \frac{1}{\pi}\int_{-1}^{1}\frac{e^{-zt}(2t^2-1)}{\sqrt{1-t^2}}dt. \quad (68)$$

Using the above, the expectation of $\alpha_k^3\cos^3\theta_{ek}$ can be simplified as

$$\begin{aligned}\mathbb{E}\{\alpha_k^3\cos^3\theta_{ek}\} &= \sqrt{\gamma_p\pi}\,\mathbb{E}\Bigg\{\alpha_k^4 e^{-\frac{\gamma_p\alpha_k^2}{2}} \\ &\Big[\frac{3}{8\pi}\int_{-1}^{1}\frac{e^{-s\alpha_k^2 t}}{\sqrt{1-t^2}}dt \\ &+\frac{1}{2\pi}s\alpha_k^2\int_{-1}^{1}e^{-s\alpha_k^2 t}\sqrt{1-t^2}dt \\ &+\frac{1}{8\pi}\int_{-1}^{1}e^{-s\alpha_k^2 t}\frac{2t^2-1}{\sqrt{1-t^2}}dt\Big]\Bigg\}. \quad (69)\end{aligned}$$

The expectations in the above equation can be computed for the Rayleigh fading channel as follows:

$$\mathbb{E}\{\gamma_k^2 e^{-u\gamma_k}\} = \frac{\partial^2}{\partial u^2}\mathcal{L}_{\gamma_k}(u) = \frac{2\Omega^2}{(1+u\Omega)^3} \quad (70)$$

$$\mathbb{E}\{\gamma_k^3 e^{-u\gamma_k}\} = -\frac{\partial^3}{\partial u^3}\mathcal{L}_{\gamma_k}(u) = \frac{6\Omega^3}{(1+u\Omega)^4}. \quad (71)$$

The following results are used for further simplification of (69) [37]:

$$\int_{-1}^{1}\frac{1}{(a+t)^3}\frac{1}{\sqrt{1-t^2}}dt = \frac{\pi}{2}\frac{1+2a^2}{(a^2-1)^{\frac{5}{2}}} \quad (72)$$

$$\int_{-1}^{1}\frac{\sqrt{1-t^2}}{(a+t)^4}dt = \frac{\pi}{2}\frac{a}{(a^2-1)^{\frac{5}{2}}} \quad (73)$$

$$\text{and} \quad \int_{-1}^{1}\frac{\sqrt{1-t^2}}{(a+t)^3}dt = \frac{\pi}{2}\frac{1}{(a^2-1)^{\frac{3}{2}}}. \quad (74)$$

After some algebraic manipulations, we arrive at

$$\mathbb{E}\{\alpha_k^3\cos^3\theta_{ek}\} = \frac{3\Omega^2\sqrt{\gamma_p\pi}}{4\sqrt{1+\gamma_p\Omega}}. \quad (75)$$

4) Derivation of $\mathbb{E}\{\alpha_k^3\cos\theta_{ek}\}$: From [20], we have

$$\mathbb{E}\{\cos\theta_{ek}|\alpha_k\} = \alpha_k\sqrt{\frac{\gamma_p\pi}{4}}e^{-\frac{\gamma_p\alpha_k^2}{2}}\left[I_0\left(\frac{\gamma_p\alpha_k^2}{2}\right)+I_1\left(\frac{\gamma_p\alpha_k^2}{2}\right)\right] \quad (76)$$

$$\begin{aligned}\mathbb{E}\{\alpha_k^3\cos\theta_{ek}\} &= \mathbb{E}\{\alpha_k^3\,\mathbb{E}\{\cos\theta_{ek}|\alpha_k\}\} \\ &= \mathbb{E}\left\{\alpha_k^4\sqrt{\frac{\gamma_p\pi}{4}}e^{-\frac{\gamma_p\alpha_k^2}{2}}\left[I_0\left(\frac{\gamma_p\alpha_k^2}{2}\right)+I_1\left(\frac{\gamma_p\alpha_k^2}{2}\right)\right]\right\} \\ &= \sqrt{\frac{\gamma_p}{4\pi}}\left[\int_{-1}^{1}\frac{1}{\sqrt{1-t^2}}\mathbb{E}\left\{\gamma_k^2 e^{-s\gamma_k(t+1)}\right\}dt \right. \\ &\left. +s\int_{-1}^{1}\sqrt{1-t^2}\mathbb{E}\left\{\gamma_k^3 e^{-s\gamma_k(t+1)}\right\}dt\right]. \quad (77)\end{aligned}$$

The integrals in the above equation have already been evaluated in the previous subsection. In particular, using (72) and (73) and simplifying the resulting expressions, we get

$$\mathbb{E}\{\alpha_k^3\cos\theta_{ek}\} = \frac{\sqrt{\gamma_p\pi}}{4}\frac{\Omega^2(4+3\gamma_p\Omega)}{(1+\gamma_p\Omega)^{3/2}}. \quad (78)$$

5) Derivation of $\mathbb{E}\{\alpha_k^4\sin^2\theta_{ek}\cos^2\theta_{ek}\}$

$$\begin{aligned}\mathbb{E}\{\alpha_k^4\sin^2\theta_{ek}\cos^2\theta_{ek}\} &= \frac{1}{8}\mathbb{E}\left\{\alpha_k^4-\alpha_k^4\cos 4\theta_{ek}\right\} \\ &= \frac{1}{4}\left[\frac{2\Omega}{\gamma_p}-\frac{3}{\gamma_p^2}+\frac{3}{\gamma_p^2(1+\gamma_p\Omega)}\right. \\ &\left.+\frac{\Omega}{\gamma_p(1+\gamma_p\Omega)^2}\right] \quad (79)\end{aligned}$$

where the second term in the first step follows from (60).

6) Derivation of $\mathbb{E}\{\alpha_k^3\sin^2\theta_{ek}\cos\theta_{ek}\}$

$$\mathbb{E}\{\alpha_k^3\sin^2\theta_{ek}\cos\theta_{ek}\} = \mathbb{E}\left\{\alpha_k^3\cos\theta_{ek}-\alpha_k^3\cos^3\theta_{ek}\right\}. \quad (80)$$

These terms are derived in (75) and (78). It follows that

$$\mathbb{E}\{\alpha_k^3\sin^2\theta_{ek}\cos\theta_{ek}\} = \frac{\sqrt{\gamma_p\pi}}{4}\frac{\Omega^2}{(1+\gamma_p\Omega)^{3/2}}. \quad (81)$$

REFERENCES

[1] A. Manesh, C. Murthy, and R. Annavajjala, "Design and analysis of distributed co-phasing with arbitrary constellations," in *Proc. ICC*, pp. 1–6, 2013.

[2] J. Chamberland and V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, 2007.

[3] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, 2007.

[4] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Distributed transmit beamforming using feedback control," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 411–426, 2010.

[5] Y. Tu and G. Pottie, "Coherent cooperative transmission from multiple adjacent antennas to a distant stationary antenna through AWGN channels," in *Proc. VTC*, vol. 1, pp. 130–134, 2002.

[6] M. Rahman, R. Mudumbai, and S. Dasgupta, "Consensus based carrier synchronization in a two node network," in *IFAC World Congress*, pp. 10038–10043, 2011.

[7] I. Brown, D.R. and H. Poor, "Time-slotted round-trip carrier synchronization for distributed beamforming," *IEEE Trans. Sig. Proc.*, vol. 56, pp. 5630–5643, Nov. 2008.

[8] R. Mudumbai, D. Brown, U. Madhow, and H. Poor, "Distributed transmit beamforming: challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, pp. 102–110, Feb. 2009.

[9] I. Thibault, G. Corazza, and L. Deambrogio, "Random, deterministic, and hybrid algorithms for distributed beamforming," in *Adv. Satellite Multimedia Syst. Conf. (ASMA) and Sig. Proc. for Space Commun. Workshop (SPSC)*, pp. 221–225, Sept. 2010.

[10] I. Thibault, G. Corazza, and L. Deambrogio, "Phase synchronization algorithms for distributed beamforming with time varying channels in wireless sensor networks," in *Wireless Commun. and Mobile Comput. Conf. (IWCMC)*, pp. 77–82, July 2011.

[11] F. Quitin, M. Mahboob Ur Rahman, R. Mudumbai, and U. Madhow, "Distributed beamforming with software-defined radios: Frequency synchronization and digital feedback," in *IEEE GLOBECOM*, pp. 4787–4792, Dec. 2012.

[12] A. Kumar, S. Dasgupta, and R. Mudumbai, "Distributed nullforming without prior frequency synchronization," in *3rd Australian Control Conference (AUCC)*, pp. 207–211, Nov. 2013.

[13] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser mimo," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1815–1831, Apr. 2014.

[14] L. Dong, A. Petropulu, and H. Poor, "A cross-layer approach to collaborative beamforming for wireless ad hoc networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2981–2993, 2008.

[15] L. Dong, A. Petropulu, and H. Poor, "Weighted cross-layer cooperative beamforming for wireless networks," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3240–3252, 2009.

[16] H. Ochiai, P. Mitran, H. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4110–4124, 2005.

[17] M. Ahmed and S. Vorobyov, "Collaborative beamforming for wireless sensor networks with gaussian distributed sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 638–643, 2009.

[18] A. Ekbal and J. Cioffi, "Distributed transmit beamforming in cellular networks-a convex optimization perspective," in *Proc. ICC*, vol. 4, pp. 2690–2694, 2005.

[19] S. Zhou and G. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over rayleigh mimo channels?," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, 2004.

[20] K. Chaythanya, R. Annavajjala, and C. Murthy, "Comparative analysis of pilot-assisted distributed co-phasing approaches in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, pp. 3722–3737, Aug. 2011.

[21] J. Holtzman, "A simple, accurate method to calculate spread-spectrum multiple-access error probabilities," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 461–464, 1992.

[22] T. Cover and J. Thomas, *Elements of information theory*. J. Wiley & Sons, Inc., 1991.

[23] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Info. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.

[24] S. Z. Selim and M. A. Ismail, "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 1, pp. 81–87, 1984.

[25] T. Zhao, A. Nehorai, and B. Porat, "K-means clustering-based data detection and symbol-timing recovery for burst-mode optical receiver," *IEEE Trans. Commun.*, vol. 54, no. 8, pp. 1492–1501, 2006.

[26] X. Ma, H. Kobayashi, and S. Schwartz, "An EM-based estimation of OFDM signals," in *Proc. WCNC*, vol. 1, pp. 228–232, 2002.

[27] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1297–1302, 1999.

[28] C. Bishop, *Pattern recognition and machine learning*. Springer, New York, 2006.

[29] J. Proakis, *Digital communications*. McGraw-Hill, 1987.

[30] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*. McGraw-Hill, New York, 1966.

[31] N. Beaulieu and C. Cheng, "Efficient nakagami-m fading channel simulation," *IEEE Trans. Veh. Technol.*, vol. 54, no. 2, pp. 413–424, 2005.

[32] A. Manesh, "Distributed co-phasing for wireless sensor networks," Master's thesis, Indian Institute of Science, Bangalore, India, 560012, June 2012.

[33] H. Shin and J. Lee, "On the error probability of binary and m-ary signals in nakagami-m fading channels," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 536–539, 2004.

[34] F. Beukers, "Gauss hypergeometric function," *Arithmetic and geometry around hypergeometric functions*, pp. 23–42, 2007.

[35] M. Simon and M.-S. Alouini, *Digital Communications over Generalized Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley, 2000.

[36] K. Quirk and L. Milstein, "The effects of phase estimation errors on rake receiver performance," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 669–682, 2002.

[37] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. Academic Publishers, New York, 1980.

Chandra R. Murthy (S'03–M'06–SM'11) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1998, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, in 2000 and 2006, respectively.

From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, where he is currently working as an Associate Professor. His research interests are in the areas of Cognitive Radio, Energy Harvesting Wireless Sensors, MIMO systems with channel-state feedback, and sparse signal recovery techniques applied to wireless communications. He is currently serving as the Chair of the IEEE Signal Processing Society, Bangalore Chapter and as an associate editor for the IEEE Signal Processing Letters. He is an elected member of the IEEE SPCOM Technical Committee for the years 2014-16.

Ramesh Annavajjala is with the Charles Stark Draper Laboratory, Cambridge, MA, USA. Earlier, he served as the Distinguished Member of Technical Staff at the Altiostar Networks Inc., a Tewksbury, MA, based start-up company developing cloud-RAN optimized 4G LTE base-station products. Prior to that, he was a Principal Member of Research Staff at the Mitsubishi Electric Research Labs (MERL), in Cambridge, MA. He has also held industry positions at ArrayComm LLC (San Jose, CA, USA), Synopsys Inc., (Bangalore, INDIA) and CMC R&D Center (Hyderabad, INDIA). He received Bachelors in Electronics and Communication Engineering from the National Institute of Technology (NIT, Warangal, INDIA, May 1998), Masters in Telecommunications from the Indian Institute of Science (IISc, Bangalore, INDIA, Jan. 2001), and Ph.D. in Electrical Engineering from the University of California at San Diego (UCSD, La Jolla, CA, June 2006).

Dr. Annavajjala was a recipient of the Purkayastha/TimeLine Ventures graduate fellowship (2002-2003), a co-recipient of the best paper award from the IEEE WPMC 2009 conference, and was a guest editor for the special issue **Wireless Cooperative Networks** of the EURASIP Journal on Advanced Signal Processing. He is a Senior Member of the IEEE, and was nominated for the MIT TR-35. He has published more than 60 papers in international journals and conferences, made numerous contributions to commercial wireless standards, and is a co-inventor of 10 US patents (granted).