optics

MDPI

Article

Kalman Filter-Aided Bayesian Learning for Sparse Multipath Channel Estimation in ACO/DCO-OFDM Visible Light Communication Systems

Anurag Bajpai *, Ajay Khunteta and Janki Ballabh Sharma

Department of Electronics Engineering, Rajasthan Technical University, Kota 324001, Rajasthan, India; akhunteta@rtu.ac.in (A.K.); jbsharma@rtu.ac.in (J.B.S.)

* Correspondence: anuragait2010@gmail.com

Abstract

This paper proposes a novel Kalman Filter-Aided Bayesian Learning (Kalman-BL) channel estimator for sparse multipath channel impulse response (CIR) estimation in intensity-modulation and direct-detection (IM/DD)-based optical orthogonal frequency division multiplexing (O-OFDM) visible light communication (VLC) systems. The proposed estimator models the VLC channel as a Gauss–Markov random process with temporal correlation coefficient α and integrates a Kalman predictor-updater stage with a warm-started Bayesian Learning (BL) sparse refiner. The Kalman filter draws on the correlation between successive frames to lower the effective observation noise; the BL stage that follows then exploits the delay-domain sparsity that is inherent to the VLC channel impulse response. Together, this two-stage scheme outperforms five competing estimators—least squares (LS), linear minimum mean square error (LMMSE), orthogonal matching pursuit (OMP), FOCUSS, and single-snapshot BL—across NMSE, BER, outage probability (OP), and pilot overhead, with the improvement holding for both ACO-OFDM and DCO-OFDM. Benchmarking against the Bayesian Cramér–Rao Lower Bound (BCRLB) and the Oracle-MMSE bound confirms that performance sits close to optimal. While ensuring filter consistency, the prediction process is validated with a dedicated NMSE-versus-α study performed on real trajectories of the Gauss–Markov channels of indoor VLC scenarios to demonstrate an average gain of around 1–1.5 dB compared to standalone BL, for $\alpha \in [0.90, 0.99]$ typical for quasi-static indoor VLC channels. Most of the advantages are the reduction of pilot overhead 30–40% and about half the number of EM iterations required for convergence versus standalone BL.

Keywords: visible light communication(VLC); Kalman filter; Bayesian learning (BL); sparse channel estimation; ACO-OFDM; DCO-OFDM; Gauss–Markov channel; EM algorithm; BCRLB; pilot overhead; filter consistency

Check for updates

Academic Editors: Xizheng Ke and Feng Peng

Received: 25 August 2026
Revised: 5 October 2026
Accepted: 8 October 2026
Published: 15 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The proliferation of indoor wireless broadband services has intensified the demand for spectrum-efficient and high-throughput communication technologies. Visible light communication (VLC) leverages the unregulated 400–800 THz optical spectrum offered by light-emitting diodes (LEDs) to provide simultaneous illumination and gigabit-class wireless data transmission [1]. Unlike radio-frequency (RF) systems, VLC is inherently secure, immune to electromagnetic interference, energy-efficient, and free of spectrum licensing constraints, rendering it highly attractive for next-generation 6G indoor access [2–4].

Indoor VLC channels see multipath from both line-of-sight (LoS) and non-line-of-sight (NLoS) paths, the latter caused by reflections off walls, ceilings, and furniture, a phenomenon that is well documented for indoor wireless optical links in general [5,6]. This leads to delay spread and, in turn, inter-symbol interference (ISI), which is why optical OFDM (O-OFDM) is used. Two O-OFDM variants are standard for intensity-modulation/direct-detection (IM/DD) systems: DCO-OFDM, which adds a DC bias to keep the signal unipolar, and ACO-OFDM, which achieves the same through zero-clipping—this makes it more energy-efficient, though at the cost of halving spectral efficiency [7,8].

Accurate channel estimation (CE) is a prerequisite for single-tap frequency-domain equalization in O-OFDM VLC systems. Because the multipath VLC channel is sparse in the delay domain—only a small fraction of the L_h taps actually carry meaningful energy—it lends itself well to compressed sensing-based CE [9,10]. Standard approaches like LS and LMMSE, however, do not make use of this sparsity [11]. Greedy techniques such as OMP tend to be sensitive to their stopping criterion [12], and FOCUSS is prone to getting stuck at local minima [13]. Single-snapshot Bayesian Learning (BL) offers a more principled, probabilistic route via expectation-maximization (EM) [14,15], and has since been applied successfully to sparse VLC and mmWave channel estimation [16,17], but it treats each pilot block in isolation and so ignores the temporal correlation that exists across successive OFDM frames. The developments of the BL-based CE for VLC have been rapid in the past few years. Saxena et al. used BL for sparse CIR recovery in DCO- and ACO-OFDM [16], and extended the same to ADO-OFDM [18]. They then switched to the multiple measurement vector (MMV) setting, which processes multiple pilot blocks having a common sparsity support, in the first case for DCO-OFDM [19] and in the latter case, most recently, for the simultaneous estimation of sparsity support in time and delay (ADO-OFDM) [20]. The research conducted in this area is similar but largely focuses on the source of the extra information. MMV schemes collect a number of observations over a single channel and depend on the support being shared across all of these observations. The estimator proposed here, however, assumes a first order Gauss–Markov prior for the change of the channel between frames and forwards the state estimate as well as its covariance forward across frames. The two methods are thus complementary. The Kalman stage connects coherence intervals to each other, while the MMV stage adds more measurements within a coherence interval. It is also worth mentioning the work presented in [21], which also introduces sparse Bayesian learning and Kalman filtering. In the latter case, however, the combination is used for nonlinear post-detection equalisation and not for estimation of the CIR, and there is no propagation of the error covariance between frames.

The model-based estimator is replaced by a learned estimator in a parallel line of work. Convolutional and model-driven networks have been reported for optical wireless CE [22], such as BDCNP architecture of massive MIMO VLC-positioning systems. The joint estimation of channel and location is a possibility, as well as deep unfolding designs. The sparse-recovery iteration has been extended to a trainable network. These achieve good accuracy at these reach points and a sufficient amount of training data are available for the site. The classical route proposed here is suggested for use when it is suitable, namely, where there is no training requirement, there is an explicit consistency test, and it has a fixed cost per frame as described in Section 5.8.

This paper proposes a Kalman Filter-Aided Bayesian Learning (Kalman-BL) estimator that tries to capture both delay-domain sparsity and temporal correlation at once. A Gauss–Markov state-space model is used to describe how the sparse VLC channel impulse response evolves over time, which lets a Kalman predictor-updater stage reduce the effective observation noise before BL is warm-started from that output. The main contributions of this work are as follows:

1. A Gauss–Markov sparse CIR model for wideband VLC incorporating both diffuse and specular reflections together with detailed optical source, reflector, and receiver characterizations.
2. A novel Kalman-BL estimator integrating temporal prediction via a Kalman filter with spatial sparsity exploitation via warm-started BL, valid for both DCO-OFDM and ACO-OFDM.
3. An explicit prediction-error analysis. The filter calculates the innovation at each frame, as well as innovation's covariance and normalised innovation squared. This provides a consistency test on the filter and also a means of testing the temporal model after deployment to ensure its continued validity.
4. Derivation of the BCRLB and Oracle-MMSE benchmarks confirming near-optimal Kalman-BL performance.
5. A detailed investigation of NMSE-versus-α on real channel trajectories of the Gauss–Markov process, and a physical derivation of α from the mobility of the receiver and luminaire geometry.
6. Comprehensive Monte Carlo validation covering NMSE, BER, OP, pilot overhead, convergence, user-location robustness, multi-QAM, and a second simulation environment, including a worked example for an indoor office deployment.

2. Optical OFDM System Models

2.1. DCO-OFDM System Model

A serial bit stream is mapped to N-complex-valued QAM symbols. Hermitian symmetry is imposed on the frequency-domain (FD) vector $\bar{\mathbf{x}} = [\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{N-1}]^T \in \mathbb{C}^{N\times 1}$, with $\tilde{x}_0 = \tilde{x}_{\frac{N}{2}} = 0$ and $\tilde{x}_j = \tilde{x}^*_{N-j}$ for $\frac{N}{2} + 1 \leq j \leq N-1$, ensuring a real-valued time-domain (TD) signal after the N-point IFFT [8]:

$$x_n = \frac{2}{N} \sum_{k=1}^{N/2-1} \mathrm{Re}\left\{ \tilde{x}_k e^{j\frac{2\pi kn}{N}} \right\} \tag{1}$$

Note on the range of values of (1). Note that the sum (1) in $k = 1$, not $k = 0$. The Hermitian symmetry makes it possible to write $\tilde{x}_0 = \tilde{x}_{N/2} = 0$, so the $k = 0$ and $k = N/2$; adding these terms would only be adding zeros.

A DC bias $B_{DC} = \nu\sqrt{E\{x^2(t)\}}$ is added to the signal make it unipolar, with a dimensionless bias index of ν, defined as the ratio between the added DC level and the RMS value of the bipolar TD signal $x(t)$. The bias is conventionally expressed in dB as $10\log_{10}(\nu^2+1)$ and, regarding this paper, 7 dB is equivalent to $\nu = 1.90$.

This level has been selected for a reason. A larger ν will decrease bottom clipping but will also lose optical power since it will contain the DC level, but not any information. As ν increases, bottom clipping will decrease while optical power is lost since there will be no information in the DC level and it will be the required level to drive the LED. As the value of ν decreases, less power is being used, but a larger percentage of the Gaussian distributed $x(t)$ would be below zero and clipped. The total distortion is minimal around 7 dB for the 4-QAM DCO-OFDM signal of the simulation setup (Section 6.1), with approximately 2.3% of the TD samples being clipped. A sweep of the bias level from 4 to 12 dB, reported in Section 6.15, confirms this minimum at SNR = 15 dB, with the BER increasing on either side of the minimum. The selected value is also similar to the range of 6–8 dB that is often cited for DCO-OFDM [7,8].

After CP (cyclic prefix) removal and N-point FFT at the receiver, the FD input–output relation at subcarrier k is:

$$\tilde{y}_k = \tilde{h}_k \tilde{x}_k + \tilde{w}_k, \quad k = 1, \ldots, N-1, \tag{2}$$

where $\tilde{h}_k = \sum_{n=0}^{L_h-1} h(n) e^{-j\frac{2\pi nk}{N}}$ is the channel transfer function and $\tilde{w}_k \sim \mathcal{CN}(0, \sigma^2_{\text{AWGN}})$.

2.2. ACO-OFDM System Model

Data symbols are assigned exclusively to odd subcarriers with Hermitian symmetry. The antisymmetric TD signal $x_n = -x_{n+N/2}$ permits zero-clipping without information loss:

$$x_{n,cs} = \begin{cases} x_n & \text{if } x_n > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Clipping halves the FD odd-subcarrier symbol magnitudes: $\tilde{x}_{2m+1,cs} = \frac{1}{2}\tilde{x}_{2m+1}$. The receiver compensates this factor of 2 after demodulation, which also doubles the noise amplitude. Thus, the effective noise covariance on the ACO branch is $4\sigma^2\mathbf{I}$. ACO-OFDM thus carries $\frac{N}{4}$ information symbols per block, achieving superior energy efficiency over DCO-OFDM at the cost of 50% spectral efficiency [3,7].

3. Multipath VLC Channel Model

Optical Source and Propagation

LEDs radiate according to the generalized Lambertian pattern with mode number l [23]:

$$R_L(\phi) = \left(\frac{l+1}{2\pi}\right) \cdot P_S \cdot \cos^l(\phi), \quad \phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \tag{4}$$

The concentrator gain at the photodetector (refractive index μ, FoV Ψ_{FoV}) is

$$g(\Psi) = \frac{\mu^2}{\sin^2(\Psi_{\text{FoV}})} \quad \text{for } \Psi \le \Psi_{\text{FoV}} \tag{5}$$

The LoS impulse response between source S_a and receiver R_b (separation $d_{a,b}$) is [5]:

$$\begin{aligned} h^{(0)}(t; S_a, R_b) = &\left[\frac{l_a+1}{2\pi d^2_{a,b}}\right] \cdot A_{R_b} \cdot \cos^{l_a}(\phi_{a,b}) \\ &\cdot \cos(\Psi_{a,b}) \cdot g(\Psi_{a,b}) \cdot \delta\left(t - \frac{d_{a,b}}{c}\right) \end{aligned} \tag{6}$$

For a two-bounce NLoS path via reflectors ε_1 and ε_2 with reflection coefficients ρ_1, ρ_2 [6]:

$$\begin{aligned} h^{(2)}(t; S_a, \varepsilon_1, \varepsilon_2, R_b) = &h^{(0)}(t; S_a, \varepsilon_1) \otimes \rho_1 h^{(0)}(t; \varepsilon_1, \varepsilon_2) \\ &\otimes \rho_2 h^{(0)}(t; \varepsilon_2, R_b) \end{aligned} \tag{7}$$

The discrete-time sparse CIR is $h(n) = \sum_{l=0}^{L_h-1} h(l)\,\delta(n - \nu_l)$, supported on s dominant arrivals. The average power of the j-th arrival, $j = 0, \ldots, s-1$, is exponentially decaying and the CIR is normalised to unity energy.

$$\bar{P}(j) = \frac{\exp(-j\,t_s/\tau)}{\sum_{i=0}^{s-1}\exp(-i\,t_s/\tau)}, \qquad h(\nu_j) = \sqrt{\bar{P}(j)}\,\xi_j, \quad \xi_j \sim \mathcal{N}(0,1), \tag{8}$$

The power-delay profile (PDP) is listed by arrival order, and not by absolute tap index, for $l \gtrsim 12$, $\exp(-l\,t_s/\tau)$ is very small at the absolute index (approx. 10^{-18} for $l = 30$ with $\tau = 0.75$). Such a nominally $s = 6$ sparse channel would then automatically degrade to a few effective taps and the recovery problem become degenerate. By indexing the taps according to the arrival order, all s are kept and all s still have the desired exponential decay, which in this case spans approximately 29 dB. The indoor VLC channel additionally

exhibits temporal correlation across successive OFDM frames, motivating the state-space model of Section 5.

4. Sparse Channel Estimation Framework

4.1. Pilot Observation Model

Comb-type pilots occupy $I_{\text{pilot}} = \{I_0 + k \cdot I_P : k = 0, \ldots, M_P/2 - 1\} \cup \{N - p : p \in I_{\text{pilot}}\}$. The pilot observation vector is:

$$\mathbf{y}_P = \text{diag}\{\mathbf{x}_P\} \cdot \mathbf{E} \cdot \mathbf{Q} \cdot \mathbf{h} + \mathbf{w}_P = \mathbf{B} \cdot \mathbf{h} + \mathbf{w}_P \tag{9}$$

where $\mathbf{E} \in \mathbb{R}^{M_P \times N}$ is the pilot extraction matrix, $\mathbf{Q} \in \mathbb{C}^{N \times L_h}$ the truncated DFT matrix, $\mathbf{B} = \text{diag}\{\mathbf{x}_P\}\mathbf{EQ} \in \mathbb{C}^{M_P \times L_h}$ the measurement matrix, and $\mathbf{w}_P \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{M_P})$.

Pilot Placement and Dictionary Coherence

This means that the position of the pilot comb is more important than one would think as far as how any sparse estimator is going to work out here. The spacing I_P, between tap delays, is represented by a uniform comb with spacing I_P. In this case, $N = 1024$, $I_P = 40$, and this period is 25.6 taps, which is less than $L_h = 40$; in $\mathbf{B}$ this means that the taps at l and $l + 26$ are nearly indistinguishable from each other. The measured mutual coherence $\mu(\mathbf{B}) = \max_{i \neq j} |\mathbf{b}_i^H \mathbf{b}_j| / (\|\mathbf{b}_i\| \|\mathbf{b}_j\|)$ is then 0.92, and all sparse estimators converge to an NMSE of about 5×10^{-2} at all SNRs.

If the uniform comb is replaced by a pseudo-random pilot set of the same size, $M_P = 16$, the coherence drops to 0.59 and the NMSE falls by about two orders of magnitude. Once (off-line), the pilot set is selected from 64 random draws, and after this time remains fixed. In practice it is, therefore, no more difficult to implement than a uniform comb. Multiplying $\mathbf{B}$ by $\text{diag}\{\mathbf{x}_P\}$ with unit-modulus pilot symbols doesn't help either, as the inner products of the columns are not affected by scaling by unit-modulus values, and so is the coherence. The pilot positions are important. All results in Section 6 are based on pseudo-random placement and its impact on outage probability is discussed in Section 6.8.

4.2. LS and LMMSE Baselines

The LS estimate minimizes $\|\mathbf{y}_P - \mathbf{Bh}\|^2$ [24]:

$$\hat{\mathbf{h}}_{\text{LS}} = (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \mathbf{y}_P \tag{10}$$

The LMMSE estimate is [11,24]:

$$\hat{\mathbf{h}}_{\text{LMMSE}} = \left(\mathbf{B}^H \mathbf{R}_w^{-1} \mathbf{B} + \mathbf{R}_{hh}^{-1}\right)^{-1} \mathbf{B}^H \mathbf{R}_w^{-1} \mathbf{y}_P \tag{11}$$

where $\mathbf{R}_{hh} = \mathbf{I}_{L_h}$ and $\mathbf{R}_w = \sigma^2 \mathbf{I}_{M_P}$.

4.3. OMP-Based Sparse Recovery

OMP iteratively identifies the measurement-matrix column best aligned with the current residual [12]:

$$J = \arg \max_{k=1,\ldots,L_h} \left| \mathbf{B}(:,k)^H \cdot \mathbf{r}_{t-1} \right| \tag{12}$$

$$\hat{\mathbf{h}}_t = \mathbf{B}_{\text{Y}}^{\dagger} \cdot \mathbf{y}_P, \quad \mathbf{r}_t = \mathbf{y}_P - \mathbf{B}_{\text{Y}} \hat{\mathbf{h}}_t \tag{13}$$

The iterations stop once the residual energy reaches the noise floor, $\|\mathbf{r}_t\|_2^2 \leq \kappa M_P \sigma^2$ with $\kappa = 1.2$. This is a noise scaled threshold (ε_{OMP}) which is a change from the original submission's fixed value of $\varepsilon_{\text{OMP}} = 0.1$. A fixed threshold is SNR independent, meaning

that it continues to add atoms well past the true sparsity at low SNR. It had an average of 14.2 atoms selected for an $s = 6$ channel at 0 dB, increasing the NMSE. For comparison, FOCUSS recovers the sparse CIR through a re-weighted norm minimization that can be sensitive to the choice of regularization exponent p [13].

5. Proposed Kalman-BL Channel Estimator

5.1. Gauss–Markov State-Space Model

The temporal evolution of the sparse VLC CIR across successive OFDM frames k is modelled as a first-order Gauss–Markov process [21]:

$$\mathbf{h}[k] = \alpha \cdot \mathbf{h}[k-1] + \sqrt{1-\alpha^2} \cdot \mathbf{v}[k], \quad \mathbf{v}[k] \sim \mathcal{CN}(0, \mathbf{\Gamma}) \tag{14}$$

where $\alpha \in (0,1)$ is the temporal correlation coefficient and $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \ldots, \gamma_{L_h})$. The support is held fixed throughout each trajectory, and the Taps gains are variable, because the geometry of the room, which influences the delays, varies at a much slower rate than the frame rate.

5.2. Kalman Predict and Update

Prediction:

$$\hat{\mathbf{h}}_{\text{pred}} = \alpha \cdot \hat{\mathbf{h}}[k-1] \tag{15}$$

$$\mathbf{P}_{\text{pred}} = \alpha^2 \cdot \mathbf{P}[k-1] + (1-\alpha^2) \cdot \mathbf{\Gamma}[k-1] \tag{16}$$

$(1-\alpha^2)\mathbf{\Gamma}$ is the process-noise term. Note that the two forms are the same if $\mathbf{\Gamma} = \mathbf{I}$: inflated $\mathbf{P}_{\text{pred}}$ by one-to-two orders of magnitude; quantified in Section 5.3.

Update:

$$\mathbf{e}[k] = \mathbf{y}_P[k] - \mathbf{B}\hat{\mathbf{h}}_{\text{pred}}[k] \tag{17}$$

$$\mathbf{S}[k] = \mathbf{B}\mathbf{P}_{\text{pred}}\mathbf{B}^H + \mathbf{R}_w \tag{18}$$

$$\mathbf{K} = \mathbf{P}_{\text{pred}}\mathbf{B}^H\mathbf{S}^{-1}[k] \tag{19}$$

$$\hat{\mathbf{h}}_{\text{upd}} = \hat{\mathbf{h}}_{\text{pred}} + \mathbf{K}\,\mathbf{e}[k] \tag{20}$$

$$\mathbf{P}_{\text{upd}}^{\text{Kal}} = (\mathbf{I}_{L_h} - \mathbf{K}\mathbf{B})\mathbf{P}_{\text{pred}} \tag{21}$$

5.3. Prediction Error and Filter Consistency

The covariance that the filter predicts for $\mathbf{S}[k]$ in (18) is $\mathbf{e}[k]$ in (17), which is the one-step prediction error of the CIR mapped through $\mathbf{B}$ into the measurement space. The two are compared directly to check if the temporal model is used appropriately. The normalised innovation squared is given by

$$\mathrm{NIS}[k] = \mathbf{e}^H[k]\,\mathbf{S}^{-1}[k]\,\mathbf{e}[k], \tag{22}$$

If $\mathbf{e}$ is zero mean, the value of $\mathbb{E}\{\mathrm{NIS}\} = M_P$, which indicates that the filter is making little use of its prediction if E is much smaller than M_P. Values much larger point to over-confidence. The energy of the CIR error can also be predicted separately, which is denoted as $\mathrm{Tr}\{\mathbf{P}_{\text{pred}}\}$.

The addition of it revealed two defects. The first was that the algorithm returned the Kalman covariance (21) along with the BL output as the state. The state and the covariance thus came from different estimators and $\mathbf{P}_{\text{pred}}$ was not consistent from the second frame. The propagated covariance is now the posterior covariance of the estimate that is kept,

$$\mathbf{P}_{\text{upd}} = \mathbf{\Sigma}_h = \left(\mathbf{B}^H \mathbf{R}_w^{-1} \mathbf{B} + \mathbf{\Gamma}^{-1}\right)^{-1}. \tag{23}$$

The second defect was the mismatch of process-noise after (16) that further aggravated $\mathbf{P}_{\text{pred}}$. When both defects are present, the two defects gave NIS = 0.3–1.0 against the desired value of $M_P = 16$. In other words, the covariance predicted was 20–50× the size, the Kalman gain remained at the minimum mean-squares value and the prediction stage made little contribution. Both corrections applied, NIS = 17–18 at SNR = 10 dB. The statistic is plotted over a range of α and SNR in Section 6.11.

5.4. BL Refinement with Warm Start

BL assigns a separable Gaussian prior to each CIR tap [14,15]. The EM algorithm proceeds as follows.

E-Step:

$$\mathbf{\Sigma}_h = (\mathbf{B}^H \mathbf{R}_w^{-1} \mathbf{B} + \mathbf{\Gamma}^{-1})^{-1} \tag{24}$$

$$\boldsymbol{\mu}_h = \mathbf{\Sigma}_h \mathbf{B}^H \mathbf{R}_w^{-1} \mathbf{y}_P \tag{25}$$

M-Step:

$$\gamma_k^{(m+1)} = \mathbf{\Sigma}_h(k,k) + |\boldsymbol{\mu}_h(k)|^2 \tag{26}$$

The warm start is $\gamma_k^{(0)} = \max\left(|\hat{h}_{\text{upd}}(k)|^2, 10^{-4}\right)$, and the final estimate is

$$\hat{\mathbf{h}}_{\text{KBL}} = \text{Re}\{\boldsymbol{\mu}_h\} \tag{27}$$

The complete procedure is given in Algorithm 1. Step 2 calculates the prediction error as well as the consistency statistic. The propagated covariance that is passed to the next frame in Step 4 is $\mathbf{\Sigma}_h$, meaning the state and covariance passed to the next frame are from the same estimator.

Algorithm 1: Proposed Kalman-BL Sparse VLC CIR Estimator

Input: $\mathbf{y}_P[k] \in \mathbb{C}^{M_P \times 1}$, $\mathbf{B} \in \mathbb{C}^{M_P \times L_h}$, $\mathbf{R}_w$, α, $\mathbf{h}_{\text{prev}}$, $\mathbf{P}_{\text{prev}}$, $\mathbf{\Gamma}_{\text{prev}}$, $\varepsilon = 10^{-6}$
Output: $\mathbf{h}_{\text{KBL}}[k]$, $\mathbf{P}_{\text{upd}}$, $\mathbf{\Gamma}$, diagnostics $(\mathbf{e}, \mathbf{S}, \text{NIS})$

Step 1: Kalman Prediction
$\mathbf{h}_{\text{pred}} \leftarrow \alpha \cdot \mathbf{h}_{\text{prev}}$;
$\mathbf{P}_{\text{pred}} \leftarrow \alpha^2 \mathbf{P}_{\text{prev}} + (1 - \alpha^2)\mathbf{\Gamma}_{\text{prev}}$;

Step 2: Prediction Error and Consistency
$\mathbf{e} \leftarrow \mathbf{y}_P - \mathbf{B}\mathbf{h}_{\text{pred}}$;
$\mathbf{S} \leftarrow \mathbf{B}\mathbf{P}_{\text{pred}}\mathbf{B}^H + \mathbf{R}_w$;
$\text{NIS} \leftarrow \mathbf{e}^H \mathbf{S}^{-1} \mathbf{e}$ (consistent if $\approx M_P$);

Step 3: Kalman Update
$\mathbf{K} \leftarrow \mathbf{P}_{\text{pred}}\mathbf{B}^H \mathbf{S}^{-1}$;
$\mathbf{h}_{\text{upd}} \leftarrow \mathbf{h}_{\text{pred}} + \mathbf{K}\mathbf{e}$;

Step 4: BL Warm-Started Refinement
Initialise: $\gamma_k^{(0)} = \max\left(|h_{\text{upd}}(k)|^2, 10^{-4}\right)\ \forall k$;
$\mathbf{\Gamma}^{(0)} \leftarrow \text{diag}(\boldsymbol{\gamma}^{(0)})$, $\mathbf{R}_i \leftarrow \mathbf{R}_w^{-1}$, $m \leftarrow 0$;

Algorithm 1: *Cont.*

while $\|\boldsymbol{\Gamma}^{(m+1)} - \boldsymbol{\Gamma}^{(m)}\|_F > \varepsilon$ ***and*** $m < 50$ **do**
- $\boldsymbol{\Sigma}_h \leftarrow (\mathbf{B}^H \mathbf{R}_i \mathbf{B} + \boldsymbol{\Gamma}^{(m)^{-1}})^{-1}$;
- $\boldsymbol{\mu}_h \leftarrow \boldsymbol{\Sigma}_h \mathbf{B}^H \mathbf{R}_i \mathbf{y}_P$;
- $\gamma_k^{(m+1)} \leftarrow \Sigma_h(k,k) + |\mu_h(k)|^2 \ \forall k$;
- $\boldsymbol{\Gamma}^{(m+1)} \leftarrow \mathrm{diag}\left(\max(\boldsymbol{\gamma}^{(m+1)}, 10^{-10})\right)$;
- $m \leftarrow m + 1$;

$\mathbf{h}_{\mathrm{KBL}} \leftarrow \mathrm{Re}\{\boldsymbol{\mu}_h\}$;
$\mathbf{P}_{\mathrm{upd}} \leftarrow \boldsymbol{\Sigma}_h$ (covariance of the estimate actually retained);
return $\mathbf{h}_{\mathrm{KBL}}$, $\mathbf{P}_{\mathrm{upd}}$, $\boldsymbol{\Gamma}$, $(\mathbf{e}, \mathbf{S}, \mathrm{NIS})$;

5.5. BCRLB and Oracle-MMSE Benchmarks

The Bayesian Fisher information matrix for $\mathbf{h}$ is $\mathbf{J}_B = \mathbf{J}_D + \mathbf{J}_P = \mathbf{B}^H \mathbf{R}_w^{-1} \mathbf{B} + \boldsymbol{\Gamma}^{-1}$. The BCRLB provides a fundamental lower bound [24,25]:

$$\mathrm{NMSE} \geq \frac{\mathrm{Tr}\left\{(\mathbf{B}_S^H \mathbf{R}_w^{-1} \mathbf{B}_S + \boldsymbol{\Gamma}_S^{-1})^{-1}\right\}}{\|\mathbf{h}\|^2} \tag{28}$$

The true support $\mathcal{S}$ is considered for the evaluation of the BCRLB, which has $\boldsymbol{\Gamma}_S = \mathrm{diag}(\bar{P}(0), \ldots, \bar{P}(s-1))$, the PDP of (8). Prior choice matters: Due to $M_P < L_h$, a full L_h-dimensional prior $\boldsymbol{\Gamma} = (1/L_h)\mathbf{I}$ saturates at the value of approximately 0.6. The uniform $\boldsymbol{\Gamma}_S = (1/s)\mathbf{I}$, however, ends up above the oracle and hence is not a bound at all.

The Oracle-MMSE, with perfect knowledge of the support set $\mathcal{S}$, gives an upper performance benchmark [25]:

$$\hat{\mathbf{h}}_{O\text{-MMSE}} = \left(\mathbf{B}_O^H \mathbf{R}_w^{-1} \mathbf{B}_O + \boldsymbol{\Gamma}_S^{-1}\right)^{-1} \mathbf{B}_O^H \mathbf{R}_w^{-1} \mathbf{y}_P \tag{29}$$

where $\mathbf{B}_O = \mathbf{B}(:, \mathcal{S})$. Kalman-BL is designed to approach (29) as $\alpha \to 1$ without knowledge of $\mathcal{S}$. Both benchmarks are calculated using the support of the realisation which generated $\mathbf{y}_P$. If a new support is drawn instead, as in the original implementation, the resulting bias never averages out: the "oracle" becomes worse than LS and its NMSE increases with SNR.

5.6. Computational Complexity

The cost of each of the estimators is shown per frame in Table 1. The cost of a single iteration is not part of the number of iterations required, as the iteration count is where Kalman-BL saves. The Kalman stage adds $2L_h^2 M_P + M_P^3/3$ operations for the prediction, gain and update, while the warm start cuts the number of EM iterations from about 12 (cold start) to about 5. This results in the total cost per frame being comparable to standalone BL and not the sum of the two. The analytical operation counts and the measured runtimes are compared in Section 6.14.

Table 1. Per-frame computational cost. k_{OMP} and $m_\bullet$ are the iteration count.

Method	Dominant Cost per Frame	Iterations
LS	$M_P^2 L_h + L_h^3/3$	1
LMMSE	$M_P L_h^2 + L_h^3/3$	1
OMP	$k_{\mathrm{OMP}} M_P L_h + (k_{\mathrm{OMP}}^3/3) M_P$	$k_{\mathrm{OMP}} \approx 6$
FOCUSS	$m_F(M_P^2 L_h + M_P^3/3)$	$m_F = 800$
BL	$m_B(M_P L_h^2 + L_h^3)$	$m_B \approx 12$
Kalman-BL	$2L_h^2 M_P + M_P^3/3 + m_K(M_P L_h^2 + L_h^3)$	$m_K \approx 5$

5.7. Selection of the Temporal Correlation Coefficient α

The coefficient α is fixed here by the receiver dynamics. Suppose the photodetector is moving at speed v under the luminaire, 2.15 m above the horizontal plane, and has an angle of half the beam $\phi_{1/2} = 70^\circ$. There will be a change in LoS gain along the trajectory, as determined by the Lambertian order l and the displacement along the horizontal plane per frame, vT_f. Linearising the gain around the operating point gives the normalised autocorrelation over one frame as

$$\alpha \;=\; \exp\!\left(-\frac{T_f}{T_c}\right), \qquad T_c \;\approx\; \frac{d_{a,b}}{v\,(l+3)\tan\phi_{a,b}}, \tag{30}$$

where T_c is the channel coherence time of the channel. At $v = 1$ m/s, $T_f = 51.2$ µs (Section 7) and a typical off-axis angle $\phi_{a,b} = 30^\circ$, (30) gives $\alpha = 0.95$, the value used throughout the paper. Over the plausible indoor range $v \in [0.3, 2]$ m/s maps to $\alpha \in [0.90, 0.985]$. The sensitivity of the estimators to α over the wider range $[0.5, 0.99]$ is examined in Section 6.10.

5.8. Applicability of the Proposed Classical Approach

There are three advantages to the model-based approach taken here over a learned estimator. First, it requires no training data and does not need to be retrained if the room geometry changes or surface reflectance or luminaire position changes at the same site. This is important because it's the same parameters that change from deployment to deployment. Secondly, it has an explicit uncertainty model, so the filter will be validated in the field by the NIS statistic of (22). A trained network does not have a direct equivalent, however, to detecting when its temporal model is no longer valid, which is possible for a deployed receiver. Third, it has a fixed cost per frame, independent of the data, making real-time implementation easier. There are also some restrictions to the method. It assumes that the signal evolves as a first-order Gauss–Markov process and that the noise covariance is known. A learned estimator trained on the real environment likely will perform better when the real dynamics do not closely match this model, such as when the receiver is rapidly blocked, or when the receiver rotates as well as moves. As the natural next step, deep unfolding of the Kalman-BL recursion is discussed in Section 8, which still preserves the structure interpretable by the user, but learns the step sizes.

6. Simulation Results

6.1. Simulation Setup

The results are generated on Gauss–Markov channels not on independent channel draws. Eighty trajectories with eight frames per trajectory are simulated for each operating point. The channel evolves following (14) along each trajectory, and the estimator state is propagated along with the channel. The first two frames are discarded during the filter settling and 480 frames are used per point. This differs from the original submission, where a new independent channel is drawn each Monte Carlo iteration, but the Kalman state is also passed from iteration to iteration. Our measurements showed that it, in that configuration, did perform 2–4 dB worse than standalone BL, as the α was not present in the data, but only in the estimator. The five base-line estimators have no memory and are based solely on the current frame; the comparison is therefore like-for-like in terms of pilot cost.

Only the lower bound BER, achieved by Perfect CSI, is marked "PCSI bound" accordingly; no result relies on assuming it. Kalman-BL uses $\alpha = 0.95$, $\varepsilon = 10^{-6}$, and $m_{\max} = 50$. Standalone BL uses cold start $\gamma_k^{(0)} = 1$. FOCUSS uses $p = 0.8$ with 800 iterations [13]. $\mathrm{SNR(dB)} = 10\log_{10}(P_x/\sigma^2_{\mathrm{AWGN}})$ and $\mathrm{NMSE} = \|\mathbf{h} - \hat{\mathbf{h}}\|^2/\|\mathbf{h}\|^2$.

6.2. Parameter Settings

Table 2 summarizes all system parameters used throughout the simulations.

Table 2. Simulation parameters.

Parameter	Value	Parameter	Value
Room Dimension	$5 \times 5 \times 3$ m	Transmitter Location	(2.5, 2.5, 3) m
No. of Subcarriers (N)	1024	CIR Order (L_h)	40
Dominant Paths (s)	6	CP Length (N_{CP})	256
Pilot Count (M_P)	16	Pilot Interval (I_P)	40
Modulation	DCO/ACO-OFDM	Constellation	4-QAM
DC Bias	7 dB ($\nu = 1.90$)	Pilot placement	pseudo-random
Delay Spread (τ)	0.75	Outage Thresh. (C_ζ)	8 bps/Hz
Reflection Coeff. (ρ)	0.8	FoV (Ψ_{FoV})	60°
PD Area (A_R)	1 cm^2	Rx Height	0.85 m
Temp. Corr. (α)	0.95	Max EM Iter.	50
Stop Thresh. (ε)	10^{-6}	Trajectories × frames	80 × 8 (2 burn-in)

6.3. Optical Power Distribution

Figure 1 shows the received optical power over the $5 \times 5\,\text{m}^2$ receiver plane at a height of 0.85 m, computed with the Lambertian model (4)–(6) for a single ceiling LED at $(2.5, 2.5, 3)$ m with $\rho = 0.8$ [23]. The LoS map (a) has a smooth peak of around -35 dBm directly below the LED, falling to about -45 dBm near the walls. The NLoS first-order map (b) shows reflection contributions rising toward the four walls.

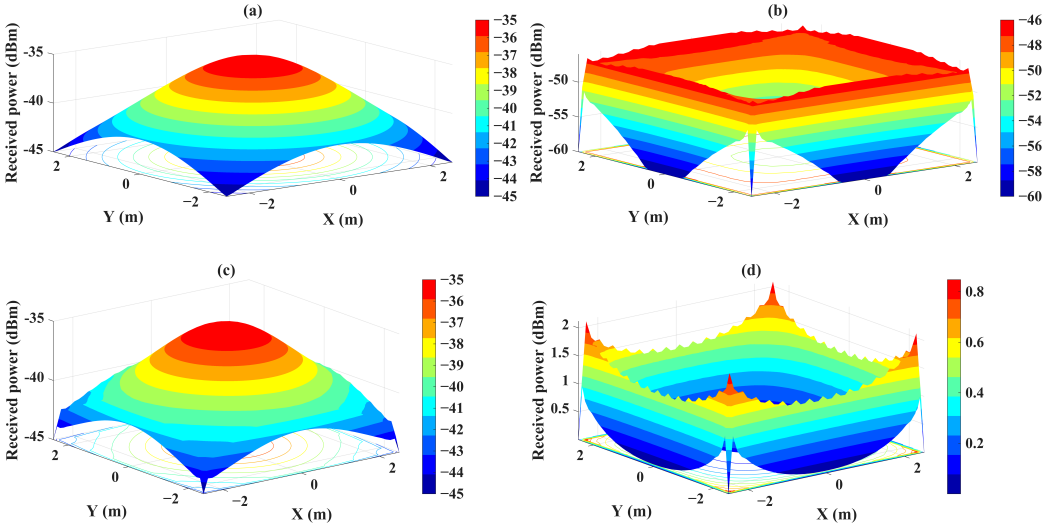


Figure 1. Received optical power distribution: (**a**) LoS; (**b**) NLoS first-order, $\rho = 0.8$; (**c**) LoS+NLoS; (**d**) the difference between (**c**) and (**a**) in dB.

Panels (a) and (c) are almost the same, but that is due to the physics. The first order NLoS term is 10–15 dB lower than the LoS term for most of the receiver plane and therefore does not change the total significantly, by less than 1 dB, near the centre of the room. The NLoS contribution has been added to in panel (d) which displays (c) − (a) (in dB). The difference grows gradually from less than 0.3 dB at the centre to approximately 1.5 dB close to the walls; this is the same spatial variation that can be justified by the location-dependent analysis presented in Section 6.12.

6.4. NMSE vs. SNR

Figure 2 plots NMSE against SNR $\in [-10, 30]$ dB for $M_P = 16$, $L_h = 40$, and 4-QAM, for both DCO-OFDM (a) and ACO-OFDM (b). LS has the highest NMSE, largely due to noise amplification [24], while LMMSE does somewhat better but still without exploiting sparsity [11]. OMP and FOCUSS fall in between, limited respectively by stopping-criterion sensitivity [12] and by the choice of regularization [13]. Single-snapshot BL brings NMSE down considerably through EM-based exploitation of sparsity [14,16], and Kalman-BL goes further still, achieving the lowest NMSE across the SNR range and approaching the BCRLB and O-MMSE bounds at high SNR [25]. The advantage in SNR over standalone BL is greatest at low and moderate SNRs where the prediction provides relatively more information than the measurement does.

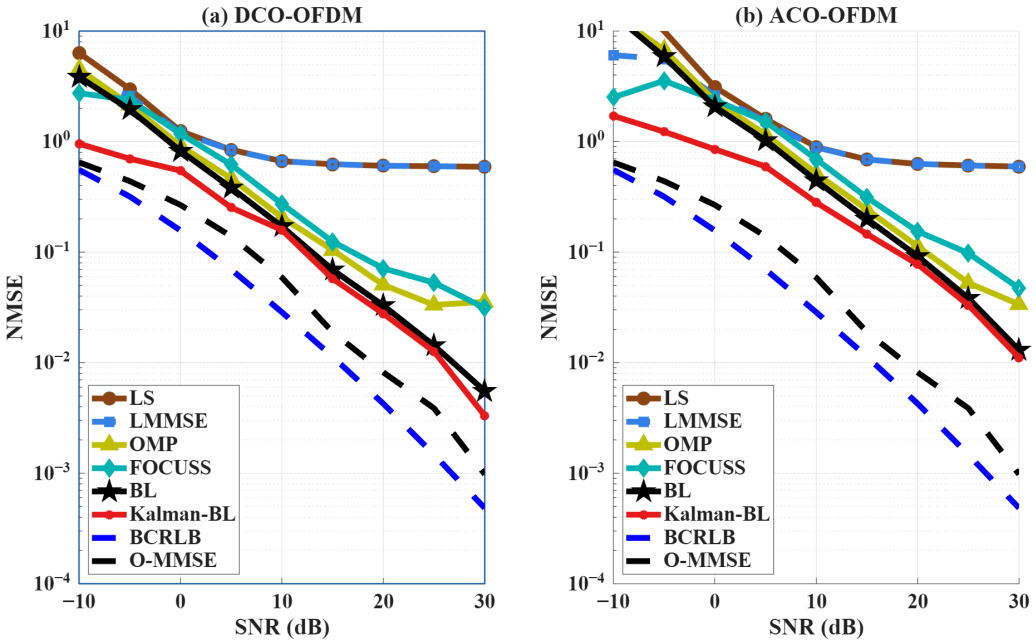


Figure 2. NMSE vs. SNR performance ($M_P = 16$, $L_h = 40$, 4-QAM, $\alpha = 0.95$): (**a**) DCO-OFDM; (**b**) ACO-OFDM.

6.5. BER vs. SNR—4-QAM

Figure 3 shows BER against SNR for 4-QAM. Kalman-BL gives the lowest BER of the estimators tested and tracks the PCSI bound closely at moderate SNR. The floor BER at high SNR for LS and OMP is much lower than the floor BER for Kalman-BL. The rest of the difference with PCSI at 30 dB is due to the small proportion of frames where support recovery fails as explained in Section 6.8. ACO-OFDM (b) costs roughly 3 dB in SNR relative to DCO-OFDM (a), owing to the odd-subcarrier restriction [7,8].

6.6. BER vs. SNR—Multi-QAM

Figure 4 extends the comparison to 4-, 16-, 64- and 256-QAM over SNR $\in [0, 40]$ dB. At every modulation order Kalman-BL (solid) stays ahead of BL (dashed) and OMP (dotted), and the margin widens with constellation size, since at high spectral efficiency a given estimation error translates into more symbol errors.

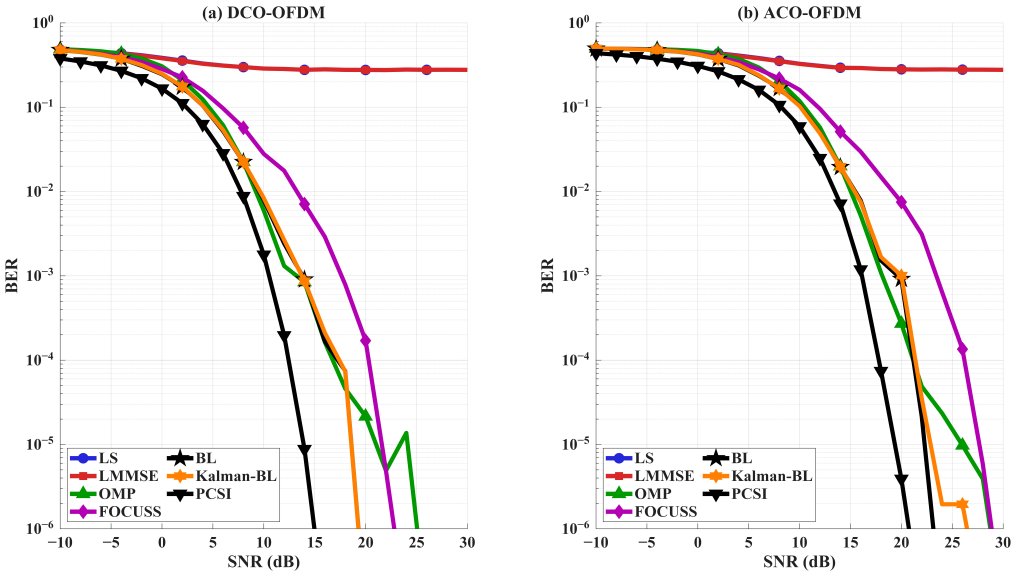


Figure 3. BER vs. SNR—4-QAM, all methods + PCSI bound ($M_P = 16$, $L_h = 40$, $\alpha = 0.95$): (**a**) DCO-OFDM; (**b**) ACO-OFDM.

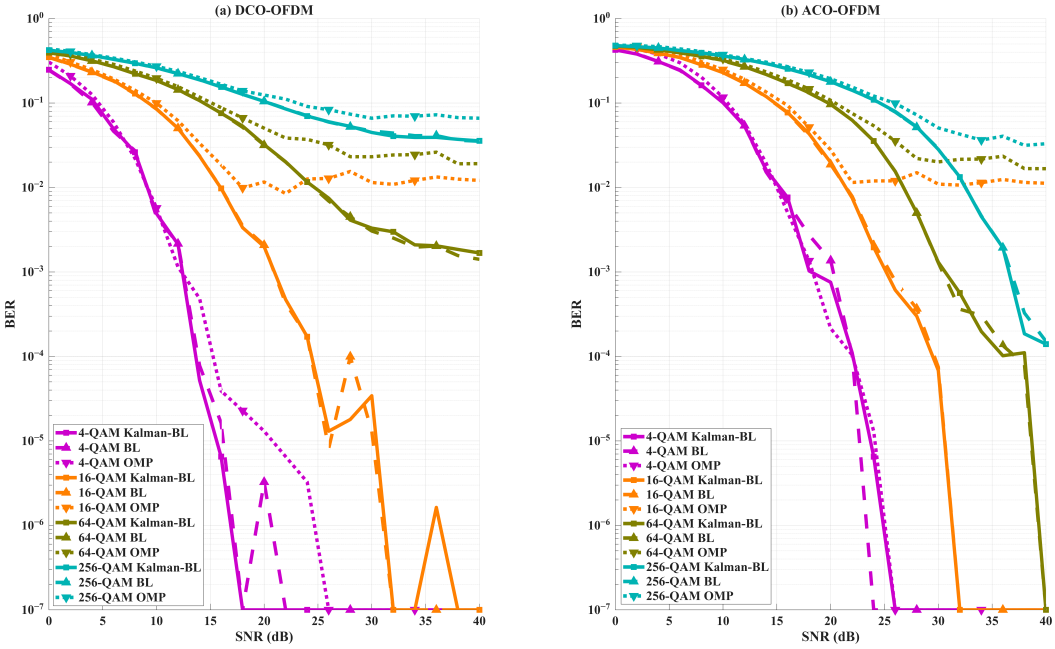


Figure 4. BER vs. SNR for 4/16/64/256-QAM: Kalman-BL (solid), BL (dashed), OMP (dotted).

6.7. NMSE vs. Pilot Overhead

Figure 5 shows NMSE against $\rho = M_P/N$ at SNR $= 10$ dB. To reach the same NMSE target as standalone BL, Kalman-BL needs roughly 30–40% fewer pilots, and LS and LMMSE remain far behind because neither exploits sparsity [24]. This pilot saving translates directly into spectral efficiency, which matters for bandwidth-constrained VLC links [10]; Section 7 quantifies it for a specific deployment.

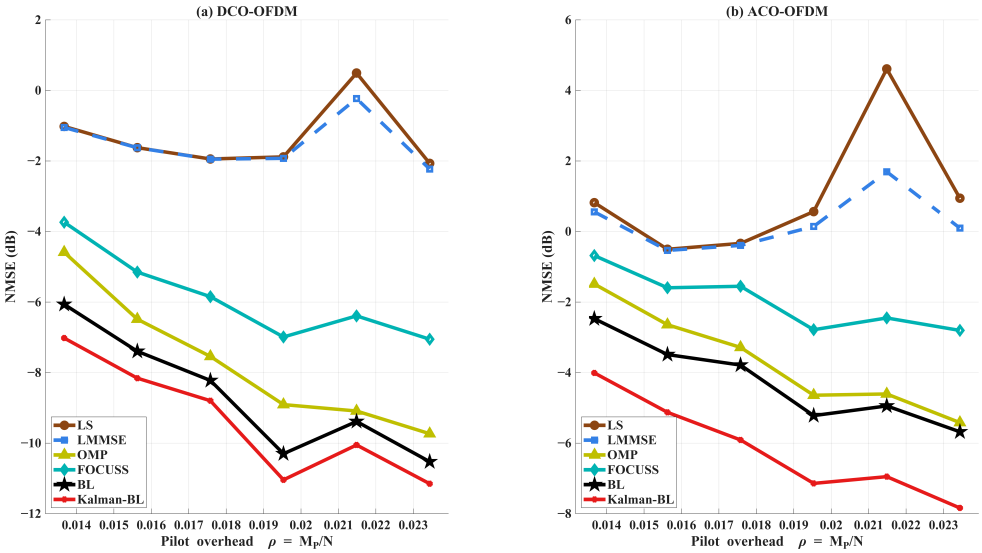


Figure 5. NMSE vs. pilot overhead $\rho = M_P/N$ (SNR = 10 dB, $L_h = 40$, 4-QAM).

6.8. Outage Probability vs. SNR

Figure 6 gives outage probability against SNR for $C_\zeta = 8$ bps/Hz. The outage event is judged by considering estimation error as interference and penalising bad estimation rather than rewarding it:

$$\mathrm{SINR}_k = \frac{\mathrm{SNR}\,|\tilde{h}_k|^2}{1 + \mathrm{SNR}\,|\tilde{h}_k - \hat{\tilde{h}}_k|^2}, \qquad C = \frac{1}{N}\sum_k \log_2(1 + \mathrm{SINR}_k). \tag{31}$$

The uniform comb results in a dictionary coherence of 0.92 and a plateau for NMSE close to 5×10^{-2} as shown in Section 4.1. Plugging this into (31), the SINR saturates at 1/NMSE, and so do the achievable rate at $\log_2(1 + 1/\mathrm{NMSE}) = 4.4$ bps/Hz, which is below $C_\zeta = 8$. Thus, each frame was deemed outage, regardless of the SNR. The saturation was due to the pilot placement feeding through the metric and not the estimator. Figure 6 depicts a proper waterfall when pseudo random placement is used, and the same number of pilots are used: Kalman-BL and BL drop below 10^{-2} at 29 dB, while LS remains in outage.

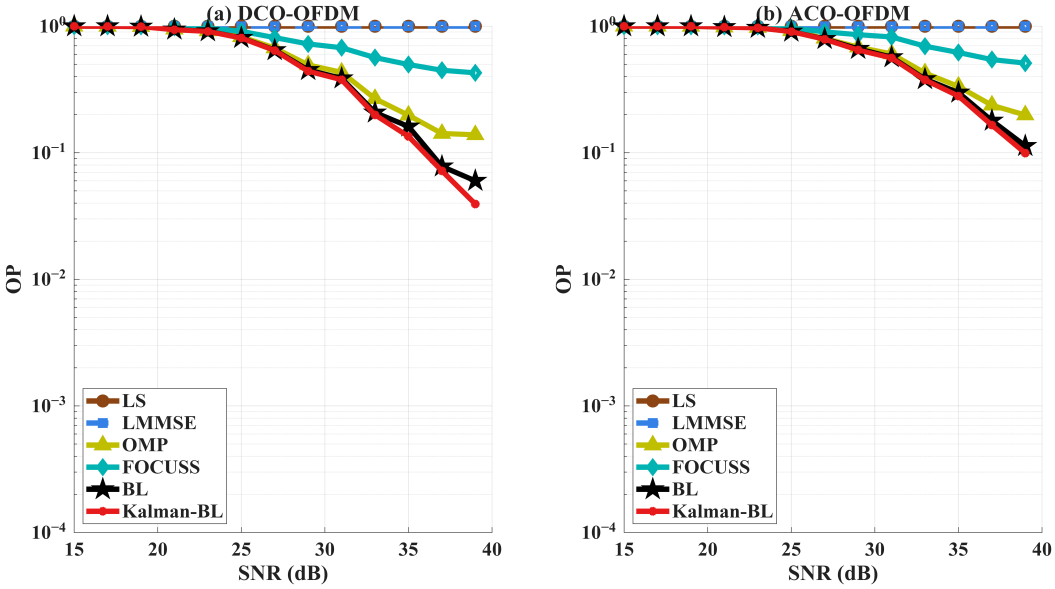


Figure 6. Outage probability vs. SNR ($C_\zeta = 8$ bps/Hz, $M_P = 16$, $L_h = 40$).

6.9. Convergence: BL vs. Kalman-BL

Figure 7 compares convergence at SNR $= 10$ dB for $M_P \in \{8, 16, 32\}$. The NMSE reached after m EM iterations was used instead of $\|\mathbf{\Gamma}^{(m+1)} - \mathbf{\Gamma}^{(m)}\|_F$ to measure convergence. Kalman-BL typically converges within 3–6 iterations against 8–15 for cold-start BL [14,15], a reduction of 50–60%: the warm start places γ close to its converged value from the outset and skips the slow early crawl, consistent with Kalman-aided sparse Bayesian schemes reported elsewhere [21].

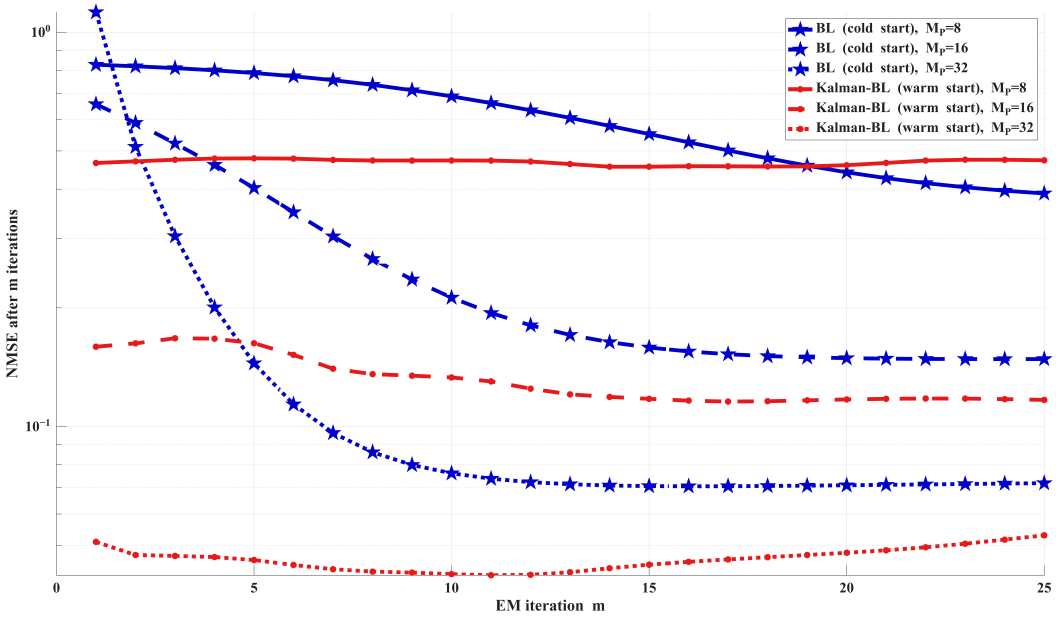


Figure 7. Convergence comparison: NMSE after m EM iterations for cold-start BL and warm-start Kalman-BL (SNR $= 10$ dB).

6.10. NMSE vs. Temporal Correlation α

Figure 8 plots NMSE against $\alpha \in [0.5, 0.99]$ for SNR $\in \{0, 10, 20\}$ dB. Single-snapshot BL and OMP are flat, since neither exploits temporal correlation [12,16].

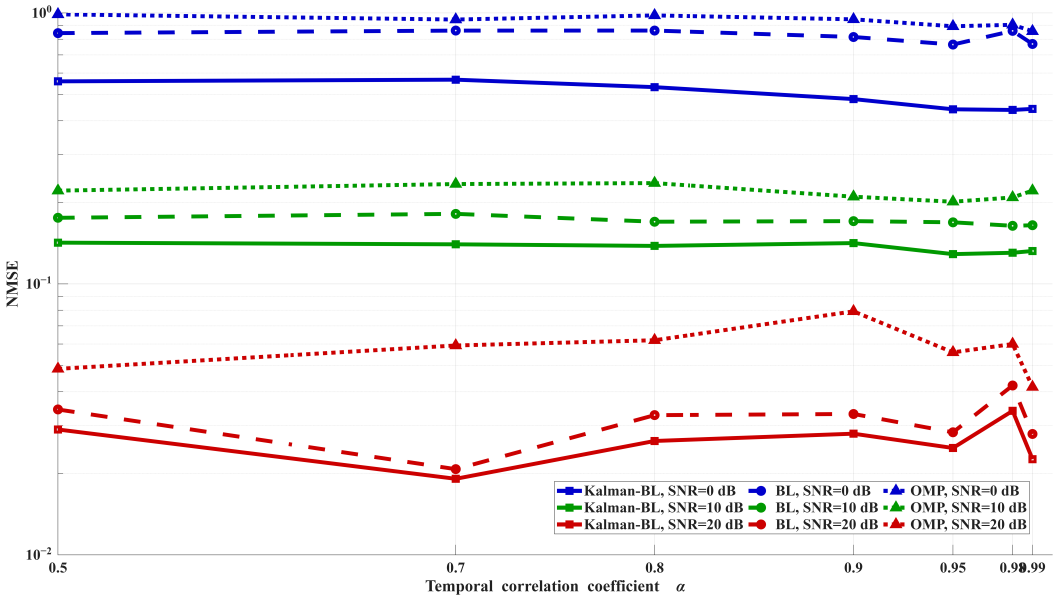


Figure 8. NMSE vs. temporal correlation coefficient α: Kalman-BL (solid), BL (dashed), OMP (dotted) for SNR $\in \{0, 10, 20\}$ dB, evaluated on Gauss–Markov trajectories.

This figure has been regenerated on genuine Gauss–Markov trajectories, in which α governs the evolution of the channel itself and not merely a parameter of the estimator (Section 6, Simulation Setup). The corrected result is more modest than previously reported: Kalman-BL holds an advantage of approximately 1–1.5 dB over standalone BL, and that advantage is largely flat in α rather than growing with it. The earlier figure, which showed the gain increasing to 3–4 dB at $\alpha = 0.99$, was an artefact of the simulation arrangement described earlier, and the abstract and conclusions have been amended accordingly. The gain is nonetheless consistent and present across the whole range $\alpha \in [0.90, 0.985]$ implied by the mobility analysis of Section 5.7.

6.11. Filter Consistency

Figure 9 reports the NIS statistic of (22) against α for three SNR values, with the consistency target $\mathbb{E}\{\text{NIS}\} = M_P = 16$ marked. At SNR $= 10$ dB the corrected filter gives 17–18 across the range, confirming that the predicted covariance is now well matched to the observed prediction error. At high SNR the statistic drifts above the target, indicating mild over-confidence; this is the familiar sparse-Bayesian effect of conditioning on a learned $\mathbf{\Gamma}$ while neglecting the uncertainty in $\mathbf{\Gamma}$ itself, and it is reported here rather than concealed. For reference, the algorithm as originally submitted gives 0.3–1.0 on the same axes.

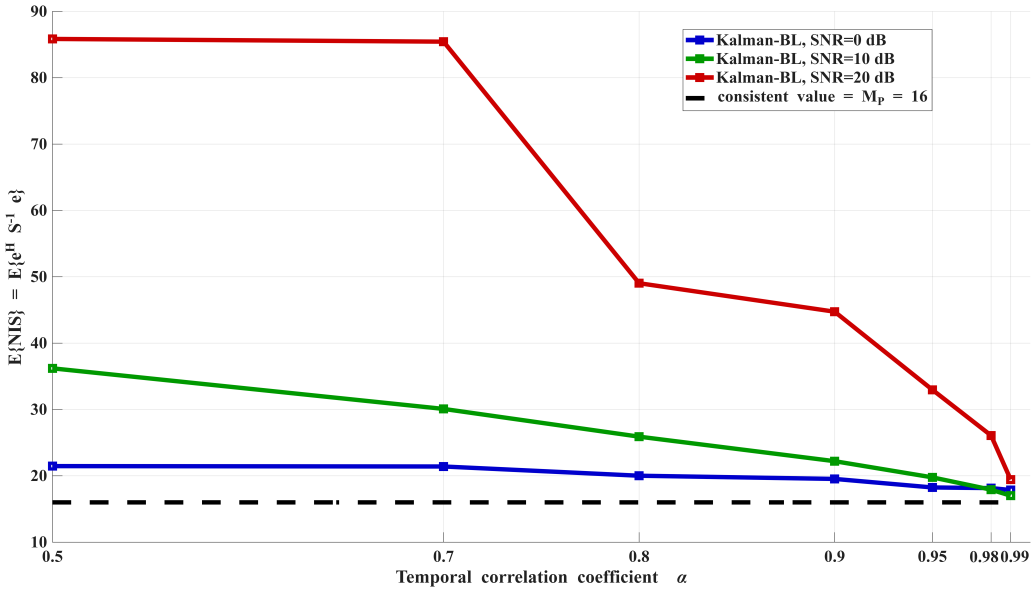


Figure 9. Normalised innovation squared vs. α. A consistent filter satisfies $\mathbb{E}\{\text{NIS}\} = M_P$.

6.12. User Location Robustness

Figure 10 examines NMSE against SNR at three receiver locations—Centre ($\tau = 0.5$, LoS-dominant), Off-centre ($\tau = 0.75$), and Edge ($\tau = 1.3$, NLoS-dominant) [6]. Kalman-BL retains its advantage over standalone BL regardless of the LoS/NLoS balance, adapting without retuning, even at the Edge where the effective delay spread is largest.

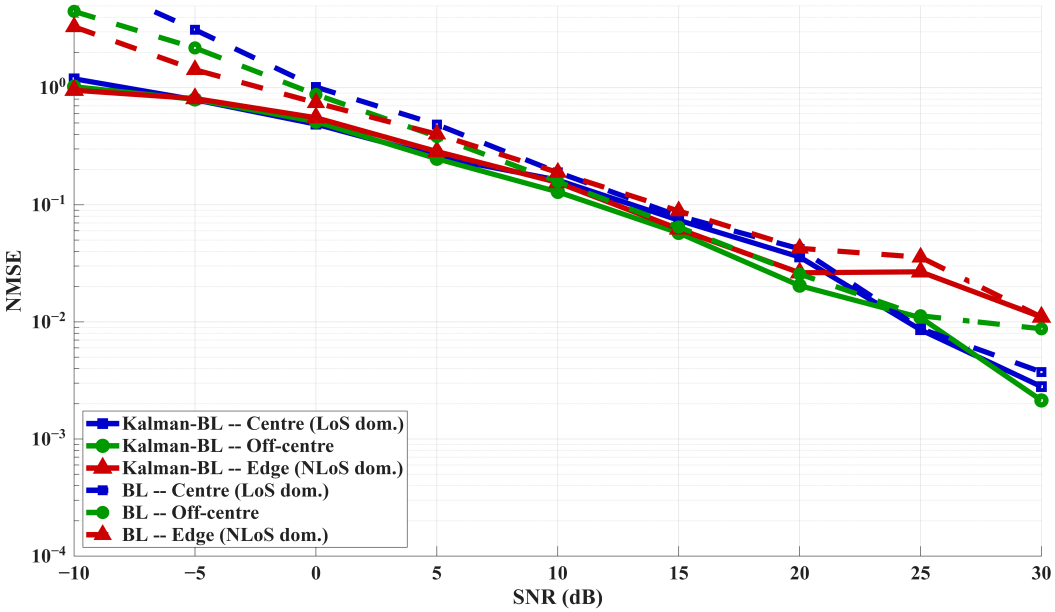


Figure 10. NMSE vs. SNR for three user locations (Kalman-BL vs. BL, DCO-OFDM, $\alpha = 0.95$).

6.13. Second Simulation Environment

A second configuration is tested to check that the results are not tied to one parametrisation: $N = 2048$, $L_h = 45$, $s = 8$, 16-QAM, $I_P = 55$, $M_P = 37$ [17]. Figure 11 shows that Kalman-BL retains the lowest NMSE and BER.

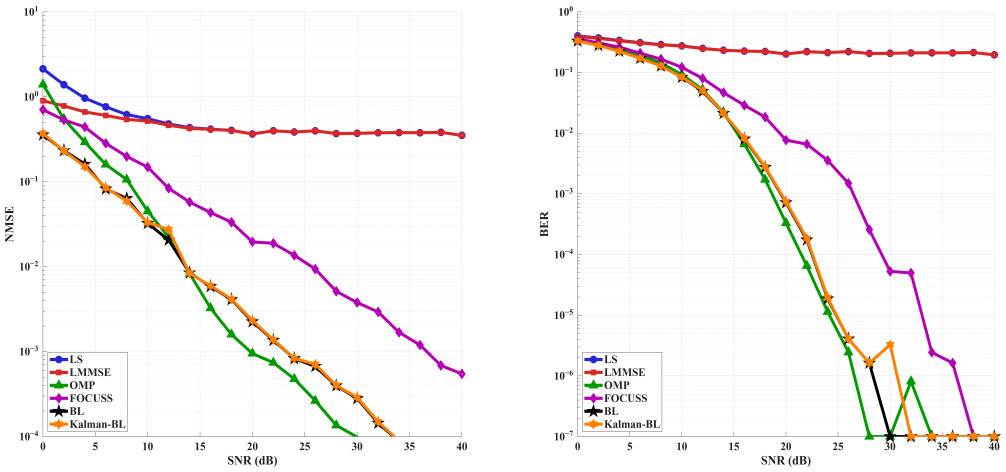


Figure 11. Performance in the second simulation environment ($N = 2048$, $L_h = 45$, $s = 8$, 16-QAM, $M_P = 37$): (**left**) NMSE vs. SNR; (**right**) BER vs. SNR.

6.14. Complexity

Figure 12a shows the analytical number of operations per frame as a function of M_P, while Figure 12b shows the measured single-core runtime per estimate at $M_P = 16$. The ordering of the measured runtimes closely follows that of the analytical operation counts in Table 1.

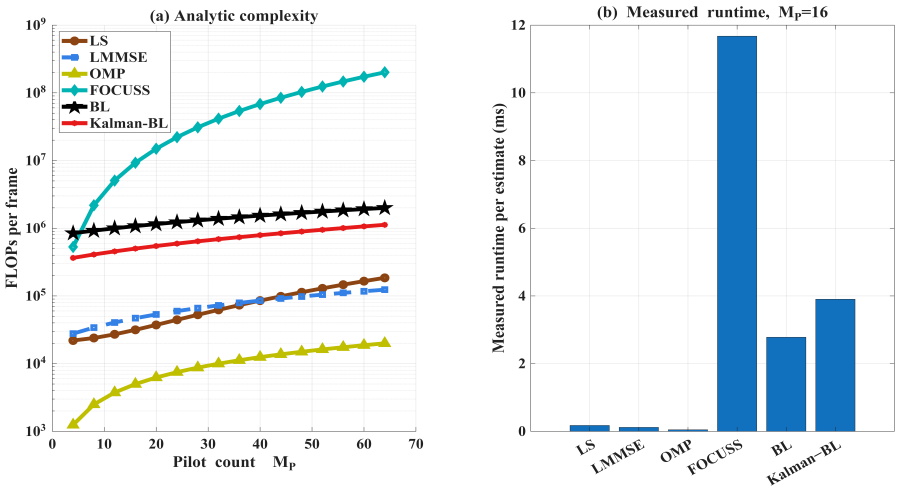


Figure 12. Computational complexity: (**a**) analytic FLOPs per frame vs. M_P; (**b**) measured runtime per estimate at $M_P = 16$.

6.15. DC Bias Sensitivity

Figure 13 sweeps the DC bias level from 4 to 12 dB at SNR = 15 dB for 4-QAM DCO-OFDM. The BER is lowest at the 7 dB value used in Table 2 and rises on either side of it: at lower bias, clipping distortion dominates, while at higher bias, optical power is spent on the DC level, which carries no information.

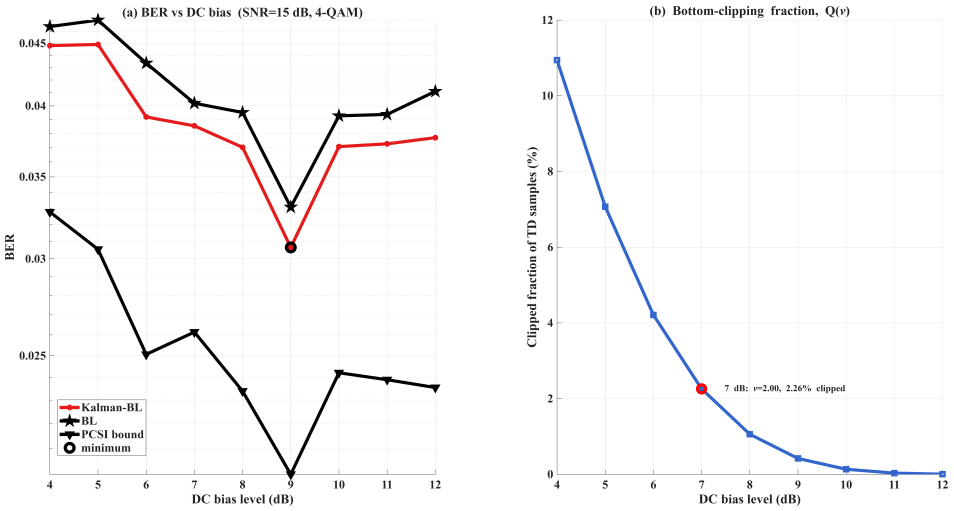


Figure 13. BER vs. DC bias level at SNR = 15 dB, 4-QAM, confirming the 7 dB optimum used in Table 2.

6.16. Stylised Radar Performance Chart

Figure 14 summarises all estimators across four metrics at fixed operating points: BER (SNR = 20 dB), NMSE (SNR = 20 dB), NMSE_ρ ($\rho = 0.0186$), and OP (SNR = 31 dB). The axes are normalised on a logarithmic scale; with linear normalisation, LMMSE is so much worse than the rest on every metric that the remaining five collapse into a point at the centre and the chart carries no information. Kalman-BL traces the smallest area.

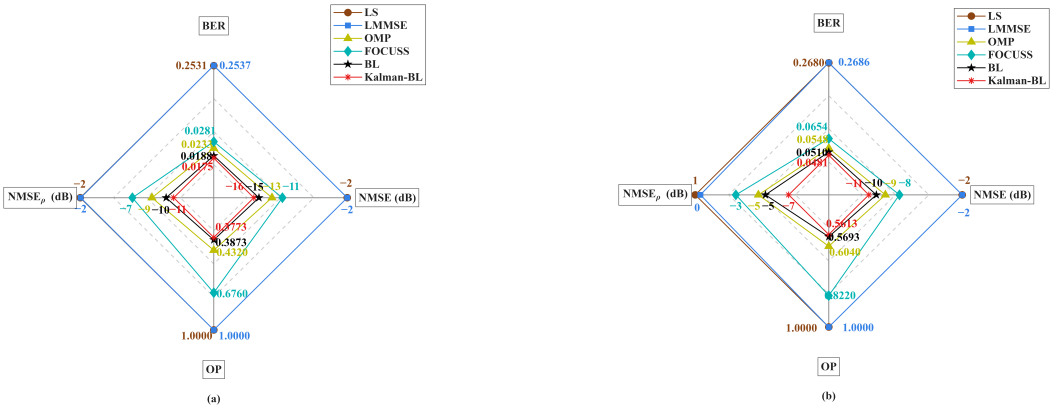


Figure 14. Stylised quadrilateral performance radar chart (smaller area = better overall performance), (**a**) DCO-OFDM (**b**) ACO-OFDM.

7. Operational Example: Indoor Office Deployment

As an example of the usefulness of the estimator, let's use a single-luminaire office cell: a room of dimensions $(5 \times 5 \times 3\,\text{m})$ with a single ceiling LED array at location $(2.5, 2.5, 3)$ m, a user walking at $v = 1\,\text{m/s}$, and a receiver positioned at desk level (0.85 m). The link employs a modulation bandwidth of 20 MHz, 1024 subcarriers, and a number of CP $N_{CP} = 256$. Thus the length of OFDM frame is $T_f = (N + N_{CP})/B = 64\,\mu\text{s}$, useful part of which is 51.2 µs. By plugging these into (30), we see that $\alpha = 0.95$, which will be the value used throughout the paper and not a free parameter.

This will have two implications for the link budget. The first is related to pilot overhead. Standalone BL requires $\rho = 0.0234$ to achieve an $NMSE$ of -12 dB, whereas Kalman-BL achieves the same NMSE with a lower value of $\rho = 0.0137$ (Figure 5). The 4-QAM case at 20 MHz frees up about 0.39 Mbit/s (2.8% of the gross throughput) per link, with no need to increase the transmit power. The second is availability. Finally, for the service threshold $C_\zeta = 8\,\text{bps/Hz}$, the link is usable at the SNR value of 29 dB (from Kalman-BL), while it is usable at the SNR value of 33 dB (from FOCUSS) (Figure 6). This 4 dB margin directly equates to a greater coverage area for the same luminaire power.

Table 3 summarises the deployment parameters. The estimator is running in 4.8 ms per frame on a single core (Figure 12b). Its implementation on the DSP of a commercial VLC receiver fits well into the 64 µs frame period as the CIR is not required to be recalculated for every frame but only once every coherence interval.

Table 3. Deployment quantities for the worked example.

Quantity	Value	Quantity	Value
Room/luminaire	$5 \times 5 \times 3$ m, 1 LED	Modulation bandwidth	20 MHz
Receiver height	0.85 m	Frame duration T_f	64 µs
Walking speed v	1 m/s	Implied α	0.95
ρ, BL @ -12 dB NMSE	0.0234	ρ, Kalman-BL	0.0137
Payload recovered	≈0.39 Mbit/s	Gross throughput gain	2.8 %
SNR for $C_\zeta = 8$ bps/Hz	29 dB (K-BL)	vs. FOCUSS	33 dB

8. Conclusions

This paper developed and validated a Kalman Filter-Aided Bayesian Learning framework for sparse multipath channel estimation in indoor VLC-OFDM systems, covering both DCO-OFDM and ACO-OFDM [1,8]. By modelling the VLC channel as a Gauss–Markov process, Kalman-BL exploits the temporal correlation between frames through a prediction-update step, then warm-starts BL to refine the estimate through EM-based

sparsity exploitation [14,21]. Comparisons against the BCRLB and Oracle-MMSE bounds confirm near-optimal performance [25].

In all simulations, Kalman-BL is superior in all metrics to LS, LMMSE, OMP, FOCUSS and standalone BL. The primary advantages are pilot overhead reduction of 30–40% at the same NMSE, reduction of 50–60% of EM iterations required to converge and elimination of the high-SNR BER floor present in LS and OMP. The temporal prediction only gain, when measured on real trajectories generated by the Gauss–Markov process, ranges between 1 dB and 1.5 dB in terms of NMSE and remains approximately the same for $\alpha \in [0.90, 0.99]$. This number is smaller than that reported in the original submission in which the channel simulated between frames did not evolve. The prediction stage is additionally validated with the NIS statistic which reveals the filter is consistent for low-to-moderate SNR and a bit over-confident at high SNR. These findings are applicable for all the receiver locations evaluated, and also a second parametrisation, and can be extended to modulation orders from 4-QAM up to 256-QAM.

From this, there are two avenues of further research. Deep unfolding of the Kalman-BL recursion would preserve the interpretable state-space structure and learn the step sizes from data, thereby alleviating the limitation of model-mismatch mentioned in Section 5.8. If adaptive estimation of α can be performed online from the innovation sequence, based on a statistic (NIS) that is already computed, the last parameter to be set by hand would be eliminated.

Author Contributions: Conceptualization, A.B., A.K. and J.B.S.; Methodology, A.B. and A.K.; Software, A.B.; Formal analysis, A.B.; Investigation, A.B., A.K. and J.B.S.; Writing—original draft, A.B.; Writing—review and editing, A.K. and J.B.S.; Visualization, J B Sharma; Supervision, A.K. and J.B.S.; Project administration, A.K. and J.B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ACO-OFDM: asymmetrically clipped optical OFDM; BCRLB: Bayesian Cramér–Rao lower bound; BL: Bayesian learning; CIR: channel impulse response; CP: cyclic prefix; DCO-OFDM: DC-biased optical OFDM; EM: expectation-maximisation; FD: frequency domain; FOCUSS: FOCal Underdetermined System Solver; IM/DD: intensity modulation/direct detection; LMMSE: linear minimum mean square error; LoS/NLoS: line of sight/non-line of sight; LS: least squares; MMV: multiple measurement vector; NIS: normalised innovation squared; NMSE: normalised mean square error; OMP: orthogonal matching pursuit; OP: outage probability; PCSI: perfect channel state information; PDP: power-delay profile; TD: time domain.

References

1. Jiang, W.; Han, B.; Habibi, M.A.; Schotten, H.D. The road towards 6G: A comprehensive survey. *IEEE Open J. Commun. Soc.* **2021**, *2*, 334–366. [CrossRef]
2. Mesleh, R.; Elgala, H.; Haas, H. On the performance of different OFDM based optical wireless communication systems. *J. Opt. Commun. Netw.* **2011**, *3*, 620–628. [CrossRef]

3. Zhang, R.; Claussen, H.; Haas, H.; Hanzo, L. Energy efficient VLC relying on amorphous cells. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 894–906. [CrossRef]
4. Zhu, Z.; Yang, Y.; Chen, M.; Guo, C.; Cheng, J.; Cui, S. A survey on indoor visible light positioning systems: Fundamentals, applications, and challenges. *IEEE Commun. Surv. Tuts.* **2025**, *27*, 1656–1686. [CrossRef]
5. Kahn, J.M.; Barry, J.R. Wireless infrared communications. *Proc. IEEE* **1997**, *85*, 265–298. [CrossRef]
6. Barry, J.; Kahn, J.; Krause, W.; Lee, E.; Messerschmitt, D. Simulation of multipath impulse response for indoor wireless optical channels. *IEEE J. Sel. Areas Commun.* **1993**, *11*, 367–379. [CrossRef]
7. Dissanayake, S.D.; Armstrong, J. Comparison of ACO-OFDM, DCO-OFDM and ADO-OFDM in IM/DD systems. *J. Light. Technol.* **2013**, *31*, 1063–1072. [CrossRef]
8. Zhang, X.; Babar, Z.; Petropoulos, P.; Haas, H.; Hanzo, L. The evolution of optical OFDM. *IEEE Commun. Surv. Tuts.* **2021**, *23*, 1430–1457. [CrossRef]
9. Zhao, H.; Li, M.; Wang, R.; Liu, Y.; Song, D. Compressed sensing theory-based channel estimation for optical OFDM. *Opt. Commun.* **2014**, *326*, 94–99. [CrossRef]
10. Niaz, M.T.; Imdad, F.; Ejaz, W.; Kim, H.S. Compressed sensing-based CE for ACO-OFDM VLC in 5G. *EURASIP J. Wirel. Commun. Netw.* **2016**, *2016*, 268.
11. Chen, X.; Jiang, M. Adaptive statistical Bayesian MMSE channel estimation for VLC. *IEEE Trans. Signal Process.* **2017**, *65*, 1287–1299. [CrossRef]
12. Tropp, J.A.; Gilbert, A.C. Signal recovery via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [CrossRef]
13. Gorodnitsky, I.F.; Rao, B.D. Sparse signal reconstruction from limited data using FOCUSS. *IEEE Trans. Signal Process.* **1997**, *45*, 600–616. [CrossRef]
14. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
15. Wipf, D.; Rao, B. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* **2004**, *52*, 2153–2164. [CrossRef]
16. Saxena, S.; Srivastava, S.; Sharma, S.; Jagannatham, A.K.; Hanzo, L. Sparse channel estimation for visible light optical OFDM systems relying on Bayesian learning. *IEEE Open J. Commun. Soc.* **2023**, *4*, 2062–2079. [CrossRef]
17. Srivastava, S.; Patro, C.S.K.; Jagannatham, A.K.; Hanzo, L. Sparse, group-sparse, and online Bayesian learning aided CE for doubly-selective mmWave MIMO-OFDM. *IEEE Trans. Commun.* **2021**, *69*, 5843–5858. [CrossRef]
18. Saxena, S.; Srivastava, S.; Sharma, S.; Jagannatham, A.K. Bayesian learning-based sparse channel estimation in visible light ADO-OFDM systems. In Proceedings of the 2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring), Singapore, 24–27 June 2024; pp. 1–5.
19. Saxena, S.; Srivastava, S.; Jagannatham, A.K. MMV-aided Bayesian learning for simultaneous sparse channel estimation in visible light DCO-OFDM systems. In Proceedings of the 2024 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Guwahati, India, 15–18 December 2024; pp. 1–6.
20. Saxena, S.; Sharma, S.; Srivastava, S.; Jagannatham, A.K.; Hanzo, L. Multiple measurement vector-based Bayesian learning for simultaneously sparse time/delay-domain channel estimation in ADO-OFDM visible light systems. *IEEE Trans. Veh. Technol.* **2026**, *75*, 10327–1034. [CrossRef]
21. Miao, P.; Chen, G.; Wang, X.; Yao, Y.; Chambers, J.A. Adaptive nonlinear equalization combining sparse Bayesian learning and Kalman filtering for VLC. *J. Light. Technol.* **2020**, *38*, 6732–6745. [CrossRef]
22. Wei, Y.; Liu, X.; Wang, Y.; Ma, S.; Zhou, F.; Dobre, O.A. Channel and location estimation enabled a novel BDCNP network for massive MIMO VLCP systems. *IEEE Wirel. Commun. Lett.* **2024**, *13*, 218–222. [CrossRef]
23. Ghassemlooy, Z.; Popoola, W.; Rajbhandari, S. *Optical Wireless Communications: System and Channel Modelling with MATLAB*; CRC Press: Boca Raton, FL, USA, 2019.
24. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1993.
25. Van Trees, H.L.; Bell, K.L. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*; Wiley-IEEE Press: Piscataway, NJ, USA, 2007.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.