

Adaptive Statistical Bayesian MMSE Channel Estimation for Visible Light Communication

Xianyu Chen and Ming Jiang*, *Senior Member, IEEE*

Abstract—Visible light communication (VLC) is considered to be one of the promising technologies for future wireless systems and has attracted an increasing number of research interests in recent years. Optical orthogonal frequency division multiplexing (O-OFDM) has been proposed for VLC systems to eliminate the multi-path interference, while also facilitating frequency domain equalisation (FDE). In comparison with the conventional radio frequency (RF) based wireless communications, there has been limited considerations on channel estimation for VLC, where the indoor optical wireless channel model differs from the traditional RF case. In this paper, we present a new channel estimation (CE) algorithm for indoor downlink (DL) VLC systems, referred to as the adaptive statistical Bayesian minimum mean square error channel estimation (AS-BMMSE-CE). Furthermore, a so-called variable statistic window (VSW) mechanism is designed for exploiting past channel information within a window of adaptively optimised size, such that the CE performance can be significantly improved. Detailed theoretical analysis is provided and verified by extensive numerical results, demonstrating the superior performance of the proposed AS-BMMSE-CE scheme.

Index Terms—Bayesian estimation, channel estimation, variable statistic window (VSW), visible light communication (VLC).

I. INTRODUCTION

IN recent years, visible light communication (VLC) [1] has emerged as a promising technology for complementing conventional radio frequency (RF) based wireless communication systems. In comparison to the RF scenario, there has been limited considerations on channel estimation (CE) for VLC. The principles of conventional CE technologies, for example the pilot-aided channel estimation (PACE) schemes [2], [3], may also be applicable to VLC scenarios. Depending on the domain where the estimators operate, we have frequency-domain (FD) or time-domain (TD) based CEs. Conventional FD CEs employ methods such as minimum mean square error (MMSE) [2], [3], genetic algorithm (GA) [4], adaptive polar linear interpolation (APLI) [5], etc., which either assume ideal conditions or suffer from notable residual error floors. On the other hand, TD CEs [6]–[8] utilise channel impulse response (CIR) for estimating channel state information (CSI) by invoking MMSE, recursive least squares (RLS) [9] or other algorithms [7], [10]. Nonetheless, they often rely on specific *a priori* information that may not be available in practical systems, or on parameters for example forgetting factors with fixed values, which therefore may not adapt to CSI variations.

Furthermore, the indoor channel for VLC [11], [12] is different from the traditional wireless radio channels. Due to the intensity modulation/direct detection (IM/DD) mechanism invoked by VLC systems, the transmitted optical signal has non-negative real values and so does the CIR. Additionally, another significant difference between the RF and VLC channels resides in their time-varying characteristics. In a typical indoor VLC system, when the user moves around within the VLC environment, the variation of the channel taps' envelopes and the path delay no longer obey the traditional Doppler spectrum [11], [12]. Moreover, compared with the sparse taps of many popular RF channel models, the taps of VLC channels are denser due to many reflections from the walls and the ceiling, thus resulting in specific design constraints from the CE perspective. In this case, algorithms designed for channels with sparsity characteristics, for example the technique of [13], may not be suitable for CE in VLC systems. Therefore, although some of the traditional CE algorithms might still be directly applicable, only those tailored for VLC scenarios may become optimum solutions.

Inspired by the CEs designed for RF channels, some CEs for optical channels [14]–[16] have been developed, where maximum likelihood sequence detection (MLSD) [17] is adopted for mitigating inter-symbol interference (ISI). In [18], the authors propose the implementation of linear decision feedback and artificial neural network (ANN) based equalisation for VLC, where equalisers are performed in real-time, though at the cost of increased complexity. As ISI can be effectively eliminated with the aid of orthogonal frequency division multiplexing (OFDM), which also has other merits and has been adopted by many modern wireless standards such as the long-term evolution (LTE), it has been suggested to extend OFDM to the VLC domain for supporting ISI-free high-rate transmissions [19]–[22]. Nonetheless, only recently, a few CEs were introduced for OFDM-aided VLC systems [23]–[25], where the authors tended to simply reuse traditional CE schemes originally proposed for RF OFDM. Furthermore, these schemes only consider simple channel models rather than the more sophisticated ones [11], [12].

Against this background, in this paper we propose a new CE scheme for optical OFDM (O-OFDM) aided VLC systems, which is capable of achieving a superior CE performance in terms of both mean square error (MSE) and bit error rate (BER) at a modest computational complexity. The novelty of this work mainly includes:

- 1) A new CE scheme referred to as adaptive statistical Bayesian minimum mean square error channel estimation (AS-BMMSE-CE) is designed. It exploits a so-

called variable statistic window (VSW) with a theoretically optimised size. Furthermore, the proposed per-tap optimisation process is suitable for the VLC channel, which is constituted by dense taps that have different statistical characteristics, thus provides high robustness and stability in terms of CE performance.

- 2) Comprehensive theoretical derivations are provided to prove that the upper MSE bound of AS-BMMSE-CE is lower than the Cramér-Rao lower bound (CRLB), and that the lower MSE bound of AS-BMMSE-CE may also be lower than the traditional Bayesian lower bound (TBLB) [7], [26] under some circumstances. Particularly, to cope with O-OFDM and the real-valued VLC channel, most derivations are developed in the real domain, which is different from the RF scenario, where derivations are based on complex numbers.
- 3) New algorithms called covariance coefficient update algorithm (CCUA) and covariance matrix update algorithm (CMUA), are designed based on a theoretically optimised pilot pattern exploiting the O-OFDM properties in the real domain. They together help to reduce the computational complexity of AS-BMMSE-CE.

The organisation of this paper is as follows. The system model is briefly reviewed in Section II, followed by an overview of the proposed VSW-aided AS-BMMSE-CE scheme in Section III. The details of AS-BMMSE-CE are provided in Section IV, where various design aspects including a complexity reduction option are discussed. Simulation results are offered and analysed in Section V, before we finally conclude our findings in Section VI.

Notations: Bold variables denote matrices or vectors; $\text{Tr}\{\cdot\}$ stands for the trace operation; $(\cdot)^T$ and $(\cdot)^H$ refer to the transpose and Hermitian transpose operations, respectively; $(\cdot)^*$ is the conjugation of (\cdot) ; $[\cdot]_i$ and $[\cdot]_{i,j}$ indicate the selection of the i^{th} element of a vector and the $(i, j)^{\text{th}}$ element of a matrix, respectively; $E\{\cdot\}$ is the expectation operation; $D\{\cdot\}$ is the variance operation; \mathbf{I}_L denotes an $L \times L$ identity matrix; $\text{diag}\{\cdot\}$ declares a diagonal matrix; and $\hat{(\cdot)}$ defines the estimate of the variable concerned.

II. SYSTEM MODEL

As an example, we consider a general VLC system based on direct-current-biased optical OFDM (DCO-OFDM) [22], as shown in Fig. 1 [27]. However, it is also worth pointing out that other popular optical OFDM (O-OFDM) schemes are also applicable with minimum modifications. For simplicity, we assume that the environmental conditions, such as for example ambient light, reflective objects, etc. remain the same in the room. Under this assumption, the indoor VLC channel may be viewed as *position-varying* rather than time-varying, implying that it fluctuates in the space domain when the user equipment (UE) moves around in the room. Moreover, it is a slow-varying case due to the low mobility of the UE.

Define the subcarrier indices of pilot symbols as a set $I_p = \{P_0 + i \cdot N_d, i = 0, 1, \dots, N_p/2 - 1\}$, where N_d is the pilot interval, N_p is the total number of pilots required for one O-OFDM symbol and P_0 is the smallest subcarrier

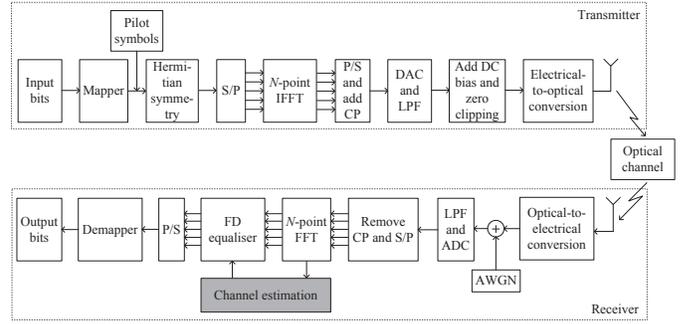


Fig. 1. Schematic of a typical DCO-OFDM system.

index among all pilots. For the transmission towards the UE at the n^{th} position in the room, pilot symbols of the same constant amplitude are multiplexed with data symbols at an equal-distance of N_d to produce a FD signal vector $\mathbf{X}_n = [X[n, 0], \dots, X[n, N-1]]^T \in \mathbb{C}^{N \times 1}$, where the sets of pilot subcarrier indices and data subcarrier indices may be expressed as $P_{\text{pilot}} = \{k | k \in I_p \text{ or } N - k \in I_p\}$ and $P_{\text{data}} = \{0, \dots, N-1\} \setminus P_{\text{pilot}}$ [28], respectively, while N is the size of inverse fast Fourier transform (IFFT) and \mathbb{C} denotes the set of complex numbers. Since IM-based optical signals have non-negative real values, \mathbf{X}_n is constrained to be Hermitian symmetric as

$$X[n, k] = X^*[n, N - k] \quad \text{for } 0 < k < \frac{N}{2}, \quad (1)$$

where $X[n, 0] = X[n, N/2] = 0$. Then, after the serial-to-parallel (S/P) and IFFT operations seen in Fig. 1, we have a real vector $\mathbf{x}_n = \mathbf{F}_I \mathbf{X}_n$, where $\mathbf{F}_I = \{f_{n,k}\} \in \mathbb{C}^{N \times N}$, $f_{n,k} = \frac{1}{N} e^{j \frac{2\pi n k}{N}}$ for $0 \leq \{n, k\} \leq N-1$. The generated electrical DCO-OFDM signal \mathbf{s}_n is then converted to its optical version and transmitted in the VLC channel of a discrete form

$$\mathbf{h}_n = [h[n, 0], \dots, h[n, L_c - 1]]^T \in \mathbb{R}_+^{L_c \times 1}, \quad (2)$$

where L_c is the maximum number of CIR taps and \mathbb{R}_+ denotes the set of positive real numbers.

In the electrical domain of the receiver, after cyclic prefix (CP) removal, S/P conversion and fast Fourier transform (FFT), the received FD signal \mathbf{Y}_n at the k^{th} subcarrier is

$$Y[n, k] = H[n, k]X[n, k] + N[n, k], \quad k = 0, \dots, N-1, \quad (3)$$

where $H[n, k]$ is the channel transfer function (CTF), and $N[n, k]$ is the complex additive white Gaussian noise (AWGN) with zero mean and variance σ^2 . Note, however that the VLC system is affected by a few noise sources, typically including the shot noise and the thermal noise. More specifically, the variance of the combined TD noise, which can be approximated as AWGN, is denoted as [29], [30]

$$\sigma_{\text{TD}}^2 = \sigma_{\text{Shot}}^2 + \sigma_{\text{Thermal}}^2, \quad (4)$$

where σ_{Shot}^2 and $\sigma_{\text{Thermal}}^2$ respectively denote the variances of the shot noise and the thermal noise formulated by [29]

$$\begin{cases} \sigma_{\text{Shot}}^2 = 2qR [P_{\text{Signal}}(t) + P_{\text{Daylight}}] \\ \sigma_{\text{Thermal}}^2 = \frac{4}{r} \cdot k_b B T \end{cases}, \quad (5)$$

where q is the charge on electron, R is the responsivity of the photo-detector (PD), $P_{\text{Signal}}(t)$ is the instantaneous received power, P_{Daylight} is the mean power received from the diffuse sunlight in indoor environment, k_b is the Boltzmann's constant, B is the bandwidth and \mathcal{T} is the temperature of the noise equivalent input resistance r . It is worth noting that although the variance σ_{TD}^2 of the combined TD noise contains a TD shot noise with a time-varying variance σ_{Shot}^2 , its equivalent FD version can be approximated as an AWGN with a constant variance of $\sigma^2 = 2qRN(P_{\text{Rx}} + P_{\text{Daylight}}) + N\sigma_{\text{Thermal}}^2$, where P_{Rx} is the average optical receive power across the room.

Then, with the aid of the CE block in Fig. 1, the estimated channel coefficients $\hat{H}[n, k]$ can be obtained. Briefly speaking, the AS-BMMSE-CE scheme estimates the mean value of the tap coefficient vector $\hat{\mathbf{h}}_{\text{h}}^n$ of length L_c , whose l^{th} ($l \in \{0, \dots, L_c - 1\}$) element is the mean tap coefficient averaged within an optimised statistic window size $\omega_{l,\text{opt}}^n$, and n refers to the UE's current position. Similarly, the covariance matrix of the CIR, denoted by the $L_c \times L_c$ matrix \mathbf{C}_{h}^n , can also be obtained through linearly smoothing its values corresponding to the UE's past and current positions within the predefined maximal statistic window size ω_{max} . More details will be revealed in Section III and Section IV.

III. VSW-AIDED AS-BMMSE-CE: AN OVERVIEW

The proposed VSW-aided AS-BMMSE-CE scheme is implemented in the CE block seen in Fig. 1, while its flowchart is portrayed in Fig. 2, where the variables are defined in relevant contexts of the paper. We assume that a comb-type pilot pattern with subcarrier indices defined by P_{pilot} is used, where the specific pilot arrangement is provided in Section IV-C. The least squares (LS) based CE is first invoked to obtain the CTF estimates at the N_p pilot subcarriers, yielding

$$\hat{H}[n, k] = \frac{Y[n, k]}{X[n, k]} = H[n, k] + \frac{X^*[n, k]}{|X[n, k]|^2} N[n, k], \quad k \in P_{\text{pilot}}. \quad (6)$$

Next, the maximum likelihood estimation (MLE) [7] process seen in Fig. 2 is used to get the estimated TD CIR vector, namely $\hat{\mathbf{h}}_{\text{ML}}^n$, which approaches the CRLB without *a priori* knowledge on CIR [7], [26].

Based on $\hat{\mathbf{h}}_{\text{ML}}^n$, a procedure referred to as CCUA, whose details are to be revealed in Algorithm 1 of Section IV-D, is invoked for achieving the covariance coefficient of the channel taps. Then for the l^{th} ($l = 0, \dots, L_c - 1$) tap, the results generated by CCUA are used to determine the variation of the coefficient as well as the associated variance, based on which a comparison with ω_{max} is conducted. According to the comparison result, we can decide whether an exhaustive search over the candidate window sizes $\omega_l^n \in [1, \omega_{\text{max}}]$ has to be launched, for identifying the optimal VSW size $\omega_{l,\text{opt}}^n$ associated with the l^{th} tap. If such an exhaustive search is needed, the proposed CMUA procedure described by Algorithm 2 in Section IV-D will be activated, which facilitates the construction of a covariance matrix to be exploited by the following optimisation on the tap-specific VSW sizes. Then, each tap's most recent coefficients within the optimal VSW are averaged. With the aid of the optimised means and variances

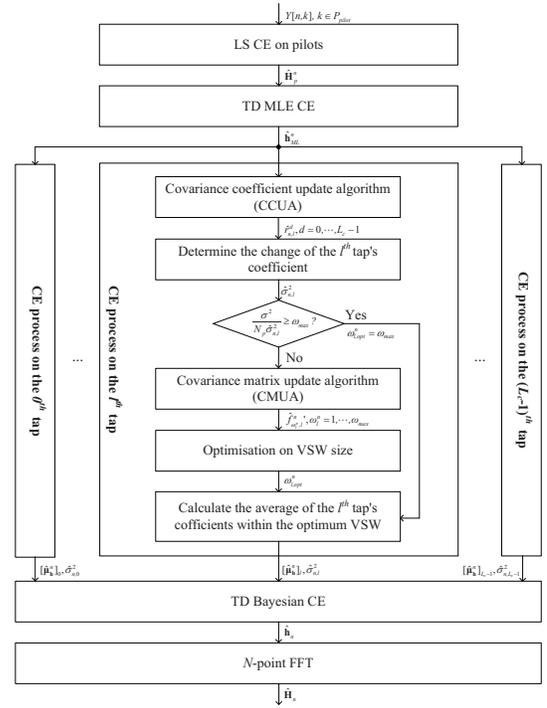


Fig. 2. The flowchart of the proposed AS-BMMSE-CE scheme.

of the CIR coefficients, the conventional BMMSE-CE procedure [7] can be used, resulting in improved CIR estimates. Finally, FD CTF estimates are obtained after applying N -point FFT on the estimated CIR.

IV. DETAILED CE DESIGN

In this section, we will elaborate on the design of the proposed AS-BMMSE-CE scheme illustrated in Fig. 2.

A. TD PACE Process

As indicated in Fig. 2, the MLE-aided TD CE function is invoked to get the initial estimates of \mathbf{h}_n in (2), utilising the LS-based channel estimates at pilot subcarriers. Assuming \mathbf{h}_n is deterministic but unknown, the MLE-based CE is capable of approaching the CRLB [7], [26]. To elaborate a little further, first note that the FD CTF vector \mathbf{H}_n can be calculated through

$$\mathbf{H}_n = \mathbf{B}\mathbf{h}_n, \quad (7)$$

where $\mathbf{B} = \{B_{k,l}\} \in \mathbb{C}^{N \times L_c}$, $B_{k,l} = e^{-j\frac{2\pi kl}{N}}$ for $0 \leq k \leq N-1$, $0 \leq l \leq L_c-1$. We denote the FD noise after FFT as

$$\mathbf{N}_n = \mathbf{D}\mathbf{n}_n \in \mathbb{C}^{N \times 1}, \quad (8)$$

where \mathbf{n}_n is the TD real-valued electrical AWGN with zero mean and covariance $\bar{\sigma}^2 \mathbf{I}_N$, \mathbf{N}_n is complex-valued AWGN with zero mean and covariance $\sigma^2 \mathbf{I}_N$, and $\mathbf{D} = \mathbf{F}_I^{-1} = \{D_{n,k}\} \in \mathbb{C}^{N \times N}$, $D_{n,k} = e^{-j\frac{2\pi nk}{N}}$ for $0 \leq \{n, k\} \leq N-1$.

Define \mathbf{H}_p^n as the CTF vector corresponding to pilot subcarriers, formulated by

$$\mathbf{H}_p^n = \mathbf{S}\mathbf{H}_n, \quad (9)$$

where \mathbf{S} is an $N_p \times N$ selecting matrix that helps to extract the pilots' indices. More specifically, the i^{th} ($i = 0, \dots, N_p - 1$) row of \mathbf{S} is constituted by zeros except the $([P_{\text{pilot}}]_i)^{th}$ element, which has a value of 1. It implies that $[\mathbf{S}]_{i,[P_{\text{pilot}}]_i} = 1$ and $\mathbf{S}\mathbf{S}^H = \mathbf{I}_{N_p}$. We also define an $N_p \times L_c$ matrix

$$\mathbf{W}_P = \mathbf{S}\mathbf{B}, \quad (10)$$

where the elements of \mathbf{W}_P are $[\mathbf{W}_P]_{k,l} = e^{-\frac{j2\pi \cdot [P_{\text{pilot}}]_k \cdot l}{N}}$ ($0 \leq k \leq N_p - 1$, $0 \leq l \leq L_c - 1$). According to [7], [26], the MLE estimate of the CIR is

$$\hat{\mathbf{h}}_{\text{ML}}^n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \hat{\mathbf{H}}_P^n, \quad (11)$$

where $\hat{\mathbf{H}}_P^n$ is the LS estimates of \mathbf{H}_P^n in (9), formulated by

$$\hat{\mathbf{H}}_P^n = \mathbf{W}_P \mathbf{h}_n + \boldsymbol{\varrho}_n^{-1} \mathbf{S} \mathbf{N}_n = \mathbf{W}_P \mathbf{h}_n + \mathbf{V}_n, \quad (12)$$

where we define

$$\mathbf{V}_n = \boldsymbol{\varrho}_n^{-1} \mathbf{S} \mathbf{N}_n, \quad (13)$$

while $\boldsymbol{\varrho}_n = \text{diag}\{p_0, \dots, p_{N_p-1}\}$ and p_i is the i^{th} ($i = 0, \dots, N_p - 1$) pilot symbol. Without loss of generality, we assume that $p_i = \pm 1$, $i = 0, \dots, N_p - 1$. Note that by using pilot symbols with constant amplitude, each element in \mathbf{V}_n is AWGN with zero mean and variance σ^2 , yielding $E\{\mathbf{V}_n\} = \mathbf{0}_{N_p \times 1}$ and $E\{\mathbf{V}_n \mathbf{V}_n^H\} = E\{\boldsymbol{\varrho}_n^{-1} \mathbf{S} \mathbf{N}_n \mathbf{N}_n^H \mathbf{S}^H \boldsymbol{\varrho}_n^{-1H}\} = \sigma^2 \mathbf{I}_{N_p}$.

Different from the MLE-based CE that assumes no information of \mathbf{h}_n , the so-called BMMSE estimator [7] assumes that the mean value and the covariance matrix of the tap-specific coefficients at the UE's n^{th} position, which are respectively denoted by an $L_c \times 1$ vector $\boldsymbol{\mu}_{\mathbf{h}}^n$ and an $L_c \times L_c$ matrix $\mathbf{C}_{\mathbf{h}}^n$, are known. The BMMSE version of the CIR estimate is [7]

$$\hat{\mathbf{h}}_n = \boldsymbol{\mu}_{\mathbf{h}}^n + \Phi_n \mathbf{W}_P^H (\hat{\mathbf{H}}_P^n - \mathbf{W}_P \boldsymbol{\mu}_{\mathbf{h}}^n), \quad (14)$$

where we define $\Phi_n = [\mathbf{W}_P^H \mathbf{W}_P + \sigma^2 (\mathbf{C}_{\mathbf{h}}^n)^{-1}]^{-1}$. Note that the BMMSE estimates of (14) are more accurate than their MLE counterparts of (11), thanks to the knowledge of $\boldsymbol{\mu}_{\mathbf{h}}^n$ and $\mathbf{C}_{\mathbf{h}}^n$. However, in practical VLC systems the values of $\boldsymbol{\mu}_{\mathbf{h}}^n$ and $\mathbf{C}_{\mathbf{h}}^n$ are typically difficult to obtain or unavailable, thus greatly restricting the applicability of the conventional BMMSE-CE method. Hence, one key issue is that how to derive a method for estimating these parameters in an efficient and robust way, such that the practicality of BMMSE-CE for VLC systems can be improved. We will show the solution to this issue in the remaining sections.

B. VSW-based Optimisation

In this section, we show how $\boldsymbol{\mu}_{\mathbf{h}}^n$ can be estimated, together with the derivation of the objective function for our CE problem. By inserting (12) into (14), we have

$$\begin{aligned} \hat{\mathbf{h}}_n &= \boldsymbol{\mu}_{\mathbf{h}}^n + \Phi_n \mathbf{W}_P^H (\mathbf{W}_P \mathbf{h}_n + \mathbf{V}_n) - \Phi_n \mathbf{W}_P^H \mathbf{W}_P \boldsymbol{\mu}_{\mathbf{h}}^n \\ &= (\mathbf{I}_{L_c} - \Phi_n \mathbf{W}_P^H \mathbf{W}_P) \boldsymbol{\mu}_{\mathbf{h}}^n + \Phi_n \mathbf{W}_P^H \mathbf{W}_P \mathbf{h}_n + \Phi_n \mathbf{W}_P^H \mathbf{V}_n \\ &= \mathbf{h}_n + \boldsymbol{\varepsilon}_n, \end{aligned} \quad (15)$$

where $\boldsymbol{\varepsilon}_n$ denotes the estimation error for the TD CIR and is formulated by

$$\boldsymbol{\varepsilon}_n = \Phi_n \mathbf{W}_P^H \mathbf{V}_n - (\mathbf{I}_{L_c} - \Phi_n \mathbf{W}_P^H \mathbf{W}_P) \Delta \mathbf{h}_n, \quad (16)$$

where

$$\Delta \mathbf{h}_n = \mathbf{h}_n - \boldsymbol{\mu}_{\mathbf{h}}^n \quad (17)$$

denotes the difference vector between the CIR \mathbf{h}_n and its mean $\boldsymbol{\mu}_{\mathbf{h}}^n$ at the UE's n^{th} position. Furthermore, (16) may be rewritten as

$$\boldsymbol{\varepsilon}_n = \Psi_1^n \mathbf{V}_n - \Psi_2^n \Delta \mathbf{h}_n, \quad (18)$$

where we define

$$\Psi_1^n = \Phi_n \mathbf{W}_P^H, \quad \Psi_2^n = \mathbf{I}_{L_c} - \Phi_n \mathbf{W}_P^H \mathbf{W}_P, \quad (19)$$

Since $\boldsymbol{\mu}_{\mathbf{h}}^n$ in (17) is not obtainable in practical systems, we may instead use its *a priori* estimate $\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n$, yielding the estimated CIR difference

$$\Delta \hat{\mathbf{h}}_n = \mathbf{h}_n - \hat{\boldsymbol{\mu}}_{\mathbf{h}}^n. \quad (20)$$

Note that the VLC channel model [11], [12] usually contains one light-of-sight (LOS) tap and a few higher-order reflective taps, where different taps may have different statistical characteristics. Thus, in order to improve the accuracy of $\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n$, we propose the so-called VSW mechanism, which exploits the tap-specific past channel information in a given statistic window with an optimised size. In this scheme, each element of $\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n$ is the tap coefficient averaged over the specific statistic window size ω_l^n , $l \in \{0, \dots, L_c - 1\}$, formulated as

$$[\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n]_l = \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} [\hat{\mathbf{h}}_{\text{ML}}^{n-k}]_l, \quad l \in \{0, \dots, L_c - 1\}, \quad (21)$$

where based on (11), the MLE-based estimate is given by [7]

$$\hat{\mathbf{h}}_{\text{ML}}^n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \hat{\mathbf{H}}_P^n = \mathbf{h}_n + \mathbf{v}_n = \boldsymbol{\mu}_{\mathbf{h}}^n + \Delta \mathbf{h}_n + \mathbf{v}_n, \quad (22)$$

while the superscript $(\cdot)^{n-k}$ in (21) denotes the $(n-k)^{th}$ position. Inserting (12) into (22), the equivalent TD noise \mathbf{v}_n can be calculated as

$$\mathbf{v}_n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \hat{\mathbf{H}}_P^n - \mathbf{h}_n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \mathbf{V}_n. \quad (23)$$

Utilising (21) and (22), we may further develop (20) as

$$\begin{aligned} [\Delta \hat{\mathbf{h}}_n]_l &= [\boldsymbol{\mu}_{\mathbf{h}}^n]_l + [\Delta \mathbf{h}_n]_l - \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} ([\boldsymbol{\mu}_{\mathbf{h}}^{n-k}]_l + [\mathbf{v}_{n-k}]_l + [\Delta \mathbf{h}_{n-k}]_l) \\ &= \frac{\omega_l^n - 1}{\omega_l^n} [\Delta \mathbf{h}_n]_l - \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} [\mathbf{v}_{n-k}]_l - \frac{1}{\omega_l^n} \sum_{k=1}^{\omega_l^n - 1} [\Delta \mathbf{h}_{n-k}]_l. \end{aligned} \quad (24)$$

If we define the FD MSE associated with the k^{th} subcarrier at the UE's n^{th} position as $\gamma^n(k) = E\{|\hat{H}[n, k] - H[n, k]|^2\}$, then the MSE averaged over one OFDM symbol can be denoted by $\Gamma_n = \frac{1}{N} \sum_{k=0}^{N-1} \gamma^n(k)$. Using (7), (15) and (18), Γ_n may be transformed to

$$\begin{aligned} \Gamma_n &= \frac{1}{N} \text{Tr}\{E\{(\hat{\mathbf{H}}_n - \mathbf{H}_n)(\hat{\mathbf{H}}_n - \mathbf{H}_n)^H\}\} \\ &= \frac{1}{N} \text{Tr}\{E\{[\mathbf{B}(\mathbf{h}_n + \boldsymbol{\varepsilon}_n) - \mathbf{B}\mathbf{h}_n][\mathbf{B}(\mathbf{h}_n + \boldsymbol{\varepsilon}_n) - \mathbf{B}\mathbf{h}_n]^H\}\} \\ &= \text{Tr}\{E\{\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^H\}\} \\ &= \sigma^2 \text{Tr}\{\Psi_1^n \Psi_1^{nH}\} + \text{Tr}\{E\{\Psi_2^n \Delta \mathbf{h}_n \Delta \mathbf{h}_n^H \Psi_2^{nH}\}\} \\ &\quad - \text{Tr}\{E\{\Psi_1^n \mathbf{V}_n \Delta \mathbf{h}_n^H \Psi_2^{nH}\}\} - \text{Tr}\{E\{\Psi_2^n \Delta \mathbf{h}_n \mathbf{V}_n^H \Psi_1^{nH}\}\}, \end{aligned} \quad (25)$$

which constitutes the objective function of the proposed AS-BMMSE-CE technique. Naturally, the estimated CIR difference denoted by (24) may be inserted into (25), forming a function of ω_l^n , $l \in \{0, \dots, L_c - 1\}$. Hence in AS-BMMSE-CE, we are interested in finding the optimum values $\omega_{l,\text{opt}}^n$, $l \in \{0, \dots, L_c - 1\}$ that minimise Γ_n of (25)

$$\omega_{l,\text{opt}}^n = \underset{\{\omega_l^n\} \in \mathbb{N}_+}{\text{argmin}} \Gamma_n, \quad l \in \{0, \dots, L_c - 1\}, \quad (26)$$

where \mathbb{N}_+ denotes the set of positive integers.

Nonetheless, as the complicated expression of (25) involves multiple coupled parameters, it may be difficult to solve (26) directly. It is therefore desirable to simplify (25), as to be discussed next.

C. Pilot Pattern and Covariance Matrices

Aiming to simplify (25), let us first cast a deeper insight into it. Note that Ψ_1^n and Ψ_2^n in (25) contain a common term of $\mathbf{W}_P^H \mathbf{W}_P$, where \mathbf{W}_P is defined in (10). Since \mathbf{W}_P is related to the pilot index, it is beneficial to optimise the pilot pattern such that $\mathbf{W}_P^H \mathbf{W}_P$ becomes a diagonal matrix, which then facilitates the simplification of (25). On the other hand, as suggested by [6], the pilots should be equally spaced in the FD to achieve the best CE performance and to achieve the minimal CRLB [7], [26].

Furthermore, recall that in O-OFDM-aided VLC systems, the transmitted data symbols are Hermitian symmetric with respect to the $(N/2)^{\text{th}}$ subcarrier [22]. Thus, \mathbf{W}_P satisfies the semi-orthogonality of

$$\mathbf{W}_P^H \mathbf{W}_P = N_p \mathbf{I}_{L_c}, \quad (27)$$

iff an uniform pilot interval of N_d is adopted and the pilot subcarriers are symmetrically allocated with respect to the $(N/2)^{\text{th}}$ subcarrier, too. In other words, the smallest pilot index P_0 should satisfy

$$P_0 + \left(\frac{N_p}{2} - 1\right) \times N_d + N_d = N - [P_0 + \left(\frac{N_p}{2} - 1\right) \times N_d], \quad (28)$$

where we have $N_p \times N_d = N$. Solving (28) yields

$$P_0 = N_d/2. \quad (29)$$

This is the unique optimised condition that P_0 must fulfil for O-OFDM-VLC systems subject to the above-mentioned design target of (27). Based on (27) and (29), we transform (25) to

$$\Gamma_n = \text{Tr}\{E\{\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^H\}\} = \Theta(f_{\omega_l^n, l}^n, \mathbf{C}_{\mathbf{h}}^n), \quad l = 0, \dots, L_c - 1, \quad (30)$$

where $\Theta(f_{\omega_l^n, l}^n, \mathbf{C}_{\mathbf{h}}^n)$ is a function of $f_{\omega_l^n, l}^n$ and $\mathbf{C}_{\mathbf{h}}^n$.

Let us now calculate the values of $f_{\omega_l^n, l}^n$ and $\mathbf{C}_{\mathbf{h}}^n$. More specifically, $f_{\omega_l^n, l}^n$ represents the l^{th} diagonal element of the diagonal covariance matrix $E\{\Delta \hat{\mathbf{h}}_n \Delta \hat{\mathbf{h}}_n^H\}$, and can be viewed as a function

$$\begin{aligned} f_{\omega_l^n, l}^n(r_{n,l}^d) &= \frac{\sigma^2}{N_p \omega_l^n} + \frac{1}{(\omega_l^n)^2} \sum_{j=1}^{\omega_l^n - 1} \sum_{k=1}^{\omega_l^n - 1} r_{n,l}^{|j-k|} + \frac{(\omega_l^n - 1)^2}{(\omega_l^n)^2} r_{n,l}^0 \\ &\quad - 2 \cdot \frac{\omega_l^n - 1}{(\omega_l^n)^2} \sum_{j=1}^{\omega_l^n - 1} r_{n,l}^j, \quad l = 0, \dots, L_c - 1, \end{aligned} \quad (31)$$

where $d = |j - k|$, $\{j, k\} = 0, \dots, \omega_l^n - 1$ and we define

$$r_{n,l}^d = r_{n,l}^{|j-k|} = E\{([\mathbf{h}_{n-j}]_l - [\boldsymbol{\mu}_{\mathbf{h}}^n]_l)([\mathbf{h}_{n-k}]_l - [\boldsymbol{\mu}_{\mathbf{h}}^n]_l)^*\}, \quad (32)$$

while $r_{n,l}^j$ in (31) is obtained by setting $k = 0$ in (32). The full derivations of (30) and (31) are provided in Appendix I. Note that $r_{n,l}^d$ of (32) are the elements of the UE position covariance matrix $\mathbf{R}_{n,l}$ associated with the l^{th} tap at the n^{th} position, where $\mathbf{R}_{n,l}$ is a real symmetric Toeplitz matrix formulated by

$$\mathbf{R}_{n,l} = \begin{bmatrix} r_{n,l}^0 & r_{n,l}^1 & \cdots & r_{n,l}^{\omega_{\max} - 1} \\ r_{n,l}^1 & r_{n,l}^0 & \cdots & r_{n,l}^{\omega_{\max} - 2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,l}^{\omega_{\max} - 1} & r_{n,l}^{\omega_{\max} - 2} & \cdots & r_{n,l}^0 \end{bmatrix}. \quad (33)$$

According to [31], the estimate of $r_{n,l}^d$ can be expressed as

$$\hat{r}_{n,l}^d = \frac{1}{\omega_{\max}} \sum_{j=0}^{\omega_{\max} - d - 1} ([\hat{\mathbf{h}}_{\text{ML}}^{n-j}]_l - [\bar{\boldsymbol{\mu}}_n]_l)([\hat{\mathbf{h}}_{\text{ML}}^{n-(j+d)}]_l - [\bar{\boldsymbol{\mu}}_n]_l), \quad (34)$$

where $\hat{\mathbf{h}}_{\text{ML}}^{(\cdot)}$ is given by (22), and $[\bar{\boldsymbol{\mu}}_n]_l$ is the mean of the l^{th} tap's coefficients, which is averaged over the maximal statistic window utilising MLE as $\bar{\boldsymbol{\mu}}_n = \frac{1}{\omega_{\max}} \sum_{k=0}^{\omega_{\max} - 1} \hat{\mathbf{h}}_{\text{ML}}^{n-k}$. Moreover, $\omega_{\max} \geq \omega_l^n$, $l \in \{0, \dots, L_c - 1\}$ is the maximum length of the statistic windows, and its value should be carefully selected. If it is too large, the accuracy of $\hat{r}_{n,l}^d$ may be biased by more distanced and thus less relevant channel information. In contrast, if it is too small, the result of $\hat{r}_{n,l}^d$ may be dominated by residual noise which is not effectively mitigated due to insufficient past channel information.

After obtaining $\bar{\boldsymbol{\mu}}_n$, we can use it to calculate (34) for generating $\mathbf{R}_{n,l}$ defined in (33). Note that the MLE estimate, namely $\hat{\mathbf{h}}_{\text{ML}}^{(\cdot)}$ in (34), is contaminated by noise. We show in Appendix I that the expectation of $\hat{r}_{n,l}^d$ in (34) contains TD noise items of

$$E\{\hat{r}_{n,l,\text{noise}}^d\} = \begin{cases} \frac{\omega_{\max} - 1}{\omega_{\max}} \sigma_0^2, & d = 0 \\ -\frac{\omega_{\max} - d}{\omega_{\max}} \cdot \frac{1}{\omega_{\max}} \sigma_0^2, & d = 1, \dots, \omega_{\max} - 1 \end{cases}, \quad (35)$$

where $\sigma_0^2 = \frac{\sigma^2}{N_p}$ is the TD residual noise variance under the specific pilot pattern designed earlier in this section.

After replacing $r_{n,l}^d$ in (31) with $\hat{r}_{n,l}^d$ in (34), we have $f_{\omega_l^n, l}^n(r_{n,l}^d) \rightarrow \hat{f}_{\omega_l^n, l}^n(\hat{r}_{n,l}^d)$. Utilising (35), we can therefore obtain the expectation of the introduced noise item as

$$E\{\hat{f}_{\omega_l^n, l,\text{noise}}^n(\hat{r}_{n,l}^d)\} = -\frac{2(\omega_l^n)^2 - 3\omega_l^n(\omega_{\max}^2 + 1) + 3\omega_{\max}^2 + 1}{3(\omega_l^n)^2 \omega_{\max}^2} \sigma_0^2, \quad (36)$$

where more details can be found in Appendix I. Then, in order to eliminate the impact from the noise specified by (36), we may use

$$\hat{f}_{\omega_l^n, l}' = \hat{f}_{\omega_l^n, l}^n - E\{\hat{f}_{\omega_l^n, l,\text{noise}}^n\} \quad (37)$$

to replace $f_{\omega_l^n, l}^n$ in (30) and (31).

Next, we proceed to calculate $\mathbf{C}_{\mathbf{h}}^n$ specified in (30). Assuming that the variations of coefficients associated with different channel taps, which are represented by the elements of $\Delta \mathbf{h}_n$, are uncorrelated [7], we have

$$\mathbf{C}_{\mathbf{h}}^n = E\{\Delta \mathbf{h}_n \Delta \mathbf{h}_n^H\} = \text{diag}\{\sigma_{n,0}^2, \dots, \sigma_{n,L_c-1}^2\}, \quad (38)$$

where $\sigma_{n,l}^2$ ($l = 0, \dots, L_c - 1$) denote the variance of $[\Delta \mathbf{h}_n]_l$ in (17) that corresponds to the l^{th} tap at the UE's n^{th} position. In order to obtain \mathbf{C}_n^n , a forgetting factor λ is exploited to calculate the estimate of $\sigma_{n,l}^2$, namely $\hat{\sigma}_{n,l}^2$. More explicitly, we define [32]

$$\bar{\sigma}_{n,l}^2 = \lambda \hat{\sigma}_{n-1,l}^2 + (1 - \lambda) \left(\hat{r}_{n,l}^0 - \frac{\omega_{\max} - 1}{\omega_{\max}} \sigma_0^2 \right). \quad (39)$$

We will discuss how to select the value of λ in Section V. Noting that $\hat{\sigma}_{n,l}^2$ should be a positive value, we may apply a small covariance constant σ_{const}^2 to (39), resulting in

$$\hat{\sigma}_{n,l}^2 = \begin{cases} \bar{\sigma}_{n,l}^2, & \bar{\sigma}_{n,l}^2 > 0 \\ \sigma_{\text{const}}^2, & \bar{\sigma}_{n,l}^2 \leq 0 \end{cases}, \quad (40)$$

which is the estimate of the l^{th} diagonal element of \mathbf{C}_n^n .

Based on (37) and (40), we therefore simplify the objective function (25) to (30) and (31), which involve a number of L_c target variables to be optimised, namely the statistic window sizes ω_l^n , $l = 0, \dots, L_c - 1$. More details on the optimisation procedure will be provided in Section IV-E.

D. Considerations on Complexity Reduction

After the operations conducted in Section IV-C, we manage to derive a simplified objective function (30). However, the calculation of (34) and (37) requires a relatively high complexity. For instance, a computational complexity of $O(\omega_{\max}^2 L_c)$ is required for thoroughly searching through $d = 0, \dots, \omega_{\max} - 1$ and $l = 0, \dots, L_c - 1$ in (34). Such a complexity, however, may be reduced by the algorithms proposed in this section.

Let us first expand (34) to

$$\hat{r}_{n,l}^d = \frac{\hat{\phi}_{l,1}^{n,d} - (\hat{\phi}_{l,2}^{n,d} + \hat{\phi}_{l,3}^{n,d})[\bar{\boldsymbol{\mu}}_n]_l + (\omega_{\max} - d)[\bar{\boldsymbol{\mu}}_n]_l^2}{\omega_{\max}}, \quad (41)$$

where

$$\begin{cases} \hat{\phi}_{l,1}^{n,d} = \sum_{j=0}^{\omega_{\max}-d-1} [\hat{\mathbf{h}}_{\text{ML}}^{n-j}]_l [\hat{\mathbf{h}}_{\text{ML}}^{n-(j+d)}]_l \\ \hat{\phi}_{l,2}^{n,d} = \sum_{j=0}^{\omega_{\max}-d-1} [\hat{\mathbf{h}}_{\text{ML}}^{n-j}]_l \\ \hat{\phi}_{l,3}^{n,d} = \sum_{j=0}^{\omega_{\max}-d-1} [\hat{\mathbf{h}}_{\text{ML}}^{n-(j+d)}]_l \end{cases}, \quad (42)$$

which may be further reformulated as

$$\begin{cases} \hat{\phi}_{l,1}^{n,d} = \hat{\phi}_{l,1}^{n-1,d} + [\hat{\mathbf{h}}_{\text{ML}}^{n-d}]_l [\hat{\mathbf{h}}_{\text{ML}}^n]_l - [\hat{\mathbf{h}}_{\text{ML}}^{n-\omega_{\max}}]_l [\hat{\mathbf{h}}_{\text{ML}}^{n-\omega_{\max}+d}]_l \\ \hat{\phi}_{l,2}^{n,d} = \hat{\phi}_{l,2}^{n-1,d} + [\hat{\mathbf{h}}_{\text{ML}}^n]_l - [\hat{\mathbf{h}}_{\text{ML}}^{n-\omega_{\max}+d}]_l \\ \hat{\phi}_{l,3}^{n,d} = \hat{\phi}_{l,3}^{n-1,d} + [\hat{\mathbf{h}}_{\text{ML}}^{n-d}]_l - [\hat{\mathbf{h}}_{\text{ML}}^{n-\omega_{\max}}]_l \end{cases}. \quad (43)$$

Thanks to the recursive form of (43), the computational complexity of (34) can be reduced to $O(\omega_{\max} L_c)$. We summarised the proposed covariance coefficient update algorithm (CCUA) in Algorithm 1.

On the other hand, a computational complexity of $O(\omega_{\max}^3)$ is imposed by (37) for fully testing $\omega_l^n = 1, \dots, \omega_{\max}$ for the l^{th} tap. We may expand (37) to

$$\hat{f}_{\omega_l^n, l}^n = \frac{\sigma^2}{N_p \omega_l^n} + \frac{1}{\omega_l^2} \hat{\phi}_{\omega_l^n, l}^{n,1} + \frac{(\omega_l^n - 1)^2}{(\omega_l^n)^2} \hat{r}_{n,l}^0 - \frac{\omega_l^n - 1}{(\omega_l^n)^2} \hat{\phi}_{\omega_l^n, l}^{n,2} + \frac{2(\omega_l^n)^2 - 3\omega_l^n(\omega_{\max}^2 + 1) + 3\omega_{\max}^2 + 1}{3\omega_l^n \omega_{\max}^2} \sigma_0^2, \quad (44)$$

Algorithm 1 Covariance Coefficient Update Algorithm (CCUA)

```

1: Initialisation: Obtain  $\hat{\phi}_{l,1}^{n-1,d}$ ,  $\hat{\phi}_{l,2}^{n-1,d}$ ,  $\hat{\phi}_{l,3}^{n-1,d}$ ,  $\bar{\boldsymbol{\mu}}_{n-1}$  and set  $l = 1$ .
2: repeat
3:    $[\bar{\boldsymbol{\mu}}_n]_l = [\bar{\boldsymbol{\mu}}_{n-1}]_l + \frac{[\hat{\mathbf{h}}_{\text{ML}}^n]_l - [\hat{\mathbf{h}}_{\text{ML}}^{n-\omega_{\max}}]_l}{\omega_{\max}}$ 
4:    $d = 0$ 
5:   repeat
6:     Calculate (43) and (41)
7:      $d = d + 1$ 
8:   until  $d > \omega_{\max}$ 
9:    $l = l + 1$ 
10: until  $l > L_c - 1$ 
11: Return:  $\hat{r}_{n,l}^d$ ,  $l = 0, \dots, L_c - 1$ ,  $d = 0, \dots, \omega_{\max} - 1$ .

```

where

$$\begin{cases} \hat{\phi}_{\omega_l^n, l}^{n,1} = \sum_{j=1}^{\omega_l^n-1} \sum_{k=1}^{\omega_l^n-1} \hat{r}_{n,l}^{|j-k|} \\ \hat{\phi}_{\omega_l^n, l}^{n,2} = 2 \sum_{j=1}^{\omega_l^n-1} \hat{r}_{n,l}^j \end{cases}. \quad (45)$$

In order to reduce the complexity, we rewrite (45) as

$$\begin{cases} \hat{\phi}_{\omega_l^n+1, l}^{n,1} = \hat{\phi}_{\omega_l^n, l}^{n,1} + \hat{\phi}_{\omega_l^n, l}^{n,1} + \hat{r}_{n,l}^0 \\ \hat{\phi}_{\omega_l^n+1, l}^{n,2} = \hat{\phi}_{\omega_l^n, l}^{n,2} + 2\hat{r}_{n,l}^{\omega_l^n} \end{cases}. \quad (46)$$

Using (46), the complexity of (37) can be reduced to $O(\omega_{\max})$. The proposed covariance matrix update algorithm (CMUA) is summarised in Algorithm 2.

Algorithm 2 Covariance Matrix Update Algorithm (CMUA)

```

1: Initialisation: Obtain  $\hat{r}_{n,l}^d$ ,  $l = 0, \dots, L_c - 1$ ,  $d = 0, \dots, \omega_{\max} - 1$ . Set  $\hat{\phi}_{1,l}^{n,1} = 0$ ,  $\hat{\phi}_{1,l}^{n,2} = 0$  with given  $l$  and  $\omega_l^n = 1$ .
2: repeat
3:   Calculate (44) and (46)
4:    $\omega_l^n = \omega_l^n + 1$ 
5: until  $\omega_l^n > \omega_{\max}$ 
6: Return:  $\hat{f}_{\omega_l^n, l}^n$ ,  $\omega_l^n = 1, \dots, \omega_{\max}$ .

```

E. Optimum VSW Size and MSE Bound

Recall that the optimum solution for the objective function Γ_n defined in (25) or (30) is given by (26), which is an integer programming problem since the variables $\omega_{l,\text{opt}}^n$ to be optimised are integers, and thus a traditional NP-complete problem [33]. Since there are a total number of ω_{\max} candidate window sizes for each of the L_c taps, the optimisation of (26) results in a high computational complexity of $O[(\omega_{\max})^{L_c}]$.

Nonetheless, note that (30), which is further developed in (56) of Appendix I, may be reformulated as

$$\begin{aligned} \hat{\Gamma}_n &= \sum_{l=0}^{L_c-1} \left\{ \frac{N_p \sigma^2}{(N_p + \frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2} + \frac{(\frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2 \hat{f}_{\omega_l^n, l}^n}{(N_p + \frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2} + \frac{2\sigma^2}{\omega_l^n} \frac{\frac{\sigma^2}{\hat{\sigma}_{n,l}^2}}{(N_p + \frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2} \right\} \\ &= \sum_{l=0}^{L_c-1} \hat{M}_{\omega_l^n, l}^n, \end{aligned} \quad (47)$$

where we define

$$\hat{M}_{\omega_l^n, l}^n = \frac{N_p \sigma^2}{(N_p + \frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2} + \frac{(\frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2 \hat{f}_{\omega_l^n, l}^n}{(N_p + \frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2} + \frac{2\sigma^2}{\omega_l^n} \frac{\frac{\sigma^2}{\hat{\sigma}_{n,l}^2}}{(N_p + \frac{\sigma^2}{\hat{\sigma}_{n,l}^2})^2} \quad (48)$$

and $\hat{\sigma}_{n,l}^2$ is given in (40). Note that the corresponding estimated version of Γ_n and $f_{\omega_{n,l}^n}^n$ are used in (47). Therefore, we can see that $\hat{\Gamma}_n$ can be effectively decoupled into independent items $\hat{M}_{\omega_{n,l}^n}^n$, $l \in \{0, \dots, L_c - 1\}$, which are associated with ω_l^n . Hence, with the aid of (48), we may solve $\hat{\Gamma}_n$ through exhaustively searching for each tap-specific $\omega_{l,\text{opt}}^n$ in the candidate solution set of $\{1, \dots, \omega_{\max}\}$, yielding

$$\omega_{l,\text{opt}}^n = \underset{\omega_l^n \in \{1, \dots, \omega_{\max}\}}{\text{argmin}} \hat{M}_{\omega_l^n}^n. \quad (49)$$

In this case, the resultant complexity required by (26) can be significantly reduced from $O[(\omega_{\max})^{L_c}]$ to $O(\omega_{\max} L_c)$.

Moreover, the exhaustive search required by (49) may be further simplified under certain conditions. More specifically, we have the following theorem:

Theorem 1: For the l^{th} tap of the CIR, there exists a condition, under which the solution of $\omega_{l,\text{opt}}^n = \omega_{\max}$ can be achieved.

The proof of Theorem 1 is given in Appendix II, where we show that one of such conditions is

$$\psi = \frac{\sigma^2}{N_p \sigma_{n,l}^2} \geq \kappa = 4\omega_{\max} - 10 + \frac{4}{\omega_{\max}}. \quad (50)$$

In this case, when given a predefined ω_{\max} , we may avoid the exhaustive search procedure required in (49) by directly setting $\omega_{l,\text{opt}}^n = \omega_{\max}$, if the above-mentioned condition of $\psi \geq \kappa$ is satisfied. However, this condition may not be easily fulfilled in a typical indoor VLC environment under a large κ . Nonetheless, if we set $\omega_{\max} > \frac{10}{3}$, then based on the definition of κ in (50), we arrive at $\kappa \geq \omega_{\max}$, which thus results in a relaxed condition of $\psi \geq \omega_{\max}$. Such a condition is easier to satisfy than (50), because the random variable ψ falls more likely into $(0, \omega_{\max})$ than into $(0, \kappa)$ due to $\omega_{\max} < \kappa$. Using the relaxed condition of $\psi \geq \omega_{\max}$, the optimisation procedure of (49) may be largely simplified. Based on the above analysis, we summarise the proposed VSW optimisation algorithm (VOA) in Algorithm 3.

Algorithm 3 VSW Optimisation Algorithm (VOA)

```

1: Initialisation: Set  $l = 0$ .
2: repeat
3:   if  $\frac{\sigma^2}{N_p \sigma_{n,l}^2} \geq \omega_{\max}$  then
4:      $\omega_{l,\text{opt}}^n = \omega_{\max}$ 
5:     goto: 14
6:   end if
7:   Execute Algorithm 2
8:    $\omega_l^n = 1$ 
9:   repeat
10:    Calculate (48)
11:     $\omega_l^n = \omega_l^n + 1$ 
12:    until  $\omega_l^n > \omega_{\max}$ 
13:     $\omega_{l,\text{opt}}^n = \underset{\omega_l^n \in \{1, \dots, \omega_{\max}\}}{\text{argmin}} \hat{M}_{\omega_l^n}^n$ 
14:     $l = l + 1$ 
15: until  $l > L_c - 1$ 
16: Return:  $\omega_{l,\text{opt}}^n$ ,  $l = 0, \dots, L_c - 1$  are the optimal VSW sizes.

```

Next, it is worth pointing out that the upper and lower bounds of the AS-BMMSE-CE scheme, termed respectively as AS-BMMSE upper bound (ASB-UB) and AS-BMMSE lower

bound (ASB-LB), are better than some existing ones. More explicitly, we have the following theorem:

Theorem 2: The upper bound of $M_{\omega_{n,l}^n}^n$ is lower than the CRLB [26], and the lower bound of $M_{\omega_{n,l}^n}^n$ is lower than the traditional Bayesian lower bound (TBLB) [7], [26], where $M_{\omega_{n,l}^n}^n$ is the ideal version of $\hat{M}_{\omega_{n,l}^n}^n$ in (48).

The proof of Theorem 2 is given in Appendix III. Based on the above discussions, we finally outline the proposed AS-BMMSE-CE scheme in Algorithm 4, whose visual illustration is provided in Fig. 2.

Algorithm 4 The AS-BMMSE Algorithm

```

1: Initialisation: Obtain  $\lambda$ ,  $\omega_{\max}$  and  $L_c$ . Set  $\hat{\mathbf{C}}_{\mathbf{h}}^n = \text{diag}(\frac{1}{L_c}, \dots, \frac{1}{L_c})$ ,  $\hat{\mathbf{h}}_{\text{ML}}^k = \mathbf{0}$ ,  $k = 0, -1, \dots, -\omega_{\max} + 2$  and  $n = 1$ .
2: repeat
3:    $\hat{H}[n, k] = \frac{\hat{Y}[n, k]}{X[n, k]}$ ,  $k \in P_{\text{pilot}}$ 
4:   Calculate (11)
5:   Execute Algorithm 1
6:    $l = 0$ 
7:   repeat
8:     Calculate (39) and (40)
9:      $l = l + 1$ 
10:  until  $l > L_c - 1$ 
11:  Execute Algorithm 3
12:   $l = 0$ 
13:  repeat
14:     $[\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n]_l = \frac{1}{\omega_{l,\text{opt}}^n} \sum_{i=0}^{\omega_{l,\text{opt}}^n - 1} [\hat{\mathbf{h}}_{\text{ML}}^{n-i}]_l$ 
15:     $l = l + 1$ 
16:  until  $l > L_c - 1$ 
17:  Calculate (14)
18:  Apply  $N$ -point FFT to get  $\hat{\mathbf{H}}_n$ 
19:   $n = n + 1$ 
20: until  $n$  approaches a predefined maximal value.

```

V. NUMERICAL RESULTS AND ANALYSIS

In this section, simulation results are provided for demonstrating the effectiveness of the proposed AS-BMMSE-CE scheme. Assuming a general indoor scenario, a room model with a size of $5 \times 5 \times 4\text{m}^3$ is adopted, where the maximal reflection order of the VLC channel model [12] is set to three, while the centre of the room is located at $(0, 0)$. Four rooftop LEDs, each assuming a fixed transmit power, form a square-shaped coverage area for both illumination and communication services. The UE employs a single PD to receive the same signal transmitted from all LEDs and moves around in the room. Naturally, the instantaneous CIR varies as soon as UE's position changes. Note that the field of view (FOV) of the PD may have an impact on the performance of VLC systems. As an example, we set the FOV to 85° as Configuration A in [11]. The parameters in Table I apply to most scenarios tested in this section, unless otherwise stated.

As the first test, in Fig. 3(a), we evaluate the theoretical MSE performance of AS-BMMSE-CE using (25) with different sizes of the statistic window for a single channel tap. For simplicity, σ^2 is normalised to N_p . Without loss of generality, we show two example cases associated with two randomly selected UE positions, namely the 6^{th} and 8^{th} taps at the positions of $(-1.6, -0.7)$ and $(-1.0, 0.5)$, respectively. From

TABLE I
MAJOR PARAMETERS FOR SIMULATIONS.

Parameter	Value
Reflection coefficient (wall/floor/ceiling)	0.8
DC bias	13dB
Sampling rate	500MHz
LED position	$(-1.5, -1.5, 4)$, $(1.5, 1.5, 4)$, $(-1.5, 1.5, 4)$, $(1.5, -1.5, 4)$
Semi-half power angle	60°
Field of view (FOV)	85°
Modulation scheme	DCO-OFDM
Number of subcarriers, N	1024
Cycle prefix length, N_{cp}	64
Maximum tap delay	62ns
Smallest FD pilot subcarrier index	16
FD pilot interval	32
Route of UE's movement	$(2.5, 2.5) \rightarrow (0, -2.5)$
Distance from floor to UE	1.0m
Maximal statistic window size, ω_{\max}	50
Forgetting factor, λ	0.6
σ_{const}^2 in (40)	σ^2/N_p

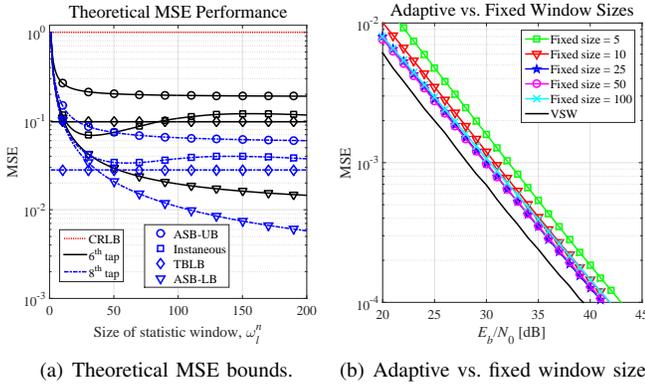


Fig. 3. The theoretical MSE performance and the simulation MSE performances of AS-BMMSE-CE exploiting fixed-size window or VSW.

the figure, we can see that the achievable MSE performance of the AS-BMMSE-CE scheme depends on the tap/position-specific statistic window size ω_l^n , where there exists a different optimal value for each case. Furthermore, we also plot the various MSE performance bounds associated with the two cases, respectively. It can be seen from Fig. 3(a) that both the ASB-UB of (72) and the ASB-LB of (70) are lower than the CRLB [7], [26], implying that AS-BMMSE-CE outperforms MLE of [7] in terms of MSE performance. Moreover, our scheme may also be capable of breaking the TBLB of [7] with the aid of an appropriately selected window size ω_l^n , as observed for instance in the case of the 6th tap in Fig. 3(a).

Next, for demonstrating the impact from the optimum value of $\omega_{l,\text{opt}}^n$ indicated by (49), we investigate the MSE versus E_b/N_0 performance of AS-BMMSE-CE under adaptive or fixed-size statistic windows in Fig. 3(b), where E_b denotes energy per bit and $N_0 = \sigma^2$. Under the adaptive option, whenever the UE moves to a different position in the room, the system calculates the optimal values $\omega_{l,\text{opt}}^n$, $l = 0, \dots, L_c - 1$ based on (26), hence the so-called VSW mechanism. It can be inferred from Fig. 3(b) that the VSW-aided scheme achieves the lowest possible MSE, as compared with its counterparts

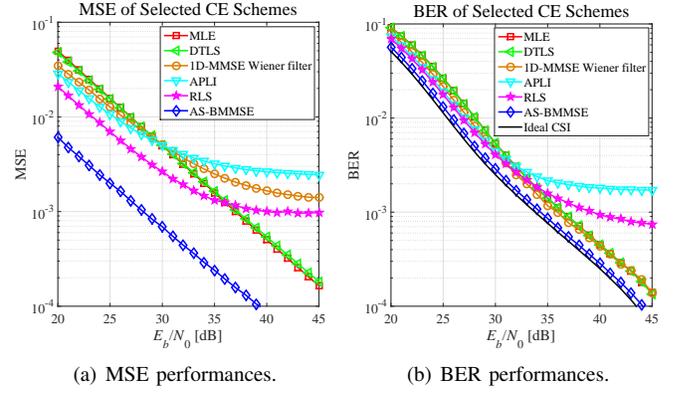


Fig. 4. The MSE and BER versus E_b/N_0 performances of AS-BMMSE-CE and other CE schemes.

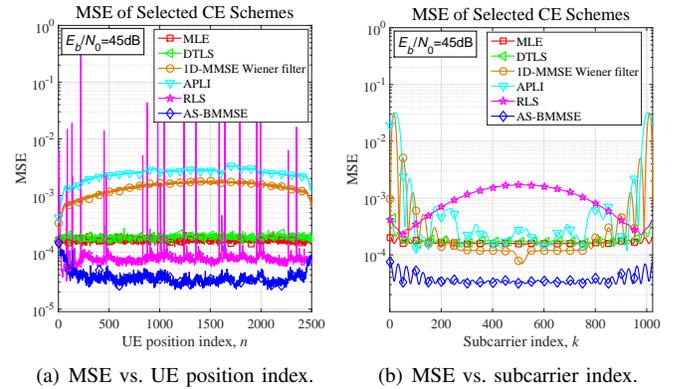


Fig. 5. The MSE versus UE position index and subcarrier index performances of AS-BMMSE-CE and other CE schemes, assuming $E_b/N_0 = 45\text{dB}$.

using a fixed-size statistic window. In the sequel, we assume that the VSW function is always enabled for AS-BMMSE.

In Fig. 4, we compare the MSE and BER performances of AS-BMMSE-CE with selected existing CE schemes, such as MLE [7], one-dimensional (1D) MMSE Wiener filtering [3], APLI [5], domain-transform least squares (DTLS) [8] and RLS [9]. The reference schemes were such configured, that they fitted into the common system platform and the channel model under comparable conditions. From Fig. 4, we can see that our method has the best MSE and BER performances among the schemes investigated. Moreover, it has only 0.5dB loss compared with the benchmark with ideal CSI, as seen in Fig. 4(b).

Fig. 5(a) shows the MSE performances of various CE schemes versus the UE position index, which corresponds to the consecutive positions of the UE when it moves along the route specified in Table I. On the other hand, Fig. 5(b) plots the various schemes' MSE performances versus the subcarrier index. From Fig. 5, we can see that while other CE methods yield worse and/or fluctuant performances at different UE positions or subcarrier indices, the proposed AS-BMMSE-CE scheme offers the best yet stable performance in both the UE position or subcarrier domain. This property is desirable, since it eventually translates to a near-uniform quality of data service

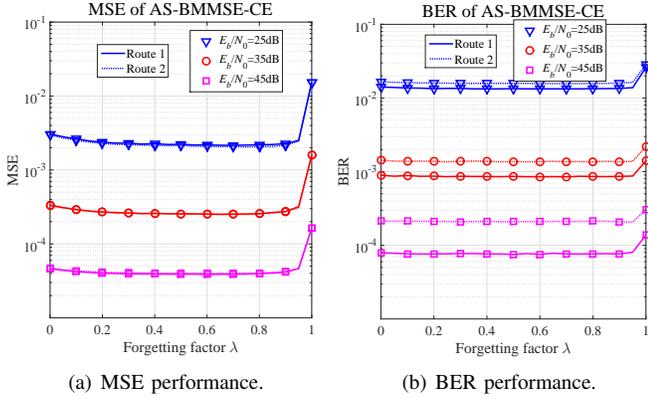


Fig. 6. Impact of the forgetting factor λ on the MSE and BER performances at $E_b/N_0 = \{25, 35, 45\}$ dB.

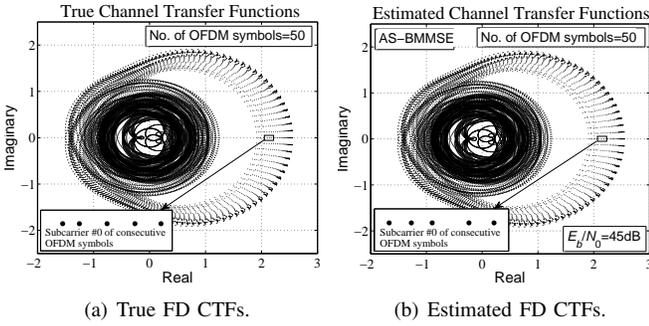


Fig. 7. An example of FD CTF estimation using AS-BMMSE-CE.

across the room.

As a further investigation, in Fig. 6, the impact from the forgetting factor λ mentioned in (39) is investigated. More specifically, the value of λ was tested in the full range of $[0, 1]$ under two example routes of UE movement, namely ROUTE1 : $(-2.5, 2.5) \rightarrow (0, 2.5)$ and ROUTE2 : $(-2.5, 2.5) \rightarrow (0, 0) \rightarrow (2.5, 0)$, respectively. From the MSE and BER performances shown in Fig. 6, we note that the value of λ does not have a significant impact on AS-BMMSE-CE, except when it becomes larger than about 0.95. This helps to simplify the implementation of AS-BMMSE-CE dispensing with the need of adapting λ , whose value may otherwise have to be acquired by complicated methods, such as some adaption to the exponential weighting factor [34].

Last but not least, Fig. 7 exhibits a visualised example for demonstrating the achievable performance of the proposed CE scheme. More explicitly, the true and estimated FD CTFs $\hat{H}[n, k]$ ($n = 71 + 40j$; $j = 0, \dots, 49$; $k = 0, \dots, N - 1$) associated with 50 consecutive UE positions starting from position #71 under a spatial measurement resolution of 40 intervals or approximately 9cm, are extracted from the full set of CE results collected along the route defined in Table I. The CTF samples are plotted on the complex plane at $E_b/N_0 = 45$ dB. As seen from Fig. 7, one set of $N = 1024$ small solid dots, where each dot denotes one CTF sample at its associated subcarrier, forms one round-shaped contour which represents one OFDM symbol. There are totally 50 such OFDM-symbol-

related contours that gradually shift from one to the next on the complex plane, reflecting the adjacent spatial positions that they correspond to. The magnified subfigures in Fig. 7 illustratively capture the CTFs at subcarrier #0 of a few consecutive OFDM symbols. The contours are symmetric with respect to the real axis, since the TD CIR of the VLC channel is real-valued. We can see that the FD CTF estimates closely match the contours of the true channel, which demonstrates that AS-BMMSE-CE is capable of capturing the instantaneously-varying fading envelop regardless of the UE's position. This illustrates the accuracy and robustness of the proposed CE approach, as exemplified in Fig. 7.

As a further remark, in Table II we summarise the computational complexity required by the various CE schemes for the processing during one OFDM symbol, where N_{tap} denotes the filter order of 1D-MMSE [3] and RLS [9] CEs, W is the sliding window size in the APLI-CE [5], and $\{\alpha, \beta\} \in [0, 1]$ are complexity-contributing probabilities associated with the proposed complexity reduction techniques, namely Algorithms 1-3 as well as Theorem 1. According to Table II, taking configurations of $L_c = 32$ and $N_{tap} = 25$ as an example, the worst-case computational complexities of additions required by AS-BMMSE are about 4.67-, 0.26- and 0.0033-fold of MLE [7], RLS [9] and 1D-MMSE Wiener filter [3], respectively. In the best case, these numbers become 2.33, 0.13 and 0.0016, respectively. Therefore, we may conclude that the proposed AS-BMMSE-CE scheme can achieve an excellent performance at the cost of a modest computational complexity.

VI. CONCLUSIONS

In this paper, a so-called AS-BMMSE-CE technique is designed for indoor DCO-OFDM-VLC systems. The proposed scheme is equipped with an efficient mechanism referred to as VSW, which offers an accurate yet robust way for tracking the instantaneous indoor optical channel. Through the VSW function, the achievable channel MSE can be minimised, hence becoming lower than the CRLB and sometimes even lower than the TBLB, thanks to the past channel information collected in the statistic window with an adaptively optimised size. Furthermore, we also devise efficient algorithms that help to reduce the computational complexity of the proposed scheme. Extensive theoretical and simulation results are provided to demonstrate the benefits of the new CE method. Our future work will be to consider the extension of AS-BMMSE to multiple-input multiple-output (MIMO) VLC systems.

APPENDIX I

DERIVATIONS OF (30) AND (35)

A. Derivation of (30)

Firstly, recall that (30) is derived from (25), which contains four items. We now expand these items individually as follows.

Observing that Φ_n in (14) is a diagonal matrix, and utilising (19) as well as the specific pilot pattern designed in Section IV-C, we may expand the first item of (25) to

$$\sigma^2 \text{Tr}\{\Psi_1^n \Psi_1^{nH}\} = \sigma^2 \sum_{l=0}^{L_c-1} \frac{N_p}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2}, \quad (51)$$

TABLE II
COMPUTATIONAL COMPLEXITY OF THE VARIOUS CE SCHEMES INVESTIGATED.

CE scheme	Number of complex additions	Number of complex multiplications
Linear interpolation	$(N_d - 1)N_p$	$(N_d - 1)N_p$
APLI [5]	$(W + \frac{15}{4}N_d - \frac{1}{4})N_p$	$(\frac{5}{2}N_d - \frac{13}{4})N_p$
1DMMSE Wiener filter [3]	$(N_{tap}^3 + N_{tap}^2 + 2N_{tap})N$	$(N_{tap}^3 + N_{tap}^2 + N_{tap})N + N_p$
MLE [7], TDLS [8]	$L_c N_p + N \log N$	$N_p + L_c N_p + \frac{1}{2}N \log N$
RLS [9]	$(5N_{tap}^2 + 3N_{tap} + 2 + L_c)N_p + N \log N$	$(5N_{tap}^2 + 6N_{tap} + 3 + L_c)N_p + \frac{1}{2}N \log N$
AS-BMMSE	$(10 + 17\alpha + \beta)L_c \omega_{\max} + 2\omega_{\max} + 6L_c$ $+ L_c N_p + N \log N$	$(7 + 20\alpha)L_c \omega_{\max} + \omega_{\max} + 6L_c + (2L_c + 1)N_p$ $+ \frac{1}{2}N \log N$

where $\sigma_{n,l}^2$ ($l = 0, \dots, L_c - 1$) are defined in (38). Similarly, the second item of (25) can be reformulated to

$$\text{Tr}\{E\{\Psi_2^n \Delta \mathbf{h} \Delta \mathbf{h}^H \Psi_2^{nH}\}\} = \sum_{l=0}^{L_c-1} \frac{(\frac{\sigma^2}{\sigma_{n,l}^2})^2}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2} f_{\omega_l^n, l}^n, \quad (52)$$

where we define $f_{\omega_l^n, l}^n$ as

$$f_{\omega_l^n, l}^n = E\left\{\left(\frac{\omega_l^n - 1}{\omega_l^n} [\Delta \mathbf{h}_n]_l - \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} [\mathbf{v}_{n-k}]_l - \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} [\Delta \mathbf{h}_{n-k}]_l\right)^2\right\}. \quad (53)$$

Since \mathbf{v}_n and $\Delta \mathbf{h}_n$, as well as \mathbf{V}_n and $\Delta \mathbf{h}_n$ are uncorrelated, we have $\text{Tr}\{E\{\mathbf{v}_n \Delta \mathbf{h}_n^H\}\} = \mathbf{0}_{L_c \times L_c}$ and $\text{Tr}\{E\{\mathbf{V}_n \Delta \mathbf{h}_n^H\}\} = \mathbf{0}_{N_p \times L_c}$, where \mathbf{V}_n , $\Delta \mathbf{h}_n$ and \mathbf{v}_n are defined in (13), (17) and (23), respectively. Furthermore, we also have $\text{Tr}\{E\{\mathbf{v}_i \mathbf{v}_j^H\}\} = \mathbf{0}_{L_c \times L_c}$, for $\forall i, j, i \neq j$. Using these conditions, (53) can be simplified to (31). Similarly, we can expand the third and forth items of (25) to

$$-\text{Tr}\{E\{\Psi_1^n \mathbf{V}_n \Delta \mathbf{h}^H \Psi_2^{nH}\}\} = \sum_{l=0}^{L_c-1} \frac{\sigma^2}{\omega_l^n} \frac{\frac{\sigma^2}{\sigma_{n,l}^2}}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2}, \quad (54)$$

and

$$-\text{Tr}\{E\{\Psi_2^n \Delta \mathbf{h} \mathbf{V}_n^H \Psi_1^{nH}\}\} = \sum_{l=0}^{L_c-1} \frac{\sigma^2}{\omega_l^n} \frac{\frac{\sigma^2}{\sigma_{n,l}^2}}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2}, \quad (55)$$

respectively.

Exploiting (51), (52), (54) and (55), the objective function (30) becomes

$$\Gamma_n = \text{Tr}\{E\{\varepsilon_n \varepsilon_n^H\}\} = \sum_{l=0}^{L_c-1} \left[\frac{N_p \sigma^2}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2} + \frac{(\frac{\sigma^2}{\sigma_{n,l}^2})^2 f_{\omega_l^n, l}^n}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2} + \frac{2\sigma^2}{\omega_l^n} \cdot \frac{\frac{\sigma^2}{\sigma_{n,l}^2}}{(N_p + \frac{\sigma^2}{\sigma_{n,l}^2})^2} \right]. \quad (56)$$

Based on (31) and (56), we note that Γ_n of (30) is decoupled to a function of three parameters, namely \mathbf{C}_h^n , $r_{n,l}^d$ and ω_l^n . Since \mathbf{C}_h^n and $r_{n,l}^d$ can be estimated by (40) and (34), respectively, ω_l^n becomes the only variable that remains to be optimised. Then we can use (56) to obtain (48) for decoupled optimisation of ω_l^n , as suggested by Algorithm 3.

Furthermore, it is worth mentioning that under the wide sense stationary uncorrelated scattering (WSSUS) channel model and exploiting the pilots' semi-orthogonal property

of (27), if $\omega_l^n \rightarrow +\infty$, $l \in \{0, \dots, L_c - 1\}$, Γ_n of (30) reduces to the traditional Bayesian estimation result of [7], [26] as

$$\lim_{\omega_l^n \rightarrow +\infty, l \in \{0, \dots, L_c - 1\}} \Gamma_n = \frac{\sigma^2 L_c}{N_p} \frac{1}{L_c} \sum_{l=0}^{L_c-1} \frac{1}{1 + \sigma^2 / (\sigma_{n,l}^2 N_p)}. \quad (57)$$

B. Derivation of (35)

We define the noise item existing in $\hat{r}_{n,l}^d$, which are the estimated elements of the UE position covariance matrix $\mathbf{R}_{n,l}$ defined in (33), as $E\{\hat{r}_{n,l}^{d, \text{noise}}\} = E\{\hat{r}_{n,l}^d\} - r_{n,l}^d$. Next, utilising (22), (32), (34) as well as $\text{Tr}\{E\{\mathbf{v}_n \Delta \mathbf{h}_n^H\}\} = \mathbf{0}_{L_c \times L_c}$, we may expand $E\{\hat{r}_{n,l}^{d, \text{noise}}\}$ to

$$\begin{aligned} & E\{\hat{r}_{n,l}^{d, \text{noise}}\} \\ &= \frac{1}{\omega_{\max}} E\left\{ \sum_{j=0}^{\omega_{\max} - d - 1} ([\hat{\mathbf{h}}_{\text{ML}}^{n-j}]_l - [\bar{\mu}_n]_l)([\hat{\mathbf{h}}_{\text{ML}}^{n-(j+d)}]_l - [\bar{\mu}_n]_l) \right\} - r_{n,l}^d \\ &= \frac{1}{\omega_{\max}} E\left\{ \sum_{j=0}^{\omega_{\max} - d - 1} ([\mathbf{h}_{n-j} + \mathbf{v}_{n-j}]_l - [\bar{\mathbf{h}}_n + \bar{\mathbf{v}}_n]_l) \right. \\ &\quad \cdot ([\mathbf{h}_{n-(j+d)} + \mathbf{v}_{n-(j+d)}]_l - [\bar{\mathbf{h}}_n + \bar{\mathbf{v}}_n]_l) \left. \right\} - r_{n,l}^d \\ &= \frac{1}{\omega_{\max}} \sum_{j=0}^{\omega_{\max} - d - 1} E\left\{ ([\mathbf{h}_{n-j}]_l - [\bar{\mathbf{h}}_n]_l)([\mathbf{h}_{n-(j+d)}]_l - [\bar{\mathbf{h}}_n]_l) \right\} \\ &\quad + \frac{1}{\omega_{\max}} \sum_{j=0}^{\omega_{\max} - d - 1} \left\{ E\{[\mathbf{v}_{n-j}]_l [\mathbf{v}_{n-(j+d)}]_l\} + E\{[\bar{\mathbf{v}}_n]_l [\bar{\mathbf{v}}_n]_l\} \right. \\ &\quad \left. - E\{[\mathbf{v}_{n-j}]_l [\bar{\mathbf{v}}_n]_l\} - E\{[\bar{\mathbf{v}}_n]_l [\mathbf{v}_{n-(j+d)}]_l\} \right\} - r_{n,l}^d \\ &= \frac{1}{\omega_{\max}} \sum_{j=0}^{\omega_{\max} - d - 1} \left\{ E\{[\mathbf{v}_{n-j}]_l [\mathbf{v}_{n-(j+d)}]_l\} + E\{[\bar{\mathbf{v}}_n]_l [\bar{\mathbf{v}}_n]_l\} \right. \\ &\quad \left. - E\{[\mathbf{v}_{n-j}]_l [\bar{\mathbf{v}}_n]_l\} - E\{[\bar{\mathbf{v}}_n]_l [\mathbf{v}_{n-(j+d)}]_l\} \right\}, \end{aligned} \quad (58)$$

where we define $\bar{\mathbf{v}}_n = \frac{1}{\omega_{\max}} \sum_{k=0}^{\omega_{\max} - 1} \mathbf{v}_{n-k}$ and $\bar{\mathbf{h}}_n = \frac{1}{\omega_{\max}} \sum_{k=0}^{\omega_{\max} - 1} \hat{\mathbf{h}}_{\text{ML}}^{n-k}$. Then, we finally arrive at (35).

APPENDIX II

PROOF OF THEOREM 1

Based on (48), we define the following function by replacing the estimates with their ideal versions as

$$\hat{M}_{\omega_l^n, l}^n = \Omega(\hat{f}_{\omega_l^n, l}^n, \hat{\sigma}_{n,l}^2) \rightarrow M_{\omega_l^n, l}^n = \Omega^n(f_{\omega_l^n, l}^n, \sigma_{n,l}^2). \quad (59)$$

Then, $M_{\omega_l^n, l}^n$ in (59) can be developed as

$$M_{\omega_l^n, l}^n = \frac{\sigma^2 / N_p}{[1 + \sigma^2 / (\sigma_{n,l}^2 N_p)]^2} \left(1 + \frac{\sigma^2}{\sigma_{n,l}^2 N_p} g_{\omega_l^n, l}^n\right), \quad (60)$$

where $g_{\omega_l^n, l}^n$ is given by

$$g_{\omega_l^n, l}^n = \frac{\sigma^2}{N_p \sigma_{n,l}^2 \omega_l^n} + \frac{(\omega_l^n)^2 + 1}{(\omega_l^n)^2} + \frac{\omega_l^n - 1}{(\omega_l^n)^2} \frac{r_{n,l}^0}{\sigma_{n,l}^2} + \sum_{k=1}^{\omega_l^n - 2} \frac{-2k}{(\omega_l^n)^2} \frac{r_{n,l}^k}{\sigma_{n,l}^2} - \frac{2(\omega_l^n - 1)}{(\omega_l^n)^2} \frac{r_{n,l}^{\omega_l^n - 1}}{\sigma_{n,l}^2}. \quad (61)$$

Note that $g_{\omega_l^n, l}^n$ in (61) is a function of the independent variables $\omega_l^n, r_{n,l}^0, \dots, r_{n,l}^{\omega_l^n - 1}$, where $r_{n,l}^d$ ($d = 0, \dots, \omega_l^n - 1$) are defined in (32). Furthermore, observing (60), we can see that $M_{\omega_l^n, l}^n$ is a linear function of $g_{\omega_l^n, l}^n$. Thus, the optimisation problem of (49) can be translated to the problem of (61).

In order to prove Theorem 1, we need to find at least one condition, under which we have $\omega_{l, \text{opt}}^n = \omega_{\max}$ for the l^{th} tap. Based on (61), we have

$$g_{\omega_l^n, l}^n - g_{\omega_l^{n+1}, l}^n = \frac{\sigma^2}{N_p \sigma_{n,l}^2} \cdot \frac{1}{\omega_l^n (\omega_l^n + 1)} + \frac{2\omega_l^n + 1}{(\omega_l^n)^2 (\omega_l^n + 1)^2} - \frac{-(\omega_l^n)^2 + \omega_l^n + 1}{(\omega_l^n)^2 (\omega_l^n + 1)^2} \cdot \frac{r_{n,l}^0}{\sigma_{n,l}^2} + \frac{2\omega_l^n}{(\omega_l^n + 1)^2} \cdot \frac{r_{n,l}^{\omega_l^n}}{\sigma_{n,l}^2} - \sum_{k=1}^{\omega_l^n - 1} 2k \left[\frac{1}{(\omega_l^n)^2} - \frac{1}{(\omega_l^n + 1)^2} \right] \cdot \frac{r_{n,l}^k}{\sigma_{n,l}^2}. \quad (62)$$

According to [35], the definition of correlation coefficients can be represented by $\rho_{XY} = \frac{\text{COV}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \in [-1, 1]$, where $\text{COV}(X, Y)$ is the covariance function, while $D(X)$ and $D(Y)$ denote the variances of X and Y , respectively. Then exploiting (32), the correlation coefficient for the $(n-j)^{\text{th}}$ and $(n-k)^{\text{th}}$ taps can be written as

$$\rho_{n,l}^{|j-k|} = \frac{E\{([\mathbf{h}_{n-j}]_l - [\boldsymbol{\mu}_{\mathbf{h}}^n]_l)([\mathbf{h}_{n-k}]_l - [\boldsymbol{\mu}_{\mathbf{h}}^n]_l)^*\}}{\sqrt{D([\mathbf{h}_{n-j}]_l)}\sqrt{D([\mathbf{h}_{n-k}]_l)}} = \frac{r_{n,l}^{|j-k|}}{\sigma_{n-j,l} \sigma_{n-k,l}}, \quad (63)$$

where $\sigma_{n-j,l} = \sqrt{D([\mathbf{h}_{n-j}]_l)}$ and $\sigma_{n-k,l} = \sqrt{D([\mathbf{h}_{n-k}]_l)}$ are the variances of the $(n-j)^{\text{th}}$ and $(n-k)^{\text{th}}$ positions, respectively. By selecting a not-too-large statistic window size, we have $\sigma_{n-j,l} \approx \sigma_{n-k,l} \approx \sigma_{n,l}$ and hence (63) reduces to

$$\rho_{n,l}^{|j-k|} = \frac{r_{n,l}^{|j-k|}}{\sigma_{n,l}^2} \in [-1, 1], \quad d = 0, \dots, \omega_l - 1. \quad (64)$$

Thus, considering the value range of $\rho_{n,l}^{|j-k|}$ given in (64), we may simplify (62) to

$$g_{\omega_l^n, l}^n - g_{\omega_l^{n+1}, l}^n \geq \frac{-4(\omega_l^n)^2 + (\frac{\sigma^2}{N_p \sigma_{n,l}^2} + 2)(\omega_l^n + 1)}{\omega_l^n (\omega_l^n + 1)^2}, \quad (65)$$

where the equality sign holds, iff $r_{n,l}^d = \alpha_d$ ($d = 0, \dots, \omega_l^n - 1$), where $\alpha_0 = \sigma_{n,l}^2$ indicates the auto-correlation coefficient of the l^{th} tap, while we set $\alpha_d = \sigma_{n,l}^2$ ($d = 1, \dots, \omega_l^n - 2$) and $\alpha_{\omega_l^n - 1} = -\sigma_{n,l}^2$. In this case, we define the numerator of (65) as

$$\Upsilon(\omega_l^n) = -4(\omega_l^n)^2 + (\frac{\sigma^2}{N_p \sigma_{n,l}^2} + 2)(\omega_l^n + 1) = -4(\omega_l^n)^2 + (\frac{\sigma^2}{N_p \sigma_{n,l}^2} + 2)\omega_l^n + (\frac{\sigma^2}{N_p \sigma_{n,l}^2} + 2), \quad (66)$$

which is a quadratic function in one unknown. Noting that $\Upsilon(0) > 0$, we have $\Upsilon(\omega_l^n) \geq 0$ ($\omega_l^n = 1, \dots, \omega_{\max} - 1$), iff $\Upsilon(\omega_{\max} - 1) \geq 0$. Thus, if we let

$$\Upsilon(\omega_{\max} - 1) = -4(\omega_{\max} - 1)^2 + (\frac{\sigma^2}{N_p \sigma_{n,l}^2} + 2)\omega_{\max} \geq 0, \quad (67)$$

which is equivalent to

$$\frac{\sigma^2}{N_p \sigma_{n,l}^2} \geq 4\omega_{\max} - 10 + \frac{4}{\omega_{\max}}, \quad (68)$$

then the condition of $\Upsilon(\omega_l^n) \geq 0$ ($\omega_l^n = 1, \dots, \omega_{\max} - 1$) is satisfied. This translates to the fulfillment of the condition $g_{\omega_l^n, l}^n \geq g_{\omega_l^{n+1}, l}^n$ ($\omega_l^n = 1, \dots, \omega_{\max} - 1$), which implies that $g_{\omega_l^n, l}^n$ is a monotonic decreasing sequence subject to $\forall \omega_l^n \in \{1, \dots, \omega_{\max} - 1\}$. This indicates that under the condition of (68), we will have the optimal statistic window size of $\omega_{l, \text{opt}}^n = \omega_{\max}$. The proof of Theorem 1 completes.

APPENDIX III PROOF OF THEOREM 2

Based on (61) and (64), we have

$$g_{\omega_l^n, l}^n \geq \frac{\sigma^2}{N_p \sigma_{n,l}^2 \omega_l^n} + \frac{2}{\omega_l^n}, \quad (69)$$

where the equality sign holds, iff $r_{n,l}^d = \sigma_{n,l}^2$ ($d = 0, \dots, \omega_l^n - 1$). Then, using (60) and (69), we arrive at the ASB-LB as

$$M_{\omega_l^n, l}^n \geq \frac{\sigma^2/N_p}{[1 + \sigma^2/(\sigma_{n,l}^2 N_p)]^2} \left[1 + (\frac{\sigma^2}{\sigma_{n,l}^2 N_p})^2 \frac{1}{\omega_l^n} + \frac{2}{\omega_l^n} \frac{\sigma^2}{\sigma_{n,l}^2 N_p} \right]. \quad (70)$$

On the other hand, based on (61) and (64), we can obtain

$$g_{\omega_l^n, l}^n \leq \frac{\sigma^2}{N_p \sigma_{n,l}^2 \omega_l^n} + 2, \quad (71)$$

where the equality sign holds, iff $r_{n,l}^d = \alpha_d$ ($d = 0, \dots, \omega_l^n - 1$), where $\alpha_0 = \sigma_{n,l}^2$ and $\alpha_d = -\sigma_{n,l}^2$ ($d = 1, \dots, \omega_l^n - 1$). Using (60) and (71), we have the ASB-UB as

$$M_{\omega_l^n, l}^n \leq \frac{\sigma^2/N_p}{[1 + \sigma^2/(\sigma_{n,l}^2 N_p)]^2} \left[1 + (\frac{\sigma^2}{\sigma_{n,l}^2 N_p})^2 \frac{1}{\omega_l^n} + 2 \frac{\sigma^2}{\sigma_{n,l}^2 N_p} \right]. \quad (72)$$

By inserting $\omega_l^n = +\infty$ and $\omega_l^n = 1$ into (70) and (72), we can get

$$M_{\min} \leq M_{\omega_l^n, l}^n \leq M_{\max}, \quad (73)$$

where $M_{\min} = \frac{\sigma^2}{N_p} \frac{1}{[1 + \sigma^2/(\sigma_{n,l}^2 N_p)]^2}$ and $M_{\max} = \frac{\sigma^2}{N_p}$. Since M_{\max} is the CRLB [26], the upper bound of the proposed AS-BMMSE-CE scheme is guaranteed to be lower than CRLB. Furthermore, we have $M_{\min} \leq M_B = \frac{\sigma^2}{N_p} \frac{1}{[1 + \sigma^2/(\sigma_{n,l}^2 N_p)]}$, proving that the lower bound of AS-BMMSE-CE is lower than M_B , which is the TBLB [7]. The proof of Theorem 2 completes.

IV. ACKNOWLEDGEMENTS

The funding supports from the Science and Technology Program Project (No. 2014B090901063) of Guangdong Province, and the Innovation Team Project (No. 20150401) of SYSU-CMU Shunde International Joint Research Institute, are gratefully acknowledged.

REFERENCES

- [1] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.
- [2] Y. Li, L. J. Cimini and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 902–915, Jul. 1998.
- [3] L. Hanzo, M. Münster, B. J. Choi and T. Keller, *OFDM and MC-CDMA for Broadband Multi-user Communications, WLANs and broadcasting*. Reading, Massachusetts: Wiley, 2003.
- [4] M. Jiang, J. Akhtman, and L. Hanzo, "Iterative Joint Channel Estimation and Multi-User Detection for Multiple-Antenna Aided OFDM Systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2904–2914, Aug. 2007.
- [5] M. Jiang, S. Huang and W. Wen, "Adaptive Polar-Linear Interpolation Aided Channel Estimation for Wireless Communication Systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 920–926, Mar. 2012.
- [6] R. Negi and J. Cioff, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.
- [7] M. Morelli and U. Mengali, "A Comparison of Pilot-Aided Channel Estimation Methods for OFDM Systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.
- [8] M. Yu and P. Sadeghi, "A study of pilot-assisted OFDM channel estimation methods with improvements for DVB-T2," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 5, pp. 2400–2405, Jun. 2012.
- [9] P. S. R. Diniz, *Adaptive filtering: Algorithms and practical Implementation*. Springer, 1997.
- [10] J. Akhtman and L. Hanzo, "Decision directed channel estimation aided OFDM employing sample-spaced and fractionally-spaced CIR estimators," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1171–1175, Apr. 2007.
- [11] J. R. Barry, J. M. Kahn, W. J. Krause, E. A. Lee and D. G. Messerschmitt, "Simulation of multipath impulse response for indoor wireless optical channels," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 3, pp. 367–379, Apr. 1993.
- [12] B. C. Jeffrey and K. Prasanna, "Iterative Site-Based Modeling for Wireless Infrared Channels," *IEEE Transaction on Antennas and Propagation*, vol. 50, no. 5, pp. 759–765, May 2002.
- [13] J. Meng, W. Yin, Y. Li, N. T. Nguyen and Z. Han, "Compressive sensing based high-resolution channel estimation for OFDM system," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 15–25, Feb. 2012.
- [14] T. Foggi, G. Colavolpe, E. Forestieri and G. Prati, "Channel Estimation Algorithms for MLSD in Optical Communication Systems," *IEEE Photonics Technology Letters*, vol. 18, no. 19, pp. 1984–1986, Oct. 2006.
- [15] W. Chung, "Channel Estimation Methods Based on Volterra Kernels for MLSD in Optical Communication Systems," *IEEE Photonics Technology Letters*, vol. 22, no. 4, pp. 224–226, Feb. 2010.
- [16] C. Gong and Z. Xu, "Channel estimation and signal detection for optical wireless scattering communication with inter-symbol interference," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5326–5337, Oct. 2015.
- [17] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
- [18] P. A. Haigh, Z. Ghassemlooy, S. Rajbhandari, I. Papanikolaou and W. Popoola, "Visible Light Communications: 170 Mb/s Using an Artificial Neural Network Equalizer in a Low Bandwidth White Light Configuration," *Journal of Lightwave Technology*, vol. 32, no. 9, pp. 1807–1813, May 2014.
- [19] J. Armstrong and B. J. C. Schmidt, "Comparison of asymmetrically clipped optical OFDM and DC-biased optical OFDM in AWGN," *IEEE Communications Letters*, vol. 12, no. 5, pp. 343–345, May 2008.
- [20] J. Armstrong, "OFDM for optical communications," *Journal of Lightwave Technology*, vol. 27, no. 3, pp. 189–204, Feb. 2009.
- [21] S. D. Dissanayake and Armstrong J, "A novel technique to simultaneously transmit ACO-OFDM and DCO-OFDM in IM/DD systems," *GLOBECOM Workshops*, pp. 782–786, Dec. 2011.
- [22] S. D. Dissanayake and J. Armstrong, "Comparison of ACO-OFDM, DCO-OFDM and ADO-OFDM in IM/DD Systems," *Journal of Lightwave Technology*, vol. 31, no. 7, pp. 1063–1072, Apr. 2013.
- [23] X. Yang, M. Zhang, X. Tang, J. Wu and D. Han, "A post-processing channel estimation method for DCO-OFDM Visible Light Communication," in *2012 8th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, Jul. 2012, pp. 1–4.
- [24] C. L. Bai, S. Zhang, S. C. Bai and Q. Luo, "Development of discrete Fourier transform-based channel estimation algorithms for a coherent optical orthogonal frequency division multiplexing transmission system," *IET Communications*, vol. 8, no. 14, pp. 2528–2534, Sep. 2014.
- [25] T. Zhang, S. Guo, H. Chen, F. Zhong and M. Chunyang, "Enhancing the bit error rate of indoor visible light communication systems using adaptive channel estimation algorithm," *IET Communications*, vol. 9, no. 4, pp. 501–507, Mar. 2015.
- [26] M. K. Steven, *Fundamentals of statistical signal processing - Estimation theory*. Prentice Hall PTR: University of Rhode Island, 1993.
- [27] X. Chen and M. Jiang, "Enhanced Bayesian MMSE Channel Estimation for Visible Light Communication," in *Proceedings of the 27th IEEE International Symposium on Personal, Indoor, Mobile Radio Communications (PIMRC' 16)*, 4-7 Sep. 2016.
- [28] J. C. Lin, "Least-squares channel estimation for mobile OFDM communication on time-varying frequency-selective fading channels," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3538–3550, Nov. 2008.
- [29] R. Singh, T. O'Farrell, and J. P. R. David, "An Enhanced Color Shift Keying Modulation Scheme for High-Speed Wireless Visible Light Communications," *Journal of Lightwave Technology*, vol. 32, no. 14, pp. 153–163, Jul. 2014.
- [30] Z. Ghassemlooy, W. Popoola and S. Rajbhandari, *Optical Wireless Communications - System and Channel Modelling with Matlab*. CRC Press, 2012.
- [31] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [32] Y. Zheng, "A novel channel estimation and tracking method for wireless OFDM systems based on pilots and Kalman filtering," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 2, pp. 275–283, Jan. 2003.
- [33] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.
- [34] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.
- [35] R. E. Walpole, R. Myers, S. L. Myers and E. Y. Keying, *Essentials of Probability & Statistics for Engineers & Scientists*. Pearson, 2012.