Enhanced Bayesian MMSE Channel Estimation for Visible Light Communication

Xianyu Chen* and Ming Jiang*#
*Sun Yat-sen University, Guangzhou, China
#SYSU-CMU Shunde International Joint Research Institute (JRI), Shunde, China
Email: jiangm7@mail.sysu.edu.cn

***Abstract*— Visible light communication (VLC) is considered as a potential candidate of five generation (5G) technologies and thus attracts increasing research interest recently. 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 to conventional radio frequency (RF) based wireless communication, there has been limited considerations on channel estimation (CE) for VLC, where the indoor optical wireless channel model differs from the traditional RF case. In this paper, we present a new channel estimation algorithm for indoor downlink (DL) VLC systems that exploits historical information in an adaptive and efficient way. The proposed scheme is capable of offering superior performance at the cost of modest complexity under practical scenarios. Detailed theoretical analysis is provided along with extensive numerical results, demonstrating the effectiveness of the proposed CE approach.**

***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. FD-CEs [2]–[4] employ methods such as minimum mean square error (MMSE) [2], [3], adaptive polar linear interpolation (APLI) [4], etc., which either assume idealistic assumptions or suffer notable residual error floors. On the other hand, TD-CEs [5]–[7] utilise channel impulse response (CIR) for estimating channel state information (CSI) by invoking MMSE, recursive least squares (RLS) [8] or other algorithms [6], [9]. 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.

Since VLC systems typically operate in indoor optical wireless channels [10] that differ from traditional wireless radio channels, a direct migration of RF CE techniques into VLC may not be optimal. 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. Furthermore, another significant difference between the two types of channels is the time-varying characteristics. More specifically, when the user equipment (UE) moves under the VLC channel, the variation of the channel taps' envelopes and the path delay no longer obey the traditional Doppler spectrum [10]. Moreover, compared with the sparse taps of many popular RF channel models, the taps of VLC channels are denser and thus impose specific design requirements from the CE perspective.

Against this background, in this paper we propose a new CE method, termed as adaptive statistical Bayesian minimum mean square error channel estimation (AS-BMMSE-CE), for orthogonal frequency division multiplexing (OFDM) aided VLC systems. The proposed new CE scheme achieves a superior performance in terms of both mean square error (MSE) and bit error rate (BER), with the aid of a new algorithm that optimises the so-called variable statistic window (VSW). By effectively tracking real-time channel variations, the AS-BMMSE-CE scheme does not have to first predict the CSI and then to modify parameters for the next prediction period, as proposed by some decision-directed CE (DDCE) techniques [9], thus helping to minimise error propagations. Furthermore, it has a fast convergence, high accuracy and high robustness, as verified by extensive simulation results and comparisons with various existing CE techniques.

The organisation of this paper is as follows. The overall system model is introduced in Section II, followed by the detailed design of the AS-BMMSE-CE scheme in Section III. Simulation results are offered and analysed in Section IV, before we finally conclude our findings in Section V.

Notations: Bold variables denote matrices or vectors; $\mathrm{Tr}\{\cdot\}$ 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 $(i,j)^{th}$ 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; $\mathrm{diag}\{\cdot\}$ 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), as

shown in Fig. 1. However, it is also worth pointing out that other popular optical OFDM (O-OFDM) schemes are also applicable with minimum modifications.

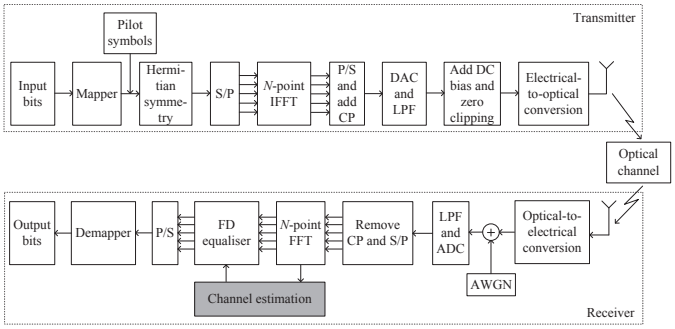


Fig. 1. Schematic of a typical DCO-OFDM system.

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, the pilot symbols are multiplexed with data symbols at an equal-distance of N_d to produce a FD signal vector $\mathbf{X}_n = \big[X[n,0], \ldots, X[n,N-1]\big]^T \in \mathbb{C}^{N\times 1}$, 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 $P_{data} = \{0, \ldots, N-1\} \backslash P_{\text{pilot}}$, 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

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

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 $\mathbf{F}_I = \{f_{n,k}\} \in \mathbb{C}^{N\times N}$, $f_{n,k} = \frac{1}{N} e^{j\frac{2\pi nk}{N}}$ for $0 \le \{n,k\} \le N-1$. The generated electrical DCO-OFDM signal $\mathbf{s}_n$ is then converted to its optical version and transmitted through the VLC channel with a discrete form

$$\mathbf{h}_n = \big[h[n,0], \ldots, h[n,L_c-1]\big]^T \in \mathbb{R}_+^{L_c\times 1}, \tag{2}$$

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 can be generated as

$$Y[n,k] = H[n,k]X[n,k] + N[n,k],\ k = 0, \ldots, N-1, \tag{3}$$

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. With the aid of the CE block in Fig. 1, whose details will be revealed in Section III, the estimated channel coefficients $\hat{H}[n,k]$ can be obtained.

III. The Design of VSW-aided AS-BMMSE-CE

The proposed VSW-aided AS-BMMSE-CE scheme is implemented in the CE block seen in Fig. 1. We assume that a comb-type pilot pattern with subcarrier indices defined by P_{pilot} is used, where the specific pilot arrangement is designed in Section III-C.

A. TD PACE Process

As the first step, the least squares (LS) based CE is employed to obtain the channel estimates at the N_p pilot subcarriers, yielding

$$\hat{H}[n,k] = \frac{Y[n,k]}{X[n,k]} = H[n,k] + \frac{X^*[n,k]}{|X[n,k]|^2} N[n,k], \tag{4}$$

where $k \in P_{\text{pilot}}$. The TD CE function based on maximum likelihood estimation (MLE) [6] is then invoked for generating the initial estimates of $\mathbf{h}_n$ in (2) with the aid of (4). Assuming $\mathbf{h}_n$ is deterministic but unknown, the MLE-based CE is capable of approaching the Cramér-Rao lower bound (CRLB) [6], [11]. To elaborate a little further, first note that the FD CTF vector $\mathbf{H}_n$ can be calculated through

$$\mathbf{H}_n = \mathbf{B}\mathbf{h}_n, \tag{5}$$

where $\mathbf{B} = \{B_{k,l}\} \in \mathbb{C}^{N\times L_c}$, $B_{k,l} = e^{-j\frac{2\pi kl}{N}}$ for $0 \le k \le N-1$, $0 \le l \le L_c - 1$. Denote the FD noise after FFT as $\mathbf{N}_n = \mathbf{D}\mathbf{n}_n \in \mathbb{C}^{N\times 1}$, which 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 $0 \le \{n,k\} \le N-1$. Define $\mathbf{H}_P^n$ as the CTF vector corresponding to pilot subcarriers, formulated by

$$\mathbf{H}_P^n = \mathbf{S}\mathbf{H}_n, \tag{6}$$

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

$$\mathbf{W}_P = \mathbf{S}\mathbf{B}, \tag{7}$$

where $[\mathbf{W}_P]_{k,l} = e^{-\frac{j2\pi\cdot[P_{\text{pilot}}]_k \cdot l}{N}}$ $(0 \le k \le N_p - 1,\ 0 \le l \le L_c - 1)$. According to [6], [11], the MLE estimate of the CIR can be written as

$$\hat{\mathbf{h}}_{\text{ML}}^n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \hat{\mathbf{H}}_P^n, \tag{8}$$

where $\hat{\mathbf{H}}_P^n$ is the LS estimates of $\mathbf{H}_P^n$ in (6), formulated by

$$\hat{\mathbf{H}}_P^n = \mathbf{W}_P \mathbf{h}_n + \boldsymbol{\varrho}_n{}^{-1}\mathbf{S}\mathbf{N}_n = \mathbf{W}_P\mathbf{h}_n + \mathbf{V}_n, \tag{9}$$

where $\mathbf{V}_n = \boldsymbol{\varrho}_n{}^{-1}\mathbf{S}\mathbf{N}_n$, $\boldsymbol{\varrho}_n = \text{diag}\{p_0, \ldots, p_{N_p-1}\}$ and p_i is the i^{th} $(i = 0, \ldots, N_p - 1)$ pilot symbol. 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 [6] 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 expressed by [6]

$$\hat{\mathbf{h}}_n = \boldsymbol{\mu}_{\mathbf{h}}^n + \boldsymbol{\Phi}_n \mathbf{W}_P^H (\hat{\mathbf{H}}_P^n - \mathbf{W}_P \boldsymbol{\mu}_{\mathbf{h}}^n), \tag{10}$$

where we define $\boldsymbol{\Phi}_n = [\mathbf{W}_P^H \mathbf{W}_P + \sigma^2 (\mathbf{C}_{\mathbf{h}}^n)^{-1}]^{-1}$. Note that the BMMSE estimates of (10) are more accurate than their MLE counterparts of (8), 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, how to derive a method that can help to estimate these parameters in an efficient and robust way, becomes a key issue in improving the practicality of BMMSE-CE for VLC systems, as to be resolved in the sequel.

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 (9) into (10), we have

$$\begin{aligned} \hat{\mathbf{h}}_n &= \boldsymbol{\mu}_{\mathbf{h}}^n + \boldsymbol{\Phi}_n \mathbf{W}_P^H (\mathbf{W}_P \mathbf{h}_n + \mathbf{V}_n) - \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{W}_P \boldsymbol{\mu}_{\mathbf{h}}^n \\ &= (\mathbf{I}_{L_c} - \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{W}_P) \boldsymbol{\mu}_{\mathbf{h}}^n + \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{W}_P \mathbf{h}_n + \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{V}_n \\ &= \mathbf{h}_n + \boldsymbol{\varepsilon}_n, \end{aligned} \tag{11}$$

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

$$\boldsymbol{\varepsilon}_n = \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{V}_n - (\mathbf{I}_{L_c} - \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{W}_P) \Delta \mathbf{h}_n, \tag{12}$$

where

$$\Delta \mathbf{h}_n = \mathbf{h}_n - \boldsymbol{\mu}_{\mathbf{h}}^n \tag{13}$$

denotes the change between the CIR $\mathbf{h}_n$ and its mean $\boldsymbol{\mu}_{\mathbf{h}}^n$ at the UE's n^{th} position. Furthermore, (12) may be rewritten as

$$\boldsymbol{\varepsilon}_n = \boldsymbol{\Psi}_1^n \mathbf{V}_n - \boldsymbol{\Psi}_2^n \Delta \mathbf{h}_n, \tag{14}$$

where we define $\boldsymbol{\Psi}_1^n = \boldsymbol{\Phi}_n \mathbf{W}_P^H$ and $\boldsymbol{\Psi}_2^n = \mathbf{I}_{L_c} - \boldsymbol{\Phi}_n \mathbf{W}_P^H \mathbf{W}_P$. Since $\boldsymbol{\mu}_{\mathbf{h}}^n$ in (13) is not obtainable in practical systems, we may instead use its *a priori* estimate $\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n$, yielding the estimated change of the channel

$$\Delta \hat{\mathbf{h}}_n = \mathbf{h}_n - \hat{\boldsymbol{\mu}}_{\mathbf{h}}^n. \tag{15}$$

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 historical 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 $\omega_l^n,\ l \in \{0, \ldots, L_c - 1\}$, formulated as

$$[\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n]_l = \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} [\hat{\mathbf{h}}_{\mathrm{ML}}^{n-k}]_l,\ l \in \{0, \ldots, L_c - 1\}, \tag{16}$$

where based on (8), the MLE-based estimate is given by [6]

$$\hat{\mathbf{h}}_{\mathrm{ML}}^n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \hat{\mathbf{H}}_P^n = \mathbf{h}_n + \mathbf{v}_n = \boldsymbol{\mu}_{\mathbf{h}}^n + \Delta \mathbf{h}_n + \mathbf{v}_n, \tag{17}$$

while the superscript $(\cdot)^{n-k}$ in (16) denotes the $(n-k)^{th}$ position. Inserting (9) into (17), the equivalent TD noise $\mathbf{v}_n$ can be calculated as

$$\mathbf{v}_n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \hat{\mathbf{H}}_P^n - \mathbf{h}_n = (\mathbf{W}_P^H \mathbf{W}_P)^{-1} \mathbf{W}_P^H \mathbf{V}_n. \tag{18}$$

Utilising (16) and (17), we may further develop (15) as

$$[\Delta \hat{\mathbf{h}}_n]_l = \frac{\omega_l^n - 1}{\omega_l^n} [\Delta \mathbf{h}_n]_l - \frac{1}{\omega_l^n} \sum_{k=0}^{\omega_l^n - 1} [\mathbf{v}_{n-k}]_l - \frac{1}{\omega_l^n} \sum_{k=1}^{\omega_l^n - 1} [\Delta \mathbf{h}_{n-k}]_l. \tag{19}$$

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 (5), (11) and (14), Γ_n may be transformed to

$$\begin{aligned} \Gamma_n =& \frac{1}{N} \mathrm{Tr}\{E\{(\hat{\mathbf{H}}_n - \mathbf{H}_n)(\hat{\mathbf{H}}_n - \mathbf{H}_n)^H\}\} = \mathrm{Tr}\{E\{\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^H\}\} \\ &= \sigma^2 \mathrm{Tr}\{\boldsymbol{\Psi}_1^n \boldsymbol{\Psi}_1^{nH}\} + \mathrm{Tr}\{E\{\boldsymbol{\Psi}_2^n \Delta \mathbf{h}_n \Delta \mathbf{h}_n^H \boldsymbol{\Psi}_2^{nH}\}\} \\ &\quad - \mathrm{Tr}\{E\{\boldsymbol{\Psi}_1^n \mathbf{V}_n \Delta \mathbf{h}_n^H \boldsymbol{\Psi}_2^{nH}\}\} - \mathrm{Tr}\{E\{\boldsymbol{\Psi}_2^n \Delta \mathbf{h}_n \mathbf{V}_n^H \boldsymbol{\Psi}_1^{nH}\}\}, \end{aligned} \tag{20}$$

which constitutes the objective function of the proposed AS-BMMSE-CE technique. Naturally, the estimated change of the channel denoted by (19) may be inserted into (20), forming a function of $\omega_l^n,\ l \in \{0, \ldots, L_c - 1\}$. Hence in AS-BMMSE-CE, we are interested in finding the optimum values $\omega_{l,\mathrm{opt}}^n,\ l \in \{0, \ldots, L_c - 1\}$ that minimise Γ_n of (20)

$$\omega_{l,\mathrm{opt}}^n = \underset{\{\omega_l^n\} \in \mathbb{N}_+}{\operatorname{argmin}}\ \Gamma_n,\ l \in \{0, \ldots, L_c - 1\}, \tag{21}$$

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

Nonetheless, as the complicated expression of (20) involves multiple coupled parameters, it may be difficult to solve (21) directly. It is therefore desirable to simplify (20), as to be discussed next.

C. Pilot Pattern and Covariance Matrices

Aiming to simplify (20), let us first cast a deeper insight into it. Note that $\boldsymbol{\Psi}_1^n$ and $\boldsymbol{\Psi}_2^n$ in (20) contain a common term of $\mathbf{W}_P^H \mathbf{W}_P$, where $\mathbf{W}_P$ is defined in (7). 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 (20). On the other hand, as suggested by [5], the pilots should be equally spaced in the FD to achieve the best CE performance and to achieve the minimal CRLB [6], [11].

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

$$\mathbf{W}_P^H \mathbf{W}_P = N_p \mathbf{I}_{L_c}, \tag{22}$$

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

$$P_0 + (\frac{N_p}{2} - 1) \times N_d + N_d = N - [P_0 + (\frac{N_p}{2} - 1) \times N_d], \tag{23}$$

where we have $N_p \times N_d = N$. With the aid of (23), we obtain

$$P_0 = N_d/2. \tag{24}$$

Based on (22) and (24), we may transform (20) to

$$\Gamma_n = \mathrm{Tr}\{E\{\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^H\}\} = \Theta(f^n_{\omega_l^n,l}, \mathbf{C}^n_{\mathbf{h}}),\ l = 0, \ldots, L_c - 1, \tag{25}$$

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

More specifically, $f^n_{\omega_l^n,l}$ 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

$$\begin{aligned} f^n_{\omega_l^n,l}(r^d_{n,l}) =& \frac{\sigma^2}{N_p \omega_l^n} + \frac{1}{(\omega_l^n)^2} \sum_{j=1}^{\omega_l^n - 1} \sum_{k=1}^{\omega_l^n - 1} r^{|j-k|}_{n,l} + \frac{(\omega_l^n - 1)^2}{(\omega_l^n)^2} r^0_{n,l} \\ & - 2 \cdot \frac{\omega_l^n - 1}{(\omega_l^n)^2} \sum_{j=1}^{\omega_l^n - 1} r^j_{n,l},\ l = 0, \ldots, L_c - 1, \end{aligned} \tag{26}$$

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

$$r^d_{n,l} = r^{|j-k|}_{n,l} = E\{([\mathbf{h}_{n-j}]_l - [\boldsymbol{\mu}^n_{\mathbf{h}}]_l)([\mathbf{h}_{n-k}]_l - [\boldsymbol{\mu}^n_{\mathbf{h}}]_l)^*\}, \tag{27}$$

while $r^j_{n,l}$ in (26) is obtained by setting $k = 0$ in (27). Note that $r^d_{n,l}$ of (27) 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

$$\mathbf{R}_{n,l} = \begin{bmatrix} r^0_{n,l} & r^1_{n,l} & \cdots & r^{\omega_{\max}-1}_{n,l} \\ r^1_{n,l} & r^0_{n,l} & \cdots & r^{\omega_{\max}-2}_{n,l} \\ \vdots & \vdots & \ddots & \vdots \\ r^{\omega_{\max}-1}_{n,l} & r^{\omega_{\max}-2}_{n,l} & \cdots & r^0_{n,l} \end{bmatrix}. \tag{28}$$

According to [12], the estimate of $r^d_{n,l}$ can be expressed as

$$\hat{r}^d_{n,l} = \frac{1}{\omega_{\max}} \sum_{j=0}^{\omega_{\max}-d-1} ([\hat{\mathbf{h}}^{n-j}_{\mathrm{ML}}]_l - [\bar{\boldsymbol{\mu}}_n]_l)([\hat{\mathbf{h}}^{n-(j+d)}_{\mathrm{ML}}]_l - [\bar{\boldsymbol{\mu}}_n]_l), \tag{29}$$

where $\hat{\mathbf{h}}^{(\cdot)}_{\mathrm{ML}}$ is given by (17), 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 $\bar{\boldsymbol{\mu}}_n = \frac{1}{\omega_{\max}} \sum_{k=0}^{\omega_{\max}-1} \hat{\mathbf{h}}^{n-k}_{\mathrm{ML}}$. Moreover, $\omega_{\max} \geq \omega_l^n,\ l \in \{0, \ldots, L_c - 1\}$ is the maximum length of the statistic windows, and its value should be carefully selected. If it is too large, the accuracy of $\hat{r}^d_{n,l}$ may be biased by more distanced and thus less relevant channel information. In contrast, if it is too small, the result of $\hat{r}^d_{n,l}$ may be dominated by residual noise which is not effectively mitigated due to insufficient historical channel information.

After obtaining $\bar{\boldsymbol{\mu}}_n$, we can use it to calculate (29) for generating $\mathbf{R}_{n,l}$ defined in (28). Note that the MLE estimate, namely $\hat{\mathbf{h}}^{(\cdot)}_{\mathrm{ML}}$ in (29), is contaminated by noise. The expectation of $\hat{r}^d_{n,l}$ in (29) contains TD noise items of

$$E\{\hat{r}^d_{n,l,\mathrm{noise}}\} = \begin{cases} \frac{\omega_{\max}-1}{\omega_{\max}} \sigma_0^2, & d = 0 \\ -\frac{\omega_{\max}-d}{\omega_{\max}} \cdot \frac{1}{\omega_{\max}} \sigma_0^2, & d = 1, \ldots, \omega_{\max} - 1 \end{cases}, \tag{30}$$

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 $r^d_{n,l}$ in (26) with $\hat{r}^d_{n,l}$ in (29), we have $f^n_{\omega_l^n,l}(r^d_{n,l}) \rightarrow \hat{f}^n_{\omega_l,l}(\hat{r}^d_{n,l})$. Utilising (30), we can therefore obtain the expectation of the introduced noise item as

$$E\{\hat{f}^n_{\omega_l^n,l,\mathrm{nosie}}(\hat{r}^d_{n,l})\} = -\frac{2(\omega_l^n)^2 - 3\omega_l^n(\omega^2_{\max}+1) + 3\omega^2_{\max} + 1}{3(\omega_l^n)^2 \omega^2_{\max}} \sigma_0^2. \tag{31}$$

Then, in order to eliminate the impact from the noise specified by (31), we may use

$$\hat{f}^n_{\omega_l^n,l}{}' = \hat{f}^n_{\omega_l^n,l} - E\{\hat{f}^n_{\omega_l^n,l,\mathrm{nosie}}\} \tag{32}$$

to replace $f^n_{\omega_l^n,l}$ in (25) and (26).

Next, we proceed to calculate $\mathbf{C}^n_{\mathbf{h}}$ specified in (25). Assuming that the changes of coefficients associated with different channel taps, which are represented by the elements of $\Delta\mathbf{h}_n$, are uncorrelated [6], we have $\mathbf{C}^n_{\mathbf{h}} = E\{\Delta\mathbf{h}_n \Delta\mathbf{h}_n^H\} = \mathrm{diag}\{\sigma^2_{n,0}, \ldots, \sigma^2_{n,L_c-1}\}$, where $\sigma^2_{n,l}\ (l = 0, \ldots, L_c - 1)$ denote the variance of $[\Delta\mathbf{h}_n]_l$ in (13) that corresponds to the l^{th} tap at the UE's n^{th} position. In order to obtain $\mathbf{C}^n_{\mathbf{h}}$, a forgetting factor λ is exploited to calculate the estimate of $\sigma^2_{n,l}$, namely $\hat{\sigma}^2_{n,l}$. More explicitly, we define [13]

$$\bar{\sigma}^2_{n,l} = \lambda \hat{\sigma}^2_{n-1,l} + (1-\lambda)(\hat{r}^0_{n,l} - \frac{\omega_{\max} - 1}{\omega_{\max}} \sigma_0^2). \tag{33}$$

Noting that $\hat{\sigma}^2_{n,l}$ should be a positive value, we may apply a small covariance constant $\sigma^2_{\mathrm{const}}$ to (33), resulting in

$$\hat{\sigma}^2_{n,l} = \begin{cases} \bar{\sigma}^2_{n,l}, & \bar{\sigma}^2_{n,l} > 0 \\ \sigma^2_{\mathrm{const}}, & \bar{\sigma}^2_{n,l} \leq 0 \end{cases}, \tag{34}$$

which is the estimate of the l^{th} diagonal element of $\mathbf{C}^n_{\mathbf{h}}$.

Based on (32) and (34), we therefore simplify the objective function (20) to (25) and (26), which involve a number of L_c target variables to be optimised, namely the statistic window sizes $\omega_l^n,\ l = 0, \ldots, L_c - 1$.

D. Optimum VSW Size and MSE Bound

Recall that the optimum solution for the objective function Γ_n defined in (20) or (25) is given by (21), which is an integer programming and thus a traditional NP-complete problem [14]. Since there are a total number of $\omega_{\max}$ candidate window sizes for each of the L_c taps, the optimisation of (21) results in a high computational complexity of $O[(\omega_{\max})^{L_c}]$.

Nonetheless, note that (25) may be reformulated as

$$\begin{aligned} \hat{\Gamma}_n &= \sum_{l=0}^{L_c-1} \left\{ \frac{N_p \sigma^2}{(N_p + \frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2} + \frac{(\frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2 \hat{f}^n_{\omega_l^n,l}{}'}{(N_p + \frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2} + \frac{2\sigma^2}{\omega_l^n} \frac{\frac{\sigma^2}{\hat{\sigma}^2_{n,l}}}{(N_p + \frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2} \right\} \\ &= \sum_{l=0}^{L_c-1} \hat{M}^n_{\omega_l^n,l}, \end{aligned} \tag{35}$$

where we define

$$\hat{M}^n_{\omega_l^n,l} = \frac{N_p \sigma^2}{(N_p + \frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2} + \frac{(\frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2 \hat{f}^n_{\omega_l^n,l}{}'}{(N_p + \frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2} + \frac{2\sigma^2}{\omega_l^n} \frac{\frac{\sigma^2}{\hat{\sigma}^2_{n,l}}}{(N_p + \frac{\sigma^2}{\hat{\sigma}^2_{n,l}})^2} \tag{36}$$

and $\hat{\sigma}^2_{n,l}$ is given in (34). Note that the corresponding estimated version of Γ_n and $f^n_{\omega_l,l}$ are used in (35).

Therefore, we can see that $\hat{\Gamma}_n$ can be effectively decoupled into independent items $\hat{M}^n_{\omega_l^n,l},\ l \in \{0, \ldots, L_c - 1\}$, which are

associated with ω_l^n. Hence, with the aid of (36), we may solve $\hat{\Gamma}_n$ through exhaustively searching for each tap-specific $\omega_{l,\text{opt}}^n$ in the candidate solution set of $\{1,\ldots,\omega_{\max}\}$, yielding

$$\omega_{l,\text{opt}}^n = \underset{\omega_l^n \in \{1,\ldots,\omega_{\max}\}}{\text{argmin}} \hat{M}_{\omega_l^n,l}^n. \tag{37}$$

In this case, the resultant complexity required by (21) can be significantly reduced from $O[(\omega_{\max})^{L_c}]$ to $O(\omega_{\max} L_c)$. We summarise the proposed VSW optimisation algorithm (VOA) in Algorithm 1. Based on the above discussions, the complete description of the proposed AS-BMMSE-CE scheme is provided by Algorithm 2.

Algorithm 1 VSW Optimisation Algorithm (VOA)

Initialisation: Set $l = 0$.
repeat
  $\omega_l^n = 1$
  repeat
    Calculate (32) and (36)
    $\omega_l^n = \omega_l^n + 1$
  until $\omega_l^n > \omega_{\max}$
  $\omega_{l,\text{opt}}^n = \text{argmin}_{\omega_l^n \in \{1,\ldots,\omega_{\max}\}} \hat{M}_{\omega_l^n,l}^n$
  $l = l + 1$
until $l > L_c - 1$
Return: $\omega_{l,\text{opt}}^n,\ l = 0,\ldots,L_c - 1$ are the optimal VSW sizes.

Algorithm 2 The AS-BMMSE Algorithm

Initialisation: Obtain λ, $\omega_{\max}$ and L_c. Set $\hat{\mathbf{C}}_{\mathbf{h}}^n = \text{diag}(\frac{1}{L_c},\ldots,\frac{1}{L_c})$, $\hat{\mathbf{h}}_{\text{ML}}^k = \mathbf{0}, k = 0,-1,\ldots,-\omega_{\max}+2$ and $n = 1$.
repeat
  $\hat{H}[n,k] = \frac{\hat{Y}[n,k]}{X[n,k]}$, $k \in P_{\text{pilot}}$
  Calculate (8) and (29)
  $l = 0$
  repeat
    Calculate (33) and (34)
    $l = l + 1$
  until $l > L_c - 1$
  Execute Algorithm 1
  $l = 0$
  repeat
    $[\hat{\boldsymbol{\mu}}_{\mathbf{h}}^n]_l = \frac{1}{\omega_{l,\text{opt}}^n} \sum_{i=0}^{\omega_{l,\text{opt}}^n - 1} [\hat{\mathbf{h}}_{\text{ML}}^{n-i}]_l$
    $l = l + 1$
  until $l > L_c - 1$
  Calculate (10)
  Apply N-point FFT to get $\hat{\mathbf{H}}_n$
  $n = n + 1$
until n approaches a predefined maximal value.

As a further remark, after some derivations we can obtain

$$M_{\omega_l^n,l}^n \le \frac{\sigma^2/N_p}{[1+\sigma^2/(\sigma_{n,l}^2 N_p)]^2}\Big[1 + (\frac{\sigma^2}{\sigma_{n,l}^2 N_p})^2 \frac{1}{\omega_l^n} + 2\frac{\sigma^2}{\sigma_{n,l}^2 N_p}\Big] \tag{38}$$

and

$$M_{\omega_l^n,l}^n \ge \frac{\sigma^2/N_p}{[1+\sigma^2/(\sigma_{n,l}^2 N_p)]^2}\Big[1 + (\frac{\sigma^2}{\sigma_{n,l}^2 N_p})^2 \frac{1}{\omega_l^n} + \frac{2}{\omega_l^n}\frac{\sigma^2}{\sigma_{n,l}^2 N_p}\Big] \tag{39}$$

as the upper and lower bounds of AS-BMMSE, respectively. Note that the AS-BMMSE upper bound (ASB-UB) of (38) is lower than the CRLB of [11], while the AS-BMMSE lower bound (ASB-LB) of (39) is lower than the traditional Bayesian lower bound (TBLB) of [6], [11], as demonstrated by our numerical results in Section IV. This proves that the proposed AS-BMMSE-CE scheme can achieve a better performance than several existing conventional arrangements.

IV. Numerical Results and Analysis

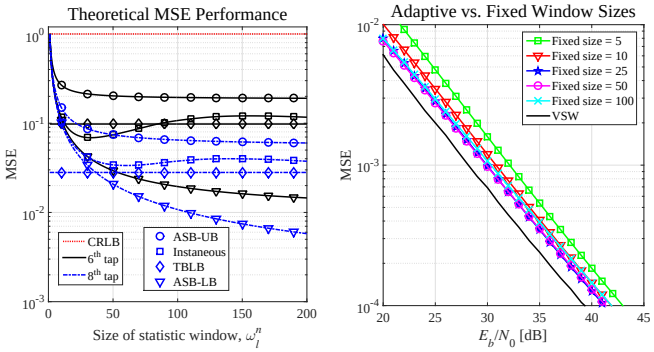
In this section, simulation results are provided for demonstrating the effectiveness of the proposed AS-BMMSE-CE scheme. Assuming a general indoor scenario, a room model with a size of $5 \times 5 \times 4\text{m}^3$ is adopted, where the maximal reflection order of the VLC channel model [10] is set 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 photodiode (PD) to receive the same signal transmitted from all LEDs and moves around in the room. Naturally, the instantaneous CIR established by [10] varies as soon as UE's position changes. The parameters in Table I apply to most scenarios tested in this section, unless otherwise stated.

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	85°
Modulation scheme	DCO-OFDM
Constellation diagram	16-QAM
Number of subcarriers, N	1024
Cycle prefix length, N_{cp}	64
Maximum tap delay	62ns
Smallest FD pilot subcarrier index	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, $\omega_{\max}$	50
Forgetting factor, λ	0.6
σ_{const}^2 in (34)	σ^2/N_p

In Fig. 2(a), we evaluate the theoretical MSE performance of AS-BMMSE-CE using (20) 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 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. 2(a) that both the ASB-UB of (38) and the ASB-LB of (39) are lower than the CRLB [6], [11], implying that AS-BMMSE-CE outperforms MLE of [6] in terms of MSE performance. Moreover, our scheme may also be capable of breaking the TBLB of [6]

with the aid of an appropriately selected window size ω_l^n, as observed for instance in the case of the 6^{th} tap. Since there are some correlations of some spacial taps between adjust positions, if these correlations are strong, the MSE of proposed method with appropriate VSW will be lower than TBLB.



(a) Theoretical MSE bounds. (b) Adaptive vs. fixed window sizes.

Fig. 2. The MSE performances of AS-BMMSE-CE exploiting fixed-size window or VSW.

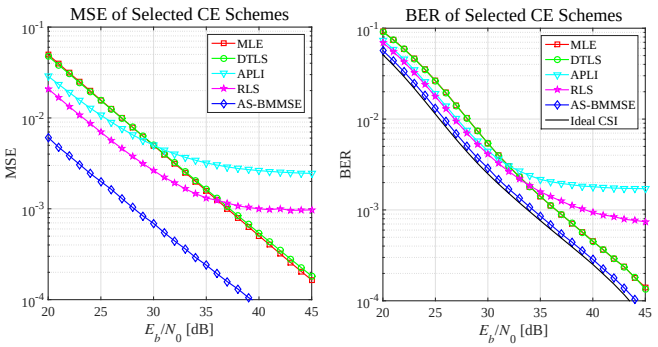
Next, for demonstrating the impact from the optimum value of $\omega_{l,\text{opt}}^n$ indicated by (37), we investigate the MSE performance of AS-BMMSE-CE under adaptive or fixed-size statistic windows in Fig. 2(b). Under the adaptive option, whenever the UE moves to a different position in the room, the system calculates the optimal values $\omega_{l,\text{opt}}^n,\ l = 0, \ldots, L_c - 1$ based on (21), hence the so-called VSW mechanism. It can be inferred from Fig. 2(b) that the VSW-aided scheme achieves the lowest possible MSE, as compared with its counterparts using a fixed-size statistic window.

In Fig. 3, we compare the MSE and BER performances of AS-BMMSE-CE with selected existing CE schemes, such as MLE [6], APLI [4], domain-transform least squares (DTLS) [7] and RLS [8]. The reference schemes were such configured, that they fitted into the common system platform and the channel model under comparable conditions. We can see that our method has the best performances among the schemes investigated. Moreover, it has only 0.5dB loss compared with the benchmark with ideal CSI, as seen in Fig. 3(b).

As a final remark, the computational complexity of AS-BMMSE-CE is relatively high. The number of complex additions and complex multiplications required for one OFDM symbol are $\mathcal{C}_{\text{add}} = 28L_c\omega_{\max} + L_c\omega_{\max}^2 + 2\omega_{\max} + 6L_c + L_cN_p + \frac{1}{3}\omega_{\max}^3 + N\log N$ and $\mathcal{C}_{\text{mul}} = 27L_c\omega_{\max} + 2L_c\omega_{\max}^2 + \omega_{\max} + 6L_c + (2L_c+1)N_p + \frac{1}{2}N\log N$, respectively. Reducing its complexity constitutes part of our future work.

V. 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 historical channel information



(a) MSE performances. (b) BER performances.

Fig. 3. The MSE and BER versus E_b/N_0 performances of AS-BMMSE-CE and other CE schemes.

collected in the statistic window with an adaptively optimised size. Extensive theoretical and simulation results are provided to demonstrate the benefits of the new CE method.

VI. 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, 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.

[5] 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.

[6] 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.

[7] 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.

[8] S. R. D. Paulo, *Adaptive filtering: Algorithms and practical Implementation*. Springer, 1997.

[9] 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.

[10] 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.

[11] M. K. Steven , *Fundamentals of statistical signal processing - Estimation theory*. Prentice Hall PTR: University of Rhode Island, 1993.

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[13] 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.

[14] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.