# Phase and Channel Estimation for High-Capacity Phase-Asynchronous Mode-Division Multiplexing Multiple-Input Multiple-Output Free-Space Optical Communications in Strong Turbulent Channels

Yiming Li, Zhaozhong Chen, Mohammed Patel, Martin P. J. Lavery, and Andrew D. Ellis, *Fellow, Optica*

*(Top-Scored Paper)*

***Abstract*—In this paper, we propose a novel pilot-aided phase and channel estimation algorithm for coherent mode-division multiplexing multiple-input multiple-output free-space optical communication system. This algorithm enables the implementation of advanced multiple-input multiple-output decoders, leading to a significantly better bit error rate performance. This algorithm also supports phase-asynchronous light sources at each transmit and receive channel, significantly reducing the hardware requirements. Moreover, it has low computational complexity and Cramér–Rao lower bound approaching estimation performance. In our proof-of-concept experiment, we employed 10 decorrelated channels to achieve a record-high 689.2 Gbit/s/wavelength line rate in strong turbulence, verifying the feasibility of our phase and channel estimation algorithm.**

***Index Terms*—Free-space optics, multiple-input multiple-output, channel estimation, atmospheric turbulence.**

## I. Introduction

THE beyond 5G/6G networks are expected to address the demanding requirements of wireless data traffic. Free-space optical (FSO) communication is a promising technology to address this demand by providing a much faster data rate, lower latency, and enhanced security [1]. To further increase the capacity of FSO transmissions, mode-division multiplexing (MDM) technology was introduced to exploit the spatial diversity (i.e. parallel spatial channels) [2]. Unlike other multiplexing technologies such as polarisation multiplexing or dense wavelength-division multiplexing (DWDM) [3], [4], MDM can also provide enhanced atmospheric turbulence resiliency in the digital domain without an adaptive optics (AO) system [5], making it a promising technology for high-capacity FSO communications with a much faster adaptation speed for turbulence.

Multiple-input multiple-output (MIMO) digital signal processing (DSP) is one of the major approaches to mitigate the negative atmospheric turbulence and improve the bit error rate (BER) performance of FSO systems [6], [7]. Most reported MIMO algorithms in optical communication systems are based on a 3-dimensional equaliser (3D-EQ) which performs minimum mean squared error (MMSE) MIMO detection. It works well in few-mode fibre (FMF) channels with approximately unitary channel matrices [8]–[10], but quantifies significant performance degradation in the presence of turbulence-induced fading [6], [7], [11]. Therefore, it is necessary to go beyond the MMSE algorithm [11] and introduce more advanced MIMO decoders (e.g. maximum-likelihood decoder (MLD) [11], sphere decoder (SD) [12] or successive interference cancellation (SIC) decoder [13]) for MDM FSO communications. To deploy advanced MIMO decoders, it is necessary to replace the 3D-EQ with an advanced DSP scheme, which separates phase and channel estimation from MIMO detection.

Phase noise is one of the major imperfections in coherent FSO systems. To date, most MDM experiments implicitly used a single narrow-linewidth high-power laser source at both the transmitter and the receiver side to obtain identical phase noise among all the modes [6], [8]. Although boost amplifiers [9] or optical injection locking [14] can be introduced in the systems to mitigate the stringent requirements on the laser source, they will still introduce extra hardware costs. In order to scale up the number of spatial modes, it is necessary to relax the hardware requirements in MDM systems by employing phase-asynchronous laser sources [10]. However, the detrimental impact of phase noise can be more pronounced due to the turbulence-induced channel matrix deterioration in MDM FSO systems. Although extensive research on phase noise estimation has been conducted in single-input single-output (SISO) systems [15], [16], they can not be directly applied to MIMO systems, where the received signal may be deteriorated by multiple phase noise parameters. Therefore, the joint phase and channel estimation for wireless MIMO systems with independent phase noise is of particular interest and has been investigated by several researchers [17]–[19]. However, previous algorithms have to estimate phase *after* channel estimation, limiting the maximum effective length of the training sequence and degrading the channel estimation accuracy. Therefore, these algorithms suffer from either significant degradation in estimation accuracy [17], [18], or unacceptably high computational complexity [19]. Therefore,

Manuscript received October 00, 2023; revised October 00, 2023. Research supported by EPSRC under Grants EP/T009047/1, EP/T009012/1, EP/S003436/1, and EP/S016171/1, and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant 713694, and Future and Emerging Technologies Open Grant SuperPixels 829116. *(Corresponding author: Zhaozhong Chen.)*

Y. Li, M. Patel and A. Ellis are with Aston Institute of Photonic Technologies, Aston University, Birmingham, B4 7ET, UK (e-mail: y.li70@aston.ac.uk; d.benton@aston.ac.uk; m.patel70@aston.ac.uk; andrew.ellis@aston.ac.uk).

Z. Chen, and M. Lavery are with James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK (e-mail: Zhaozhong.Chen@glasgow.ac.uk; Martin.Lavery@glasgow.ac.uk).

there is a pressing demand for a new phase and channel estimation algorithm with low computational complexity and high estimation accuracy.

In this paper, we propose a novel pilot-aided phase and channel estimation algorithm for MDM FSO systems with independent phase noise to address the aforementioned requirements. By moving the phase noise estimation and recovery *before* full channel estimation,

1) the algorithm supports advanced MIMO decoders;
2) the algorithm supports phase-asynchronous laser sources;
3) the algorithm has very low computational complexity;
4) the algorithm has Cramér–Rao Lower Bound (CRLB)-approaching estimation accuracy.

We also validate this algorithm in an experimental MDM FSO system, which was previously discussed in our recent European Conference on Optical Communication (ECOC) paper [20]. By employing 10 decorrelated polarization- and mode-multiplexed transmit channels, we achieved a record-high line rate of 689.2 Gbit/s in strong turbulence, indicating the feasibility of our phase and channel estimation algorithm.

The remainder of this paper is organized as follows: the system model is given in Section II. Section III proposes the novel phase and channel estimation algorithm. Section IV provides the main simulation results of the proposed algorithm. Section V depicts the proof-of-concept experimental setup. Section VI provides the experimental results to verify the effectiveness of the algorithm in the MDM MIMO FSO systems. Finally, Section VII summarizes the key advantages of the proposed algorithm.

*Notations*: $\mathbf{x}_{m,:}$ denotes the $m^{th}$ row vector of matrix $\mathbf{X}$, and $\mathbf{x}_{:,m}$ denotes the $m^{th}$ column vector of $\mathbf{X}$.

## II. System Model

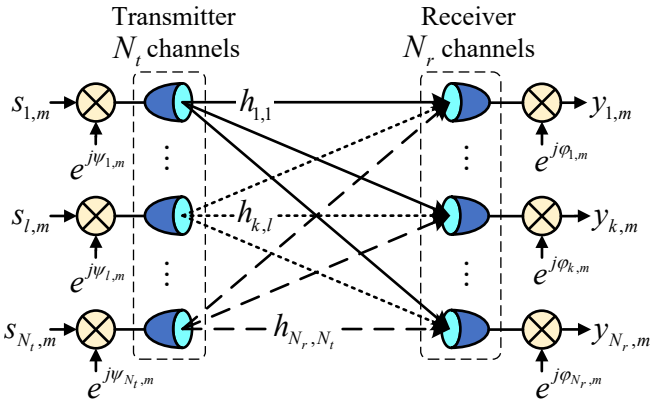


Fig. 1. Schematic diagram for MDM MIMO FSO systems.

As shown in Fig. 1, a point-to-point MDM FSO system with $N_t$ transmit channels and $N_r$ receive channels is considered. In Fig. 1, $s_{l,m}$ represents the transmitted symbol at the $l^{th}$ transmitter and the $m^{th}$ time interval. $y_{k,m}$ represents the received signal at the $k^{th}$ receiver and the $m^{th}$ time interval. $h_{k,l}$ is the channel gain between the $k^{th}$ receiver and the $l^{th}$ transmitter. $\psi_{l,m}$ represents the phase noise at the $l^{th}$ transmitter and the $m^{th}$ time interval. And $\varphi_{l,m}$ represents the phase noise at the $k^{th}$ receiver and the $m^{th}$ time interval.

As shown in Fig. 2, data symbols are transmitted as frames. The frame length is $L_f$. In each frame, there is a training sequence of length $L_{ts}$. After the training sequence, $L_p$ consecutive *pilot* symbols (a pilot group) are inserted every $L_d$ *data* symbols. The cell length ($L_c$), cell number ($N_c$), and pilot rate ($R_p$) of a frame are defined as

$$\begin{cases} L_c = L_p + L_d \\ N_c = \dfrac{L_f - L_{ts}}{L_p + L_d} \\ R_p = \dfrac{L_p}{L_p + L_d} \end{cases}. \tag{1}$$

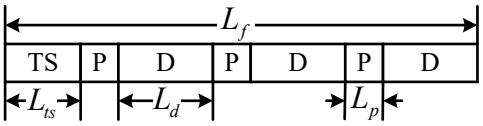


Fig. 2. Frame structure of the transmitted data in MIMO systems. TS: Training Sequence, P: Pilot, D: Data.

The channel matrix has a rank of $N_t$ in a general MIMO system when $N_r \geq N_t$. A feasible phase and channel estimation can not be obtained within one pilot group if $L_p < N_t$ (there are insufficient equations to solve the channel matrix). In this case, although the phase and channel can be estimated by several pilot groups, the inter-pilot-group phase noise significantly contaminates the channel matrix estimation. On the other hand, the impact of intra-pilot-group phase noise on channel estimation increases as $L_p$ increases (more phase wander during $L_p$), and so minimizing $L_p$ is recommended. Therefore, the optimal condition of $L_p = N_t$ is important for the phase and channel estimation.

To simplify the mathematical analysis and focus on the phase and channel estimation, the following assumptions are adopted in the algorithm derivation and analysis:

A1. The pilots are *known* at the receiver. Moreover, all the transmitters simultaneously transmit mutually orthogonal and element-wise normalized pilots of length $L_p = N_t$ in one pilot group.
A2. The laser sources for different transmit channels and local oscillators (LOs) for different receive channels are independent (Fig. 1), which is a generalized case of different kinds of practical systems [2], [10], [21].
A3. The channel matrix $\mathbf{H}$ remains constant in a frame and there is no inter-symbol interference (ISI) in the system.
A4. The phase noise is modelled by a Wiener process, which is static in one symbol period but varies from symbol to symbol [10], [15], [17].
A5. Perfect signal-to-noise ratio (SNR) estimation, timing recovery, frame and bit synchronization, and frequency estimation are assumed, which can be achieved by standard algorithms [11], [22]–[25].

In the above-mentioned system, the discrete-time baseband received signal of the system can be described as

$$\mathbf{y}_{:,m} = \mathbf{\Phi}_m \mathbf{H} \mathbf{\Psi}_m \mathbf{s}_{:,m} + \mathbf{n}_{:,m}, \tag{2}$$

where

$m$ represents the $m^{th}$ time interval.

$\mathbf{y}_{:,m} = [y_{1,m}, \cdots, y_{N_r,m}]^T$ is the received signal vector, and $(\cdot)^T$ is the transpose operator.

$\mathbf{\Phi}_m = \text{diag}\left(e^{j\varphi_{1,m}}, \cdots, e^{j\varphi_{N_r,m}}\right)$ is the diagonal phase matrix of the receiver LOs, and $\text{diag}(\mathbf{x})$ is a diagonal matrix where the diagonal elements are given by vector $\mathbf{x}$.

$\mathbf{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,N_t} \\ \vdots & \ddots & \vdots \\ h_{N_r,1} & \cdots & h_{N_r,N_t} \end{bmatrix}$ is the channel matrix.

$\mathbf{\Psi}_m = \text{diag}\left(e^{j\psi_{1,m}}, \cdots, e^{j\psi_{N_t,m}}\right)$ is the diagonal phase matrix of the transmitter laser sources.

$\mathbf{s}_{:,m} = [s_{1,m}, \cdots, s_{N_t,m}]^T$ is the transmitted signal vector with normalized power.

$\mathbf{n}_{:,m} = [n_{1,m}, \cdots, n_{N_r,m}]^T$ is the independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) vector, and $n_{k,m} \sim \mathcal{CN}\left(0, \sigma_n^2\right)$ obeys circularly-symmetric complex Gaussian distribution with mean 0 and variance $\sigma_n^2$.

Following assumption (A4), the phase of the laser sources for different transmit and receive channels can be modelled as Wiener process, and $\varphi_{k,m}$ and $\psi_{l,m}$ are then given by

$$\begin{cases} \varphi_{k,m} = \varphi_{k,m-1} + \Delta\varphi_{k,m} \\ \psi_{l,m} = \psi_{l,m-1} + \Delta\psi_{l,m} \end{cases}, \tag{3}$$

where $\Delta\varphi_{k,m} \sim \mathcal{N}\left(0, \sigma_{\Delta\varphi}^2\right)$ and $\Delta\psi_{l,m} \sim \mathcal{N}\left(0, \sigma_{\Delta\psi}^2\right)$ are the phase innovations at the $k^{th}$ LO and the $l^{th}$ transmit laser source, respectively, at the $m^{th}$ symbol period. $\mathcal{N}\left(\mu, \sigma^2\right)$ denotes real Gaussian distribution with mean $\mu$ and variance $\sigma^2$. $\sigma_{\Delta\varphi}^2 = 2\pi\Delta\nu_r T$, and $\sigma_{\Delta\psi}^2 = 2\pi\Delta\nu_t T$. $\Delta\nu_r$ and $\Delta\nu_t$ are the 3 dB linewidths of the LOs and signal lasers, respectively. $T$ is the symbol period.

## III. Estimation Algorithm

### A. Basic Principles of the Estimation Algorithm

Throughout our algorithm, *pilots* are used for estimating the phase and channel, while the training sequence is only used for frame synchronization and frequency estimation in assumption (A5). Without loss of generality, the time index of the first symbol in the first pilot group is labelled as $m = 1$ (the time indices of the training sequence are $-L_{ts} + 1 \leq m \leq 0$, which are omitted in the algorithm).

Following assumption (A3), a suitable estimation of channel matrix $\mathbf{H}$ should exploit all the pilot information throughout a frame to reduce estimation error. However, if the phase estimation is performed *after* channel estimation, there will be a trade-off between the AWGN and the phase noise, and the channel estimation can be severely contaminated by phase noise when using all the pilots. To remove the trade-off, it is necessary to estimate and recover phase *before* channel estimation.

To estimate the phase of different transmit and receive channels within a pilot group, a weighted linear least squares (WLLS) estimator is proposed to estimate the phase *before* full channel information. By extracting $N_r + N_t - 1$ phase information from $N_r N_t$ observed angular terms in the channel matrix, and assigning a smaller weight to the observed terms with larger errors, the phase estimation error is significantly reduced by employing the WLLS estimator. Moreover, Wiener filters are employed to further improve the phase estimation accuracy by exploiting the inter-pilot-group phase information.

It will be shown later that the WLLS estimator needs the information of $|\mathbf{H}|$ as weight coefficients. By applying conventional linear least squares (LLS) algorithm, the estimation of $|\mathbf{H}|$ is possible in each pilot group. Because of the element-wise absolute operator $|\cdot|$, the phase contamination phenomenon can not influence the amplitude of the channel matrix. Therefore, it is possible to calculate the element-wise average of $|\mathbf{H}|$ throughout all the pilots within a frame before phase recovery.

As a conclusion of the above mentioned reasons, the structure of our newly proposed algorithm is shown in Fig. 3.

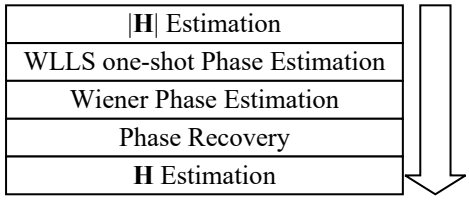


Fig. 3. Structure of estimation algorithm.

*Remark 1:* The influence of the intra-pilot-group phase noise is neglectable in the low SNR region and very subtle in the high SNR region. Therefore, it is neglected throughout this Section. The detailed influence of intra-pilot-group phase noise is discussed in [26] for interested readers.

### B. Step 1: Estimation of Channel Amplitude

It is straightforward to prove that the phase terms $\mathbf{\Phi}_m$ and $\mathbf{\Psi}_m$ do not change the amplitude of the equivalent channel matrix $\mathbf{\Phi}_m \mathbf{H} \mathbf{\Psi}_m$. Therefore, the estimation of channel amplitude at the $i^{th}$ pilot group can be given by conventional LLS algorithm as [17]

$$\left|\hat{\mathbf{H}}_i\right| = \left|\mathbf{Y}_i \mathbf{S}_i^+\right|, \tag{4}$$

where $|\cdot|$ denotes the element-wise absolute value operator, the matrices $\mathbf{Y}_i = \left[\mathbf{y}_{:,(i-1)L_c+1}, \cdots, \mathbf{y}_{:,(i-1)L_c+N_t}\right]$, $\mathbf{S}_i = \left[\mathbf{s}_{:,(i-1)L_c+1}, \cdots, \mathbf{s}_{:,(i-1)L_c+N_t}\right]$, and $(\cdot)^+$ is the Moore–Penrose inverse operator.

Following assumption (A1), $\mathbf{S}_i \mathbf{S}_i^H = N_t \mathbf{I}_{N_t}$. Therefore, the estimation of channel amplitude at the $i^{th}$ pilot group can be

$$\left|\hat{\mathbf{H}}_i\right| = \frac{1}{N_t}\left|\mathbf{Y}_i \mathbf{S}_i^H\right|, \tag{5}$$

where $(\cdot)^H$ is the Hermitian transpose operator.

When the intra-pilot-group phase noise is neglected, and the reference phase is set at the middle of the $i^{th}$ pilot group, the time index of the reference phase can be defined as $m_i \triangleq (i-1) \times L_c + \lceil N_t/2 \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Then (5) can be written as

$$\begin{aligned} \left|\hat{\mathbf{H}}_i\right| &\approx \frac{1}{N_t}\left|\left(\mathbf{\Phi}_{m_i} \mathbf{H} \mathbf{\Psi}_{m_i} \mathbf{S}_i + \mathbf{N}_i\right) \mathbf{S}_i^H\right| \\ &= \left|\mathbf{\Phi}_{m_i} \mathbf{H} \mathbf{\Psi}_{m_i} + \frac{1}{N_t} \mathbf{N}_i \mathbf{S}_i^H\right| \\ &= \left|\mathbf{\Phi}_{m_i} \mathbf{H} \mathbf{\Psi}_{m_i} + \mathbf{N}_i'\right|, \end{aligned} \tag{6}$$

where $\mathbf{N}_i = \left[\mathbf{n}_{:,(i-1)\times L_c+1}, \cdots, \mathbf{n}_{:,(i-1)\times L_c+N_t}\right]$ is the corresponding AWGN matrix. Because $\mathbf{S}_i^H/\sqrt{N_t}$ is a unitary matrix, $\mathbf{N}'_i = \frac{1}{N_t}\mathbf{N}_i\mathbf{S}_i^H$ is also an i.i.d. circularly-symmetric Gaussian matrix, and the element at the $k^{th}$ row and $l^{th}$ column of $\mathbf{N}'_i$ obeys the distribution $n'_{i,k,l} \sim \mathcal{CN}\left(0, \sigma_n^2/N_t\right)$.

Noting the i.i.d. property of $\mathbf{N}'_i$, and $\sigma_n^2$ can be perfectly estimated according to assumption (A5), the element-wise squaring of channel amplitude estimation can be calculated over the whole frame as

$$\left|\hat{\mathbf{H}}\right|_{sq} = \frac{1}{N_c}\sum_{i=1}^{N_c}\left(\left|\hat{\mathbf{H}}_i\right| \odot \left|\hat{\mathbf{H}}_i\right|\right) - \sigma_n^2\mathbf{1}_{N_r\times N_t}, \tag{7}$$

where $\odot$ is the element-wise product operator, and $\mathbf{1}_{X\times Y}$ denotes the $X \times Y$ all one matrices.

Finally, the estimation of channel amplitude over the whole frame can be calculated as

$$\left|\hat{\mathbf{H}}\right| = \sqrt{\max\left(\mathbf{0}_{N_r\times N_t}, \left|\hat{\mathbf{H}}\right|_{sq}\right)}. \tag{8}$$

The element-wise maximum operator in (8) is to prevent the estimated value in (7) from accidentally being less than 0 when $h_{i,j} \to 0$.

*Remark 2:* A conventional estimate of channel amplitude would be $\left|\hat{\mathbf{H}}\right| = \left|\frac{1}{N_c}\sum_{i=1}^{N_c}\hat{\mathbf{H}}_i\right|$ or $\left|\hat{\mathbf{H}}\right| = \frac{1}{N_c}\sum_{i=1}^{N_c}\left|\hat{\mathbf{H}}_i\right|$. However, in the first case, the estimation is corrupted by inter-pilot-group phase noise, reducing the accuracy for high phase noise, whilst for the latter the absolute value of the observable is dominated by AWGN at low SNR (consider the case of $h_{k,l} = 0$, this estimation gives a result which depends on the AWGN and is larger than 0). Both of these issues are resolved by adopting (9) and (10).

### C. Step 2: Weighted Linear Least Squares Phase Estimation

Similar to (6), the following approximation holds

$$\hat{\mathbf{H}}_i = \frac{1}{N_t}\mathbf{Y}_i\mathbf{S}_i^H \approx \mathbf{\Phi}_{m_i}\mathbf{H}\mathbf{\Psi}_{m_i} + \mathbf{N}'_i. \tag{9}$$

Define the angular term of $\hat{\mathbf{H}}_i$ and $\mathbf{H}$ as

$$\begin{cases}\mathbf{A}_i \triangleq \angle\hat{\mathbf{H}}_i \\ \mathbf{\Theta} \triangleq \angle\mathbf{H}\end{cases}, \tag{10}$$

where $\angle(\cdot)$ denotes the element-wise phase angle of complex matrices. The element at the $k^{th}$ row and $l^{th}$ column of $\mathbf{A}_i$ can be represented as

$$\begin{aligned} a_{i,k,l} &\approx \angle\left(e^{j\varphi_{k,m_i}}\left|h_{k,l}\right|e^{j\theta_{k,l}}e^{j\psi_{l,m_i}} + n'_{i,k,l}\right) \\ &= \angle\left(\left|h_{k,l}\right|e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}+\theta_{k,l}\right)} + \left|h_{k,l}\right|\frac{n'_{i,k,l}}{|h_{k,l}|}\right) \\ &\approx \varphi_{k,m_i} + \psi_{l,m_i} + \theta_{k,l} + \frac{n''_{i,k,l}}{|h_{k,l}|}, \end{aligned} \tag{11}$$

where $\theta_{k,l}$ is the element at the $k^{th}$ column and $l^{th}$ row of $\mathbf{\Theta}$, and $n''_{i,k,l} \sim \mathcal{N}\left(0, \sigma_n^2/(2N_t)\right)$ is the azimuthal component of $n'_{i,k,l}$. The last approximation is based on the small noise assumption, which is also used in [15]. On the other hand, the angular term $a_{:,k,l}$ given by (11) is within the interval of $[-\pi, \pi)$, which is referred to as phase wrapping. Therefore, a standard phase unwrapping algorithm given by [27] is necessary to unwrap each $\mathbf{a}_{:,k,l}$.

To eliminate phase ambiguity, it is intuitively satisfying to set one of the transmit laser sources as reference [18], [25]. Without loss of generality, the phase of the last transmit laser, $\psi_{N_t,m}$, is set as reference, and (11) can be rewritten as

$$\begin{aligned} &a_{i,k,l} \\ &\approx (\varphi_{k,m_i} + \psi_{N_t,m_i}) + (\psi_{l,m_i} - \psi_{N_t,m_i}) + \theta_{k,l} + \frac{n''_{i,k,l}}{|h_{k,l}|} \\ &= \beta_{k,m_i} + \beta_{N_r+l,m_i} + \theta_{k,l} + \frac{n''_{i,k,l}}{|h_{k,l}|}, \end{aligned} \tag{12}$$

where

$$\beta_{q,m} = \begin{cases}\varphi_{q,m} + \psi_{N_t,m}, (1 \leqslant q \leqslant N_r) \\ \psi_{q-N_r,m} - \psi_{N_t,m}, (N_r+1 \leqslant q \leqslant N_r+N_t-1),\end{cases} \tag{13}$$

is one of the $N_r + N_t - 1$ phase noise values correlated to the reference.

In order to enable matrix calculation, $\mathbf{A}_i$ is rearranged as

$$\boldsymbol{\alpha}_i = \left[\mathbf{a}_{i,:,1}^T, \cdots, \mathbf{a}_{i,:,N_t}^T\right]^T. \tag{14}$$

Define

$$\mathbf{C} \triangleq \begin{bmatrix}\mathbf{I}_{N_r} & \mathbf{B}_1 \\ \vdots & \vdots \\ \mathbf{I}_{N_r} & \mathbf{B}_{N_t-1} \\ \mathbf{I}_{N_r} & \mathbf{0}_{N_r\times(N_t-1)}\end{bmatrix}, \tag{15}$$

where $\mathbf{I}_X$ is the $X \times X$ identity matrix. $\mathbf{0}_{X\times Y}$ is the $X \times Y$ all zero matrix, and $\mathbf{B}_i = \left[\mathbf{0}_{N_r\times(i-1)}, \mathbf{1}_{N_r\times 1}, \mathbf{0}_{N_r\times(N_t-1-i)}\right]$. Then (12) can be rewritten in vector form as

$$\boldsymbol{\alpha}_i = \mathbf{C}\boldsymbol{\beta}_{:,m_i} + \begin{bmatrix}\boldsymbol{\theta}_{:,1} \\ \vdots \\ \boldsymbol{\theta}_{:,N_t}\end{bmatrix} + \begin{bmatrix}\mathbf{n}''_{i,:,1} \\ \vdots \\ \mathbf{n}''_{i,:,N_t}\end{bmatrix} \oslash \begin{bmatrix}|\mathbf{h}_{:,1}| \\ \vdots \\ |\mathbf{h}_{:,N_t}|\end{bmatrix}, \tag{16}$$

where $\oslash$ is the element-wise division operator. It is easy to verify that $\mathbf{C}$ is a matrix with rank $N_r + N_t - 1$ (full rank) [18], which enables the estimation of $\boldsymbol{\beta}_{:,m_i}$.

Considering the fact that the LLS algorithm requires i.i.d. noise on different observed data, if we assume (8) has perfect estimation on the channel amplitude, (16) can be modified by (8) as

$$\boldsymbol{\alpha}'_i = \mathbf{C}'\boldsymbol{\beta}_{:,m_i} + \begin{bmatrix}\left|\hat{\mathbf{h}}_{:,1}\right| \\ \vdots \\ \left|\hat{\mathbf{h}}_{:,N_t}\right|\end{bmatrix} \odot \begin{bmatrix}\boldsymbol{\theta}_{:,1} \\ \vdots \\ \boldsymbol{\theta}_{:,N_t}\end{bmatrix} + \begin{bmatrix}\mathbf{n}''_{i,:,1} \\ \vdots \\ \mathbf{n}''_{i,:,N_t}\end{bmatrix}, \tag{17}$$

where

$$\boldsymbol{\alpha}'_i = \left[\left|\hat{\mathbf{h}}_{:,1}^T\right|, \cdots, \left|\hat{\mathbf{h}}_{:,N_t}^T\right|\right]^T \odot \boldsymbol{\alpha}_i, \tag{18}$$

$$\mathbf{C}' = \underbrace{\begin{bmatrix}\left|\hat{\mathbf{h}}_{:,1}\right| & \cdots & \left|\hat{\mathbf{h}}_{:,1}\right| \\ \vdots & \cdots & \vdots \\ \left|\hat{\mathbf{h}}_{:,N_t}\right| & \cdots & \left|\hat{\mathbf{h}}_{:,N_t}\right|\end{bmatrix}}_{\text{repeat } N_r+N_t-1 \text{ times}} \odot \mathbf{C}. \tag{19}$$

Finally, the WLLS estimation is given as

$$\hat{\boldsymbol{\beta}}_{:,m_i} = (\mathbf{C}')^{+}\boldsymbol{\alpha}'_i. \tag{20}$$

*Remark 3:* By substituting (17) into (20), it can be seen that the estimation of (20) has a quasi-static phase bias of

$$(\mathbf{C}')^{+}\left(\begin{bmatrix} \left|\hat{\mathbf{h}}_{:,1}\right| \\ \vdots \\ \left|\hat{\mathbf{h}}_{:,N_t}\right| \end{bmatrix} \odot \begin{bmatrix} \boldsymbol{\theta}_{:,1} \\ \vdots \\ \boldsymbol{\theta}_{:,N_t} \end{bmatrix}\right). \tag{21}$$

However, it will be shown in Sec. III-E that the channel estimation algorithm can perfectly compensate for this bias, and result in a feasible estimation of $\boldsymbol{\Phi}_m\mathbf{H}\boldsymbol{\Psi}_m$.

### D. Step 3: Wiener Phase Estimation

In this step, the inter-pilot-group phase sequence, which can be denoted as $\left[\hat{\boldsymbol{\beta}}_{:,m_1},\cdots,\hat{\boldsymbol{\beta}}_{:,m_{N_c}}\right]$, is considered. And an element-wise inter-pilot-group Wiener phase estimator is proposed to further suppress the phase estimation error.

By substituting (17) into (20), the noise term of $\hat{\boldsymbol{\beta}}_{:,m_i}$ can be written as

$$\begin{aligned}\mathbf{n}_{\hat{\beta}(:,m_i)} =&(\mathbf{C}')^{+}\left[\left(\mathbf{n}''_{i,:,1}\right)^T,\cdots,\left(\mathbf{n}''_{i,:,N_t}\right)^T\right]^T \\ =&\begin{bmatrix} \boldsymbol{\varsigma}_{1,:} \\ \vdots \\ \boldsymbol{\varsigma}_{N_r+N_t-1,:} \end{bmatrix}\begin{bmatrix} \mathbf{n}''_{i,:,1} \\ \vdots \\ \mathbf{n}''_{i,:,N_t} \end{bmatrix},\end{aligned} \tag{22}$$

where $\boldsymbol{\varsigma}_{q,:}$ is the $q^{th}$ row of $(\mathbf{C}')^{+}$.

Note that $n''_{i,k,l} \sim \mathcal{N}\left(0,\sigma_n^2/(2N_t)\right)$ are i.i.d. AWGN variables, it is obvious that $n_{\hat{\beta}(q,m_i)} \sim \mathcal{N}\left(0,\sigma^2_{n_W(q)}\right)$, where

$$\sigma^2_{n_W(q)} = \frac{\left\|\boldsymbol{\varsigma}_{q,:}\right\|^2}{2N_t}\sigma_n^2, \tag{23}$$

and $\|\cdot\|$ is the Euclidean norm of a vector. On the other hand, as a direct result from (13), the equivalent phase noise variance between adjacent pilot groups (e.g. $\beta_{q,m_i}$ and $\beta_{q,m_{i+1}}$) can be given as

$$\sigma^2_{p_W(q)} = \begin{cases} L_c\cdot\left(\sigma^2_{\Delta\varphi}+\sigma^2_{\Delta\psi}\right) & ,(q\le N_r) \\ L_c\cdot 2\sigma^2_{\Delta\psi} & ,(q>N_r) \end{cases}, \tag{24}$$

where $\sigma^2_{\Delta\varphi}$ and $\sigma^2_{\Delta\psi}$ are the variances of the phase innovations in (3).

For the $q^{th}$ row of the phase sequence, a balanced two-sided Wiener phase estimator with tap length $L_{tap} = 2L_W+1$ can be given by [28]

$$\hat{\beta}'_{q,m_i} = \sum_{t=-L_W}^{L_W}\omega_{t,q}\hat{\beta}_{q,m_{i-t}}, \tag{25}$$

where $\boldsymbol{\omega}_{:,q} = [\omega_{-L_W,q},\cdots,\omega_{L_W,q}]^T$ are the corresponding Wiener coefficients. When $i\le t$, $\hat{\beta}_{q,m_{i-t}}$ should be set to 0 to enable the calculation of (25).

According to [15], the coefficient vector can be calculated by

$$\omega_{:,q} = \left(\mathbf{K}^{-1}\mathbf{1}_{(2L_W+1)\times 1}\right)\left(\mathbf{1}_{1\times(2L_W+1)}\mathbf{K}^{-1}\mathbf{1}_{(2L_W+1)\times 1}\right)^{-1}, \tag{26}$$

where the $(2L_W+1)\times(2L_W+1)$ matrix $\mathbf{K}$ is

$$\mathbf{K} = \mathbf{K}_p + \mathbf{K}_n, \tag{27}$$

the element at the $k^{th}$ row and $l^{th}$ column of $\mathbf{K}_p$ is

$$k_{p(k,l)} = \begin{cases} \sigma^2_{p_W(q)}\cdot\max(k,l), & (\max(k,l)\leqslant L_W) \\ \sigma^2_{p_W(q)}\cdot\min(k,l), & (\min(k,l)\geqslant L_W+2) \\ 0, & \text{otherwise} \end{cases}, \tag{28}$$

and

$$\mathbf{K}_n = \sigma^2_{n_W(q)}\cdot\mathbf{I}_{(2L_W+1)}. \tag{29}$$

The terms $\mathbf{K}_p$ and $\mathbf{K}_n$, which optimize the coefficient vector, arise from the autocorrelations of phase noise and AWGN, respectively.

### E. Step 4 and 5: Phase Recovery and Channel Estimation

Using the estimated phase given by (25), the full channel estimation (including phase) at the $i^{th}$ pilot group can be calculated as

$$\begin{aligned}\hat{\mathbf{H}}_i =&\frac{1}{N_t}\mathrm{diag}\left(e^{-j\hat{\beta}'_{1,m_i}},\cdots,e^{-j\hat{\beta}'_{N_r,m_i}}\right)\mathbf{Y}_i\mathbf{S}_i^H \\ &\times\mathrm{diag}\left(e^{-j\hat{\beta}'_{N_r+1,m_i}},\cdots,e^{-j\hat{\beta}'_{N_r+N_t-1,m_i}},1\right).\end{aligned} \tag{30}$$

Therefore, the channel estimation averaged over the whole frame can be given as

$$\hat{\mathbf{H}} = \frac{1}{N_c}\sum_{i=1}^{N_c}\hat{\mathbf{H}}_i. \tag{31}$$

Moreover, the phase at the $m^{th}$ ($m_i<m<m_{i+1}$) symbol period can be approximated by linear interpolation as

$$\hat{\beta}'_{q,m} = \hat{\beta}'_{q,m_i} + \frac{\hat{\beta}'_{q,m_{i+1}}-\hat{\beta}'_{q,m_i}}{m_{i+1}-m_i}(m-m_i). \tag{32}$$

Finally, the overall joint phase and channel estimation at the $m^{th}$ symbol period can be calculated as

$$\begin{aligned}\hat{\boldsymbol{\Phi}}_m\hat{\mathbf{H}}\hat{\boldsymbol{\Psi}}_m =&\mathrm{diag}\left(e^{j\hat{\beta}'_{1,m}},\cdots,e^{j\hat{\beta}'_{N_r,m}}\right)\hat{\mathbf{H}} \\ &\times\mathrm{diag}\left(e^{j\hat{\beta}'_{N_r+1,m}},\cdots,e^{j\hat{\beta}'_{N_r+N_t-1,m}},1\right).\end{aligned} \tag{33}$$

*Remark 4:* As discussed earlier, the phase estimation of (20) has a quasi-static bias of (21). This bias also exists (but in the negative direction) in the elements of the diagonal matrices in (30). Therefore, the positive and negative bias is cancelled when calculating (33), which guarantees a feasible estimation of $\hat{\boldsymbol{\Phi}}_m\hat{\mathbf{H}}\hat{\boldsymbol{\Psi}}_m$.

*Remark 5:* A numerical comparison of different estimation algorithms is given in Table I, while the analytical computational complexity is given in [26]. The weighing coefficient for multiplication is set at $C_M = 1$ for a fair and compatible comparison with existing research. The pilot rate is set at $R_p = 0.1$. The frame length is set at $L_f = 10^5$. For the sum-product algorithm maximum a posteriori (SPA-MAP) algorithm, we use the same parameters as [19, Table I] except $N_t$ and $N_r$ for a fair comparison. As shown in Table I, the proposed WLLS-Wiener algorithm has a much lower computational complexity when compared with the existing algorithms in [17]–[19].

TABLE I
COMPUTATIONAL COMPLEXITY OF DIFFERENT ALGORITHMS

| MIMO | $2 \times 2$ | $4 \times 4$ | $8 \times 8$ |
|---|---|---|---|
| $L_W = 5$ | 19.6 | 58.5 | 193.5 |
| $L_W = 50$ | 57.0 | 100.4 | 237.6 |
| EKF in [18] | 3.6e2 | 3.7e3 | 6.8e4 |
| EKF-EKS in [18] | 4.8e2 | 5.1e3 | 8.1e4 |
| EKF in [17] | 1.2e3 | 5.6e5 | 2.8e8 |
| SPA-MAP in [19] | 5.9e3 | 1.0e5 | 1.1e7 |
| Online MAP in [18] | 5.1e6 | 7.8e7 | 1.1e9 |
| Offline MAP in [18] | 1.0e8 | 1.5e9 | 2.2e10 |

## IV. SIMULATION RESULTS

In this section, the numerical performance of our newly proposed phase and channel estimation algorithm is evaluated against the CRLB [29]. The BER performance of the proposed algorithm is also evaluated. The transmitted data is generated by a random sequence generator in Matlab. It is assumed that $\sigma^2_{\Delta\varphi} = \sigma^2_{\Delta\psi} = \sigma^2_{\Delta}$, and $\sigma^2_n = 1/\text{SNR}$ throughout this section. Unless specified otherwise, we use a value of $\sigma^2_{\Delta} = 10^{-4}$, corresponding to a linewidth of ˜5.48 MHz for the experimental system studied in Sections V and VI, which is more than an order of magnitude broader than typical integrable tunable laser assemblies (ITLAs). The i.i.d. optical power distribution at the $l^{th}$ receiver from the $k^{th}$ transmitter is described by Gamma-Gamma turbulence model as [30, p. 510, (67)]

$$f_I\left(I_{k,l}\right) = \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} I_{k,l}^{\frac{\alpha+\beta}{2}-1} K_{\alpha+\beta}\left(2\sqrt{\alpha\beta I_{k,l}}\right), \quad (34)$$

where

$$\begin{cases} \alpha = \left\{\exp\left[0.49\sigma_R^2\left(1+1.11\sigma_R^{12/5}\right)^{-7/6}\right]-1\right\}^{-1} \\ \beta = \left\{\exp\left[0.51\sigma_R^2\left(1+0.69\sigma_R^{12/5}\right)^{-5/6}\right]-1\right\}^{-1} \end{cases}, \quad (35)$$

and $\sigma_R^2$ is the Rytov variance. To verify the performance of our algorithm in the worst channel condition, we set $\sigma_R^2 \to \infty$ to represent very very strong turbulence under the saturation regime in this section [30, p. 140, (15)], use the parameters in Table II unless otherwise specified, and discuss the main simulation results on estimation and BER performance. Further numerical details are given in [26] for interested readers.

The performance of channel matrix estimation is shown in Fig. 4. At the SNR of 10 dB, 20 dB, and 30 dB, the optimal mean squared error (MSE) of the LLS algorithm is $2.02 \times 10^{-4}$, $4.41 \times 10^{-5}$, $1.01 \times 10^{-5}$, respectively, while the MSE of the WLLS-Wiener algorithm is $3.14 \times 10^{-5}$, $3.08 \times 10^{-6}$, and $3.15 \times 10^{-7}$, respectively. In the conventional LLS algorithm, a training sequence is necessary to estimate the channel matrix, and the phase estimation is performed *after* the channel estimation. When the training sequence is too short, the AWGN significantly deteriorated the estimation accuracy. When the training sequence is too long, the intra-training-sequence phase noise significantly deteriorated the estimation accuracy. In our WLLS-Wiener algorithm, the phase estimation and recovery is performed *before* the channel matrix estimation, and the effect of phase noise deterioration on a long training sequence is cancelled out. Moreover, all the pilot groups within a frame can be utilized to obtain a much better estimation performance. As the channel is estimated from the pilots in the WLLS-Wiener algorithm, the training sequence for the channel estimation can be removed, significantly reducing the data overhead.

The phase estimation MSE performance for both WLLS one-shot estimator and WLLS-Wiener estimator are shown in Fig. 5. To eliminate the influence of the algorithm-introduced quasi-static phase bias given by Eq. (21), the phase bias was

TABLE II
SIMULATION PARAMETERS UNLESS OTHERWISE SPECIFIED

| Parameter | Value |
|---|---|
| Symbol format | QPSK |
| Type of MIMO decoder | MMSE |
| Number of transmit channels ($N_t$) | 2 |
| Number of receive channels ($N_r$) | 2 |
| Variance of phase noise ($\sigma^2_{\Delta}$) | $10^{-4}$ |
| Typical Baud rate and linewidth for $\sigma^2_{\Delta}$ | 6.28 GBaud @ 100 kHz<br>34.46 GBaud @ 5.48 MHz |
| Pilot rate ($R_p$) | $1/10$ |
| Length of training sequence ($L_{ts}$) | 0 |
| Wiener filter tap length ($L_{tap}$) | 101 |
| Number of independent Monte-Carlo trials | $\geq 10^5$ |
| Symbol length in each trial | $3 \times 10^3$ |
| Hard-decision forward error correction (HD-FEC) limit with 6.25% overhead [31] | $4.7 \times 10^{-3}$ |

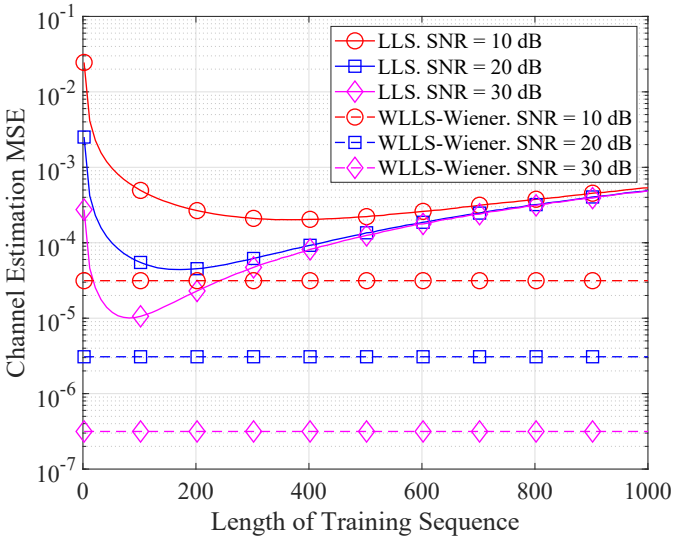


Fig. 4. MSE of channel estimation for a $2 \times 2$ MIMO system. A comparison between our proposed algorithm (dashed lines) and the conventional training-sequence-based approach with LLS estimators (lines).

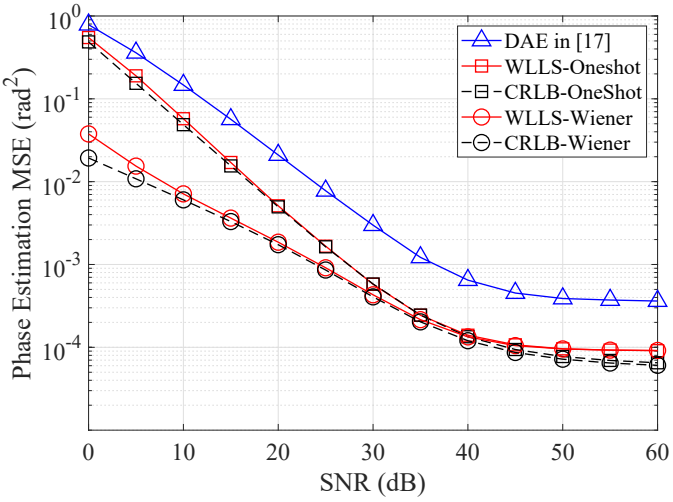


Fig. 5. MSE of phase noise estimation for a $2 \times 2$ MIMO system. Different phase estimators. Lines: MSE performance of DAE in [17] (triangles), our proposed algorithm without (squares) and with (circles) Wiener filtering; dashed lines: CRLBs without (squares) and with (circles) Wiener filtering.

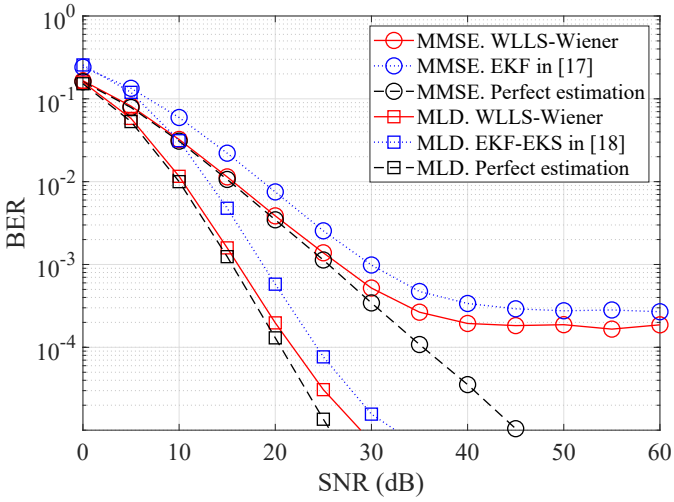


Fig. 6. BER of $2\times2$ MIMO systems with MMSE (circles) and MLD (squares) decoders (A comparison with the EKF/EKS-based algorithms in [17], [18]). Red lines: proposed algorithm; blue dotted lines: EKF/EKS-based algorithms; black dashed lines: perfect phase and channel estimation.

also added to the reference phase matrices, resulting in a fair and unbiased comparison. The corresponding CRLBs, which are derived in [26], are also depicted. The results indicate that the proposed WLLS one-shot estimator performs better than the conventional data-aided estimation (DAE) algorithm in [17]. This is mainly because the WLLS one-shot estimator extracts $N_r + N_t - 1$ phase information from $N_r N_t$ observed angular terms, and assigns a smaller weight to the observed data with larger errors, thereby reducing the phase estimation error (refer to (17)-(20) for details). Moreover, the performance of the WLLS one-shot estimator is very close to the CRLB for one-shot estimation. The slight performance degradation in the low SNR region for the WLLS one-shot estimator is mainly due to the small noise assumption. The slight performance degradation in the high SNR region is mainly due to neglecting the intra-pilot-group phase noise. On the other hand, introducing the Wiener estimator further improves the performance of the WLLS one-shot estimator in the low SNR region. This is because the Wiener filtering reduces the impact of AWGN. Moreover, the performance of the WLLS-Wiener estimator is also approaching the CRLB for Wiener Estimation. The slight performance degradation in the low SNR region for the WLLS-Wiener estimator is mainly due to the small noise assumption and the finite filter tap length. The slight performance degradation in the high SNR region is mainly due to neglecting the intra-pilot-group phase noise.

The BER performance of both MLD and MMSE decoder are shown in Fig. 6. A comparison with the extended-Kalman filter (EKF)/extended-Kalman smoother (EKS)-based algorithms in [17], [18] is also given here. Because of better phase and channel estimation accuracy, the proposed algorithm outperforms the conventional EKF/EKS-based algorithms. Moreover, a BER floor is observed when using the MMSE MIMO decoder, which is a direct consequence of the phase estimation MSE floor in Fig. 5. This problem can be suppressed by exploiting higher degree of freedom in MIMO systems (e.g.

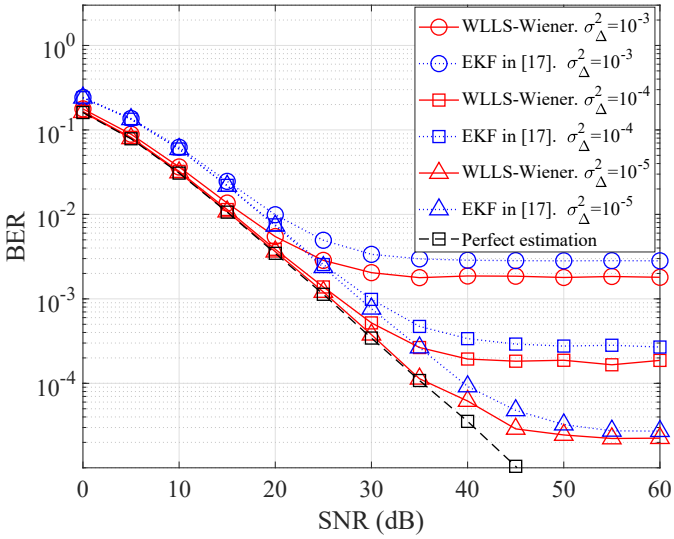


Fig. 7. BER of a $2 \times 2$ MIMO system. The normalised phase noise $\sigma^2_\Delta = 10^{-3}$ (circles), $10^{-4}$ (squares), $10^{-5}$ (triangles). Red lines: proposed algorithm; blue dotted lines: EKF in [17]; black dashed lines: perfect phase and channel estimation.

by applying MLD). When compared with the perfect phase and channel estimation scenario at the hard-decision forward error correction (HD-FEC) limit, the proposed algorithm has an SNR penalty of approximately 0.5 dB for both MLD and MMSE MIMO decoders, while the conventional algorithms has an SNR penalty of approximately 3.2 dB and 3.5 dB for MLD and MMSE MIMO decoders, respectively.

Fig. 7 compares the BER performance of the proposed phase and channel estimation algorithm for different phase noise variances, including $\sigma^2_\Delta = 10^{-3}$ (6.28 GBaud @ 1 MHz linewidth), $\sigma^2_\Delta = 10^{-4}$ (6.28 GBaud @ 100 kHz linewidth), and $\sigma^2_\Delta = 10^{-5}$ (62.8 GBaud @ 100 kHz linewidth). Similar to Fig. 6, the proposed algorithm consistently outperforms the conventional EKF-based algorithms because of better phase

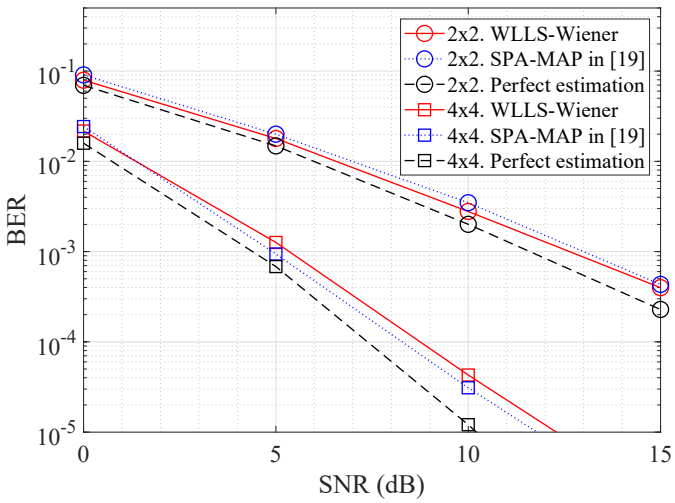


Fig. 8. BER of $2 \times 2$ and $4 \times 4$ MIMO systems with MLD decoding (A comparison with the SPA-MAP algorithm in [19]). $\sigma_{\Delta}^2 = 2.47 \times 10^{-4}$, $R_p = 1/20$, BPSK. Red lines: proposed algorithm; blue dotted lines: SPA-MAP algorithm in [19]; black dashed lines: perfect phase and channel estimation.

and channel estimation accuracy. It is noteworthy that, in the low SNR region, despite significant differences in phase variance, we observe very small BER differences in Fig. 7. This is attributed to the dominance of AWGN over phase estimation errors in this region. In the high SNR region, the BER floor is lower when the phase noise is smaller, establishing a direct correlation with the dominant phase estimation errors.

In Fig. 8, we compare the BER performance of our proposed algorithm with the SPA-MAP algorithm in [19]. To make a fair comparison, we used the same parameters as [19], including the phase noise ($\sigma_{\Delta}^2 = 2.47 \times 10^{-4}$), the pilot rate ($R_p = 1/20$), the training sequence length ($L_{ts} = 10$ for SPA-MAP), and the modulation format (binary phase-shift keying (BPSK)). In the $2 \times 2$ system, our proposed algorithm slightly outperforms the SPA-MAP algorithm, this is mainly because of the better channel estimation accuracy in our WLLS-Wiener algorithm (Fig. 4), which is also the case at the low SNR region in the $4 \times 4$ system. On the other hand, the SPA-MAP algorithm slightly outperforms our WLLS-Wiener algorithm at the high SNR region in the $4 \times 4$ system, which is mainly because of the lower BER will benefit the decision-feedback maximum a posteriori (MAP) phase estimation in SPA-MAP, while the linear interpolation in (32) will slightly degrade the phase estimation performance when $R_p < 1/10$. At the HD-FEC limit, the proposed algorithm has an SNR penalty of approximately 0.7 dB for both $2 \times 2$ and $4 \times 4$ systems, While the SPA-MAP algorithm has an SNR penalty of approximately 1.3 dB and 0.6 dB for $2 \times 2$ and $4 \times 4$ systems, respectively.

## V. Experimental Setup

The proof-of-concept experimental setup is depicted in Fig. 9. At the transmitter side, a 34.46 GBaud root-raised cosine (RRC) shaped (roll-off factor = 0.1) dual-polarization quadrature phase shift keying (DP-QPSK) signal was generated by a 39.385 GSa/s polarization multiplexing Ciena Transponder. In the transponder, the ITLA has a linewidth of <100 kHz. The signal had a frame structure of 20,000 symbols with 1,680 symbols in the training sequence, and 1 random pilot symbol for every 9 data symbols, which was generated from a $2^{15} - 1$ pseudo random binary sequence (PRBS) (Fig. 9(a)). After amplified by an erbium-doped fiber amplifier (EDFA), the symbols were passed through an acousto-optic modulator (AOM) to generate a 20 µs burst signal with a period of 160 µs (Fig. 9(b)) to emulate independent receivers with different phase noise. To emulate independent transmitters, the signal was split into 5 copies and delayed by variable fibre delay lines (FDLs) with lengths of 0, 280, 560, 840, and 1120 symbols, respectively (Fig. 9(c)). As 1120 symbols have a duration very much less than the coherence time of the laser, this corresponds to random phase offsets for each channel with strongly correlated phase evolutions. The decorrelated signals were then connected to 5 variable optical attenuators (VOAs) to compensate for the mode-dependent loss in the mode-selective photonic lanterns (MSPLs). Finally, the signals were coupled into a free-space turbulence emulator.

The turbulence emulator in Fig. 9 is detailed in Fig. 10. We first collimated the MDM beam from an FMF into free space using a transmit collimator (Thorlabs PAF2P-A10C) with a 10 mm focal length. To employ the polarization-sensitive spatial light modulator (SLM) for turbulence generation, the collimated beam was passed through a Thorlabs PBS054 polarizing beam splitter (PBS) and the polarization of the upper beam was rotated by 90° using a D-shaped half-wave plate (HWP) after a mirror (Thorlabs MRA05-P01). Due to the distributed nature of the strong turbulence, it can not be accurately described by one phase plate [32]. Therefore, by employing a square mirror (Thorlabs PFSQ05-03-P01), the polarized beams were reflected 4 times on the Santec SLM-200 $1920 \times 1080$ SLM, where 4 turbulence patterns were generated from the von Kármán spectrum, and mapped onto the SLM using the multi-plane light conversion (MPLC) method [33]. After rotating the lower beam by 90° using another D-shaped HWP and reflecting it by another mirror(Thorlabs MRA05-P01), the two beams were composed by another PBS (Thorlabs PBS054) and collected by a receive coupler (Thorlabs PAF2P-A10C) with a 10 mm focal length. By connecting the receive coupler to a single-mode fibre (SMF) and an FMF, respectively, the received power distribution was recorded and depicted in Fig. 10(c). Here the experimental results were well-fitted with the Gamma-Gamma turbulence model ($\sigma_R^2 = 9.01$ for SMF and $\sigma_R^2 = 1.45$ for FMF), indicating a successful emulation of the strong turbulence ($\sigma_R^2 > 1$).

At the receiver, another 6-mode MSPL was employed to decompose the MDM signals. To emulate independent receivers by employing a time-division multiplexing (TDM) receiver, the decomposed signals were passed through the FDLs with 5 km length increment between adjacent modes, generating a ˜24.5 µs delay, which was slightly longer than the 20 µs signal burst (Fig. 9(d)). As the delay line increment exceeds the coherence length of the lasers, this corresponds to the receivers having independent LOs, suitable to verify a phase asynchronized algorithm [10]. After adaptively amplified by 6 EDFAs, the signals were combined into one SMF

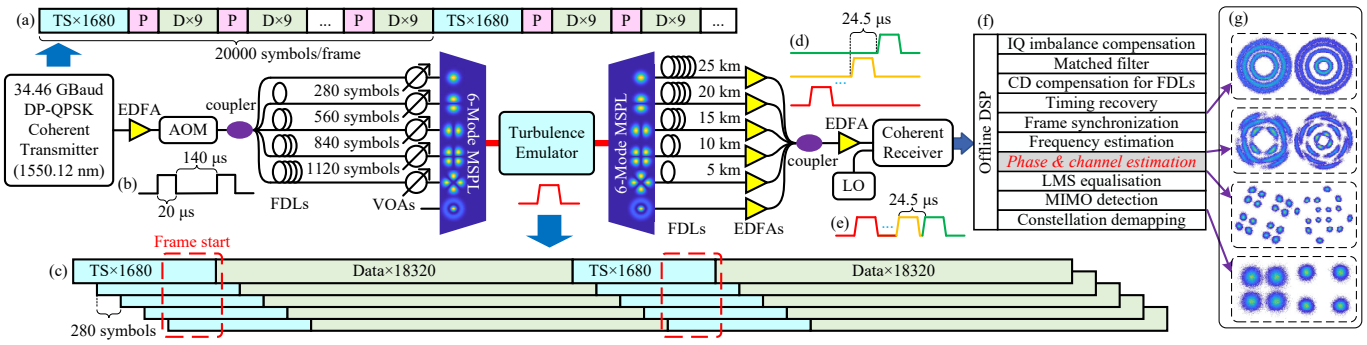


Fig. 9. The proof-of-concept experimental setup. DP-QPSK: dual-polarization quadrature phase shift keying; TS: training sequence; P: pilot; D: data; EDFA: erbium-doped fiber amplifier; AOM: acousto-optic modulator; FDLs: fibre delay lines; VOAs: variable optical attenuators; MSPL: mode-selective photonic lantern; LO: local oscillator; DSP: digital signal processing; CD: chromatic dispersion; LMS: least mean squares; MIMO: multiple-input multiple-output. (a) Frame structure at the coherent transmitter; (b) signal burst after AOM; (c) delayed signal structure after transmitter FDLs; (d) time-division multiplexing (TDM) signal after receiver FDLs; (e) TDM signal after receiver coupler; (f) offline DSP; (g) illustrative constellation of a 2×2 system.

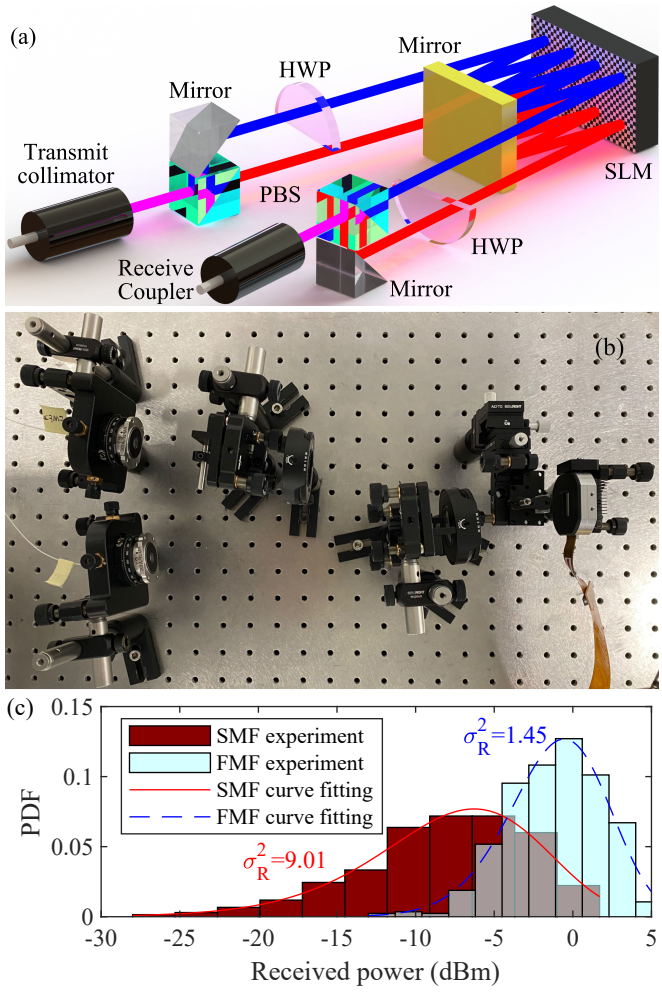


Fig. 10. The strong turbulence emulator. PBS: polarising beam splitter; HWP: half-wave plate; SLM: spatial light modulator; SMF: single-mode fibre; FMF: few-mode fibre; PDF: probability density function. (a) The schematic diagram; (b) the photo of the emulator; (c) the received power distribution.

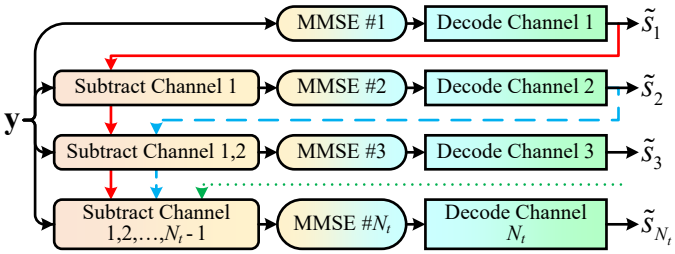


Fig. 11. The schematic diagram of the SIC decoder. MMSE: minimum mean square error decoder.

and amplified by another EDFA, generating the TDM signal for the coherent receiver. The LO for the coherent receiver has a linewidth of <100 kHz (Fig. 9(e)). The signals were decoded by an offline DSP (Fig. 9(f)), which employed our proposed phase and channel estimation algorithm. Throughout the experiment, we chose a practical tap length of 201 for the Wiener phase estimator. The estimation algorithm was decoupled from the least mean squares (LMS) ISI equalisation and the MIMO detection. As shown in Fig. 9(g), our decoupled DSP structure facilitates independent compensation for phase noise and inter-channel interference (ICI), thereby enabling the utilization of advanced MIMO decoders. This differs from the conventional 3D-EQ algorithm [6], wherein the phase estimation, channel estimation and MMSE MIMO decoding are intermingled within a single module, and advanced MIMO decoders can't be supported. It is also noteworthy that the LMS ISI equaliser was working with fractionally-spaced inputs at a rate of 2 samples per symbol, and a tap length of 21, resulting in a span of ±10 symbols, much smaller than the transmitter delay of 280 symbols. In our experiments, we chose the SIC MIMO decoder depicted in Fig. 11 [2], [13], which has a practical polynomial-level computational complexity and significant BER performance improvement through nonlinear ICI cancellation, to compare with the conventional MMSE decoder with linear ICI cancellation [11, p. 970]. Whilst in the SIC decoder the occurrence of error propagation events was minimised by employing the optimal decoding order [34], no further attempt was made to reduce the impact of the remaining events. To compensate for practical imperfections in the experimental system, compensation algorithms for IQ imbalance, pulse matching, chromatic dispersion (CD), timing, synchronization, and frequency distortions were also employed

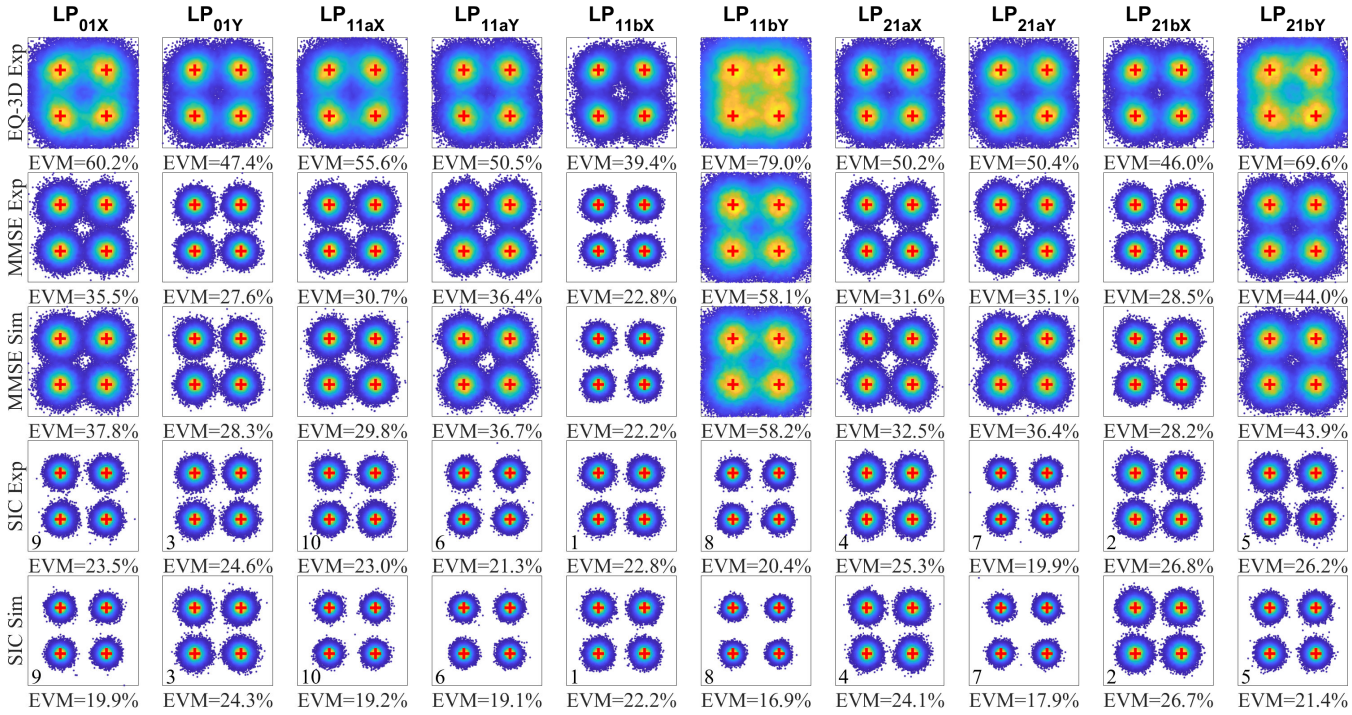


Fig. 12. The typical channel-wise constellation diagrams of different $10 \times 12$ MIMO system configurations. The SIC order is given at the bottom left of the corresponding constellation. LP: linearly polarized (modes in few-mode fibre); 3D-EQ: conventional 3-dimensional equaliser in [6]; MMSE: minimum mean square error (decoder); SIC: successive interference cancellation (decoder); Exp: experiment; Sim: reference simulation; EVM: error vector magnitude. Red crosses: reference constellation points.

before the proposed estimation algorithm.

## VI. Experimental results

Considering our absence of cognizance regarding the true phase and channel information in the experimental system, we are unable to calculate the MSE performance of our phase and channel estimation algorithm. However, we can assess the estimation performance by gauging BER performance degradation. To generate simulation signals for reference purposes without true phase and channel information, we utilised the phase and channel matrices estimated from the experiments using our proposed algorithm. The matrices were applied both to the channel and the receiver in the corresponding simulation, representing the perfect estimation scenario. By employing the first-order approximation, the discrepancy between the estimated experimental signals and reference simulation signals is proportional to the estimation error. Therefore, we can verify the effectiveness of the proposed phase and channel estimation algorithm by comparing the reference simulation BER and the experimental BER obtained by

1) 3D-EQ algorithm (conventional linear MMSE MIMO decoding coupled with phase and channel estimation) [6],
2) MMSE algorithm (linear MMSE MIMO decoding decoupled from phase and channel estimation) [11, p. 970],
3) SIC algorithm (nonlinear SIC MIMO decoding decoupled from phase and channel estimation) [2], [13].

Fig. 12 depicts the typical channel-wise experimental and reference simulation BER constellations of the $10 \times 12$ MIMO system. We first tested the conventional 3D-EQ based phase and channel estimation algorithm, which suffered from the mixture of phase and channel estimation. Here, a phase rotation is evident, which also degrades the channel estimation accuracy. Consequently, the 3D-EQ algorithm obtained a much worse error vector magnitude (EVM) performance when compared with the MMSE decoder using our proposed estimation algorithm, indicating the superior estimation performance of our proposed algorithm. Moreover, the experimental and reference simulation MMSE systems have similar constellations in Fig. 12, indicating a good phase and channel estimation accuracy. Due to the recursive interference cancellation nature of the SIC decoder (Fig. 11), the influence of practical imperfections (e.g. CD, ISI, phase noise, device nonlinearities, etc.) will accumulate during the cancellation process, leading to a higher discrepancy between the reference simulation and experimental results in the last several decoded channels (a significant discrepancy is observed from the $5^{th}$ decoded channel in Fig. 12). Despite the EVM degradation in the experimental results, the SIC channels consistently outperform the MMSE channels, indicating the necessity of employing our new phase and channel estimation algorithm to deploy advanced MIMO decoders.

Fig. 13 depicts the experimental and reference simulation BER performance of the $10 \times 12$ MIMO system under 100 independent strong turbulence patterns. Here we also compared the performance of the 3D-EQ, the MMSE and the SIC algorithms. In Fig. 13(a), the MMSE decoder obtained a significantly better BER when compared with the conventional 3D-EQ algorithm, indicating a better estimation performance of our proposed algorithm. Moreover, a notable consistency

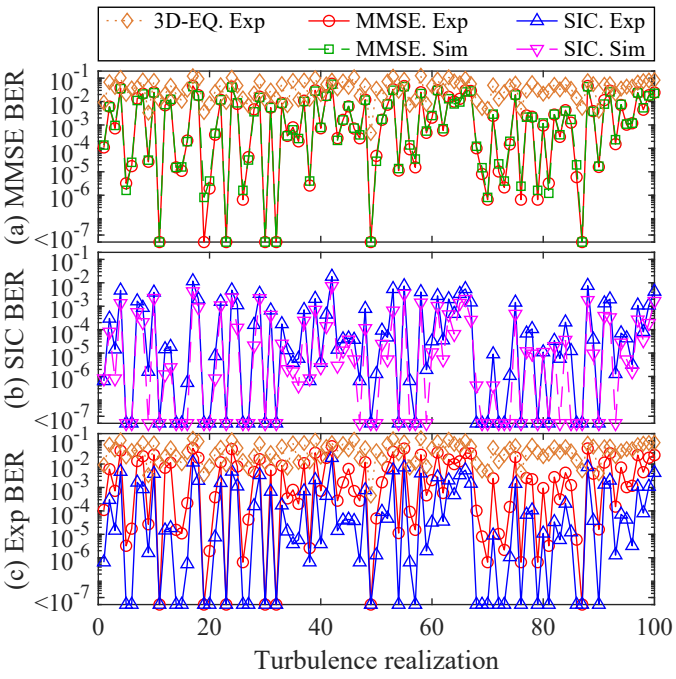


Fig. 13. The BER performance comparison of the $10 \times 12$ MIMO system under 100 independent strong turbulence patterns. 3D-EQ: conventional 3-dimensional equaliser in [6]; Exp: experiment; Sim: reference simulation. (a) MMSE decoder; (b) SIC decoder; (c) Experimental results.

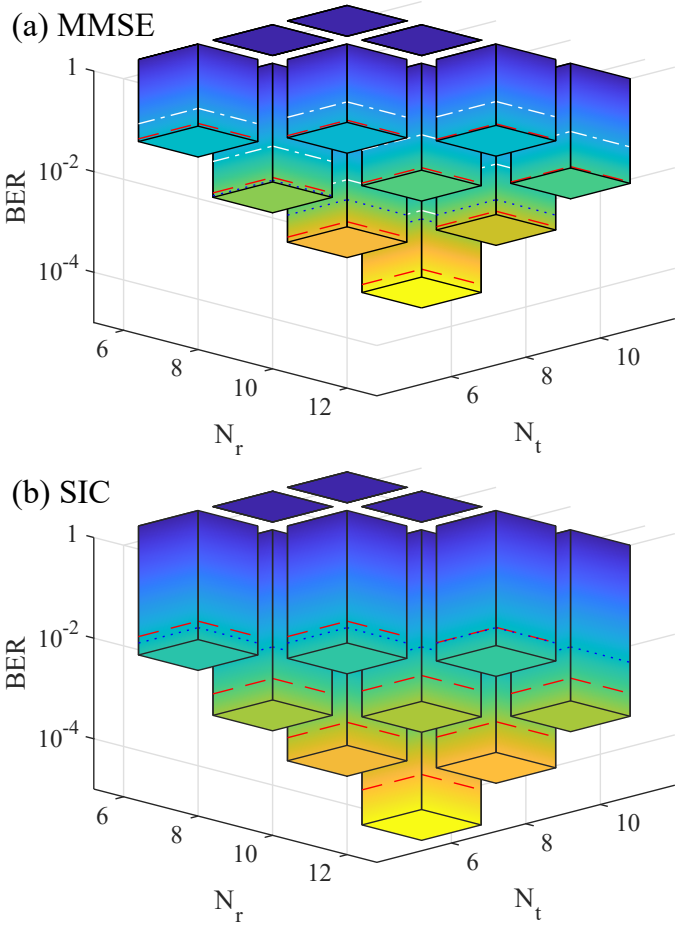


Fig. 14. A comparison between the experimental and the reference simulation BER performance of different MIMO configurations ($N_t = 6, 8, 10$, $N_r = 6, 8, 10, 12$, $N_t \leq N_r$). Average of 100 independent strong turbulence tests. Black lines: reference simulation results; red dashed lines: experimental results; white dash-dotted lines: experimental results with conventional 3-dimensional equalisers in [6]; blue dotted lines: HD-FEC limit of $4.7 \times 10^{-3}$.

between the reference simulation and experimental results is evident when employing the MMSE decoder, affirming the estimation accuracy of our proposed algorithm. In Fig. 13(b), the SIC decoder had a higher fragility to practical imperfections, leading to a higher discrepancy between the reference simulation and experimental results. This is mainly due to the recursive interference cancellation nature of the SIC algorithm. As shown in Fig. 13(c), our proposed algorithm can improve the system BER performance in different turbulence realisations by (1) improving the estimation accuracy (when comparing the 3D-EQ and the MMSE algorithms), and (2) supporting advanced MIMO decoders (when comparing the MMSE and the SIC algorithms).

Fig. 14 compares the average BER performance of different MIMO configurations ($N_t = 6, 8, 10$, $N_r = 6, 8, 10, 12$, $N_t \leq N_r$, lowest order modes connected) in 100 independent strong turbulence tests. The reference simulation BERs were depicted by black lines, while the corresponding experimental BERs were depicted by red dashed lines. We also depicted the result of the 3D-EQ algorithm by white dash-dotted lines, and the reference HD-FEC limit of $4.7 \times 10^{-3}$ by the blue dotted lines. Similar to Fig. 13, the limited estimation accuracy of the conventional 3D-EQ algorithm leads to a significant BER degradation, while the MMSE decoder employing our proposed algorithm quantifies a good consistency with the reference simulation results. Due to the limited signal length, we observed a higher discrepancy at lower average BER. A higher performance degradation was also observed when employing the SIC decoder due to its higher fragility to practical system imperfections. Notwithstanding the higher performance degradation, the SIC decoder consistently outperformed the MMSE decoder. As a result, we obtained a lower BER than the HD-FEC limit in a $8 \times 12$ MMSE system (551.4 Gbit/s/wavelength) and a $10 \times 12$ SIC system (689.2 Gbit/s/wavelength), respectively. By (1) improving the estimation accuracy (when comparing the 3D-EQ and the MMSE algorithms), and (2) supporting advanced MIMO decoders (when comparing the MMSE and the SIC algorithms), our proposed estimation algorithm can improve the system BER performance in MDM systems with different numbers of transmit and receive channels.

## VII. Conclusion

We proposed a novel pilot-aided phase and channel estimator in this paper. The proposed estimator has low computational complexity and CRLB-approaching estimation accuracy. It can also support phase-asynchronous laser sources and advanced MIMO decoders, significantly reducing the stringent laser requirements and improving the BER performance of MDM MIMO FSO communications in turbulent channels.

A proof-of-concept experiment was carried out to verify the feasibility of the proposed phase and channel estimation

algorithm. A notable consistency is observed when applying the MMSE decoder, affirming the estimation accuracy of our proposed algorithm. Notwithstanding the BER degradation from practical system imperfections, the advanced SIC decoder consistently outperformed the conventional MMSE decoder. By exploiting 10 decorrelated channels in the SIC-based MDM system, we achieved a record-high line rate of 689.2 Gbit/s with a BER lower than the HD-FEC limit in strong turbulence, indicating the necessity of employing our proposed algorithm for advanced MIMO decoders.

Although this algorithm is developed for MDM FSO systems, it can also benefit MDM fibre systems, multicore fibre systems, and radio-frequency MIMO systems. Considering the estimation performance, computational complexity, and adaptability of the newly proposed algorithm, it will provide useful guidelines for designing MIMO systems, and will facilitate deploying practical commercial systems in the future.

## ACKNOWLEDGMENT

We thank Dr David Benton, Dr Chao Gao and Dr Long Jian for fruitful discussions. We also thank Dr Zhouyi Hu for meticulous proofreading. Finally, we thank Ciena and Dr. Charles Laperle for kindly providing the WaveLogic 3 transponder used in our experiments.

## REFERENCES

[1] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE communications surveys & tutorials*, vol. 16, no. 4, pp. 2231–2258, 2014.

[2] Y. Li, Z. Chen, Z. Hu, D. M. Benton, A. A. Ali, M. Patel, M. P. Lavery, and A. D. Ellis, "Enhanced atmospheric turbulence resiliency with successive interference cancellation DSP in mode division multiplexing free-space optical links," *Journal of Lightwave Technology*, 2022.

[3] N. Cvijetic, D. Qian, J. Yu, Y.-K. Huang, and T. Wang, "Polarization-multiplexed optical wireless transmission with coherent detection," *Journal of Lightwave Technology*, vol. 28, no. 8, pp. 1218–1227, 2010.

[4] J. Poliak, R. M. Calvo, and F. Rein, "Demonstration of 1.72 tbit/s optical data transmission under worst-case turbulence conditions for ground-to-geostationary satellite communications," *IEEE Communications Letters*, vol. 22, no. 9, pp. 1818–1821, 2018.

[5] N. K. Fontaine, R. Ryf, Y. Zhang, J. C. Alvarado-Zacarias, S. van der Heide, M. Mazur, H. Huang, H. Chen, R. Amezcua-Correa, G. Li *et al.*, "Digital turbulence compensation of free space optical link with multimode optical amplifier," in *45th European Conference on Optical Communication (ECOC 2019)*. IET, 2019, pp. 1–4.

[6] H. Huang, Y. Cao, G. Xie, Y. Ren, Y. Yan, C. Bao, N. Ahmed, M. A. Neifeld, S. J. Dolinar, and A. E. Willner, "Crosstalk mitigation in a free-space orbital angular momentum multiplexed communication link using 4×4 MIMO equalization," *Optics letters*, vol. 39, no. 15, pp. 4360–4363, 2014.

[7] Y. Ren, Z. Wang, G. Xie, L. Li, A. J. Willner, Y. Cao, Z. Zhao, Y. Yan, N. Ahmed, N. Ashrafi *et al.*, "Atmospheric turbulence mitigation in an OAM-based MIMO free-space optical link using spatial diversity combined with MIMO equalization," *Optics letters*, vol. 41, no. 11, pp. 2406–2409, 2016.

[8] R. G. Van Uden, C. M. Okonkwo, H. Chen, H. de Waardt, and A. M. Koonen, "Time domain multiplexed spatial division multiplexing receiver," *Optics express*, vol. 22, no. 10, pp. 12 668–12 677, 2014.

[9] G. Rademacher, B. J. Puttnam, R. S. Luís, T. A. Eriksson, N. K. Fontaine, M. Mazur, H. Chen, R. Ryf, D. T. Neilson, P. Sillard *et al.*, "Peta-bit-per-second optical communications system using a standard cladding diameter 15-mode fiber," *Nature Communications*, vol. 12, no. 1, p. 4238, 2021.

[10] K. Shibahara, T. Mizuno, and Y. Miyamoto, "MIMO carrier phase recovery for carrier-asynchronous SDM-MIMO reception based on the extended Kalman filter," *Optics Express*, vol. 29, no. 11, pp. 17 111–17 124, 2021.

[11] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[12] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639–1642, 1999.

[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[14] R. Slavík, Z. Liu, and D. J. Richardson, "Optical injection locking for carrier phase recovery and regeneration," in *Optical Fiber Communication Conference*. Optical Society of America, 2017, pp. Th4I–3.

[15] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *Journal of Lightwave Technology*, vol. 25, no. 9, pp. 2675–2692, 2007.

[16] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for $M$-QAM constellations," *Journal of Lightwave Technology*, vol. 27, no. 8, pp. 989–999, 2009.

[17] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4790–4807, 2012.

[18] A. A. Nasir, H. Mehrpouyan, R. Schober, and Y. Hua, "Phase noise in MIMO systems: Bayesian Cramér–Rao bounds and soft-input estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2675–2692, 2013.

[19] R. Krishnan, G. Colavolpe, A. G. i Amat, and T. Eriksson, "Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3360–3375, 2015.

[20] Y. Li, Z. Chen, D. M. Benton, M. Patel, M. P. J. Lavery, and A. D. Ellis, "689 Gbps single-wavelength mode-division multiplexing free-space optical transmission in strong turbulence," in *European Conference on Optical Communication (ECOC)*, 2023, p. We.A.2.2.

[21] A. F. Alfredsson, E. Agrell, and H. Wymeersch, "Iterative detection and phase-noise compensation for coded multichannel optical transmission," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5532–5543, 2019.

[22] L. Zhao and W. Namgoong, "A novel phase-noise compensation scheme for communication receivers," *IEEE Transactions on Communications*, vol. 54, no. 3, pp. 532–542, 2006.

[23] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613–1621, 1997.

[24] V. Simon, A. Senst, M. Speth, and H. Meyr, "Phase noise estimation via adapted interpolation," in *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, vol. 6. IEEE, 2001, pp. 3297–3301.

[25] N. Hadaschik, M. Dorpinghaus, A. Senst, O. Harmjanz, U. Kaufer, G. Ascheid, and H. Meyr, "Improving MIMO phase noise estimation by exploiting spatial correlations," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3. IEEE, 2005, pp. iii–833.

[26] Y. Li, "Supplementary materials for phase and channel estimation," 2024, https://doi.org/10.17036/researchdata.aston.ac.uk.00000626.

[27] H. Meyr, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. John Wiley & Sons, 1998.

[28] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms, and Applications*, 4th ed. Prentice Hall: Pearson Education, 2007.

[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.

[30] L. C. Andrews, *Laser Beam Propagation through Random Media*, 2nd ed. Bellingham, WA: SPIE Press Monograph, 2005.

[31] L. M. Zhang and F. R. Kschischang, "Staircase codes with 6% to 33% overhead," *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 1999–2002, 2014.

[32] M. P. Lavery, "Vortex instability in turbulent free-space propagation," *New Journal of Physics*, vol. 20, no. 4, p. 043023, 2018.

[33] Y. Li, Z. Chen, D. M. Benton, M. Patel, M. P. Lavery, and A. D. Ellis, "Single-wavelength polarization-and mode-division multiplexing free-space optical communication at 689 Gbit/s in strong turbulent channels," *Optics Letters*, vol. 48, no. 13, pp. 3575–3578, 2023.

[34] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1722–1730, 2003.