Blind and Semi-Blind Channel Estimation/ Equalization for Poisson Channels in Optical Wireless Scattering Communication Systems

Beiyuan Liu, *Associate Member, IEEE*, Chen Gong, *Senior Member, IEEE*, Julian Cheng, *Senior Member, IEEE*, Zhengyuan Xu, *Senior Member, IEEE*, and Jiajia Liu, *Senior Member, IEEE*

***Abstract*— The communication systems for Poisson channels require long pilot for channel estimation and may result in large percentage of overhead due to the signal-dependent noise of Poisson-distributed signal, i.e., both signal part and noise part experience random processes. In this paper, both blind and semi-blind channel estimation methods are studied to shorten the overhead and increase the transmission efficiency for Poisson channels. First, fractionally spaced equalizers are proposed based on modified constant modulus algorithm (CMA) and subspace (SS). Second, a data-aided iterative channel estimation (ICE) is designed and analyzed in terms of its asymptotic unbiasedness and convergence. The proposed methods are evaluated based on both a constant channel scenario and a varying channel scenario. Numerical simulation results show that the modified CMA has the worst bit-error rate performance but requires the lowest computational complexity. Besides, both SS based channel estimation and ICE have negligible overhead and comparable bit-error rate performances with respect to the conventional periodic pilot based channel estimation having 50% overhead.**

***Index Terms*— Blind channel estimation, constant modulus algorithm, data-aided iterative channel estimation, semi-blind channel estimation, subspace based channel estimation, poisson channel.**

Manuscript received 4 July 2021; revised 5 November 2021 and 11 January 2022; accepted 15 January 2022. Date of publication 26 January 2022; date of current version 12 August 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1801904, in part by the Youth Program of National Natural Science Foundation of China under Grant 62101447, in part by the Key Program of National Natural Science Foundation of China under Grant 61631018, and in part by the Key Research Program of Frontier Sciences of China Academy of Science under Grant QYZDY-SSWJSC003. An earlier version of this paper was presented in part at ICC 2020 [DOI: 10.1109/ICC40277.2020.9148723]. The associate editor coordinating the review of this article and approving it for publication was D. W. K. Ng. *(Corresponding author: Chen Gong.)*

Beiyuan Liu and Jiajia Liu are with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China (e-mail: lby@nwpu.edu.cn; liujiajia@nwpu.edu.cn).

Chen Gong and Zhengyuan Xu are with the Key Laboratory of Wireless-Optical Communications, Chinese Academy of Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China (e-mail: cgong821@ustc.edu.cn; xuzy@ustc.edu.cn).

Julian Cheng is with the School of Engineering, The University of British Columbia, Kelowna, BC V1V 1V7, Canada (e-mail: julian.cheng@ubc.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3144360.

Digital Object Identifier 10.1109/TWC.2022.3144360

I. Introductions

POISSON channel model has been widely used in the communication systems with particle-type received signals, such as optical wireless scattering communication (OWSC) systems [2] and diffusion-based molecular communication (MC) systems [3]. For the OWSC system, the received signals are typically characterized as the photoelectrons inspired by the received photons due to the wave-particle duality of optical signals and large path loss of a non-line-of-sight (NLOS) link. For the diffusion-based MC system, the particle-distributed molecules are transmitted through the medium via Brownian motion mechanism. These emerging transmission technologies make Poisson channel an important research topic.

Existing works on Poisson channel have focused on the channel capacity [4], [5] and signal processing [6]–[10]. Most receiver-side signal processing techniques assume short delay spread of the channel impulse response (CIR) where the inter-symbol interference (ISI) is not considered [11], [12]. However, for OWSC systems, the delay spread can be much longer than a symbol slot for large transmission distance due to different arrival times of the photons. Such a phenomenon can be verified by Monte-Carlo based channel simulation where channel state information can be obtained by the transmission medium parameters and the geometry parameters of the transceiver architecture.

Although it is believed that Monte-Carlo simulation can provide precise channel impulse response, it requires accurate transmission medium parameters, which are difficult to measure and obtain in real time. Therefore, we are motivated to develop instantaneous channel estimation techniques for Poisson channel with ISI. Instantaneous channel estimation and signal detection were first studied for Poisson channels by considering ISI [10]. This study showed that, due to stochastic property of the signal part, the pilot length of the unbiased least square (LS) channel estimation typically requires 256-length pilot to achieve a relatively accurate channel estimation. Such a long pilot requires large overhead in a fast fading channel. To address the long pilot issue, the correlation of channel state information among different wavelengths was applied to reduce the overhead

of the entire multi-wavelength parallel channels by pilot allocation [13].

It is worthwhile to mention that in MC, the blind channel estimation, or the so-called non-coherent signal detection, has been studied. A general class of threshold decoders was proposed for memmory-limited diffusion-based Poisson channel considering oversampling [14]. Then, a maximum-likelihood multi-symbol detector as well as a decision-feedback detector were proposed for the scenario with negligible ISI [15]. Besides, a low complexity non-coherent detector was proposed using the knowledge of the diffusive channel response [16]. However, most works in MC consider either memory-limited or ISI negligible case, which is reasonable in MC field since the complexity at the receiver is limited. While in this work, we try to give more general channel estimation for Poisson channels with ISI, where the signal detection complexity is affordable at the receiver, for instance, in optical wireless scattering communication.

In this paper, we consider blind and semi-blind channel estimation to solve the long pilot issue of Poisson channel. Blind channel estimation (or equalization) method aims to estimate the channel impulse response without a training sequence. In RF communication, constant modulus algorithm (CMA) is a popular approach for blind channel estimation due to its linear structure and low computational complexity [17]–[20]. Another high data efficiency blind channel estimation is the subspace-based blind channel estimation using the second-order statistics of the oversampled received signal [21]–[25], where oversampling can help construct a "tall" matrix, which is essential when applying the subspace-based blind channel estimation. Although there have been extensive works conducted on blind channel estimation in radio frequency communications [26]–[28], few studies on Poisson channel in optical wireless scattering communication have been reported, which motivates us to study the blind channel estimation for Poisson channel. It is worth mentioning that artificial neural network (ANN) has been applied to blind detection for Poisson channel [29] and achieves good performance, but the method is analytically intractable. In this paper, we mainly focus on the fundamental principles and algorithms of blind channel estimation, and some of the proposed blind channel estimation methods can approach the optimal bit-error rate performance. The main contributions of this paper can be summarized as follows:

- We show that Poisson channel does not have blind channel estimation ambiguity due to the intensity modulation. This fact reveals that Poisson channel can achieve full blind channel estimation without requiring preambles or coding-based methods like differential encoding to remove absolute phase information.
- We propose a modified constant modulus (CM) criterion for Poisson channel by adding a mean constaint to the constant modulus cost, and show the preference of using baud-spaced equalizer BSE) and fractionally spaced equalizer (FSE) under gradient descent structure.
- We propose a subspace-based channel estimation (SS-CE) for Poisson channel based on the second-order statistics. It is shown that the proposed subspace-based method does not require the knowledge of background radiation noise, and most importantly, does not require any preambles. Simulation results show that, unlike the Gaussian channel where the SS-CE has high data efficiency, this blind channel estimation demands a large amount of received symbols due to Poisson stochastic process. Moreover, the channel estimation accuracy is sensitive to the eigenvalue threshold. In particular, the eigenvalue threshold should match the data size of the received symbols.
- An iterative channel estimation (ICE) method is proposed and proved to have high data efficiency and self-adaption for channel variation. This method uses a short initial pilot to obtain a rough initial channel estimate and updates the channel estimate by applying the blockwise detected data symbols iteratively. The effectiveness of this data-aided channel estimation is evaluated by analyzing its asymptotic unbiasedness and convergence.

Compared with existing works, our paper considers more general channel estimation methods without any prior knowledge on the channel propagation function. Our goal does not focus on the trade-off between computational complexity and estimation performance, but focus on the estimation performance assessd by a benmark selected to be the optimal bound (achieved by the maximum-likelihood dynamic programming).

The reminder of this paper is organized as follows. In Section II, we prove that Poisson channel with on-off keying (OOK) modulation does not have blind channel estimation ambiguity. In Section III, we propose fractionally spaced channel estimation methods for Poisson channel, including a modified CMA and the SS-CE using the second-order statistics. In Section IV, the data-aided ICE is proposed and evaluated. In Section V, we numerically evaluate the proposed blind and semi-blind channel estimation approaches. Finally, Section VI concludes this paper.

II. Blind-Estimability of Poisson Channel

A. System Model

We only consider OOK modulation is this paper, since it is a simple but promising modulation especially in weak signal intensity regime, which is capacity-approaching under continuous-time Poisson channel model [30]. As shown in Fig. 1, we consider an OOK modulation over a multi-tap Poisson channel. The channel can be divided into two parts. The first part is a standard linear time-invariant (LTI) channel that determines the mean of the output signal, and the second part is a memoryless stochastic Poisson process. Let $\{x_k\} \in \{0, 1\}$ denote the transmission symbols and $\{r_k\} \in \mathbb{Z}^+$ denote the received signal characterized by discrete photons; let $\boldsymbol{h} = [h_0, h_1, \ldots, h_{D-1}]^T$ denote the LTI channel impulse response; The received signal subjects to Poisson distribution, given by

$$r_k \sim \mathcal{P}\left(\sum_{d=0}^{D-1} P x_{k-d} h_d + \beta\right) \tag{1}$$

where $\mathcal{P}(\cdot)$ denotes Poisson random process function with respect to its mean; P is the transmission power of symbol "1"; D is the channel tap length; and β is the background noise

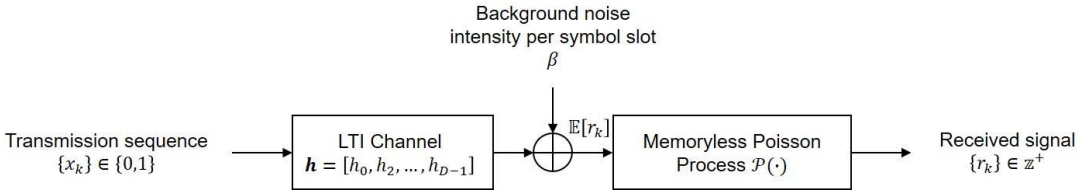

Fig. 1. The multi-tap Poisson channel model.

intensity per symbol slot, which is the mean of Poisson-distributed noise.

In the memoryless stochastic Poisson process, the output signal is a Poisson stochastic process, where the probability mass function of the received signal is given by

$$\mathbb{P}(r_k = n) = \frac{\left(\sum_{d=0}^{D-1} P x_{k-d} h_d + \beta\right)^n}{n!} \cdot \exp\left(-\sum_{d=0}^{D-1} P x_{k-d} h_d - \beta\right). \quad (2)$$

B. Blind Estimation Ambiguity

A fundamental question on blind channel estimation for Poisson channel is whether there exists estimation ambiguity, i.e., whether the channel and the transmission sequence can be uniquely determined. For coherent RF communication, typically if the modulation constellation shows rotational symmetric properties, another legal transmission sequence that differs from the true sequence can be obtained by rotating the true modulation symbols in one direction and channel impulse response in the opposite direction. However, when OOK modulation is adopted for Poisson channel, we can prove that the ambiguity does not exist.

Definition 1 (Distinguishable and Undistinguishable): For two nonzero channels $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$, and for any nonzero symbol sequence $\boldsymbol{x}_a$, if we can find another nonzero symbol sequence $\boldsymbol{x}_b$ satisfying $\boldsymbol{x}_b \neq \boldsymbol{x}_a$, such that $\boldsymbol{h}_a \star \boldsymbol{x}_a = \boldsymbol{h}_b \star \boldsymbol{x}_b$ where "$\star$" denotes the convolution operation, we say $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$ are *undistinguishable*. Otherwise, we say that $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$ are *distinguishable*.

Remark 1: It is seen that the undistinguishablity is strict according to this definition, since for any nonzero symbol sequence its conterpart should exist. Otherwise, if we find even one channel coefficient vector such that its conterpart does not exist, the two channels are distinguishable. The reason why we made such definition is that, if a channel coefficient vector without its conterpart exists, we can use it to distinguish the two channels.

Theorem 1 (Inexistence of Channel Estimation Ambiguity): For the Poisson channel under OOK modulation, any two nonzero legitimate channel coefficient vectors $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$ are *undistinguishable*. In other words, the Poisson channel under OOK modulation does not have channel estimation ambiguity.

Proof: We prove via contradiction. Suppose that the channel estimation ambiguity exists, which implies for an arbitrary transmission sequence $\boldsymbol{x}_a$, we can always find a channel $\boldsymbol{h}_b \neq \boldsymbol{h}_a$ such that $\boldsymbol{h}_a \star \boldsymbol{x}_a = \boldsymbol{h}_b \star \boldsymbol{x}_b$.

Let the actual channel be $\boldsymbol{h}_a$. Let $\boldsymbol{x}_a = \delta(k) = [1]$, i.e., a Kronecker delta sequence. According to the assumption that channel estimation ambiguity exists, there exists another pair $(\boldsymbol{h}_b, \boldsymbol{x}_b)$ such that $\boldsymbol{h}_a \star \boldsymbol{x}_a = \boldsymbol{h}_b \star \boldsymbol{x}_b$ with $\boldsymbol{h}_a \neq \boldsymbol{h}_b$ and $\boldsymbol{x}_a \neq \boldsymbol{x}_b$.

We assume that all legitimate sequences should begin and end with symbol "1". Let m be the length of $\boldsymbol{x}_b$. Consider another sequence $\boldsymbol{x}_m = \delta(k) + \delta(k-m+1) = \underbrace{[1, 0, \ldots, 0, 1]}_{m}$ with $m \geq 2$. We have that

$$\begin{aligned} \boldsymbol{h}_a \star \boldsymbol{x}_m &= \boldsymbol{h}_a \star (\boldsymbol{x}_a + \boldsymbol{x}_a(m-1)) \\ &= \boldsymbol{h}_b \star \boldsymbol{x}_b + \boldsymbol{h}_b \star \boldsymbol{x}_b(m-1) \\ &\triangleq \boldsymbol{h}_b \star \boldsymbol{x}'_m \end{aligned} \quad (3)$$

where $\boldsymbol{x}_a(m-1)$ is the shifted sequence that starts with the mth symbol of $\boldsymbol{x}_a$. Note that the mth symbol of $\boldsymbol{x}'_m$ is 2, which contradicts to the fact that $\boldsymbol{x} \in \{0, 1\}$. The only argument left to prove is that for $\boldsymbol{h}_b \star \boldsymbol{x} = 0$ implies that $\boldsymbol{x} = 0$. This can be readily proved based on the fact that all taps of $\boldsymbol{h}_b$ are non-negative. Therefore, Theorem 1 is proved. ■

Remark 2: Theorem 1 implies that, for any suspicious channel vector, we can transmit at least one corresponding sequence to remove the suspect. Since the length of suspicious channel vectors are restricted, the length of those delicately chosen sequences also have restricted length. Hence, in the worst case, we can transmit a finite-length sequence that covers all the delicately chosen sequences to make the real channel uniquely determined.

According to Theorem 1, we can perform full-blind channel estimation on Poisson channels under OOK modulation. While for Gaussian channel with phase-shift keying modulation, only semi-blind channel estimation can be performed by adding several reference symbols to determine the actual phases of symbol sequence and the channel.

III. Fractionally Spaced Blind Channel Estimation

The blind channel estimation can be realized by constructing redundent channel state information, which can be achieved by fractionally spaced signals. In this section, two representative methods are proposed.

A. Modified Constant Modulus Algorithm

A linear equalizer can be classified into FSE and BSE [20], where FSE corresponds to the oversampled signal and BSE

corresponds to the baud-spaced signal. For BSE, the equalized output is given by

$$y_k = \sum_{i=0}^{F-1} (r_{k-i} - \beta) f_i \tag{4}$$

where r_k is the received signal in symbol slot k; $\boldsymbol{f} = [f_0, f_1, \ldots, f_{F-1}]^T$ is the BSE tap vector; and F is the tap length of $\boldsymbol{f}$.

For FSE, the equalized output is given by

$$y_k = \sum_{i=0}^{F-1} \left(\boldsymbol{r}_{k-i}^{\mathrm{FS}} - \frac{\beta}{M}\mathbf{1} \right)^T \boldsymbol{f}_i^{\mathrm{FS}} \tag{5}$$

where $\boldsymbol{r}_k^{\mathrm{FS}} \triangleq [r_{kM+M-1}^{\mathrm{FS}}, r_{kM+M-2}^{\mathrm{FS}}, \ldots, r_{kM}^{\mathrm{FS}}]^T$ is the fractionally spaced symbol vector sampled from the baud-spaced symbol r_k; $\boldsymbol{f}^{\mathrm{FS}}$ is the FSE tap vector; $\boldsymbol{f}_i^{\mathrm{FS}}$ is the ith subvector of $\boldsymbol{f}^{\mathrm{FS}}$ with M entries; M is the space time or oversampling order of FSE.

It can be readily obtained that the BSE is equivalent to an elaborately designed FSE where each tap coefficient of the FSE corresponding to a oversampled symbol is chosen to be that of the BSE, which yields

$$\boldsymbol{f}_*^{\mathrm{FS}} = [\underbrace{f_0, \ldots, f_0}_{M}, \underbrace{f_1, \ldots, f_1}_{M}, \ldots, \underbrace{f_{F-1}, \ldots, f_{F-1}}_{M}]. \tag{6}$$

This property holds based on the fact that the summation of all fractionally spaced samples of a symbol equals the baud-spaced sample due to the Poisson channel nature. Since BSE can be realized by FSE, the optimal solution of the FSE cannot be worse than that of the BSE theoretically. However, oversampling in Poisson channel reduces the accuracy of fractionally spaced symbols and meanwhile increases searching complexity of $\boldsymbol{f}$, which may lead to convergence issues. Hence, the limitation of searching complexity may result in worse performance of FSE compared with BSE. The preference of the BSE and the FSE will be discussed according to the numerical simulation results in Section V.

CMA is a well-known blind equalization algorithm for its low complexity and ease of implementation. CMA is designed for the signals with constant modulus and zero mean. If the output of the equalizer also satisfies constant modulus, then the output can be regarded as an estimate of the transmission signal [16]. The CM cost function is defined to show the deviation of the output from the transmission signal, and it is defined as

$$J_{CM} = \mathbb{E}\left[\left(1 - |y_k|^2\right)^2 \right]. \tag{7}$$

Note that for OOK modulation, symbol $x_k \in \{0, 1\}$ no longer has a constant squared-modulus or a zero mean. However, the distances to $1/4$ for symbols zero and one are constant. Let y_k denote the equalizer output. To construct a constant squared-modulus, we consider the following constant modulus term

$$J_{\mathrm{CM}} = \mathbb{E}\left[\left(\frac{1}{4} - \left| y_k - \frac{1}{2} \right|^2 \right)^2 \right] = \mathbb{E}\left[\left(y_k^2 - y_k \right)^2 \right]. \tag{8}$$

However, eq. (8) still needs to be improved since $y_k = 0$ is a legitimate detected symbol and can be achieved by a null equalizer. To avoid this, we add the following mean constraint

$$J_{\mathrm{Ave}} = \mathbb{E}\left[\left(\frac{1}{W} \sum_{k=w}^{w+W-1} y_n - \mu \right)^2 \right] \tag{9}$$

where W is the observation window and μ is the duty cycle, i.e., the percentage of symbol "1" among the transmission OOK symbols.

The optimization function is the CM cost plus the weighted average constraint, and the optimization problem becomes

$$\min_{\boldsymbol{f}^{\mathrm{FS}}} J_{\mathrm{CM}} + J_{\mathrm{Ave}} \tag{10}$$

where $\boldsymbol{f}^{\mathrm{FS}}$ is the FSE tap coefficient vector.

For the realization of CMA, we consider uniform step gradient descent to obtain the optimal equalizer. The derivatives of J_{CM} and J_{Ave} are given, respectively, by

$$\frac{\partial J_{\mathrm{CM}}}{\partial \boldsymbol{f}^{\mathrm{FS}}} = 2y_k(y_k - 1)(2y_k - 1)\left(\tilde{\boldsymbol{r}}_k^{\mathrm{FS}} - \frac{\beta}{M}\mathbf{1} \right)^T \tag{11}$$

and

$$\frac{\partial J_{\mathrm{Ave}}}{\partial \boldsymbol{f}^{\mathrm{FS}}} = \left[\frac{1}{W} \sum_{i=0}^{W-1} \left(\tilde{\boldsymbol{r}}_{k+i}^{\mathrm{FS}} - \frac{\beta}{M}\mathbf{1} \right)^T \boldsymbol{f}^{\mathrm{FS}} - \mu \right] \cdot \left(\frac{1}{W} \sum_{i=0}^{W-1} \tilde{\boldsymbol{r}}_{k+i}^{\mathrm{FS}} - \frac{\beta}{M}\mathbf{1} \right)^T \tag{12}$$

where $\tilde{\boldsymbol{r}}_k^{\mathrm{FS}} \triangleq \left[\left(\boldsymbol{r}_k^{\mathrm{FS}}\right)^T, \left(\boldsymbol{r}_{k-1}^{\mathrm{FS}}\right)^T, \ldots, \left(\boldsymbol{r}_{k-F+1}^{\mathrm{FS}}\right)^T \right]^T$, and W is the observation window on calculating the average of the received signal.

The equalizer tap update is given by

$$\boldsymbol{f}_{k+1}^{\mathrm{FS}} = \boldsymbol{f}_k^{\mathrm{FS}} - \left(\frac{\partial J_{\mathrm{CM}}}{\partial \boldsymbol{f}^{\mathrm{FS}}} + \frac{\partial J_{\mathrm{Ave}}}{\partial \boldsymbol{f}^{\mathrm{FS}}} \right) \cdot \Delta \tag{13}$$

where Δ is the stepsize.

B. Second-Order Statistics Based Blind Channel Estimation

Inspired by the idea for the Gaussian channel, we propose an SS-CE using the second-order statistics, which also requires fractionally spaced signal [21]. The key idea is to use the second-order statistics of the oversampled received sequence, since oversampling can form a cyclostationary sequence that provides phase information. Similar to the FSE structure proposed in Sec. III-A, the subspace-based method also processes fractionally spaced signals.

The equivalent channel impulse response in the matrix form under M-order oversampling can be written as

$$\boldsymbol{H}_M = \begin{pmatrix} h_{D-1,1} & h_{D-2,1} & \cdots & h_{0,1} \\ \vdots & \vdots & & \vdots \\ h_{D-1,M} & h_{D-2,M} & \cdots & h_{0,M} \end{pmatrix}. \tag{14}$$

Then, the mean of the fractionally spaced received signal sequence $\mathbb{E}[\boldsymbol{r}_{\mathrm{FS}}]$ can be written as

$$\mathbb{E}[\boldsymbol{r}_{\mathrm{FS}}] = P\boldsymbol{H}\boldsymbol{x} + \frac{\beta}{M}\mathbf{1} \tag{15}$$

where $\boldsymbol{x}$ is the transmission sequence vector; β is the mean of the Poisson background noise intensity in a symbol slot; and $\boldsymbol{H} \in \mathbb{R}^{MK\times(D+K-1)}$ is the equivalent channel in matrix form, given by

$$\boldsymbol{H} = \begin{pmatrix} h_{D-1,1} & h_{D-2,1} & \cdots & h_{0,1} & & \\ \vdots & \vdots & & \vdots & & \\ h_{D-1,M} & h_{D-2,M} & \cdots & h_{0,M} & & \\ & h_{D-1,1} & h_{D-2,1} & \cdots & h_{0,1} & \\ & \vdots & \vdots & & \vdots & \\ & h_{D-1,M} & h_{D-2,M} & \cdots & h_{0,M} & \\ & & & & & \ddots \end{pmatrix} \tag{16}$$

where K, called the smoothing factor, is a positive parameter that guarantees more rows than columns of $\boldsymbol{H}$. This is because $\boldsymbol{H}$ should be a tall matrix to avoid loss of information of $\boldsymbol{H}\boldsymbol{H}^T$ when applying SS-CE. Thus, the following constraint should be satisfied

$$MK \geq D + K - 1. \tag{17}$$

Constraint (17) implies $M \geq 2$. It is seen that (17) always holds for $K = D$, and thus for simplicity we choose $K = D$, where $\boldsymbol{H}$ is now an $MD \times (2D-1)$ matrix.

We have the following results on the second-order statistics of the received signal.

Theorem 2: Let $\boldsymbol{x}(k) \triangleq [x_k, x_{k+1}, \ldots, x_{k+2D-2}]^T$. Let $\boldsymbol{r}(k)$ denote the received sequence corresponding to $\boldsymbol{x}(k)$ through (15). We have

$$\boldsymbol{H}\boldsymbol{H}^T = \frac{1}{(\mu-\mu^2)P^2}\left(\boldsymbol{R}(0) - \boldsymbol{R}(2D) - \text{diag}^{\frac{1}{2}}\left(\boldsymbol{R}(2D)\right)\right), \tag{18a}$$

$$\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^T = \frac{1}{(\mu-\mu^2)P^2}\Big(\boldsymbol{R}(1) - \boldsymbol{R}(2D) - \text{diag}^{\frac{1}{2}}\left(\boldsymbol{R}(2D)\right)\cdot \boldsymbol{J}^M\Big) \tag{18b}$$

where $\boldsymbol{R}(0) = \mathbb{E}[\boldsymbol{r}_{\text{FS}}(k)\boldsymbol{r}_{\text{FS}}^T(k)]$; $\boldsymbol{R}(1) = \mathbb{E}[\boldsymbol{r}_{\text{FS}}(k-1)\boldsymbol{r}_{\text{FS}}^T(k)]$; $\boldsymbol{R}(2D) = \mathbb{E}[\boldsymbol{r}_{\text{FS}}(k-2D)\boldsymbol{r}_{\text{FS}}^T(k)]$ and $\boldsymbol{J}$ is a shifting matrix given by

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{0}^T & 0 \\ \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}. \tag{19}$$

Proof: See Appendix A. ∎

Once the two terms $\boldsymbol{H}\boldsymbol{H}^T$ and $\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^T$ are obtained, equivalent channel matrix $\boldsymbol{H}$ and channel taps $\boldsymbol{H}_M$ can be obtained by the standard subspace-based channel estimation method proposed in [21, eqs. (66)-(76)].

It is seen form (18) that the second-order statistics have a positive real scalar of ambiguity $\frac{1}{P^2}$ for unknown P. However, such ambiguity does not influence the signal detection since the product of the transmission power P and the channel taps $\boldsymbol{h}$ is fixed.

An important issue is how to select the dominant subspace according to eigenvalue decomposition of $\boldsymbol{H}\boldsymbol{H}^T$ by choosing eigenvalue threshold λ_{th}. Only the eigenvectors with the corresponding eigenvalues larger than threshold λ_{th} belong to the dominant subspace, which will be used to recover $\boldsymbol{H}$. The selection of λ_{th} depends on the received symbol number used for one-time channel estimation and will be discussed in Section V-A.

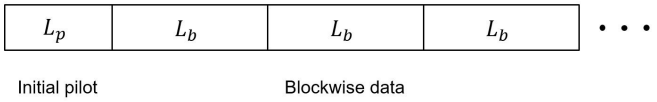

Fig. 2. The schematic diagram of the proposed data-aided based iterative channel estimation and signal detection.

C. Complexity Analysis

For the Modified CMA in Poisson channel, the computation mainly lies in the calculation of the two derivatives in each iteration, which yields the computational complexity of approximately $O(WFL)$. Assume that L is the sequence length. For the SS-CE, the computation requires calculating three expectations $R(0)$, $R(1)$ and $R(2D)$, which yields the computational complexity of approximately $O(12D^2L)$.

IV. Iterative Channel Estimation

The methods in Section III do not need pilots and thus can reduce the overhead. Another typical semi-blind channel estimation only applies quite few pilots to reduce the overhead, namely, the iterative channel estimation (ICE). It applies an iterative structure to update the channel estimate and initialize using a short pilot sequence, which is also known as data-aided channel estimation.

In this section, we propose an ICE method that iteratively updates the channel estimate using blockwise data symbols, as shown in Fig. 2, where L_p is the initial pilot length and L_b is the data block length. In this method, each iteration has two phases: channel estimation phase and signal detection phase. In the signal detection phase, the ICE estimator first detects the new arrival L_b-length data symbol block using the current channel estimate by the maximum-likelihood dynamic programming (MLDP). Then, in the channel estimation phase, the ICE estimator treats the previous detected data symbols as additional pilots to update the channel estimate by applying the unbiased channel estimator proposed in [10]. It should be mentioned that the unbiased LS channel estimator adopts all detected symbols in the previous blocks to increase the estimation accuracy. During the iteration, an observation window W_b (the maximum data symbol block number for channel estimation in the channel estimation phase) can be introduced to handle the time varying channel and reduce the amount of calculation. Such method is also known as data-aided channel estimation.

A. Bias of ICE

A fundamental problem is to study the channel estimation performance when unbiased LS channel estimation uses a contaminated pilot, i.e., a pilot contains error symbols and is different from what we expect.

Let $\{x_k\}_{k=0}^{L-1}$ be the true pilot and $\{\hat{x}_k\}_{k=0}^{L-1}$ be the contaminated pilot having bit-error rate ϵ, where L is the pilot length. Let $\boldsymbol{r} \triangleq [r_{D-1}, r_D, \ldots, r_{L-1}]^T$ be the received signal sequence. The channel estimates based on the true pilot and based on the contaminated pilot are, respectively, given by

$$\hat{\boldsymbol{h}} = \frac{1}{P}\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}(\boldsymbol{r} - \beta\mathbf{1}) \tag{20}$$

and

$$\hat{\boldsymbol{h}}_d = \frac{1}{P}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\right)^{-1}\hat{\boldsymbol{X}}(\boldsymbol{r} - \beta\mathbf{1}) \tag{21}$$

where $\boldsymbol{X}$ is a Toeplitz Matrix of the pilot symbols $\{x_k\}_{k=0}^{L-1}$ and it is given by

$$\boldsymbol{X} = \begin{pmatrix} x_{D-1} & x_D & \cdots & x_{L-1} \\ x_{D-2} & x_{D-1} & \cdots & x_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ x_0 & x_1 & \cdots & x_{L-D} \end{pmatrix}; \tag{22}$$

$\hat{\boldsymbol{X}}$ is a Toeplitz matrix defined in the same manner as $\boldsymbol{X}$ from $\{\hat{x}_k\}_{k=0}^{L-1}$.

We study first the expectation of the iterative channel estimation.

Lemma 1: For $\forall\ \varepsilon > 0$, we have (23), shown at the bottom of the page, where $\boldsymbol{I}$ is a D by D identity matrix; and $\boldsymbol{E}$ is a D by D all-one square matrix (see Appendix B for proof).

Remark 3: When L approaches infinity, both $(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T)^{-1}$ and its approximation $\frac{1}{(\mu-\mu^2)(L-D+1)}(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E})$ approach null matrix. However, the difference between them can be ignored. With the analytic expression of $(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T)^{-1}$, we can formulate the expectation of the LS channel estimation using contaminated pilot.

Theorem 3: The unbiased LS channel estimation using a contaminated pilot becomes biased, where the bias is a function of channel taps $\boldsymbol{h}$ and bit-error rate ϵ of the detected data symbols. When the pilot length L is sufficiently long, we have the expectation of the channel estimate given by (24), shown at the bottom of the page.

Proof: See Appendix C. ∎

Remark 4: The bias is given by

$$\begin{aligned}\text{bias} &= \boldsymbol{h} - \mathbb{E}[\hat{\boldsymbol{h}}_d] \\ &= \frac{\epsilon}{1-\mu}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right) \\ &\quad \times ((2-2\mu)\boldsymbol{I} + (2\mu-1)\boldsymbol{E})\cdot\boldsymbol{h}\end{aligned} \tag{25}$$

which is a function of the real channel $\boldsymbol{h}$ and bit-error rate ϵ of the detected data symbols.

Remark 5: When bit-error rate ϵ approaches zero, the channel estimation using the contaminated pilot can approach the actual channel $\boldsymbol{h}$. The channel estimation bias is essentially caused by non-zero mean OOK signal $\{x_k\} \in \{0, 1\}$, which would be zero if the input signal was zero mean. In case of $\mu = 0.5$, eq. (24) can be simplified to

$$\mathbb{E}[\hat{\boldsymbol{h}}_d] \approx \left(\boldsymbol{I} - 2\epsilon\left(\boldsymbol{I} - \frac{1}{D+1}\boldsymbol{E}\right)\right)\boldsymbol{h}. \tag{26}$$

The square of bias is then given by

$$\begin{aligned}\text{bias}^2 &= \left\|2\epsilon\left(\boldsymbol{I} - \frac{1}{D+1}\boldsymbol{E}\right)\boldsymbol{h}\right\|_2^2 \\ &= 4\epsilon^2\boldsymbol{h}^T\left(\boldsymbol{I} - \frac{D+2}{(D+1)^2}\boldsymbol{E}\right)\boldsymbol{h}.\end{aligned} \tag{27}$$

B. Channel Estimation MSE of ICE

Under the proposed iterative channel estimation and signal detection, we guarantee that the channel estimate becomes more accurate and the bit-error rate decreases after each iteration, i.e., the iterative channel estimation and signal detection converge. For simplicity, we consider $\mu = 0.5$.

To study the convergence of the iterative channel estimation and signal detection, we need to obtain the mean squared error (MSE) of each step in the channel estimation process. We first introduce the channel estimation MSE of the unbiased LS channel estimation without pilot contamination.

For perfect pilot without contamination, the expectation of the unbiased LS channel estimation MSE under Poisson channel with $\mu = 0.5$ is given by

$$\begin{aligned}\mathbb{E}\left[\text{MSE}_U\right] \cong{}& \frac{2(D^3+D^2+2)\|\boldsymbol{h}\|_1}{P\,(L-D+1)\,(D+1)^2} \\ &+ \frac{4\beta D(D^2+D-1)}{P^2(L-D+1)(D+1)^2}.\end{aligned} \tag{28}$$

Please see Appendix D for the justification of (28).

Furthermore, given contaminated pilot with error probability ϵ, the expectation of the LS channel estimation MSE under Poisson channel with $\mu = 0.5$ is given by

$$\mathbb{E}\left[\text{MSE}_C\right] = \mathbb{E}\left[\text{MSE}_U\right] + \mathbb{E}\left[\text{MSE}_S\right] + \text{bias}^2 \tag{29}$$

where $\mathbb{E}\left[\text{MSE}_U\right]$ is the unbiased LS estimation MSE given by (28); bias is the bias of LS estimation using contaminated

$$\lim_{L\to\infty}\mathbb{P}\left(\frac{\left\|\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\right)^{-1} - \frac{1}{(\mu-\mu^2)(L-D+1)}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)\right\|_F}{\left\|\frac{1}{(\mu-\mu^2)(L-D+1)}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)\right\|_F} \le \varepsilon\right) = 1 \tag{23}$$

$$\mathbb{E}[\hat{\boldsymbol{h}}_d] \to \left(\boldsymbol{I} - \frac{\epsilon}{1-\mu}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)((2-2\mu)\boldsymbol{I} + (2\mu-1)\boldsymbol{E})\right)\boldsymbol{h} \quad \text{in probability.} \tag{24}$$

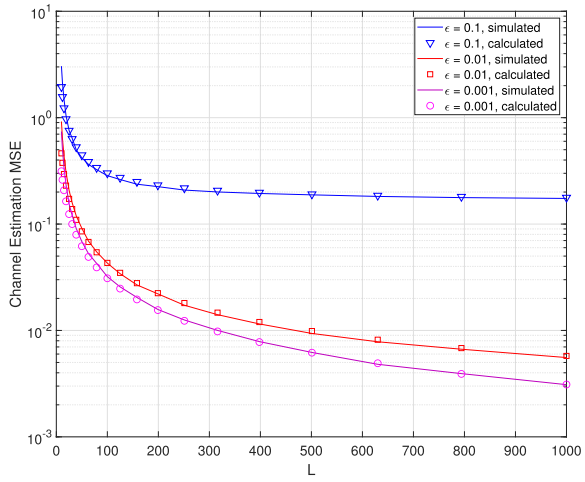

Fig. 3. Channel estimation MSE versus length of the contaminated pilot for both simulation and calculation results. The channel tap is $\boldsymbol{h} = [2.539, 1.561, 0.378, 0.053]^T$. The transmission power of symbol "1" $P = 10$ W. The background radiation intensity is 0.01 photons per symbol slot. The simulated results is averaged over 500 realizations. The calculation uses (29).

pilot given by (27); and MSE_S is the part caused by the random position of the contaminated pilot symbols, given by

$$\mathbb{E}\left[\text{MSE}_S\right] \approx \frac{4\epsilon(1-\epsilon)}{L-D+1} \cdot \left(\frac{D}{D+1}\sum_{i=0}^{D-1} h_i^2 - \frac{1}{(D+1)^2}\sum_{i\neq j}|i-j|h_i h_j\right). \tag{30}$$

Please refer to Appendix E for the justification of (29).

Figure 3 shows the comparison between the simulated and the calculated results of $\mathbb{E}\left[\text{MSE}_C\right]$. It is seen that the calculated MSE fits the simulated one well, especially for long L, which verifies (28) and (29).

C. Convergence Analysis

Based on the derived channel estimation MSE given by (28), (29) and (30), now we are able to analyze under what condition the proposed ICE can converge.

The channel estimation phase in the ICE is essentially a channel estimation with contaminated pilot. The convergence of the proposed ICE implies that the following inequalities hold

$$\mathbb{E}\left[\text{MSE}_C^{(k+1)}\right] \leq \mathbb{E}\left[\text{MSE}_C^{(k)}\right] \tag{31}$$

$$\epsilon_{k+1} \leq \epsilon_k \tag{32}$$

where $\text{MSE}_C^{(k)}$ is the channel estimation MSE aided by k blocks of data sequence; and ϵ_k is the bit-error rate with aid of k data blocks, i.e., the bit-error rate expectation in the $(k+1)$th MLDP phase.

Although we cannot obtain an explicit expression of bit-error rate ϵ under MLDP signal detection, we know that it becomes smaller when channel estimation accuracy increases. Thus, given $\mathbb{E}\left[\text{MSE}_C^{(k)} \leq \text{MSE}_C^{(k-1)}\right]$, we can conclude that $\epsilon_k \leq \epsilon_{k-1}$. Further, $\mathbb{E}\left[\text{MSE}_C^{(k+1)}\right]$ is based on ϵk, which implies applying the mathematical induction method for convergence justification.

Our intention is to find under what condition can we derive $\mathbb{E}\left[\text{MSE}_C^{(k+1)}\right] \leq \mathbb{E}\left[\text{MSE}_C^{(k)}\right]$ based on $\epsilon_k \leq \epsilon_{k-1}$. By the assumption that $D \ll L$ and $\tilde{\epsilon}_k \ll 1$, and according to (29), we have that

$$\mathbb{E}\left[\text{MSE}_C\right] \approx \frac{a}{L} + \frac{b\tilde{\epsilon}_k(1-\tilde{\epsilon}_k)}{L} + c\tilde{\epsilon}_k^2 \tag{33}$$

where $\tilde{\epsilon}_k \triangleq \frac{kL_b}{L_p+kL_b}\epsilon_k$; a, b and c are constant parameters given by

$$a \triangleq \frac{2(D^3+D^2+2)\|\boldsymbol{h}\|_1}{P(D+1)^2} + \frac{4\beta D(D^2+D-1)}{P^2(D+1)^2}; \tag{34}$$

$$b \triangleq \frac{4}{L-D+1}\left(\frac{D}{D+1}\sum_{i=1}^{D} h_i^2 - \frac{1}{(D+1)^2}\sum_{i\neq j}|i-j|h_i h_j\right); \tag{35}$$

$$c \triangleq 4\boldsymbol{h}^T\left(\boldsymbol{I} - \frac{D+2}{(D+1)^2}\boldsymbol{E}\right)\boldsymbol{h}. \tag{36}$$

According to the simplification (33), the recurrence condition can be written as

$$\frac{a}{L_p+(k+1)L_b} + \frac{b\frac{kL_b}{L_p+kL_b}\epsilon_k}{L_p+(k+1)L_b} + c\left(\frac{kL_b}{L_p+kL_b}\right)^2\epsilon_k^2 \leq \frac{a}{L_p+kL_b} + \frac{b\frac{(k-1)\ L_b}{L_p+(k-1)\ L_b}\epsilon_{k-1}}{L_p+kL_b} + c\left(\frac{(k-1)L_b}{L_p+(k-1)L_b}\right)^2\epsilon_{k-1}^2. \tag{37}$$

Since $\epsilon_k \leq \epsilon_{k-1}$, we have the following tighter condition when replacing ϵ_{k-1} by ϵ_k in the right hand side of the inequality

$$\frac{a}{L_p+(k+1)L_b} + \frac{b\frac{kL_b}{L_p+kL_b}\epsilon_k}{L_p+(k+1)L_b} + c\left(\frac{kL_b}{L_p+kL_b}\right)^2\epsilon_k^2 \leq \frac{a}{L_p+kL_b} + \frac{b\frac{(k-1)\ L_b}{L_p+(k-1)\ L_b}\epsilon_k}{L_p+kL_b} + c\left(\frac{(k-1)L_b}{L_p+(k-1)L_b}\right)^2\epsilon_k^2. \tag{38}$$

After some simplifications, we have

$$c\left[\left(\frac{kL_b}{L_p+kL_b}\right)^2 - \left(\frac{(k-1)L_b}{L_p+(k-1)L_b}\right)^2\right]\epsilon_k^2 \leq \frac{aL_b + b\frac{(k-1)L_b-L_p}{(k-1)L_b+L_p}\epsilon_k}{(L_p+(k+1)L_b)(L_p+kL_b)}. \tag{39}$$

Assuming $(k-1)L_b - L_p > 0$, which holds for large k and small L_p, we can eliminate the term with parameter b on the right hand side of (39). After some mathematical manipulations, we have the following tighter condition

$$\frac{L_p+(k+1)L_b}{L_p+(k-1)L_b}\cdot\left(\frac{kL_b}{L_p+kL_b} + \frac{(k-1)L_b}{L_p+(k-1)L_b}\right)L_p\epsilon_k^2 \leq \frac{a}{c}. \tag{40}$$

Note that the right hand side roughly has the same magnitude of D/λ_r, where λ_r is the mean received photon number per symbol. Whereas for the left hand side, when k approaches infinity, it converges to $2L_p\epsilon_k^2$. As long as the initial pilot length is not extremely long, the convergence condition is easy to satisfy. Empirically speaking, if the initial pilot is sufficiently long such that the initial channel estimate is sufficiently accurate, using the detected symbols as additional contaminated pilot will only decrease the channel estimation performance. However, since the intention of semi-blind channel estimation is to reduce the overhead, short L_p is desirable in real application. Therefore, we can conclude that the proposed ICE will most probably eventually converge when k is large.

D. Complexity Analysis

The computational complexity of the ICE mainly lies in the signal detection stage, where the maximum-likelihood dynamic programming is adopted. The complexity is $O(D^2L)$ when detecting an L-length sequence. Note that the entire sequence is divided into several L_b-length subsequences. When the ICE starts to detect a new subsequence, it also detects all the previous symbols. Consequently, the complexity is $O\left(\frac{D^2}{L_b}L^2\right)$.

V. Numerical Simulations

We evaluate the performance of the proposed blind and semi-blind channel estimation approach. Besides the normalized channel estimation MSE, we also evaluate the system detection error probability of the proposed channel estimation approach as the final performance evaluation metric.

A. Simulation Under a Constant Channel Realization

We consider a four-tap Poisson channel $\boldsymbol{h} = [2.539, 1.561, 0.378, 0.053]^T$ for optical wireless scattering communication under consideration, which is adopted in [10]. The $T/2$-spaced channel taps is $\boldsymbol{h}^{\mathrm{FS}} = [1.043, 1.496, 0.895, 0.666, 0.249, 0.129, 0.039, 0.015]$. The duty cycle is set to $\mu = 0.5$ and the transmission power P_m is half of that of symbol "1" $P_m = 0.5P$. The background noise intensity β is set to be 0.01 photons per symbol. Based on the received sequence generated by those parameters, we conduct the proposed blind and semi-blind channel estimation and plot the performances averaged by $1,000$ simulations.

Figure 4 plots the bit-error rate of the proposed CMA for both BSE and FSE. The observation window W is set to be 100. To increase the search capability, we perform N_{CMA} parallel CMA processes using different step sizes and different filter initials, where the final result is the one with minimum cost (10).

To increase the search capability of CMA, we perform CMA using different step size and different initial values of the filter taps, and the final result of the optimal filter is the one with the minimum CM cost. The initials of the filter $\boldsymbol{f}_{\mathrm{ini}}^{\{\mathrm{FS,BS}\}}$ are randomly chosen according to

$$\boldsymbol{f}_{\mathrm{ini}}^{\{\mathrm{FS,BS}\}} = \frac{\xi W}{F \cdot \|\tilde{\boldsymbol{r}}_k^{\mathrm{FS}}\|_1} \tag{41}$$

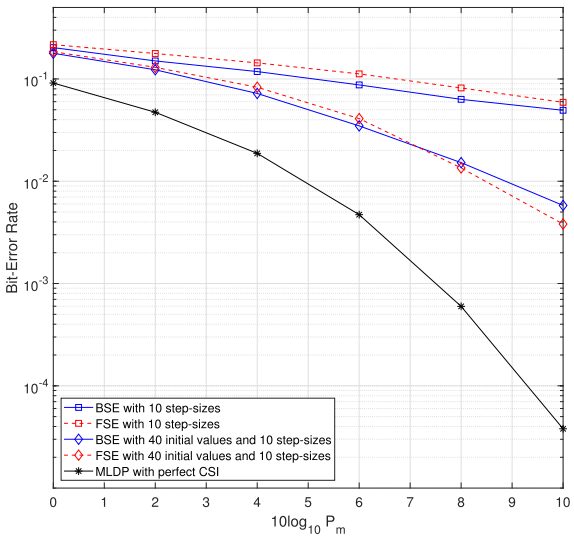

Fig. 4. Bit-error rate vs. the transmission power for both BSE and FSE structures with different number of initial values of the filter. The iteration number of CMA is set to be 10^4.

where $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_F]^T$ and $\{\xi_i\}_{i=1}^F$ are the independent and identically distributed (i.i.d.) random variables uniformly distributed in $[-1, 1]$.

For each equalizer type, we consider two scenarios: one scenario has 40 different initial values of the equalizer and the other scenario has only one initial value. Both scenarios are performed 10 times with each time using a certain step size. The results are obtained from the optimal ones with the minimum CM costs. It is seen that for only one initial value, BSE outperforms FSE; but for 40 different initial values, FSE has lower bit-error rate. This indicates that the FSE is more likely to fall into local optima and demands higher search complexity to achieve the global optima, since oversampling in Poisson channel increases the randomness of each sampled signal. On the contrary, the BSE can reach suboptimal performance with low search complexity. The black line is a benchmark that applies MLDP to signal detection with perfect CSI and has been plotted in [10, Fig. 5]. Clearly, CMA has far worse bit-error rate performance than MLDP with perfect CSI, which is reasonable since MLDP requires much higher computational complexity compared with that of the CMA.

Figure 5 plots the variations of bit-error rate and the CM cost, which is J_{CM} in (10), along with the iteration of the CMA for both BSE and FSE. It is seen that the bit-error rate remains unchanged when the number of iterations is less than 100, i.e., the observation window W, and that the bit-error rate decreases linearly with respect to the logarithm scale of the number of iterations when the number of iterations is larger than W. Therefore, it becomes more difficult to decrease the bit-error rate when the number of iterations is more than 10^4. Note that J_{CM} first increases and then decreases as the iteration progresses. Moreover, J_{CM} has smaller value at the beginning of iteration compared with that in the convergence phase. This tendency reveals that the CM cost J_{CM} cannot be directly applied as the CM criterion for Poisson channel and the average constraint of the equalized signal J_{Ave} is crucial for CMA to converge.

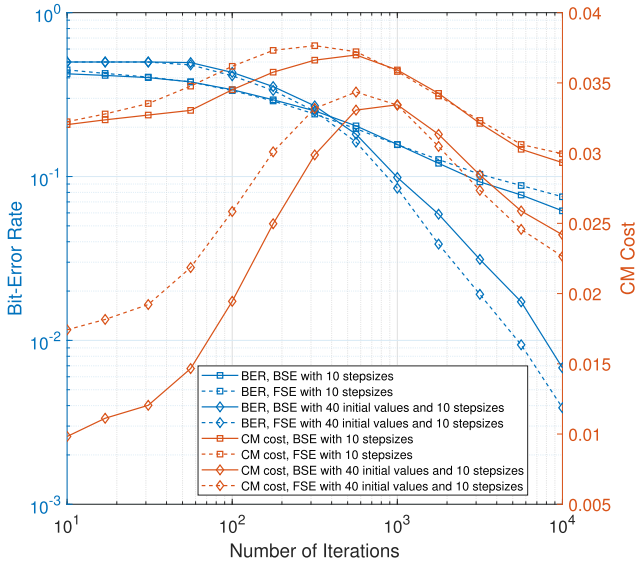

Fig. 5. The bit-error rate (blue figures) and CM cost J_{CM} (orange figures) vs. the number of iterations for both BSE and FSE structures with different number of initial values of the filter. The transmission power $P_m = 10$ W.

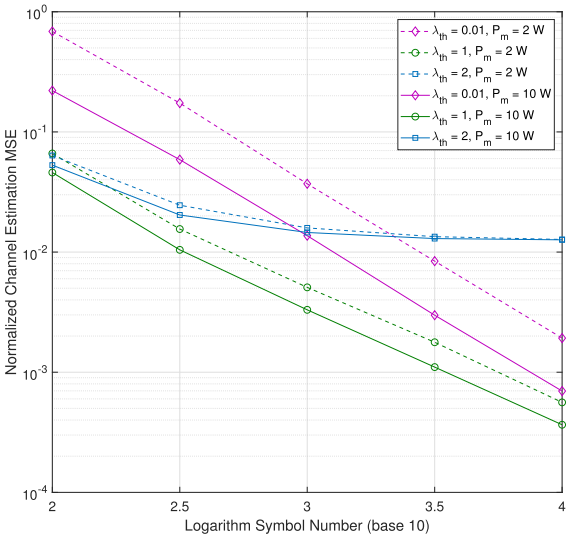

Fig. 6. The normalized channel estimation MSE versus the data length using the full-blind channel estimation algorithm.

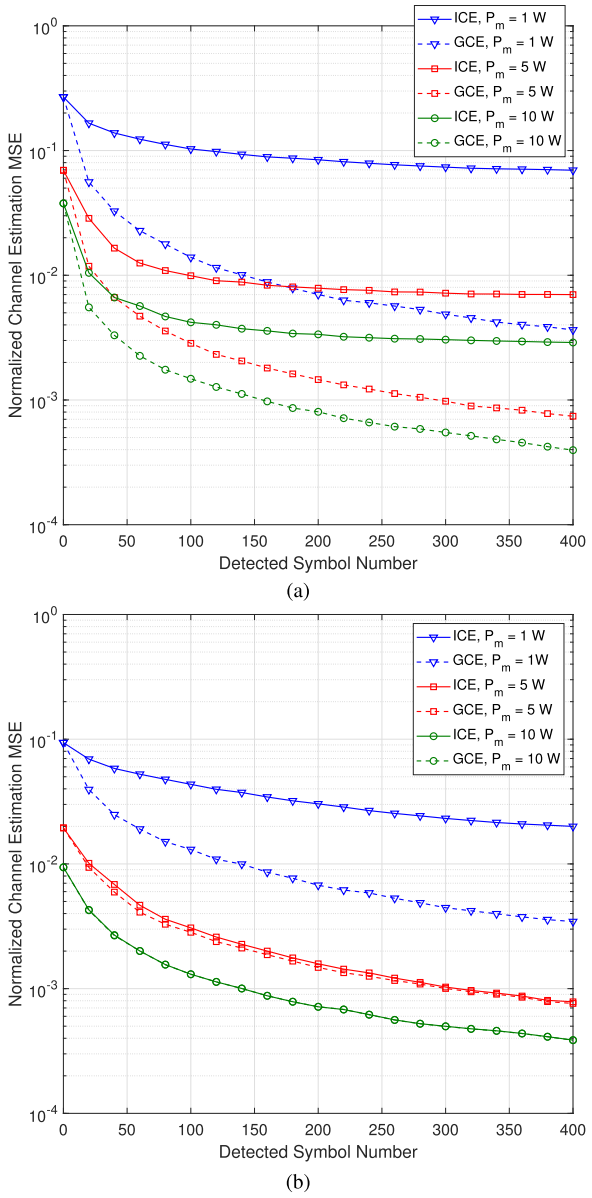

Fig. 7. Normalized channel estimation MSE versus the detected symbol number for (a) $L_p = 10$ and (b) $L_p = 20$.

Figure 6 plots the normalized channel estimation mean square error (MSE) with respect to the number of data symbols using SS-CE. The eigenvalues of the actual $\boldsymbol{H}\boldsymbol{H}^T$ sorted in descending order are 8.7375, 5.5144, 2.8204, 1.4195, 0.1062, 1.8364×10^{-5}, 1.5292×10^{-6} and 0. Thus, the thresholds are set to be 0.01, 1 and 2 so that the numbers of the eigenvalues used for reconstructing $\boldsymbol{H}$ are different. It is seen that the eigenvalue thresholds determine the achievable minimum normalized channel estimation MSE regardless of the transmission power and the number of data symbols, as shown by the curves under $\lambda_{\mathrm{th}} = 2$. Note that the curves of smaller thresholds have steeper slopes, which means lower channel estimation MSE can be achieved for sufficient data symbols. Besides, the slopes of curves having the same eigenvalue threshold and different transmission powers are approximately the same before the curves converge to the minimal MSEs. Therefore, the selection of λ_{th} depends on how many data symbols used for one time channel estimation but hardly depends on transmission power.

Figures 7(a) and 7(b) plot the normalized channel estimation MSE versus the detected symbol number using the proposed ICE method. For comparison, we introduce a genius channel estimation (GCE) that adopts all the transmitted symbols for channel estimation and pretends not to know them during signal detection. The performance of the GCE acts as an upper bound. The data block length are both set to be $L_b = 20$. Three different transmission powers are considered. It is seen that the iterative channel estimation has a large gap compared with GCE for $L_p = 10$, but is almost undistinguishable to GCE for $L_p = 20$ and $P_m \geq 5W$. This phenomenon indicates that the initial channel estimate should be accurate enough

TABLE I
A BRIEF COMPARISON OF THE PROPOSED THREE BLIND OR SEMI-BLIND CHANNEL ESTIMATION METHODS

Method	Modified CMA	SS-CE	ICE
Preambles	$\leq D+F-1$	0	$\geq 2D-1$
Iterations needed	$> 10^4$	10^4	10^2
Computational complexity	$O(WFL)$	$O(D^2L)$	$O\left(\frac{D^2}{L_b}L^2\right)$
BER Performance	Far from optimality	Near-optimality (large L)	Near-optimality
Prior knowledge	Duty cycle μ, background radiation β	Duty cycle μ	Background radiation β

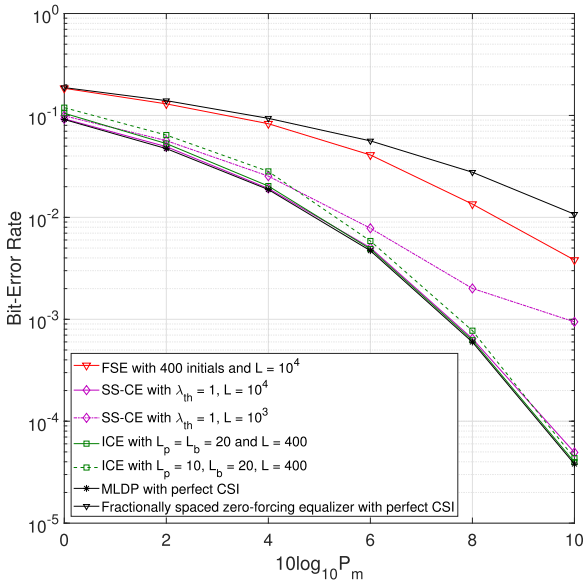

Fig. 8. Bit-error rate versus the transmission power for the proposed three blind channel estimation methods.

to initialize the iteration, which agrees with the convergence analysis in Section IV-B.

Figure 8 plots the bit-error rate performance versus the logarithm transmission power using the proposed three blind channel estimation methods. The fractionally spaced zero-forcing equalizer with perfect CSI serves as a benchmark of the proposed FSE equalizer, which is derived by letting the entire system impulse response (including the channel and the equalizer) be a delta sequence. It can be justified by the fact that the benchmark is essentially a zero-forcing equalizer where the Poisson signal-dependent noise is not considered. Note that except the modified CMA, both the SS-CE and the ICE have near optimal bit-error performances when some channel estimation parameters, for example the consumed data symbol length L of the SS-CE and the initial pilot length L_p of the ICE, are taken appropriate values. It is seen for the SS-CE, the lack of data symbols results in a large gap of bit-error rate, which becomes more significant for large transmission power. Furthermore, the ICE requires the smallest number of data symbols, therefore, it is a high data-efficient blind channel estimation method. The SS-CE, however, does not show high data efficiency since it requires large L to obtain an reliable channel estimate. A brief comparison of the proposed channel estimation methods is listed in Table I, shown on the top of this page.

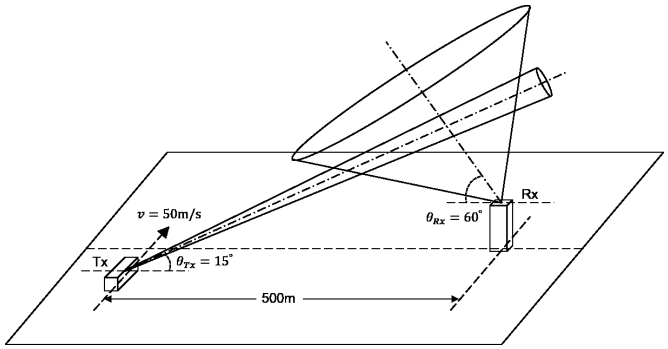

Fig. 9. The geometric diagram of the optical wireless scattering communication system.

B. Simulations in a Realistic Communication Scenario

We now evaluate three proposed blind channel estimation methods in a realistic channel-varying communication scenario and compare them with the conventional periodic pilot based channel estimation [31], [32]. The actual scene is chosen to be an optical wireless scattering communication where a mobile transmitter communicates with a fixed receiver through an NLOS link and the receiver receives photon-excited Poisson-distributed signals due to extremely large path loss.

The geometric diagram of the transceiver is shown in Fig. 9. The transmitter is assumed to move towards north at the speed of 50 m/s and the receiver is located 500 m away in the east direction of the transmitter in the beginning. The mounted UV LED array on the transmitter sends beam to the east with elevation angle $\theta_{\mathrm{Tx}} = 15°$, and the photon detector on the receiver heads due east with elevation angle set as $\theta_{Rx} = 60°$. The heights of the transmitter and the receiver are ignored. Besides the tangential movement, we also assume the transmission beam has a regular and uniform vibration in the vertical plane due to the jolt during moving. The amplitude of the vibration is $20°$ with a period $T_v = 0.5$ s.

We applied the Monte-Carlo ray tracing method to simulate the CIRs within 1 s transmission. It is seen that the channel coherent time corresponding to turbulence is around 1 s due to the multi-path of OWSC system, which is much longer than that of free space optics systems. In total, we simulated 500 CIRs during the transmission process, i.e., every 2 ms for simulating a CIR realization, assuming the CIR within 2 ms remains unchanged. Other simulation parameters are listed in Table II, shown on the top of the next page.

Figure 10 shows the bit-error rate of the three proposed blind channel estimation methods during the simulated

TABLE II
SIMULATION PARAMETERS

Symbol	Parameter representation	Value	Symbol	Parameter representation	Value
λ	Wavelength	260 nm	g	Atmosphere parameter [33]	0.72
k_r	Rayleigh scattering coefficient	0.266×10^{-3} m^{-1}	β_T	Tx beam divergence	20°
k_m	Mie scattering coefficient	0.284×10^{-3} m^{-1}	β_R	Rx FOV	90°
k_a	Absorbance coefficient	0.802×10^{-3} m^{-1}	S	Area of receiving aperture	1.77×10^{-4} m^2
γ	Atmosphere parameter [33]	0.017	τ	Symbol duration	2 μs
f	Atmosphere parameter [33]	0.5			

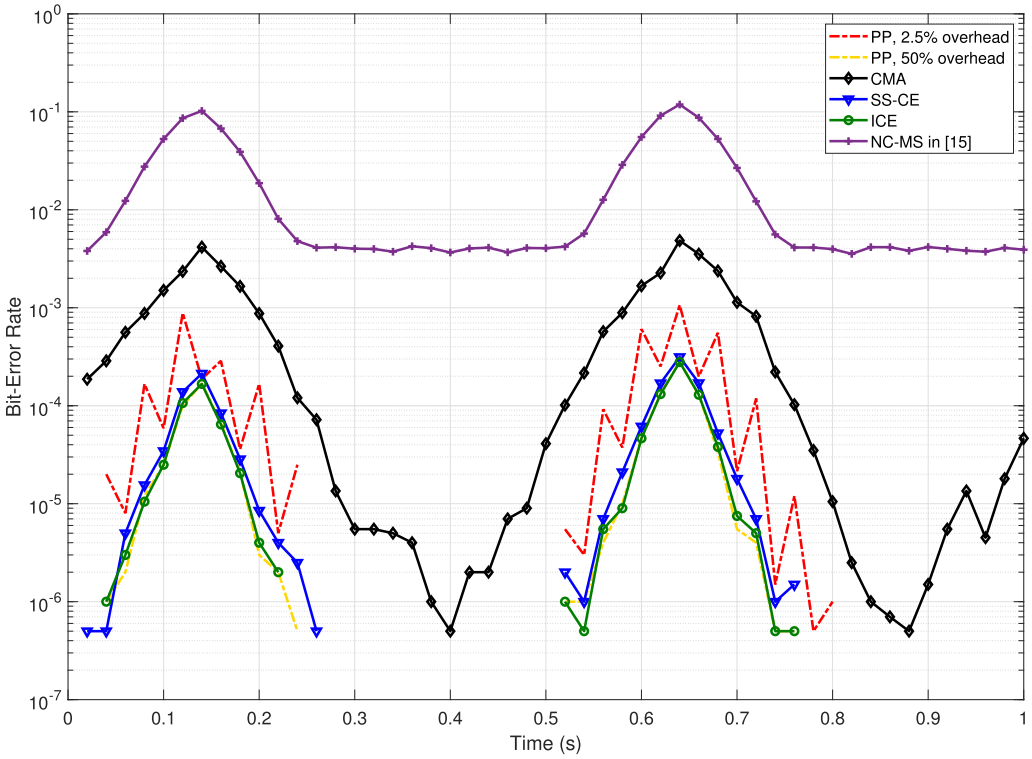

Fig. 10. The distribution of the bit-error rate over the simulated 1 second for the three proposed blind channel estimation methods. The transmission power is $P_m = 1$ W.

transmission process, and compares them with conventional periodic pilot (PP) based channel estimation and the non-coherent multiple-symbol (NC-MS) signal detection method proposed in [15]. Based on the simulation parameters, there are 500 different simulated channel realizations within the transmission process. The pilot length of PP based channel estimation is fixed at 500 symbols, which is half of one channel realization duration. The transmission process is divided into 50 intervals, and the bit-error rate is simulated for each interval. The bit-error rates are averaged over 200 trails to obtain stable results. For the modified CMA, 40 initials and 10 different stepsize are adopted; for the SS-CE, 10^4 symbols for one time channel estimation are adopted with λ_{th} set to be 0.05 times of the sum of all eigenvalues; and for the ICE, it uses 20-length initial pilot and 50-length data blocks for each iteration. It is seen that the fluctuation of bit-error rate is mainly determined by the variations of the angular parameters. The bit-error rate performances of the proposed SS-CE and ICE are close to that of the 50% overhead PP based channel estimation. When the overhead of the PP based channel estimation decreases to 2.5%, the bit-error rate performance degradation can be observed and is worse than that of SS-CE and ICE, even though the PP based channel estimation consumes more overhead. Hence, it can be concluded that the proposed SS-CE and ICE can remove at least 2.5% overhead. It seems that the reducted overhead is not observious, but for the system with higher angular variations and with multiple access, the reduced overhead can be more significant. Moreover, the bit-error rate under CMA is large due to its simple signal detection, while the other channel estimation methods have higher computational complexity. It can be seen that all the three proposed blind channel estimation methods outperform the non-coherent multiple-symbol (NC-MS) detector proposed in [15]. This is because the ISI which the authors in [15] assumed negligible is quite serious in our simulation scenario.

VI. CONCLUSION

We studied blind and semi-blind channel estimation methods for Poisson channel and proved that Poisson channel with OOK modulation does not have estimation ambiguity. Furthermore, three methods are designed against the

Poisson-distributed signal, and show different results and properties compared with those in the Gaussian channel. For the modified CMA, the mean constraint needs to be added to avoid the equalizer converging to null. For the SS-CE, more second-order statistics are required to apply the classical subspace approach. Moreover, the ICE is proposed and proved to be a biased channel estimation method where the bias is proportional to the bit-error rate.

Based on the simulation results through a constant channel scenario and a varying channel scenario, we can observe that the modified CMA has the worst performance but requires the lowest computational complexity. The SS-CE can achieve near optimal performance when the data length applied for one time channel estimation is sufficiently long. For the ICE, its bit-error rate performance can converge to genius channel estimation, and high data efficiency can be guaranteed at the cost of higher computational complexity.

APPENDIX A
PROOF OF THEOREM 2

Note that for OOK modulation $x \in \{0, 1\}$, the autocorrelation matrix of the transmission sequence is given by

$$\mathbb{E}\left[\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right] = (\mu - \mu^2)\boldsymbol{I} + \mu^2\boldsymbol{E} \tag{42}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{E}$ is an all-one square matrix.

Let $\mathcal{P}$ represent Poisson random process, where $\mathcal{P}(\boldsymbol{a})$ represents a possible value of random vector $\boldsymbol{a}$. We have

$$\begin{aligned}
\boldsymbol{R}(0) &= \mathbb{E}[\boldsymbol{r}(k)\boldsymbol{r}^T(k)] \\
&= \mathbb{E}\left[\mathcal{P}\left(P\boldsymbol{H}\boldsymbol{x}(k) + \frac{\beta}{M}\mathbf{1}\right)\mathcal{P}^T\left(P\boldsymbol{H}\boldsymbol{x}(k) + \frac{\beta}{M}\mathbf{1}\right)\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\left(P\boldsymbol{H}\boldsymbol{x}(k) + \frac{\beta}{M}\mathbf{1}\right)\left(P\boldsymbol{x}(k)\boldsymbol{H} + \frac{\beta}{M}\mathbf{1}\right)^T\right] \\
&\quad + \mathbb{E}\left[\boldsymbol{n}_x\boldsymbol{n}_x^T\right] + \mathbb{E}\left[\boldsymbol{n}_\beta\boldsymbol{n}_\beta^T\right] \\
&= P^2\mathbb{E}\left[\boldsymbol{H}\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{H}^T\right] \\
&\quad + \frac{P\beta}{M}\left(\mathbb{E}\left[\boldsymbol{H}\boldsymbol{x}(k)\mathbf{1}^T\right] + \mathbb{E}\left[\mathbf{1}\boldsymbol{x}^T(k)\boldsymbol{H}^T\right]\right) \\
&\quad + \frac{\beta^2}{M^2}\boldsymbol{E} + \mathbb{E}\left[\boldsymbol{n}_x\boldsymbol{n}_x^T\right] + \mathbb{E}\left[\boldsymbol{n}_\beta\boldsymbol{n}_\beta^T\right] \\
&\overset{(b)}{=} P^2\boldsymbol{H}\left((\mu - \mu^2)\boldsymbol{I} + \mu^2\boldsymbol{E}\right)\boldsymbol{H}^T \\
&\quad + \frac{P\beta\mu}{M}\left(\boldsymbol{H}\boldsymbol{E}_1 + \boldsymbol{E}_1^T\boldsymbol{H}^T\right) \\
&\quad + \frac{\beta^2}{M^2}\boldsymbol{E} + \mu P\mathrm{diag}(\boldsymbol{H}\mathbf{1}) + \frac{\beta}{M}\boldsymbol{I} \\
&= (\mu - \mu^2)P^2\boldsymbol{H}\boldsymbol{H}^T + \mu^2P^2\boldsymbol{H}\boldsymbol{E}\boldsymbol{H}^T \\
&\quad + \frac{P\beta\mu}{M}\left(\boldsymbol{H}\boldsymbol{E}_1 + \boldsymbol{E}_1^T\boldsymbol{H}^T\right) \\
&\quad + \frac{\beta^2}{M^2}\boldsymbol{E} + \mu P\mathrm{diag}(\boldsymbol{H}\mathbf{1}) + \frac{\beta}{M}\boldsymbol{I}
\end{aligned} \tag{44}$$

where $\boldsymbol{n}_x$ is the "noise" due to Poisson random process of the signal part; $\boldsymbol{n}_\beta$ is the "noise" of the background radiation intensity due to Poisson distribution; $\boldsymbol{E}_1$ is an all-one non-square matrix; (a) is obtained by considering Poisson distribution as its expectation plus the "Poisson noise"; and (b) holds according to (42).

Similarly, we can write the shifted autocorrelation matrix of the received sequence $\boldsymbol{R}(1)$ and $\boldsymbol{R}(2D)$, respectively, as

$$\begin{aligned}
\boldsymbol{R}(1) &= \mathbb{E}[\boldsymbol{r}(k-1)\boldsymbol{r}^T(k)] \\
&= (\mu - \mu^2)P^2\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^T + \mu^2P^2\boldsymbol{H}\boldsymbol{E}\boldsymbol{H}^T \\
&\quad + \frac{P\beta\mu}{M}\left(\boldsymbol{H}\boldsymbol{E}_1 + \boldsymbol{E}_1^T\boldsymbol{H}^T\right) + \frac{\beta^2}{M^2}\boldsymbol{E} \\
&\quad + \mu P\mathrm{diag}(\boldsymbol{H}\mathbf{1})\cdot\boldsymbol{J}^M + \frac{\beta}{M}\boldsymbol{J}^M
\end{aligned} \tag{45}$$

and

$$\begin{aligned}
\boldsymbol{R}(2D) &= \mathbb{E}[\boldsymbol{r}(k-2D)\boldsymbol{r}^T(k)] \\
&= \mu^2P^2\boldsymbol{H}\boldsymbol{E}\boldsymbol{H}^T + \frac{P\beta\mu}{M}\left(\boldsymbol{H}\boldsymbol{E}_1 + \boldsymbol{E}_1^T\boldsymbol{H}^T\right) \\
&\quad + \frac{\beta^2}{M^2}\boldsymbol{E}.
\end{aligned} \tag{46}$$

Note that we have the properties (43), shown at the bottom of the page, where $\otimes$ stands for Kronecker product.

$$\boldsymbol{H}\boldsymbol{E}\boldsymbol{H}^T = \boldsymbol{E} \otimes \begin{pmatrix} \left(\sum_{d=1}^D h_{d,1}\right)^2 & \cdots & \left(\sum_{d=1}^D h_{d,1}\right)\left(\sum_{d=1}^D h_{d,M}\right) \\ \vdots & \ddots & \vdots \\ \left(\sum_{d=1}^D h_{d,M}\right)\left(\sum_{d=1}^D h_{d,1}\right) & \cdots & \left(\sum_{d=1}^D h_{d,M}\right)^2 \end{pmatrix}_{M\times M}; \tag{43a}$$

$$\mathrm{diag}(\boldsymbol{H}\mathbf{1}) = \boldsymbol{I} \otimes \begin{pmatrix} \sum_{d=1}^D h_{d,1} & & \\ & \ddots & \\ & & \sum_{d=1}^D h_{d,M} \end{pmatrix}_{M\times M}; \tag{43b}$$

$$\mathrm{diag}(\boldsymbol{H}\boldsymbol{E}_1) = \mathrm{diag}(\boldsymbol{E}_1^T\boldsymbol{H}^T) = \mathrm{diag}(\boldsymbol{H}\mathbf{1}) \tag{43c}$$

According to (43), we have that

$$\begin{aligned}\text{diag}\left(\boldsymbol{R}(2D)\right) &= \text{diag}\left(\mu^2P^2\boldsymbol{H}\boldsymbol{E}\boldsymbol{H}^T+\frac{P\beta\mu}{M}\left(\boldsymbol{H}\boldsymbol{E}_1\right.\right.\\ &\quad\left.\left.+\boldsymbol{E}_1^T\boldsymbol{H}^T\right)+\frac{\beta^2}{M^2}\boldsymbol{E}\right)\\ &= \mu^2P^2\text{diag}^2(\boldsymbol{H}\boldsymbol{1})+\frac{2P\beta\mu}{M}\text{diag}(\boldsymbol{H}\boldsymbol{1})+\frac{\beta^2}{M^2}\boldsymbol{I}\\ &= \left(\mu P\text{diag}(\boldsymbol{H}\boldsymbol{1})+\frac{\beta}{M}\boldsymbol{I}\right)^2. \end{aligned}\tag{47}$$

Based on (44), (45) and (47), we can derive the second-order statistics of $\boldsymbol{H}$ as

$$\begin{aligned}\boldsymbol{H}\boldsymbol{H}^T &= \frac{1}{(\mu-\mu^2)P^2}\left(\boldsymbol{R}(0)-\boldsymbol{R}(2D)\right.\\ &\quad\left.-\text{diag}^{\frac{1}{2}}\left(\boldsymbol{R}(2D)\right)\right); \end{aligned}\tag{48a}$$

$$\begin{aligned}\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^T &= \frac{1}{(\mu-\mu^2)P^2}\left(\boldsymbol{R}(1)-\boldsymbol{R}(2D)\right.\\ &\quad\left.-\text{diag}^{\frac{1}{2}}\left(\boldsymbol{R}(2D)\right)\boldsymbol{J}^M\right). \end{aligned}\tag{48b}$$

Appendix B
Proof of Lemma 1

Suppose that the probabilities of false detection of symbol "0" and "1" are the same. Thus, the duty cycle of $\{\hat{x}_t\}$ is still μ. Define $w_t \triangleq \hat{x}_t-\mu$, where μ is the duty cycle. Then, $w_t\in\{-\mu,1-\mu\}$ can be regarded as a zero-mean random variable. For long data length L, according to the weak law of large numbers (W.L.L.N.), we have that

$$\begin{aligned}&\frac{1}{L-D+1}\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\\ &\quad= \frac{1}{L-D+1}[\hat{\boldsymbol{x}}_0,\hat{\boldsymbol{x}}_1,\ldots,\hat{\boldsymbol{x}}_{L-D}]\\ &\qquad\cdot[\hat{\boldsymbol{x}}_0,\hat{\boldsymbol{x}}_1,\ldots,\hat{\boldsymbol{x}}_{L-D}]^T\\ &\quad= \frac{1}{L-D+1}\sum_{k=0}^{L-D}\left[(\boldsymbol{w}_k+\mu\boldsymbol{1})(\boldsymbol{w}_k+\mu\boldsymbol{1})^T\right]\\ &\quad\overset{W.L.L.N.}{\cong} (\mathbb{E}[\boldsymbol{w}_k\boldsymbol{w}_k^T]+\mu^2\boldsymbol{E})\\ &\quad= \left[\left(\mu-\mu^2\right)\boldsymbol{I}+\mu^2\boldsymbol{E}\right] \end{aligned}\tag{49}$$

where $\boldsymbol{I}$ is a D by D identity matrix; and $\boldsymbol{E}$ is a D by D all-one matrix.

The inverse matrix is given by

$$\begin{aligned}(L-D+1)\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\right)^{-1} &\cong \left[\left(\mu-\mu^2\right)\boldsymbol{I}+\mu\boldsymbol{E}\right]^{-1}\\ &= \frac{1}{(\mu-\mu^2)}\left(\boldsymbol{I}-\frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right). \end{aligned}\tag{50}$$

According to the form of W.L.L.N., we rewrite (49) as

$$\begin{aligned}\lim_{L\to\infty}\mathbb{P}\left(\left\|(L-D+1)\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\right)^{-1}\right.\right.\\ \left.\left.-\frac{1}{(\mu-\mu^2)}\left(\boldsymbol{I}-\frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)\right\|_F\le\varepsilon\right)=1. \end{aligned}\tag{51}$$

Devide both sides of the inequality sign in (51) by $\frac{1}{(\mu-\mu^2)}\left(\boldsymbol{I}-\frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)$, eq. (23) can be obtained.

Appendix C
Proof of Theorem 3

To facilitate the analysis, denote by $\boldsymbol{r}'\triangleq[r'_{D-1},r'_D,\ldots,r'_{L-1}]^T$ a realization of the received signal if the transmitter sent the contaminated pilot $\{\hat{x}_k\}_{k=0}^{L-1}$. The channel estimate $\hat{\boldsymbol{h}}$ can also be written using the contaminated pilot

$$\hat{\boldsymbol{h}}=\frac{1}{P}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\right)^{-1}\hat{\boldsymbol{X}}(\boldsymbol{r}'-\beta\boldsymbol{1}).\tag{52}$$

The expectation of the degraded channel estimate based on the contaminated pilot can be written as

$$\begin{aligned}&\mathbb{E}[\hat{\boldsymbol{h}}_d]\\ &= \mathbb{E}\left[\frac{1}{P}\left(\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\right)^{-1}\hat{\boldsymbol{X}}(\boldsymbol{r}-\beta\boldsymbol{1})\right]\\ &\overset{(a)}{\underset{P}{\to}} \mathbb{E}\left[\frac{1}{P(\mu-\mu^2)(L-D+1)}\left(\boldsymbol{I}-\frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)\right.\\ &\quad\left.\cdot\hat{\boldsymbol{X}}(\bar{\boldsymbol{r}}-\beta\boldsymbol{1})\right]\\ &= \mathbb{E}\left[\frac{1}{P(\mu-\mu^2)(L-D+1)}\left(\boldsymbol{I}-\frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)\right.\\ &\quad\left.\cdot\hat{\boldsymbol{X}}\begin{pmatrix}\bar{r}'_{D-1}-\beta-\sum_{i=0}^{D-1}P(2\hat{x}_{D-1-i}-1)h_ie_{D-1-i}\\ \bar{r}'_D-\beta-\sum_{i=0}^{D-1}P(2\hat{x}_{D-i}-1)h_ie_{D-i}\\ \vdots\\ \bar{r}'_{L-1}-\beta-\sum_{i=0}^{D-1}P(2\hat{x}_{L-1-i}-1)h_ie_{L-1-i}\end{pmatrix}\right] \end{aligned}\tag{53}$$

$$\begin{aligned}&\overset{(b)}{=} \boldsymbol{h}-\frac{1}{(\mu-\mu^2)(L-D+1)}\left(\boldsymbol{I}-\frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)\\ &\quad\cdot\mathbb{E}\begin{pmatrix}\sum_{j=D-1}^{L-1}\hat{x}_j\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_ie_{j-i}\\ \sum_{j=D-1}^{L-1}\hat{x}_{j-1}\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_ie_{j-i}\\ \vdots\\ \sum_{j=D-1}^{L-1}\hat{x}_{j-D+1}\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_ie_{j-i}\end{pmatrix} \end{aligned}\tag{54}$$

where (a) holds due to (23); $\overset{P}{\to}$ denotes convergence based on probability; (b) holds according to (52) and $\mathbb{E}[\hat{\boldsymbol{h}}]=\boldsymbol{h}$; $\boldsymbol{h}$ is the actual channel; $\bar{\boldsymbol{r}}$ denotes the mean of $\boldsymbol{r}$; $\bar{r}_t$ is the mean of r_t; and $\{e_i\}_{i=1}^L$ is defined by

$$e_i=\begin{cases}0, & \hat{x}_i=x_i;\\ 1, & \hat{x}_i\neq x_i.\end{cases}\tag{55}$$

The mth row of the expectation term in (53) is $\sum_{j=D-1}^{L-1}\mathbb{E}\left[\hat{x}_{j-m}\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_ie_{j-i}\right]$. For $i\neq m$, we have

$$\begin{aligned}\mathbb{E}[\hat{x}_{j-m}(2\hat{x}_{j-i}-1)h_ie_{j-i}] &= \mathbb{E}[x_{j-m}]\mathbb{E}[(2\hat{x}_{j-i}-1)\cdot h_ie_{j-i}]\\ &= \frac{\mu(2\mu-1)}{D-1}\sum_{i\neq m}h_i\epsilon; \end{aligned}\tag{56}$$

and for $i = m$, we have

$$\mathbb{E}[\hat{x}_{j-m}(2\hat{x}_{j-i}-1)h_i e_{j-i}] = \mathbb{E}[(2\hat{x}_{j-m}^2 - \hat{x}_{j-m})h_m e_{j-m}] = \mu h_m \epsilon. \quad (57)$$

Therefore, eq. (53) can be further simplified to

$$\begin{aligned}\mathbb{E}[\hat{\boldsymbol{h}}_d] &= \boldsymbol{h} - \frac{1}{1-\mu}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)((2-2\mu)\boldsymbol{I} \\ &\quad + (2\mu-1)\boldsymbol{E})\,\epsilon\boldsymbol{h} \\ &= \left(\boldsymbol{I} - \frac{\epsilon}{1-\mu}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right)((2-2\mu)\boldsymbol{I} \right. \\ &\quad \left. +(2\mu-1)\boldsymbol{E})\right)\boldsymbol{h}. \quad (58)\end{aligned}$$

APPENDIX D
JUSTIFICATION OF (28)

The channel estimation MSE of the unbiased channel estimation has the following form [10]

$$\mathrm{MSE}_U = \frac{1}{P}\mathrm{Tr}\left((\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\mathrm{diag}(\boldsymbol{X}^T\boldsymbol{h})\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\right) + \frac{\beta}{P^2}\mathrm{Tr}\left((\boldsymbol{X}\boldsymbol{X}^T)^{-1}\right). \quad (59)$$

Using (23) and denoting $\boldsymbol{Q} \triangleq (\boldsymbol{I} - \frac{1}{D+1}\boldsymbol{E})(\boldsymbol{I} - \frac{1}{D+1}\boldsymbol{E})$, we have

$$\boldsymbol{Q} = \boldsymbol{I}\cdot\boldsymbol{I} - \frac{2}{D+1}\boldsymbol{E} + \frac{1}{(D+1)^2}\boldsymbol{E}\cdot\boldsymbol{E} = \boldsymbol{I} - \frac{D+2}{(D+1)^2}\boldsymbol{E}. \quad (60)$$

Based on (59) and Lemma 1, the expectation of MSE_U can be simplified to

$$\begin{aligned}\mathbb{E}[\mathrm{MSE}_U] &= \frac{1}{P}\mathbb{E}\left[\mathrm{Tr}\left((\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\mathrm{diag}(\boldsymbol{X}^T\boldsymbol{h})\boldsymbol{X}^T\right.\right. \\ &\quad \left.\left.(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\right)\right] + \mathbb{E}\left[\frac{\beta}{P^2}\mathrm{Tr}\left((\boldsymbol{X}\boldsymbol{X}^T)^{-1}\right)\right] \\ &= \frac{1}{P}\mathbb{E}\left[\mathrm{Tr}\left(\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}\right.\right. \\ &\quad \left.\left.\cdot\,\mathrm{diag}(\boldsymbol{X}^T\boldsymbol{h})\right)\right] + \frac{\beta}{P^2}\mathbb{E}\left[\mathrm{Tr}\left((\boldsymbol{X}\boldsymbol{X}^T)^{-1}\right)\right] \\ &\cong \frac{1}{P}\left(\frac{4}{L-D+1}\right)^2\mathbb{E}\left[\mathrm{Tr}\left(\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}\right.\right. \\ &\quad \left.\left.\cdot\,\mathrm{diag}(\boldsymbol{X}^T\boldsymbol{h})\right)\right] + \frac{\beta}{P^2}\mathrm{Tr}\left(\frac{4}{L-D+1}\boldsymbol{Q}\right) \\ &= \frac{1}{P}\left(\frac{4}{L-D+1}\right)^2\sum_{k=0}^{L-D}\mathbb{E}\left[\boldsymbol{x}_k^T\boldsymbol{Q}\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{h}\right] \\ &\quad + \frac{4\beta}{P^2(L-D+1)}\mathrm{Tr}(\boldsymbol{Q}). \quad (61)\end{aligned}$$

Note that $\mathbb{E}\left[\boldsymbol{x}_k^T\boldsymbol{Q}\boldsymbol{x}_k\boldsymbol{x}^T\right]$ is a row vector, denoted by $\boldsymbol{q}_x^T$. According to (60), we have the following for the mth entry of $\boldsymbol{q}_x$

$$\begin{aligned}\boldsymbol{q}_x(m) &= \mathbb{E}\left[x_{k,m}^3 Q_1 + \sum_{i\neq m}x_{k,i}^2 x_{k,m}Q_1 + 2\sum_{i\neq m}x_{k,i}x_{k,m}^2 Q_2 \right. \\ &\quad \left. + 2\sum_{i\neq j\neq m}x_{k,i}x_{k,j}x_{k,m}\right] \\ &= \frac{1}{2}Q_1 + \frac{D-1}{4}Q_1 + \frac{D-1}{2}Q_2 + \frac{(D-1)(D-2)}{8}Q_2 \\ &= \frac{D^3+D^2+2}{8(D+1)^2} \quad (62)\end{aligned}$$

where $\boldsymbol{x}_k \triangleq [x_{k,1},\ldots,x_{k,m},\ldots,x_{k,D}]^T$; $Q_1 \triangleq \frac{D^2+D-1}{(D+1)^2}$ is the value of each diagnal element of $\boldsymbol{Q}$; and $Q_2 \triangleq -\frac{D+2}{(D+1)^2}$ is the value of each nondiagonal element of $\boldsymbol{Q}$.

Then, the final expression of $\mathbb{E}[\mathrm{MSE}_U]$ can be further written as

$$\mathbb{E}[\mathrm{MSE}_U] \cong \frac{2(D^3+D^2+2)\|\boldsymbol{h}\|_1}{P(L-D+1)(D+1)^2} + \frac{4\beta D(D^2+D-1)}{P^2(L-D+1)(D+1)^2}. \quad (63)$$

APPENDIX E
JUSTIFICATION OF (29)

The MSE expectation $\mathbb{E}[\mathrm{MSE}_C]$ can be divided into three parts: the part caused by Poisson stocastic process, denoted by $\mathbb{E}[\mathrm{MSE}_U]$; the part caused by the bias, denoted by bias^2; and the part caused by the random position of the contaminated pilot symbols, denoted by $\mathbb{E}[\mathrm{MSE}_S]$, which yields

$$\mathbb{E}[\mathrm{MSE}_C] = \mathbb{E}[\mathrm{MSE}_U] + \mathbb{E}[\mathrm{MSE}_S] + \mathrm{bias}^2. \quad (64)$$

The first part has been derived as (63) and the second part has been given by (25). Note the third part is left to handle. According to (53), $\mathbb{E}[\mathrm{MSE}_S]$ is given by

$$\mathbb{E}[\mathrm{MSE}_S] = (\boldsymbol{h}_b - \bar{\boldsymbol{h}}_b)^T(\boldsymbol{h}_b - \bar{\boldsymbol{h}}_b) \quad (65)$$

where $\boldsymbol{h}_b$ is given by

$$\boldsymbol{h}_b = \frac{4}{L-D+1}\left(\boldsymbol{I} - \frac{1}{\frac{1-\mu}{\mu}+D}\boldsymbol{E}\right) \cdot \begin{pmatrix} \sum_{j=D-1}^{L-1}\hat{x}_{j-1}\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_i e_{j-i} \\ \sum_{j=D-1}^{L-1}\hat{x}_{j-2}\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_i e_{j-i} \\ \vdots \\ \sum_{j=D-1}^{L-1}\hat{x}_{j-D}\sum_{i=0}^{D-1}(2\hat{x}_{j-i}-1)h_i e_{j-i} \end{pmatrix}. \quad (66)$$

Note that $\mathbb{E}[\mathrm{MSE}_S]$ is also the summation of the variances, each of which corresponding to each entry of $\boldsymbol{h}_b$. We have that

$$\mathbb{E}[\mathrm{MSE}_S] = \mathbb{E}\left[\sum_{i=1}^{D}\mathrm{Var}[\boldsymbol{h}_b(i)]\right]. \quad (67)$$

For the mth entry of $\boldsymbol{h}_b$, we have

$$\begin{aligned}\boldsymbol{h}_b(m) &= \frac{4}{L-D+1}\sum_{n=D}^{L}\sum_{i=1}^{D}\left(\frac{D}{D+1}x_{n+i-m}\right.\\ &\qquad\left.-\frac{1}{D+1}\sum_{k\neq m}x_{n+i-k}\right)(2x_n-1)h_i e_n\\ &\triangleq \frac{4}{L-D+1}\sum_{n=D}^{L}\alpha_{m,n}e_n \end{aligned} \tag{68}$$

where

$$\begin{aligned}\alpha_{m,n} \triangleq \sum_{i=1}^{D}\left(\frac{D}{D+1}x_{n+i-m}-\frac{1}{D+1}\sum_{k\neq m}x_{n+i-k}\right)\\ \cdot(2x_n-1)h_i. \end{aligned} \tag{69}$$

Note that $\{\alpha_{m,n}\}$ is constant given $\{\hat{\boldsymbol{x}}_n\}$, and that $\mathrm{Var}[\boldsymbol{h}_b(m)]$ is the variance of the weighted summation of $L-D+1$ Binomial distributed random variables. We have

$$\begin{aligned}\mathrm{Var}[\boldsymbol{h}_b(m)] &\cong \frac{16}{L-D+1}\mathbb{E}\left[\alpha_{m,n}^2\right]\mathrm{Var}[e_n]\\ &= \frac{16}{L-D+1}\mathbb{E}\left[\alpha_{m,n}^2\right]\epsilon(1-\epsilon). \end{aligned} \tag{70}$$

It is seen that term $(2x_n-1)$ in $\alpha_{m,n}$ is either 0 or 1. Therefore, we have the following simplification for $\alpha_{m,n}$.

$$\alpha_{m,n}=\sum_{i=1}^{D}\left(\frac{D}{D+1}x_{n+i-m}-\frac{1}{D+1}\sum_{k\neq m}x_{n+i-k}\right)h_i. \tag{71}$$

Define $\xi_{n,i}\triangleq \frac{D}{D+1}x_{n+i-m}-\frac{1}{D+1}\sum_{k\neq m}x_{n+i-k}$. We have that

$$\mathbb{E}\left[\alpha_{m,n}^2\right]=\mathbb{E}\left[\left(\sum_{i=1}^{D}\xi_{n,i}h_i\right)^2\right]=\sum_{i=1}^{D}\sum_{j=1}^{D}\mathbb{E}[\xi_{n,i}\xi_{n,j}]h_ih_j. \tag{72}$$

For $i=j$, we have

$$\begin{aligned}\mathbb{E}\left[\xi_{n,i}\xi_{n,j}\right] &= \mathbb{E}\left[\left(\frac{D}{D+1}x_{n+i-m}-\frac{1}{D+1}\sum_{k\neq m}x_{n+i-k}\right)^2\right]\\ &= \frac{D^2}{(D+1)^2}\mathbb{E}\left[x_{n+i-m}^2\right]\\ &\quad-\frac{2D}{(D+1)^2}\sum_{k\neq m}\mathbb{E}\left[x_{n+i-m}x_{n+i-k}\right]\\ &\quad+\frac{1}{(D+1)^2}\sum_{k_1\neq m}\sum_{k_2\neq m}\mathbb{E}\left[x_{n+i-k_1}x_{n+i-k_2}\right]\\ &= \frac{D^2}{(D+1)^2}\cdot\frac{1}{2}-\frac{2D}{(D+1)^2}\cdot\frac{(D-1)}{4}\\ &\quad+\frac{1}{(D+1)^2}\left[\frac{D-1}{2}+\frac{(D-1)(D-2)}{4}\right]\\ &= \frac{D}{4(D+1)}. \end{aligned} \tag{73}$$

For $i\neq j$, we have that

$$\begin{aligned}\mathbb{E}\left[\xi_{n,i}\xi_{n,j}\right] &= \mathbb{E}\left[\left(x_{n+i-m}-\frac{1}{D+1}\sum_{k=0}^{D-1}x_{n+i-k}\right)\right.\\ &\qquad\left.\times\left(x_{n+j-m}-\frac{1}{D+1}\sum_{k=0}^{D-1}x_{n+j-k}\right)\right]\\ &= \mathbb{E}[x_{n+i-m}x_{n+j-m}]\\ &\quad-\frac{1}{D+1}\mathbb{E}\left[x_{n+i-m}\sum_{k=0}^{D-1}x_{n+j-k}\right]\\ &\quad-\frac{1}{D+1}\mathbb{E}\left[x_{n+j-m}\sum_{k=0}^{D-1}x_{n+i-k}\right]\\ &\quad+\frac{1}{(D+1)^2}\mathbb{E}\left[\left(\sum_{k=0}^{D-1}x_{n+i-k}\right)\left(\sum_{k=0}^{D-1}x_{n+j-k}\right)\right]. \end{aligned} \tag{74}$$

Note that the combination of m, i and j will make the second and third terms in (74) complicated to formulate. Hence we assume that there is only one autocorrelation term in either the second term or the third term. Based on this assumption, we can further write (74) as

$$\begin{aligned}\mathbb{E}\left[\xi_{n,i}\xi_{n,j}\right] &\approx \frac{1}{4}-\frac{D}{2(D+1)}-\frac{1}{D+1}\left(\frac{D-1}{4}+\frac{1}{2}\right)\\ &\quad+\frac{1}{(D+1)^2}\left((D-|i-j|)\cdot\frac{1}{2}\right.\\ &\qquad\left.+\left[D^2-(D-|i-j|)\right]\cdot\frac{1}{4}\right)\\ &= -\frac{|i-j|}{4(D+1)^2}. \end{aligned} \tag{75}$$

Finally, we can derive the expression of $\mathbb{E}\left[\mathrm{MSE}_S\right]$ as

$$\begin{aligned}\mathbb{E}\left[\mathrm{MSE}_S\right] \approx \frac{4}{L-D+1}\left(\frac{D}{D+1}\sum_{i=0}^{D-1}h_i^2\right.\\ \left.-\frac{1}{(D+1)^2}\sum_{i\neq j}|i-j|h_ih_j\right)\epsilon(1-\epsilon). \end{aligned} \tag{76}$$

REFERENCES

[1] B. Liu, C. Gong, J. Cheng, and Z. Xu, "Data-aided channel estimation for Poisson channels with inter-symbol interference," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[2] Z. Xu and B. M. Sadler, "Ultraviolet communications: Potential and state-of-the-art," *IEEE Commun. Mag.*, vol. 46, no. 5, pp. 67–73, May 2008.

[3] V. Jamali, A. Ahmadzadeh, C. Jardin, C. Sticht, and R. Schober, "Channel estimation for diffusive molecular communications," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4238–4252, Oct. 2016.

[4] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 244–256, Mar. 1991.

[5] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 303–322, Jan. 2009.

[6] A. R. Calderbank and C. N. Georghiades, "Synchronizable codes for the optical OPPM channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1097–1107, Jul. 1994.

[7] T. T. Nguyen and L. Lampe, "Coded multipulse pulse-position modulation for free-space optical communications," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1036–1041, Apr. 2010.

[8] H. Kim, B. Nachman, and A. El Gamal, "Superposition coding is almost always optimal for the Poisson broadcast channel," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1782–1794, Apr. 2016.
[9] N. D. Chatzidiamantis, G. K. Karagiannidis, and M. Uysal, "Generalized maximum-likelihood sequence detection for photon-counting free space optical systems," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3381–3385, Dec. 2010.
[10] C. Gong and Z. Xu, "Channel estimation and signal detection for optical wireless scattering communication with inter-symbol interference," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5326–5337, Oct. 2015.
[11] G. Wang, C. Gong, and Z. Xu, "Signal detection and achievable rates for multiple access optical wireless scattering communication," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
[12] C. Gong and Z. Xu, "LMMSE SIMO receiver for short-range non-line-of-sight scattering communication," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5338–5349, Oct. 2015.
[13] B. Liu, C. Gong, J. Cheng, and Z. Xu, "Correlation-based LTI channel estimation for multi-wavelength optical scattering NLOS communication," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1648–1661, Mar. 2020.
[14] R. Mosayebi, H. Arjmandi, A. Gohari, M. Nasiri-Kenari, and U. Mitra, "Receivers for diffusion-based molecular communication: Exploiting memory and sampling rate," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 12, pp. 2368–2380, Dec. 2014.
[15] V. Jamali, N. Farsad, R. Schober, and A. Goldsmith, "Non-coherent detection for diffusive molecular communication systems," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2515–2531, Jun. 2018.
[16] B. Li, M. Sun, S. Wang, W. Guo, and C. Zhao, "Local convexity inspired low-complexity noncoherent signal detector for nanoscale molecular communications," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2079–2091, May 2016.
[17] J. R. Treichler, I. Fijalkow, and C. R. Johnson, "Fractionally spaced equalizers," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 65–81, May 1996.
[18] T. N. Cao, V. Jamali, N. Zlatanov, P. L. Yeoh, J. Evans, and R. Schober, "Fractionally spaced equalization and decision feedback sequence detection for diffusive MC," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 117–121, Jan. 2021.
[19] M. G. Larimore and M. Goodman, "Implementation of the constant modulus algorithm at RF bandwidths," in *Proc. 19th Asilomar Conf. Circuits, Syst. Comput.*, 1985, pp. 626–630.
[20] R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.
[21] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.
[22] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Process.*, vol. 50, nos. 1–2, pp. 83–99, Apr. 1996.
[23] C. Shin, R. W. Heath, Jr., and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.
[24] F. Gao, Y. Zeng, A. Nallanathan, and T.-S. Ng, "Robust subspace blind channel estimation for cyclic prefixed MIMO ODFM systems: Algorithm, identifiability and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 378–388, Feb. 2008.
[25] Y. Zeng and T.-S. Ng, "A semi-blind channel estimation method for multiuser multiantenna OFDM systems," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1419–1429, May 2004.
[26] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.
[27] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 540–553, Jan. 2018.
[28] S. Marcos, "A network of adaptive Kalman filters for data channel equalization," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2620–2627, Sep. 2000.
[29] S. Ayra and Y. H. Chung, "ANN-assisted real-time blind signal detection over time-varying doubly-stochastic Poisson channel for indoor optical wireless communications," *IEEE Photon. J.*, vol. 5, no. 5, pp. 1–19, Oct. 2020.
[30] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel. I," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1461, Jun. 1988.
[31] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," *Proc. IEEE*, vol. 108, no. 2, pp. 292–307, Feb. 2020.
[32] Y. Xun, J. Liu, J. Ning, and H. Zhang, "An experimental study towards the in-vehicle network of intelligent and connected vehicles," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Jan. 2018, pp. 1–6.
[33] R. J. Drost, T. J. Moore, and B. M. Sadler, "UV communications channel modeling incorporating multiple scattering interactions," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 28, no. 4, pp. 686–695, 2011.

Beiyuan Liu (Associate Member, IEEE) received the B.E. and Ph.D. degrees in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2015 and 2020, respectively. From 2018 to 2019, he was a Visiting Ph.D. Student with The University of British Columbia, Canada. He is currently an Associate Professor with the School of Cybersecurity, Northwestern Polytechnical University. His research interests include the areas of optical wireless communications and dual-functional radar-communications.

Chen Gong (Senior Member, IEEE) received the B.S. degree in electrical engineering and mathematics (minor) from Shanghai Jiao Tong University, Shanghai, China, in 2005, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2008, and the Ph.D. degree from Columbia University, New York City, NY, USA, in 2012. He was a Senior Systems Engineer with the Qualcomm Research, San Diego, CA, USA, from 2012 to 2013. He is currently a Faculty Member with the University of Science and Technology of China. His research interests in wireless communications, optical wireless communications, and signal processing. He was selected by the Young 1000 Talent Program of China Government in 2014 and awarded the Hong Kong Qiushi Outstanding Young Researcher Award in 2016.

Julian Cheng (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1995, the M.Sc. (Eng.) degree in mathematics and engineering from Queen's University, Kingston, ON, Canada, in 1997, and the Ph.D. degree in electrical engineering from the University of Alberta, Edmonton, AB, Canada, in 2003. He was with Bell Northern Research and NORTEL Networks. He is currently a Full Professor with the Faculty of Applied Science, School of Engineering, The University of British Columbia, Kelowna, BC, Canada. His current research interests include digital communications over fading channels, statistical signal processing for wireless applications, optical wireless communications, and 5G wireless networks. He is the President of the Canadian Society of Information Theory and the Secretary for the Radio Communications Technical Committee of the IEEE Communications Society. He was the Co-Chair of the 12th Canadian Workshop on Information Theory in 2011, the 28th Biennial Symposium on Communications in 2016, and the Sixth EAI International Conference on Game Theory for Networks (GameNets 216). He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE COMMUNICATIONS LETTERS, and IEEE ACCESS. He has served as a Guest Editor for a Special Issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Optical Wireless Communications. He serves as an Area Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He is also a Registered Professional Engineer with the Province of British Columbia, Canada.

Zhengyuan Xu (Senior Member, IEEE) received the B.S. and M.S. degrees from Tsinghua University, China, and the Ph.D. degree from the Stevens Institute of Technology, USA. He was a Full Professor (tenured) at the University of California at Riverside and later at Tsinghua University before he joined the University of Science and Technology of China (USTC). He was the Founding Director of the multi-campus Center for Ubiquitous Communication by Light (UC-Light), University of California, and the Founding Director of the Wireless-Optical Communications Key Laboratory of Chinese Academy of Sciences. He was a Distinguished Expert and a Chief Scientist of the National Key Basic Research Program of China. He has published over 300 international journals and conference papers. He has coauthored a book titled *Visible Light Communications: Modulation and Signal Processing* (Wiley-IEEE Press). His research focuses on optical wireless communications, mobile networking, artificial intelligence, wireless big data, sensing, ranging, and localization. He has been on the Elsevier Annual List of Most Cited Chinese Researchers since 2014. He has received more than 5000 citations according to Google Scholar. He was the Founding Chair of IEEE Workshop on Optical Wireless Communications in 2010. He has served as an Associate Editor for different IEEE/OSA journals.

Jiajia Liu (Senior Member, IEEE) was a Full Professor at the School of Cyber Engineering, Xidian University, from 2013 to 2018. He is currently a Full Professor (the Vice Dean) with the School of Cybersecurity, Northwestern Polytechnical University. He has published more than 200 peer-reviewed papers in many high quality publications, including prestigious IEEE journals and conferences. His research interests cover a wide range of areas, including intelligent and connected vehicles, mobile/edge/cloud computing and storage, the Internet of Things security, wireless and mobile *ad-hoc* networks, and space-air-ground integrated networks. He received the Niwa Yasujiro Outstanding Paper Award in 2012, the IEEE ComSoc Asia–Pacific Outstanding Young Researcher Award in 2017, the IEEE ComSoc Asia–Pacific Outstanding Paper Award in 2019, the IEEE VTS Early Career Award in 2019, the IEEE ComSoc Best Young Professional (YP) Award in Academia in 2020, and the Best Paper Awards from many international conferences, including IEEE flagship events, such as IEEE GLOBECOM in 2016 and 2019, IEEE WCNC in 2012 and 2014, IEEE WiMob in 2019, IEEE IC-NIDC in 2018, and AICON in 2019. He was also a recipient of the Tohoku University President Award 2013. He is serving on technical program committees for numerous international conferences, like the Leading Symposium Co-Chair of AHSN Symposium for GLOBECOM 2017, CRN Symposium for ICC 2018, and AHSN Symposium for ICC 2019. He is the Vice Chair of IEEE IOT-AHSN TC. He has been actively joining the society activities, like serving as an Associate Editor for IEEE TRANSACTIONS ON COMPUTERS from October 2015 to June 2017, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from January 2016 to January 2021, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since May 2018, and IEEE TRANSACTIONS ON COMMUNICATIONS since September 2020, an Editor for *IEEE Network* since July 2015 and IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING since January 2019, and a Guest Editor for top ranking international journals, like IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, *IEEE Network*, and IEEE INTERNET OF THINGS JOURNAL. He is a Distinguished Lecturer of IEEE Communications Society and Vehicular Technology Society.