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 communications. 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 efficient 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 $(\cdot)$; $[\cdot]$, and $[\cdot]_{i,j}$ indicate the selection of the $i^{th}$ element of a vector and the $(i,j)^{th}$ element of a matrix, respectively; $E\{\cdot\}$ is the expectation operation; $D\{\cdot\}$ is the variance operation; $I_L$ denotes an $L \times L$ identity matrix; diag$(\cdot)$ declares a diagonal matrix; and $(\cdot)$ 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 $I_p = \{P_0 + i \cdot N_d, \; i = 0, 1, \ldots, N_p/2 - 1\}$, 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 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

$$X_n = \left[ X[n,0], X[n,1], \ldots, X[n,N-1] \right]^{T} \in \mathbb{C}^{N \times 1},$$

where the sets of pilot subcarrier indices and data subcarrier indices may be expressed as $P_{\text{pilot}} = \{k|k \in I_p\}$ or $N - k \in I_p\}$ and $P_{\text{data}} = \{0,1, \ldots, N-1\} \backslash P_{\text{pilot}}$ [28], respectively, while $N$ is the size of inverse fast Fourier transform (IFFT) and $C$ denotes the set of complex numbers. Since IM-based optical signals have non-negative real values, $X_n$ is constrained to be Hermitian symmetric as

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

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 $x_n = F^{-1}X_n$, where $F^{-1} = \{f_{n,k}\} \in \mathbb{C}^{N \times N}$, $f_{n,k} = \frac{1}{N}e^{j2\pi nk}$ for $0 \leq \{n,k\} \leq N - 1$. The generated electrical DCO-OFDM signal $s_n$ is then converted to its optical version and transmitted in the VLC channel of a discrete form

$$h_n = [h[n,0], \ldots, h[n,L_c-1]]^{T} \in \mathbb{R}^{L_c \times 1},$$

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 $Y_n$ at the $k^{th}$ subcarrier is

$$Y[n,k] = H[n,k]X[n,k] + N[n,k], \quad k = 0, \ldots, N-1,$$

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 $\sigma^2$. 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

$$\sigma^2 \approx \sigma^2_{\text{Shot}} + \sigma^2_{\text{Thermal}},$$

where $\sigma^2_{\text{Shot}}$ and $\sigma^2_{\text{Thermal}}$ respectively denote the variances of the shot noise and the thermal noise formulated by

$$\left\{ \begin{array}{l} \sigma^2_{\text{Shot}} = 2qR[P_{\text{Signal}}(t) + P_{\text{Daylight}}] \\ \sigma^2_{\text{Thermal}} = \frac{1}{2} \cdot k_bBT \end{array} \right..$$
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, \( P_{\text{Daylight}} \) is the mean power received from the diffuse sunlight in indoor environment, \( k_b \) is the Boltzmann’s constant, \( B \) is the bandwidth and \( T \) is the temperature of the noise equivalent input resistance \( r \). It is worth noting that although the variance \( \sigma_{\text{TD}}^2 \) of the combined TD noise contains a TD shot noise with a time-varying variance \( \sigma_{\text{Shot}}^2 \), 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 \( P_{\text{Rx}} \) 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 \( \hat{\mu}_n^l \) of length \( L_c \), whose \( l^{th} \) \((l \in \{0, \ldots, L_c - 1\}) \) element is the mean tap coefficient averaged within an optimised statistic window size \( \omega_{n, \text{opt}}^l \), 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 \( C_n^l \), 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 \( \omega_{\text{max}} \). 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 implemented in the CE block seen in Fig. 1, while its flowchart is revealed in Section III and Section IV.

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 \( \hat{\mu}_n^l \) of length \( L_c \), whose \( l^{th} \) \((l \in \{0, \ldots, L_c - 1\}) \) element is the mean tap coefficient averaged within an optimised statistic window size \( \omega_{n, \text{opt}}^l \), 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 \( C_n^l \), 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 \( \omega_{\text{max}} \). More details will be revealed in Section III and Section IV.

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

Based on \( \hat{h}_n^{\text{ML}} \), 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 \( l^{th} \) \((l = 0, \ldots, L_c - 1) \) 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 \( \omega_{\text{max}} \) is conducted. According to the comparison result, we can decide whether an exhaustive search over the candidate window sizes \( \omega_n^l \in [1, \omega_{\text{max}}, \ldots] \) has to be launched, for identifying the optimal VSW size \( \omega_{n, \text{opt}}^l \) 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 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 \( h_n \) in (2), utilising the LS-based channel estimates at pilot subcarriers. Assuming \( h_n \) 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

\[
\mathbf{H}_n = \mathbf{B} h_n, \tag{7}
\]

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

\[
\mathbf{N}_n = \mathbf{D} n_n \in \mathbb{C}^{N \times 1}, \tag{8}
\]

where \( n_n \) is the TD real-valued electrical AWGN with zero mean and covariance \( \sigma_n^2 \mathbf{I}_N \), \( n_n \) is complex-valued AWGN with zero mean and covariance \( \sigma_n^2 \mathbf{I}_N \), and \( \mathbf{D} = \mathbf{F}_l^{-1} = \{D_{n,k}\} \in \mathbb{C}^{N \times N} \), \( D_{n,k} = e^{-j \frac{2\pi kn}{N}} \) for \( 0 \leq |n,k| \leq N - 1 \). Define \( \mathbf{H}_n^p \) as the CTF vector corresponding to pilot subcarriers, formulated by

\[
\mathbf{H}_n^p = \mathbf{S} h_n, \tag{9}
\]
where $S$ is an $N_p \times N$ selecting matrix that helps to extract the pilots’ indices. More specifically, the $i^{th}$ $(i = 0, \ldots, N_p - 1)$ row of $S$ is constituted by zeros except the $(i|p_{\text{pilot}}|)$ element, which has a value of 1. It implies that $[S]_{i|p_{\text{pilot}}|} = 1$ and $SS^H = I_{N_p}$. We also define an $N_p \times L_c$ matrix

$$W_p = SB,$$

(10)

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

$$\hat{h}_{n}^\text{ML} = (W_p^H W_p)^{-1} W_p^H \hat{H}_p,$$

(11)

where $\hat{H}_p$ is the LS estimates of $H_p$ in (9), formulated by

$$\hat{H}_p = W_p h_n + \gamma_n^{-1} SN_n = W_p h_n + V_n,$$

(12)

where we define

$$V_n = \gamma_n^{-1} SN_n,$$

(13)

while $\gamma_n = \text{diag}\{p_0, \ldots, p_{N_p-1}\}$ and $p_i$ is the $i^{th}$ $(i = 0, \ldots, N_p - 1)$ pilot symbol. Without loss of generality, we assume that $p_i = \pm 1$, $i = 0, \ldots, N_p - 1$. Note that by using pilot symbols with constant amplitude, each element in $V_n$ is AWGN with zero mean and variance $\sigma^2$, yielding $E\{V_n\} = 0_{N_p \times 1}$ and $E\{V_n V_n^H\} = E\{\gamma_n^{-1} SN_n N_p^H S^H \gamma_n^{-1} H_n^H\} = \sigma^2 I_{N_p}$.

Different from the MLE-based CE that assumes no information of $h_n$, 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 $L_c 	imes 1$ vector $\mu_{n}^H$ and an $L_c \times L_c$ matrix $C_{n}^H$, are known. The BMMSE version of the CIR estimate is [7]

$$\hat{h}_{n} = \mu_{n}^H + \Phi_{n} W_p^H (\hat{H}_p - W_p \mu_{n}^H),$$

(14)

where we define $\Phi_{n} = [W_p^H W_p + \sigma^2 C_{n}^H]^{-1}$. Note that the BMMSE estimates of (14) are more accurate than their MLE counterparts of (11), thanks to the knowledge of $\mu_{n}^H$ and $C_{n}^H$. However, in practical VLC systems the values of $\mu_{n}^H$ and $C_{n}^H$ are typically difficult to obtain or unavailable, thus greatly restricting the applicability of the conventional BMMSE-CE method. Hence, one key issue is 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 $\mu_{n}^H$ can be estimated, together with the derivation of the objective function for our CE problem. By inserting (12) into (14), we have

$$\hat{h}_{n} = \mu_{n}^H + \Phi_{n} W_p^H (W_p h_n + V_n) - \Phi_{n} W_p^H W_p \mu_{n}^H = (I_{L_c} - \Phi_{n} W_p^H W_p) \mu_{n}^H + \Phi_{n} W_p^H W_p h_n + \Phi_{n} W_p^H V_n = h_n + \varepsilon_n,$$

(15)

where $\varepsilon_n$ denotes the estimation error for the TD CIR and is formulated by

$$\varepsilon_n = \Phi_{n} W_p^H V_n - (I_{L_c} - \Phi_{n} W_p^H W_p) \Delta h_n,$$

(16)

where

$$\Delta h_n = h_n - \mu_{n}^H$$

(17)

denotes the difference vector between the CIR $h_n$ and its mean $\mu_{n}^H$ at the UE’s $n^{th}$ position. Furthermore, (16) may be rewritten as

$$\varepsilon_n = \Psi_{1}^n V_n - \Psi_{2}^n \Delta h_n,$$

(18)

where we define

$$\Psi_{1}^n = \Phi_{n} W_p^H, \quad \Psi_{2}^n = I_{L_c} - \Phi_{n} W_p^H W_p,$$

(19)

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

$$\hat{\Delta h}_n = h_n - \hat{\mu}_{n}^H = \hat{h}_n - \mu_{n}^H.$$

(20)

Note that the VLC channel model [11], [12] usually contains one line-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 $\hat{\mu}_{n}^H$, 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 $\hat{\mu}_{n}^H$ is the tap coefficient averaged over the specific statistic window size $\omega_n^H$, $l \in \{0, \ldots, L_c - 1\}$, formulated as

$$[\hat{\mu}_{n}^H]_l = \frac{1}{\omega_n^H} \sum_{k=0}^{\omega_n^H-1} [\hat{h}_{n-k}^\text{ML}]_l, \quad l \in \{0, \ldots, L_c - 1\},$$

(21)

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

$$\hat{h}_{n}^\text{ML} = (W_p^H W_p)^{-1} W_p^H h_n = h_n + v_n = \mu_{n}^H + \Delta h_n + v_n,$$

(22)

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

$$\hat{v}_n = h_n - (W_p^H W_p)^{-1} W_p^H \hat{H}_p = (W_p^H W_p)^{-1} W_p^H V_n.$$

(23)

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

$$[\Delta h_{n}]_l = [\hat{\mu}_{n}^H]_l + [\Delta h_{n}]_l - \frac{1}{\omega_n^H} \sum_{k=0}^{\omega_n^H-1} ([\hat{\mu}_{n}^H]_l + [v_{n-k}]_l + [\Delta h_{n-k}]_l) = \frac{1}{\omega_n^H} [\Delta h_{n}]_l - \frac{1}{\omega_n^H} \sum_{k=0}^{\omega_n^H-1} [v_{n-k}]_l - \frac{1}{\omega_n^H} \sum_{k=0}^{\omega_n^H-1} [\Delta h_{n-k}]_l.$$

(24)

If we define the FD MSE associated with the $k^{th}$ subcarrier at the UE’s $n^{th}$ position as $\gamma_n(k) = E\{|H_n, k| - H[n, k]|^2\}$, 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), $\gamma_n$ can be transformed to

$$\gamma_n = \frac{1}{N} \text{Tr} \{E\{[H_n - \hat{H}_n](H_n - \hat{H}_n)^H\}\} = \frac{1}{N} \text{Tr} \{E\{[B(h_n + \varepsilon_n) - B_h](B(h_n + \varepsilon_n) - B_h)^H\}\} = \frac{1}{N} \text{Tr} \{E\{\varepsilon_n \varepsilon_n^H\}\} = \sigma^2 \text{Tr} \{\Psi_1^n \Psi_1^n\} + \text{Tr} \{E(\Psi_2^n \Delta h_n \Delta h_n^H \Psi_2^H)\}

- \text{Tr} \{E(\Psi_1^n V_n \Delta h_n^H \Psi_2^H)\} - \text{Tr} \{E(\Psi_2^n \Delta h_n V_n^H \Psi_1^H)\},$$

(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 $\omega_{n,l}^i$, $l \in \{0, \ldots, L_c - 1\}$. Hence in AS-BMMSE-CE, we are interested in finding the optimum values $\omega_{n,l}^{i_{\text{opt}}}$, $l \in \{0, \ldots, L_c - 1\}$ that minimise $\Gamma_n$ of (25)

$$\omega_{n,l}^{i_{\text{opt}}} = \text{argmin} \, \Gamma_n, \, l \in \{0, \ldots, L_c - 1\},$$

(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 $\mathbf{W}_P^2$ and $\mathbf{W}_2^0$ 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 $(N/2)^{th}$ subcarrier [22]. Thus, $\mathbf{W}_P$ satisfies the semi-orthogonality of

$$\mathbf{W}_P^H \mathbf{W}_P = N_p \mathbf{I}_{L_c},$$

(27)

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

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

(28)

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

$$P_0 = N_d / 2.$$

(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

$$\Gamma_n = \text{Tr} \{ \mathbf{E} \} = \Theta(f_{\omega,n,l}^0, \mathbf{C}_h^0), \quad l = 0, \ldots, L_c - 1,$$

(30)

where $\Theta(f_{\omega,n,l}^0, \mathbf{C}_h^0)$ is a function of $f_{\omega,n,l}^0$ and $\mathbf{C}_h^0$.

Let us now calculate the values of $f_{\omega,n,l}^0$ and $\mathbf{C}_h^0$. More specifically, $f_{\omega,n,l}^0$ represents the $l$th diagonal element of the diagonal covariance matrix $\mathbf{E} \{ \Delta \mathbf{h}_n \Delta \mathbf{h}_n^H \}$, and can be viewed as a function

$$f_{\omega,n,l}^0(r_{n,l}^d) = \frac{\sigma_d^2}{\omega_n^0} + \frac{1}{\omega_n^0} \sum_{j=1}^{l} \sum_{k=1}^{l-1} \left( (\omega_n^0)^{-2} \omega_n^{j-k} - 2 \omega_n^{l-1} \right) \omega_n^{j-l} r_{n,l}^d$$

$$- \frac{\omega_n^{l-1}}{\omega_n^0} \sum_{j=1}^{l} r_{n,l}^d, \quad l = 0, \ldots, L_c - 1,$$

(31)

where $d = |j - k|$, $\{ j, k \} = 0, \ldots, \omega_n^0 - 1$ and we define

$$r_{n,l}^d = r_{n,l}^{[j-k]} = E \{ ([\mathbf{h}_{n-j}]_t - [\mu_n^j])([\mathbf{h}_{n-k}]_t - [\mu_n^k])^* \},$$

(32)

while $r_{n,l}^d$ 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

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

(33)

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

$$\hat{r}_{n,l}^d = \frac{1}{\omega_{\max} - d} \sum_{j=0}^{\omega_{\max} - d - 1} (\hat{\mathbf{h}}_{ML}^{(n-j)}_\omega - [\mu_n^j]) (\hat{\mathbf{h}}_{ML}^{(n-j)(d)}_\omega - [\mu_n^j])$$

(34)

where $\hat{\mathbf{h}}_{ML}^{(n)}$ is given by (22), and $[\mu_n^j]$ is the mean of the $l$th tap’s coefficients, which is averaged over the maximal statistic window utilising MLE as $\mu_n^j = \frac{1}{\omega_{\max} - k} \sum_{k=0}^{\omega_{\max} - 1} \hat{\mathbf{h}}_{ML}^{(n-k)}$.

Moreover, $\omega_{\max} \geq \omega_n^0$, $l \in \{0, \ldots, L_c - 1\}$ 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 $\mu_n^j$, we can use it to calculate (34) for generating $\mathbf{R}_{n,l}$ defined in (33). Note that the MLE estimate, namely $\hat{\mathbf{h}}_{ML}^{(n)}$, 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

$$E \{ \hat{f}_{n,l,\text{noise}} \} = \left\{ \begin{array}{ll} \omega_{\max} - d \sigma_d^2, & d = 0 \\ \frac{\omega_{\max} - d}{\omega_{\max}} \sigma_d^2, & d = 1, \ldots, \omega_{\max} - 1 \end{array} \right.$$  

(35)

where $\sigma_d^2 = \frac{\sigma_n^2}{\omega_n^0}$ 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 $f_{\omega,n,l}^0(\hat{r}_{n,l}^d) \rightarrow \hat{f}_{\omega,n,l}^0(\hat{r}_{n,l}^d)$. Utilising (35), we can therefore obtain the expectation of the introduced noise item as

$$E \{ \hat{f}_{n,l,\text{noise}} \} = \frac{2(\omega_n^0)^2 - 3 \omega_n^0 (\omega_{\max}^n + 1) + 3 \omega_{\max}^n + 1}{3(\omega_n^0)^2 \omega_{\max}^n} \sigma_d^2,$$

(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

$$\hat{f}_{\omega,n,l}^0 = \hat{f}_{\omega,n,l}^0 - E \{ \hat{f}_{n,l,\text{noise}} \}$$

(37)

to replace $f_{\omega,n,l}^0$ in (30) and (31).

Next, we proceed to calculate $\mathbf{C}_h^0$ 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

$$\mathbf{C}_h^0 = E \{ \Delta \mathbf{h}_n \Delta \mathbf{h}_n^H \} = \text{diag} \{ \sigma_{n,0}^2, \ldots, \sigma_{n,L_c-1}^2 \}.$$  

(38)
where $\sigma^2_{n,l} (l = 0, \ldots, L_c - 1)$ denote the variance of $[\Delta h_n]_l$ in (17) that corresponds to the $l$th tap at the UE's $n$th position. In order to obtain $C^n h$, a forgetting factor $\lambda$ is exploited to calculate the estimate of $\sigma^2_{n,l}$, namely $\hat{\sigma}^2_{n,l}$. More explicitly, we define [32]

$$\hat{\sigma}^2_{n,l} = \lambda \sigma^2_{n-1,l} + (1 - \lambda) (r^2_{n,l} - \frac{\omega^2 - 1}{\omega^2} \sigma^2_0).$$  \hspace{1cm} (39)

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

$$\hat{\sigma}^2_{n,l} = \begin{cases} \hat{\sigma}^2_{n,l}, & \hat{\sigma}^2_{n,l} > 0 \\ \sigma^2_{\text{const}}, & \hat{\sigma}^2_{n,l} \leq 0 \end{cases} \hspace{1cm} (40)$$

which is the estimate of the $l$th diagonal element of $C^n h$.

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 $\omega^n_l, l = 0, \ldots, L_c - 1$. 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 $O(\omega^2 L_c)$ is required for thoroughly searching through $d = 0, \ldots, \omega^2 - 1$ and $l = 0, \ldots, L_c - 1$ in (34). Such a complexity, however, may be reduced by the algorithms proposed in this section.

Let us first expand (34) to

$$\hat{f}^{n,d}_{l,t} = \hat{\varphi}^{n,d}_{l,t} - \varphi^{n,d}_{l,t} (\hat{\mu}_n)^2, \hspace{1cm} (41)$$

where

$$\left\{ \begin{array}{l} \hat{\varphi}^{n,d}_{l,1} = \sum_{j=0}^{\omega - 1} \varphi^{n,j}_{l,ML} |h_{ML}^{n,j}|, \\
\hat{\varphi}^{n,d}_{l,2} = \sum_{j=0}^{\omega - 1} \varphi^{n,j}_{l,ML} |h_{ML}^{n,j}|, \\
\hat{\varphi}^{n,d}_{l,3} = \sum_{j=0}^{\omega - 1} \varphi^{n,j}_{l,ML} |h_{ML}^{n,j}|, \\
\end{array} \right. \hspace{1cm} (42)$$

which may be further reformulated as

$$\left\{ \begin{array}{l} \hat{\varphi}^{n,d}_{l,1} = \varphi^{n-1,d}_{l,1} + |h_{ML}^{n-1,d}| \left( |h_{ML}^{n-1,d}| - |h_{ML}^{n-\omega+1,d}| \right), \\
\hat{\varphi}^{n,d}_{l,2} = \varphi^{n-1,d}_{l,2} + |h_{ML}^{n-1,d}| \left( |h_{ML}^{n-1,d}| - |h_{ML}^{n-\omega+1,d}| \right), \\
\hat{\varphi}^{n,d}_{l,3} = \varphi^{n-1,d}_{l,3} + |h_{ML}^{n-1,d}| \left( |h_{ML}^{n-1,d}| - |h_{ML}^{n-\omega+1,d}| \right). \\
\end{array} \right. \hspace{1cm} (43)$$

Thanks to the recursive form of (43), the computational complexity of (34) can be reduced to $O(\omega^2 L_c)$. We summarised the proposed covariance coefficient update algorithm (CCUA) in Algorithm 1.

On the other hand, a computational complexity of $O(\omega^2 L_c)$ is imposed by (37) for fully testing $\omega^n_l = 1, \ldots, \omega^2$ for the $l$th tap. We may expand (37) to

$$\hat{f}^{n,1}_{l,t} = \frac{\sigma^2}{N_p^2 \sigma^2_{n,l}} + \frac{1}{\omega^2} \hat{f}^{n,1}_{l,t} + \varrho^{n-1}_{l,t} - \varrho^{n-1}_{l,t} + \frac{\omega^2 - 1}{\omega^2} \varrho^{n,2}_{l,t} + 2(\varrho^{n,2}_{l,t} - 3\varrho^{n,2}_{l,t} \omega^2 + 1 + 3\omega^2 + 1) \sigma^2_0, \hspace{1cm} (44)$$

Algorithm 1 Covariance Coefficient Update Algorithm (CCUA)

1: **Initialisation:** Obtain $\hat{\varphi}^{n-1,d}_{l,1}, \hat{\varphi}^{n-1,d}_{l,2}, \hat{\varphi}^{n-1,d}_{l,3}, \mu_{n-1}$ and set $l = 1$.
2: **repeat**
3: \hspace{1cm} $[\hat{\mu}_n]_l = [\mu_{n-1}]_l + \left( \frac{|h_{ML}^{n-\omega+1,d}|}{\omega^2} \right)$
4: \hspace{1cm} $d = 0$
5: **repeat**
6: \hspace{2cm} Calculate (43) and (41)
7: \hspace{2cm} $d = d + 1$
8: **until** $d > \omega^2$
9: \hspace{1cm} $l = l + 1$
10: **until** $l > L_c$
11: **Return:** $\hat{f}^{n,d}_{l,t}, l = 0, \ldots, L_c - 1, d = 0, \ldots, \omega^2 - 1$.

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

Algorithm 2 Covariance Matrix Update Algorithm (CMUA)

1: **Initialisation:** Obtain $\hat{f}^{n,d}_{l,t}, l = 0, \ldots, L_c - 1, d = 0, \ldots, \omega^2 - 1 - 1$. Set $\hat{f}^{n,1}_{l,t} = 0, \hat{f}^{n,2}_{l,t} = 0$ with given $l$ and $\omega^n_l = 1$.
2: **repeat**
3: \hspace{1cm} Calculate (44) and (46)
4: \hspace{1cm} $\omega^n_l = \omega^n_l + 1$
5: **until** $\omega^n_l > \omega^2$
6: **Return:** $\hat{f}^{n,d}_{l,t}, l = 1, \ldots, \omega^2$.

E. Optimum VSW Size and MSE Bound

Recall that the optimum solution for the objective function $\Gamma_n$ defined in (25) or (30) is given by (26), which is an integer programming problem since the variables $\omega^n_{\text{opt}}$ to be optimised are integers, and thus a traditional NP-complete problem [33]. Since there are a total number of $\omega^2$ candidate window sizes for each of the $L_c$ taps, the optimisation of (26) results in a high computational complexity of $O(\omega^2 L_c)$. Nonetheless, note that (30), which is further developed in (56) of Appendix I, may be reformulated as

$$\Gamma_n = \sum_{l=0}^{L_c-1} \left( \frac{N_p \sigma^2}{(N_p + \frac{\sigma^2}{\sigma_{n,l}})^2} + \frac{\frac{\sigma^2}{\sigma_{n,l}} f^{n,d}_{l,t} + \sigma^2}{\frac{\sigma^2}{\sigma_{n,l}} N_p + \frac{\sigma^2}{\sigma_{n,l}} \sigma_{n,l}} \right)$$

where we define

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

\[
\omega_{n,\text{opt}} = \arg\min_{\omega_n \in \{1, \ldots, \omega_{\text{max}}\}} \hat{M}_{n,l}.
\]  

(49)

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

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

The proof of Theorem 1 is given in Appendix II, where we show two example cases associated with two randomly selected UE positions, namely the 6th and 8th taps at the positions of \((−1.6, −0.7)\) and \((−1.0, 0.5)\), respectively. From

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 \( M_{\omega_{n,l}} \) is lower than the CRLB [26], and the lower bound of \( M_{\omega_{n,l}} \) is lower than the traditional Bayesian lower bound (TBLB) [7], [26], where \( M_{\omega_{n,l}} \) is the ideal version of \( M_{\omega_{n,l}} \) 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 \( \lambda \), \( \omega_{\text{max}} \) and \( L_c \). Set \( C_n^b = \text{diag}(\frac{1}{\tau_1}, \ldots, \frac{1}{\tau_{2M}}) \), \( h_{\text{ML}} = 0, k = 0, -1, \ldots, -\omega_{\text{max}} + 2 \) and \( n = 1 \).
2: **repeat**
3: \( \tilde{H}[n,k] = \hat{Y}[n,k] \times [\sum_{l=0}^{L_c-1} |h_{\text{ML},l}|]^{-1} \)
4: Calculate (11)
5: **Execute Algorithm** 1
6: **until** \( l = L_c - 1 \)
7: **repeat**
8: Calculate (39) and (40)
9: \( l = l + 1 \)
10: **until** \( l > L_c - 1 \)
11: **Execute Algorithm** 3
12: **until** \( l = L_c - 1 \)
13: **repeat**
14: \( \mu_{n,l} = \frac{1}{\omega_{l,\text{opt}} \sum_{l=0}^{L_c-1} |h_{\text{ML},l}|]^{-1} \)
15: \( l = l + 1 \)
16: **until** \( l > L_c - 1 \)
17: Calculate (14)
18: **Apply** \( N \)-point FFT to get \( \hat{H}_n \)
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 \( 85^\circ \) 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, \( \sigma^2 \) is normalised to \( N_p \). Without loss of generality, we show two example cases associated with two randomly selected UE positions, namely the 6th and 8th taps at the positions of \((−1.6, −0.7)\) and \((−1.0, 0.5)\), respectively. From
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 $\omega^p_{\text{opt}}$, 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 $\omega^p_{\text{opt}}$, 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 $\omega^p_{\text{opt}}$ 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 $\omega^p_{\text{opt}}$, $l = 0, \ldots, L_e - 1$ 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 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 filter [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.
across the room.

As a further investigation, in Fig. 6, the impact from the forgetting factor $\lambda$ mentioned in (39) is investigated. More specifically, the value of $\lambda$ 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 $\lambda$ 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 $\lambda$, 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 $\hat{H}[n,k]$ ($n = 71 + 40j$; $j = 0, \ldots, 49$; $k = 0, \ldots, N - 1$) 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 $N = 1024$ 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 illustrate 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] CE, $W$ 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 Algorithm I-3 as well as Theorem I. 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$\cdot$10$^6$, 0.26$\cdot$10$^6$ and 0.0033$\cdot$10$^6$-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 $\Phi_n$ 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

$$
\sigma^2 \text{Tr}\{\Psi_1 \Psi_1^H\} = \sigma^2 \sum_{l=0}^{L_c-1} \frac{N_p}{(N_p + \sigma^2_{n,l})^2}. 
$$

(51)
where $\sigma^2_{n,l} (l = 0, \ldots, L_c - 1)$ are defined in (38). Similarly, the second item of (25) can be reformulated to

$$\text{Tr}\{\Psi_2^n \Delta h \Delta h^H \Psi_2^m \} = \sum_{l=0}^{L_c-1} \left( \frac{\sigma^2 v_n}{\sigma^2_{n,l}} \right)^2 \left( \frac{\omega^0 l}{N_p + \frac{\sigma^2}{\sigma^2_{n,l}}} \right)^2 f_{\omega^0 l, t}^m,$$

(52)

where we define $f_{\omega^0 l, t}^m$ as

$$f_{\omega^0 l, t}^m = E\left( (\omega^0 l - 1) | \Delta h_n | - \frac{\omega^0 l}{\omega^0 l} \sum_{k=0}^{\omega^0 l - 1} |v_{n-k}| - \frac{\omega^0 l}{\omega^0 l} \sum_{k=0}^{\omega^0 l - 1} |\Delta h_{n-k}| \right)^2. $$

(53)

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

$$-\text{Tr}\{\Psi_2^n v_n \Delta h \Delta h^H \Psi_2^m \} = \sum_{l=0}^{L_c-1} \frac{\omega^0 l}{N_p + \frac{\sigma^2}{\sigma^2_{n,l}}} \frac{\sigma^2_{n,l}}{\omega^0 l} \left( \frac{\sigma^2}{\sigma^2_{n,l}} \right)^2,$$

(54)

and

$$-\text{Tr}\{\Psi_2^n \Delta h v_n^H \Psi_2^m \} = \sum_{l=0}^{L_c-1} \frac{\omega^0 l}{N_p + \frac{\sigma^2}{\sigma^2_{n,l}}} \frac{\sigma^2_{n,l}}{\omega^0 l} \left( \frac{\sigma^2}{\sigma^2_{n,l}} \right)^2,$$

(55)

respectively.

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

$$\Gamma_n = \text{Tr}\{\epsilon_n e_n^H\} = \sum_{l=0}^{L_c-1} \left( \frac{\sigma^2}{\sigma^2_{n,l}} \right)^2 f_{\omega^0 l, t}^m + \frac{\sigma^2}{\omega^0 l} \left( \frac{\sigma^2}{\omega^0 l} \right)^2 + \frac{2\sigma^2}{\omega^0 l} \left( \frac{\sigma^2}{\omega^0 l} \right)^2,$$

(56)

Based on (31) and (56), we note that $\Gamma_n$ of (30) is decoupled to a function of three parameters, namely $C^0_n$, $r_d l$ and $\omega^0 l$. Since $C^0_n$ and $r_d l$ can be estimated by (40) and (34), respectively, $\omega^0 l$ becomes the only variable that remains to be optimized. Then we can use (56) to obtain (48) for decoupled optimization of $\omega^0 l$, 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 $\omega^0 l \rightarrow +\infty$, $l \in \{0, \ldots, L_c - 1\}$, $\Gamma_n$ of (30) reduces to the traditional Bayesian estimation result of [7], [26] as

$$\lim_{\omega^0 l \rightarrow +\infty, l \in \{0, \ldots, L_c - 1\}} \Gamma_n = \frac{\sigma^2_{n,l}}{N_p} \sum_{l=0}^{L_c-1} \frac{1}{1 + \sigma^2/(\sigma^2_{n,l} N_p)}. $$

(57)

B. Derivation of (35)

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

$$\text{E} \{ r^d_{n,l,\text{noise}} \} = \frac{1}{\omega^0 l} \sum_{j=0}^{\omega^0 l - 1} \left( [h^0_{n-j,l} - [\mu_{n,l}]][h^0_{n-(j+d),l} - [\mu_{n,l}])] \right)^2 - r^d_{n,l},$$

where $[\mu_{n,j}] = [\mu_{n,j}] + [\bar{y}_{n-j}] - [\bar{h}_{n} + \bar{v}_{n}]$. Using these conditions, (53) can be simplified to (31). Similarly, we can expand the third and forth items of (25) to

$$-\text{Tr}\{\Psi_2^m v_n \Delta h \Delta h^H \Psi_2^m \} = \sum_{l=0}^{L_c-1} \frac{\omega^0 l}{N_p + \frac{\sigma^2}{\sigma^2_{n,l}}} \frac{\sigma^2}{\omega^0 l} \left( \frac{\sigma^2}{\omega^0 l} \right)^2,$$

(54)

and

$$-\text{Tr}\{\Psi_2^m \Delta h v_n^H \Psi_2^m \} = \sum_{l=0}^{L_c-1} \frac{\omega^0 l}{N_p + \frac{\sigma^2}{\sigma^2_{n,l}}} \frac{\sigma^2}{\omega^0 l} \left( \frac{\sigma^2}{\omega^0 l} \right)^2,$$

(55)

respectively.

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

$$\Gamma_n = \text{Tr}\{\epsilon_n e_n^H\} = \sum_{l=0}^{L_c-1} \left( \frac{\sigma^2}{\sigma^2_{n,l}} \right)^2 f_{\omega^0 l, t}^m + \frac{\sigma^2}{\omega^0 l} \left( \frac{\sigma^2}{\omega^0 l} \right)^2 + \frac{2\sigma^2}{\omega^0 l} \left( \frac{\sigma^2}{\omega^0 l} \right)^2,$$

(56)

where we define $\bar{v}_n = \frac{1}{\omega^0 l} \sum_{k=0}^{\omega^0 l - 1} v_{n-k}$ and $\bar{h}_n = \frac{1}{\omega^0 l} \sum_{k=0}^{\omega^0 l - 1} h^0_{n-k}$. Then, we finally arrive at (35).
where \( g_{\omega^n, l}^n \) is given by
\[
g_{\omega^n, l}^n = \frac{\sigma^2}{N_p \sigma_l^2} \cdot \frac{1}{\sigma_l} \cdot \left( \frac{(\omega_l^n)^2 + \omega_l^n - 1}{(\omega_l^n)^2} \right)^2
+ \sum_{k=1}^{\omega_l^n - 2} -2k \cdot r_{n,l}^k \cdot \frac{2(\omega_l^n - 1) \cdot r_{n,l}^{2k-1}}{(\omega_l^n)^2 \sigma_l^2}.
\]  
(61)

Note that \( g_{\omega^n, l}^n \) in (61) is a function of the independent variables \( \omega_l^n, r_{n,l}^0, \ldots, r_{n,l}^{\omega_l^n-1} \), where \( r_{n,l}^d \) \((d = 0, \ldots, \omega_l^n - 1)\) are defined in (32). Furthermore, observing (60), we can see that \( M_{\omega^n, l}^n \) is a linear function of \( g_{\omega^n, l}^n \). 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 \( \omega_{n, \text{opt}}^n = \omega_{\text{max}} \) for the \( l \)th tap. Based on (61), we have
\[
g_{\omega^n, l}^n - g_{\omega^n, l+1}^n = \frac{\sigma^2}{N_p \sigma_l^2} \cdot \frac{1}{\sigma_l} \cdot \left( \frac{1}{(\omega_l^n)^2} + \frac{2(\omega_l^n)^2 + 2(\omega_l^n - 1)}{(\omega_l^n)^2 (\omega_l^n + 1)^2} \right) - \frac{\omega_l^n - 2}{\sigma_l} \cdot \frac{2k^2}{(\omega_l^n + 1)^2} \cdot \frac{r_{n,l}^{2k-1}}{\sigma_l^2}.
\]  
(62)

According to [35], the definition of correlation coefficients can be represented by 
\( \rho_{XY} = \frac{\text{COV}(X,Y)}{\sqrt{\text{COV}(X,X) \cdot \text{COV}(Y,Y)}} \in [-1, 1] \), 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 \((n-j)^{th}\) and \((n-k)^{th}\) taps can be written as
\[
\rho_{n,l}^{j-k} = \frac{\sigma_{n-j,l} - (\mu_{n-j,l}^2)}{\sigma_{n-k,l} - (\mu_{n-k,l}^2)} = \frac{r_{n,l}^{j-k}}{\sigma_{n,l}^2} \in [-1, 1], \quad d = 0, \ldots, \omega_l - 1.
\]  
(63)

Thus, considering the value range of \( \rho_{n,l}^{j-k} \) given in (64), we may simplify (62) to
\[
g_{\omega^n, l}^n - g_{\omega^n, l+1}^n \geq \frac{\omega_l^n - 2}{(\omega_l^n + 1)^2} \cdot \frac{(\omega_l^n)^2 + 2(\omega_l^n - 1)}{N_p \sigma_l^2},
\]  
(65)

where the equality sign holds, if \( r_{n,l}^d = \alpha_d \) \((d = 0, \ldots, \omega_l - 1)\), where \( \alpha_d = \sigma_{n,l}^2 \) indicates the auto-correlation coefficient of the \( d \)th tap, while we set \( \alpha_d = \sigma_{n,l}^2 \) \((d = 1, \ldots, \omega_l - 2)\) and \( \alpha_{\omega_l - 1} = -\sigma_{n,l}^2 \). In this case, we define the numerator of (65) as
\[
\Upsilon(\omega_l^n) = 4(\omega_l^n)^2 + \frac{\sigma^2}{N_p \sigma_l^2} (\omega_l^n - 1)^2 + \frac{(\sigma_l^2 + 2)(\omega_l^n + 1)}{(\omega_l^n)^2 - 1}.
\]  
(66)

which is a quadratic function in one unknown. Noting that \( \Upsilon(0) > 0 \), we have \( \Upsilon(\omega_l^n) \geq 0 \) \((\omega_l^n = 1, \ldots, \omega_{\text{max}} - 1)\), iff \( \Upsilon(\omega_{\text{max}} - 1) \geq 0 \). Thus, if we let
\[
\Upsilon(\omega_{\text{max}} - 1) = -4(\omega_{\text{max}} - 1)^2 + \frac{\sigma^2}{N_p \sigma_l^2} (\omega_{\text{max}} - 1)^2 \geq 0,
\]  
(67)

which is equivalent to
\[
\frac{\sigma^2}{N_p \sigma_l^2} \geq 4 \omega_{\text{max}} - 10 + \frac{4}{\omega_{\text{max}}},
\]  
(68)

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

APPENDIX III
PROOF OF THEOREM 2

Based on (61) and (64), we have
\[
g_{\omega_l^n, l}^n \geq \frac{\sigma^2}{N_p \sigma_l^2} \cdot \frac{(\omega_l^n)^2 + 2(\omega_l^n - 1)}{\omega_l^n} + 2, \quad \text{when the equality sign holds, if } r_{n,l}^d = \alpha_d \quad (d = 0, \ldots, \omega_l - 1). \quad \text{Then, using (60) and (69), we arrive at the ASB-LB as}
\]  
\[
M_{\omega_l^n, l}^n \geq \frac{\sigma^2/N_p}{[1 + (\omega_l^n)^2/(\sigma_l^2)]^2} \left[ 1 + \frac{\sigma^2}{\sigma_l^2} \right] \frac{1}{\omega_l^n} + \frac{2}{\sigma_l^2}.
\]  
(70)

On the other hand, based on (61) and (64), we can obtain
\[
g_{\omega_l^n, l}^n \leq \frac{\sigma^2}{N_p \sigma_l^2} \cdot \frac{\omega_l^n}{\omega_l^n + 1} + 2, \quad \text{where the equality sign holds, if } r_{n,l}^d = \alpha_d \quad (d = 0, \ldots, \omega_l - 1), \quad \text{where } \alpha_0 = \sigma_{n,l}^2 \text{ and } \alpha_d = -\sigma_{n,l}^2 \quad (d = 1, \ldots, \omega_l - 1). \quad \text{Using (60) and (71), we have the ASB-UB as}
\]  
\[
M_{\omega_l^n, l}^n \leq \frac{\sigma^2/N_p}{[1 + (\omega_l^n)^2/(\sigma_l^2)]^2} \left[ 1 + \frac{\sigma^2}{\sigma_l^2} \right] \frac{1}{\omega_l^n} + \frac{2}{\sigma_l^2}.
\]  
(72)

By inserting \( \omega_l^n = +\infty \) and \( \omega_l^n = 1 \) into (70) and (72), we can get
\[
M_{\text{min}} \leq M_{\omega_l^n, l}^n \leq M_{\text{max}},
\]  
(73)

where \( M_{\text{min}} = \frac{\sigma^2}{N_p \sigma_l^2} \left[ 1 + \frac{\sigma^2}{\sigma_l^2} \right] \) and \( M_{\text{max}} = \frac{\sigma^2}{\sigma_l^2} \). Since \( M_{\text{max}} \) is the CRLB [26], the upper bound of the proposed AS-BMMSE-CE scheme is guaranteed to be lower than CRLB. Furthermore, we have \( M_{\text{min}} \leq M_B = \frac{\sigma^2}{N_p \sigma_l^2} \left[ 1 + \frac{\sigma^2}{\sigma_l^2} \right] \), 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


