# Performance Analysis of Random Fourier Features Based Unsupervised Multistage-Clustering for VLC

Rangeet Mitra, Vimal Bhatia,Sandesh Jain, Kwonhue Choi

***Abstract*—Visible light communication (VLC) has emerged as a secure, cost-effective, and green supplement to the existing radio frequency (RF) communication systems. However, the performance of a typical VLC link is degraded by the following factors: (a) inter-symbol interference (ISI), which arises due to limited modulation bandwidth of light-emitting diode (LED), (b) the dispersion due to multi-path reflections, and (c) nonlinear transfer characteristics of LED. To mitigate ISI and LED nonlinearity, existing blind post-distorters, like the Volterra modified cascaded multimodulus algorithm (MCMMA) and the Hammerstein improved multistage clustering (HIMSC) approaches, experience degradations from modeling errors due to abrupt polynomial-series truncation. To mitigate these degradations, we propose a novel random Fourier features (RFF) based improved multi-stage clustering (RFF-IMSC) algorithm for blind post-distortion. Simulations performed over standardized IEEE 802.15 PAN VLC channels indicate that the proposed RFF-RIMSC post-distorter delivers improved performance as compared to the existing blind post-distorters. Also, analytical guarantees are presented and error-rate performance are analyzed for the $\epsilon$-optimality of this approach with respect to supervised loss-functions.**

***Index Terms*—Visible Light Communications, Random Fourier Features, blind post-distortion.**

## I. Introduction

VISIBLE light communication (VLC) [1], [2] has evolved into a license-free, green and low-cost supplement to existing radio-frequency (RF) backbones, which has led to diverse VLC based applications in light-fidelity (Li-Fi) based systems, vehicle-to-vehicle (V2V) communications, and internet of things (IoT) systems. Among VLC systems, the widely used intensity-modulation/direction-detection (IM/DD) based VLC systems operate by modulating the intensity of light emitting diode (LED) with the transmitted symbols, which are consequently detected at the receiver by a photo-diode (PD).

In spite of the above benefits of VLC, the following impairments are found to degrade the performance of typical VLC links [3], [4]: (a) inter-symbol interference (ISI) arising due to dispersive nature of VLC channel and finite LED modulation-bandwidth, and (b) LED nonlinearity, which leads to an additive distortion at the receiver. While several works consider a linear LED model, the degradation in bit error rate (BER) due to nonlinear LED characteristics is non-negligible for signals with a large dynamic range and at high switching frequencies [5, Sec. 3.5]. To counter these degradations, existing works propose various pre-distortion [6], [7], and post-distortion algorithms [4], [8]–[10] for accurate symbol-detection over such impaired VLC channels. The pre-distortion algorithms deployed at the transmitter rely on the perfect feedback of detected symbols, which is an idealistic assumption in-general. To avoid dependence on this feedback, open-loop post-distortion methods are viable, which are mostly classified as [11]: (a) supervised post-distorters [4], [8], [12], and (b) unsupervised post-distorters [13], [14]. Notably, the supervised post-distorters need pilots for detection/equalization that add to the training-overhead and degrade the spectral efficiency [15]. Therefore, to mitigate the VLC channel-impairments without depending on pilots, this work focuses on unsupervised post-distortion for VLC.

Existing works for unsupervised post-distortion for VLC mostly derive from multi-modulus algorithms (MMA), e.g., the cascaded-MMA (CMMA) [16], and the Volterra modified-CMMA (MCMMA) [13], [17]. Besides, the generic MMA based approaches, multistage-clustering (MSC) [18], and improved MSC (IMSC) [19] were found to outperform the conventional MMA based approaches; however, these approaches are suited for linear systems and are well-known for their sub-optimality for nonlinear VLC channels. For nonlinear channels, the suitability of these works is enhanced by using polynomial kernels [13], [14]. Notably, a Hammerstein-IMSC (HIMSC) based blind post-distorter was derived for joint mitigation of LED nonlinearity and ISI [14], and was analytically linked to the maximium correntropy criterion (MCC) [20]. However, the HIMSC post-distorter has high computational complexity which increases with the order of the used Hammerstein kernel. Furthermore, the HIMSC post-distorter is impaired by approximation errors due to finite-order truncation of Hammerstein series.

In the context of existing supervised approaches, the random Fourier features (RFF) based post-distorters [11], [21], [22] were derived recently to mitigate degradation due to LED nonlinearity and ISI. These RFF-based explicit RKHS methods achieve similar performance as compared to the generic dictionary-based based post-distorters [4], [23] with reduced computational complexity, and are implemented under a finite memory budget. Besides, the RFF based algorithms are free from the inherent bottlenecks in the dictionary-based approaches [11] and offer guarantees of uniqueness of the optimality-points and generalized representation, and significantly lower approximation-error [22, Sec. IV].

Also, existing RFF based post-distorters for VLC depend on appropriate kernel-width initialization. In this context, the popular methods for kernel width estimation, like the Silver-

R. Mitra is with the Ecole De Technologie Superieure, Montreal-H3C1K3, Canada (e-mail: rangeet.mitra.1@ens.etsmtl.ca).

S. Jain and V. Bhatia are with the Indian Institute of Technology Indore, Indore-453552, India (e-mail: phd1601202004@iiti.ac.in, vbhatia@iiti.ac.in).

K.Choi is with the BWC Lab, Yeungnam University, South Korea, KR-712749 (e-mail: gonew@yu.ac.kr).

man's and Scott's rule of thumb [24], adaptive techniques [23], [25], and Bayesian leaning methods [26] are prone to estimation errors. Besides, adaptive estimation-methods require further computational steps for kernel-width optimization, and upon poor initialization, require more iterations to converge. Notably, recent works propose kernel-width sampling which alleviates the necessity for kernel-width initialization [27], and allows for hyperparameter-independence.

*Contributions*: Motivated by the negligible approximation error of RFFs [22, Sec. IV] and simple implementation under a finite memory-budget, we explore RFF based unsupervised post-distortion for VLC in this paper. The contributions of this letter are summarized as:

- In this paper, we propose a novel blind RFF based IMSC (RFF-IMSC) post-distorter for a bandlimited VLC link impaired by LED nonlinearity and ISI. Addtionally, to alleviate dependence on kernel-width values, we propose an unsupervised RFF-IMSC post-distorter with sampled kernel widths (RFF-SKW-IMSC).
- Furthermore, the optimality-guarantees for generic RFF-IMSC and the RFF-SKW-IMSC, are derived for blind post-distortion over a nonlinear VLC channel within the same analytical framework.
- Analytical bounds are derived for the BER of the proposed RFF-IMSC and the RFF-SKW IMSC.
- Simulations are presented for the standardized IEEE 802.15 PAN VLC channels for both memoryless and with memory LED nonlinearity. The presented computer-simulations indicate that the RFF-IMSC based approaches deliver better performance as compared to the existing Volterra-MCMMA and Hammerstein-IMSC approaches.

*Notations:* Scalars, vectors, and matrices are represented by simple, small boldface, and capital boldface letters, respectively. The subscript $(\cdot)_n$ is used to denote the quantity at the $n^{\text{th}}$ time instant/iteration. $(\cdot)^T$ denotes the transpose operator. The real part of a quantity $(\cdot)$ is denoted by $(\cdot)^R$ or $\Re(\cdot)$ while $(\cdot)^I/\Im(\cdot)$ denotes the imaginary part of a quantity $(\cdot)$, and $\{\cdot\}_{i=1}^{D}$ denotes $D$ probabilistic samples of $(\cdot)$.

## II. System model

In this Section, we describe the considered system model for the VLC link. The transmitted symbols, $s_n$, are drawn from a carrier-less amplitude and phase quadrature amplitude modulation (CAP-QAM) [28], [29]. Next, an appropriate DC bias $b$ is added to the input signal to ensure forward bias.

*LED Model*: We consider a Weiner model [10], [12] for modeling LED nonlinearity with memory, which consists of a linear time invariant (LTI) block cascaded with a memoryless nonlinearity. The LTI block is modeled as [12]:

$$\tilde{s}_n = \sum_{m=0}^{N-1} b_m s_{n-m}, \tag{1}$$

where $N$ denotes memory of the LTI block and $b_m$ denotes coefficients corresponding to the $m^{th}$ delay. The memoryless LED nonlinearity (denoted by $f_{\text{LED}}(\cdot)$) is modeled by Rapp's model given as [30]:

$$f_{\text{LED}}(\tilde{s}_n) = \begin{cases} \frac{(\tilde{s}_n - V_{\text{th}})}{\left(1+\left(\frac{(\tilde{s}_n - V_{\text{th}})}{V_{\max}}\right)^{2q}\right)^{\frac{1}{2q}}} & \tilde{s}_n \geq V_{\text{th}} \\ 0 & \tilde{s}_n < V_{\text{th}} \end{cases} \tag{2}$$

where $V_{\max}$, $V_{\text{th}}$, and $q$ are maximum saturation voltage, threshold voltage of an LED, and knee factor, respectively. The received signal at the PD is written as

$$x_n = \sum_{j=0}^{M_1-1} h_j f_{\text{LED}}(\tilde{s}_{n-j}) + v_n, \tag{3}$$

where $\{h_j \in \mathbb{R}\}_{j=0}^{M_1-1}$ is the overall channel impulse response (CIR) of length $M_1$. Notably, this is modeled by the convolution of LED low-pass CIR and the IEEE 802.15 PAN VLC CIR [31]–[33], and their updates in [3]. The LED low-pass response is modeled by a Butterworth low-pass filter with cut-off frequency $f_c = 20$ MHz [14], [31]. The overall noise $v_n$ consists of ambient light shot noise and thermal noise [22], and is modeled by a zero-mean Gaussian random variable with variance $\sigma_n^2$, denoted as $v_n \sim \mathcal{N}(0, \sigma_n^2)$. Upon optical-electrical conversion at the receiver using a photodetector, and after DC bias subtraction, the received signal is input to the proposed RFF-IMSC post-distorter for symbol-detection.

## III. Proposed RFF-IMSC blind Post-Distorter

In this Section, we propose an RFF-based improved multistage-clustering (RFF-IMSC) based blind post-distorter for a VLC link impaired by LED nonlinearity. To mitigate LED nonlinearity, received observations are first mapped to a high dimensional feature space using a randomized feature map $\hat{\Psi} : \mathbb{R}^{2d} \longrightarrow \mathbb{R}^D$ (where $d >> D$), defined as [34]:

$$\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n) = \sqrt{\frac{2}{D}} \begin{bmatrix} \cos(\boldsymbol{\omega}_1^T[\Re(\mathbf{x}_n); \Im(\mathbf{x}_n)] + \alpha_1) \\ \cos(\boldsymbol{\omega}_2^T[\Re(\mathbf{x}_n); \Im(\mathbf{x}_n)] + \alpha_2) \\ \vdots \\ \cos(\boldsymbol{\omega}_D^T[\Re(\mathbf{x}_n); \Im(\mathbf{x}_n)] + \alpha_D) \end{bmatrix}, \tag{4}$$

where $d$ denotes the memory of the post-distorter, i.e., $\mathbf{x}_n$ is a vector of past $d$ samples of incoming $x_n$, $D$ is the number of RFF dimensions with each $\{\boldsymbol{\omega}_i\}_{i=1}^{D} \sim \mathcal{N}(\mathbf{0}_D, \frac{1}{\sigma^2}\mathbf{I}_D)$ (where $\sigma$ denotes the kernel width) and each $\{\alpha_i\}_{i=1}^{D}$ is uniformly drawn from $[0 \;\; 2\pi]$ [34]. Using the RFF mapped regressors $\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)$, the output of the proposed RFF-IMSC algorithm is expressed as:

$$y_n^R = \Re\big(\boldsymbol{\Omega}_n^T \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)\big), \text{ and } y_n^I = \Im\big(\boldsymbol{\Omega}_n^T \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)\big), \tag{5}$$

where $\boldsymbol{\Omega}_n = \boldsymbol{\Omega}_n^R + j\boldsymbol{\Omega}_n^I$ denotes the post-distorter weight vector. The cost-function for the proposed unsupervised RFF-IMSC post-distorter is expressed below separately for the real and imaginary parts as [14]:

$$\mathcal{J}_{\text{RFF-IMSC}}^R = \min_{\boldsymbol{\Omega}_n^R} \;\; \frac{1}{Q}\sum_{j=1}^{Q} \exp\Big(-\frac{(y_n^R - \mu_j^R)^2}{\gamma}\Big), \tag{6}$$

$$\mathcal{J}_{\text{RFF-IMSC}}^I = \min_{\boldsymbol{\Omega}_n^I} \;\; \frac{1}{Q}\sum_{j=1}^{Q} \exp\Big(-\frac{(y_n^I - \mu_j^I)^2}{\gamma}\Big),$$

where $\gamma$ is the spread parameter, and the set of modulii are denoted as $\{\mu_j^R\}_{j=1}^{Q}$ and $\{\mu_j^I\}_{j=1}^{Q}$ with $Q = \sqrt{M}$, with $M$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2021.3089933, IEEE Communications Letters



denoting the CAP-QAM modulation-order. The gradients of the cost function formulated in (3) and (6) are expressed as:

$$\frac{\partial \mathcal{J}^R_{\text{RFF-IMSC}}}{\partial \mathbf{\Omega}^R_n} = \frac{\sum_{j=1}^{Q} \exp\Big(-\frac{(y^R_n-\mu^R_j)^2}{\gamma}\Big)(y^R_n-\mu^R_j)\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)}{Q},$$

$$\frac{\partial \mathcal{J}^I_{\text{RFF-IMSC}}}{\partial \mathbf{\Omega}^I_n} = \frac{\sum_{j=1}^{Q} \exp\Big(-\frac{(y^I_n-\mu^I_j)^2}{\gamma}\Big)(y^I_n-\mu^I_j)\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)}{Q}.$$

The weights of the proposed RFF-IMSC post-distorter are updated by a stochastic gradient descent (SGD) algorithm with step-size $\eta$. The final SGD update-equations for the weights of the proposed RFF-IMSC post-distorter are given as:

$$\begin{aligned}\mathbf{\Omega}^R_{n+1} &= \mathbf{\Omega}^R_n + \eta\Re(e^{\text{RFF-IMSC}}_n)\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n),\\ \mathbf{\Omega}^I_{n+1} &= \mathbf{\Omega}^I_n + \eta\Im(e^{\text{RFF-IMSC}}_n)\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n),\end{aligned} \tag{7}$$

where $\Re(e^{\text{RFF-IMSC}}_n) = \frac{1}{Q}\sum_{j=1}^{Q}\exp\Big(-\frac{(y^R_n-\mu^R_j)^2}{\gamma}\Big)(y^R_n-\mu^R_j)$ and $\Im(e^{\text{RFF-IMSC}}_n) = \frac{1}{Q}\sum_{j=1}^{Q}\exp\Big(-\frac{(y^I_n-\mu^I_j)^2}{\gamma}\Big)(y^I_n-\mu^I_j)$. Further, we denote equalizer lag, $\tau = \lceil\frac{d-1}{2}\rceil$. Next, through the theorem below, we motivate the optimality of the RFF-IMSC with respect to a supervised loss function $\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n), \mathbf{\Omega})$:

**Theorem 1.** *The unique fixed point of the unsupervised objective, $\mathcal{J}_{\text{RFF-IMSC}}$, lies in an infinitesimal neighborhood of the minimizer/maximizer of a convex/concave Lipschitz-continuous supervised loss-function $\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n), \mathbf{\Omega})$.*

*Proof.* Please refer Appendix A. ∎

For square M-QAM with average bit-energy $E_b$, we present the following analytical result on the BER of the RFF-IMSC:

**Theorem 2.** *The BER of the RFF-IMSC post-distorter is bounded by:*

$$\begin{aligned}&\frac{2}{\log_2 Q}\Bigg(\sqrt{\frac{Q-1}{Q}}\Bigg)\text{erfc}\Bigg(\sqrt{\frac{3E_b\log_2 Q}{2(Q-1)(1+\frac{\eta}{2D})\sigma^2_n}}\Bigg)\\ &\qquad\leq P_{e,\text{RFF-IMSC}} \leq\\ &\frac{2}{\log_2 Q}\Bigg(\sqrt{\frac{Q-1}{Q}}\Bigg)\text{erfc}\Bigg(\sqrt{\frac{3E_b\log_2 Q}{2(Q-1)\Big[\Big(1+\frac{\eta}{2D}\Big)\sigma^2_n+\frac{2\rho\log(Q)}{Q}\Big]}}\Bigg),\end{aligned} \tag{8}$$

*where $\text{erfc}(\cdot)$ denotes the complementary error-function.*

*Proof.* Please refer Appendix B. ∎

Notably, the proposed RFF-IMSC post-distorter depends on an appropriate kernel-width initialization, and this optimal value is known to change across deployment scenarios and is non-trivial to estimate generally. To address this limitation, we present the corresponding hyperparameter-free formulation (called the RFF-SKW-IMSC). For the RFF map with sampled kernel width (RFF-SKW), the frequencies $\{\boldsymbol{\omega}_i\}_{i=1}^D$ are drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0}_D, \frac{1}{\sigma^2}\mathbf{I}_D)$. Similarly, the kernel widths $\{\sigma_i^2\}_{i=1}^D$ are drawn from an inverse-Gamma distribution [27]. The RFF-SKW, denoted by $\hat{\Psi}^{\text{RFF-SKW}}_{\tilde{\sigma}}(\cdot)$, is mathematically expressed as [27]:

$$\hat{\Psi}^{\text{RFF-SKW}}_{\tilde{\sigma}}(\mathbf{x}_n) = \sqrt{\frac{2}{D}}\begin{bmatrix}\cos(\tilde{\sigma}_1\widetilde{\boldsymbol{\omega}}_1^T[\Re(\mathbf{x}_n);\Im(\mathbf{x}_n)]+\alpha_1)\\ \cos(\tilde{\sigma}_2\widetilde{\boldsymbol{\omega}}_2^T[\Re(\mathbf{x}_n);\Im(\mathbf{x}_n)]+\alpha_2)\\ \vdots\\ \cos(\tilde{\sigma}_D\widetilde{\boldsymbol{\omega}}_D^T[\Re(\mathbf{x}_n);\Im(\mathbf{x}_n)]+\alpha_D)\end{bmatrix}, \tag{9}$$

where $\{\widetilde{\boldsymbol{\omega}}_i\}_{i=1}^D \sim \mathcal{N}(\mathbf{0}_D, \mathbf{I}_D)$, and $\{\tilde{\sigma}_i\}_{i=1}^D \sim \sqrt{\Gamma(\gamma,\delta)}$, where $\gamma$, and $\delta$ denotes shape, and rate parameter of Gamma distribution, respectively. The appropriate values of the parameters for Gamma distribution, $\gamma$ and $\delta$, are given as [27]:

$$\begin{aligned}\gamma &= \frac{M}{4}, && \text{for M-PAM}\\ &= \frac{\sqrt{M}}{2}, && \text{for M-CAP,}\\ \delta &= \frac{\sigma_x^2}{4\pi^2},\end{aligned} \tag{10}$$

where $\sigma_x^2$ denotes the regressor-variance. We present the following theorem, as an optimality-proof for RFF-SKW-IMSC:

**Theorem 3.** *The unique fixed point of the unsupervised objective, $\mathcal{J}_{\text{RFF-IMSC}}$, lies in an infinitesimal neighborhood of the minimizer/maximizer of a convex/concave Lipschitz-continuous supervised loss-function $\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF-SKW}}_{\tilde{\sigma}}(\mathbf{x}_n), \mathbf{\Omega})$, subject to the use of a reasonably large number of RFFs $D$.*

These guarantees of hyperparameter-free near-optimal performance without training-overhead, makes the proposed RFF-SKW-IMSC attractive for post-distortion over VLC links. *Computational Complexity:* The computational complexity of the RFF-IMSC and the RFF-SKW-IMSC is $\mathcal{O}(D)$, and is similar to the supervised approaches in [22], [34].

## IV. Simulations

In this section, simulations are performed to validate the performance of the proposed blind post-distorters. Proposed RFF-IMSC, and RFF-SKW-IMSC algorithm is compared against the following existing blind post-distortion algorithms: (a) Volterra-MCMMA, (b) IMSC and (c) HIMSC [14]. $10^5$ symbols are considered over an ensemble of 100 Monte-Carlo runs. Parameters for the LTI block in (1) of the frequency-dependent LED nonlinearity are chosen as: $b_o = 1$, $b_1 = 0.2$, and $b_2 = 0.1$ as in [12]. 4000 symbols were used for training the post-distorter. Parameters for Rapp's model in (2) are chosen as: $V_{\text{th}} = 0.2$ V, $V_{\max} = 0.5$ V, and $q = 0.5$ [30]. 45 first order and 25 second order filter taps are chosen for Volterra-MCMMA post-distorter. By cross-validation, the order of Hammerstein expansion was chosen as 4 for HIMSC post-distorter. For the proposed RFF-IMSC post-distorter, RFF dimension is chosen as $D = 70$, and kernel width is initialized by trial and error for best performance.

The BER vs signal-to-noise ratio (SNR) and the BER vs Rate performances are depicted in Fig. 1 and Fig. 2 for "residential", and for "manufacturing cell" scenarios respectively. It is observed from Fig. 1, that the proposed RFF-IMSC algorithm delivers better BER performance over the existing polynomial series (Volterra-MCMMA, and HIMSC) based unsupervised post-distorters, and achieve $2-3$ dB SNR gain at BER of $10^{-4}$ over the existing HIMSC post-distorter. This gain is due to the modeling/approximation error for polynomial series based post-distorters that occurs due to finite-order truncation of Volterra/Hammerstein kernels. Notably, it is observed from Fig. 1 and Fig. 2, that the proposed probabilistic kernel width based RFF-SKW-IMSC post-distorter delivers equivalent BER performance as compared to fixed best kernel width (found by brute force trial and error) based RFF-IMSC

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2021.3089933, IEEE Communications Letters



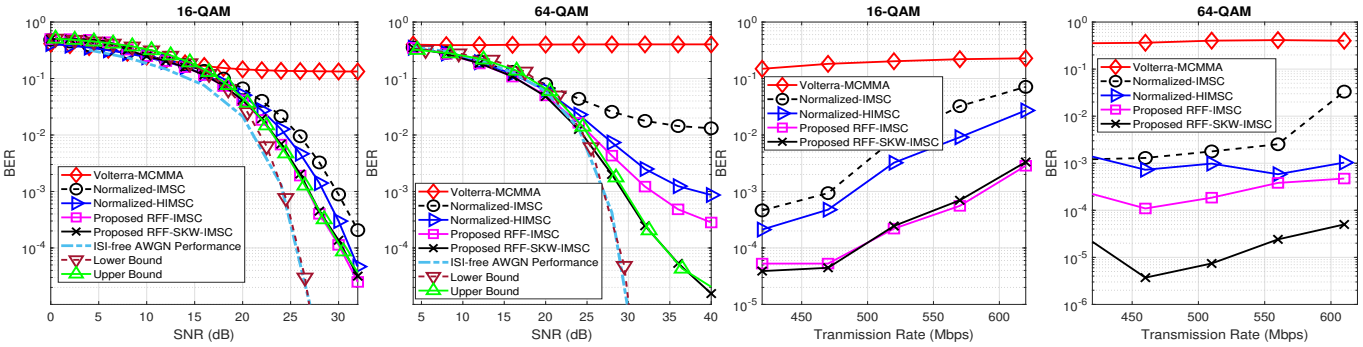


Fig. 1. BER vs SNR and BER vs Sum-Rate for the existing Volterra-MCMMA, IMSC, HIMSC and the proposed RFF-IMSC post-distorter for 16-QAM and 64-QAM over Residential scenarios.

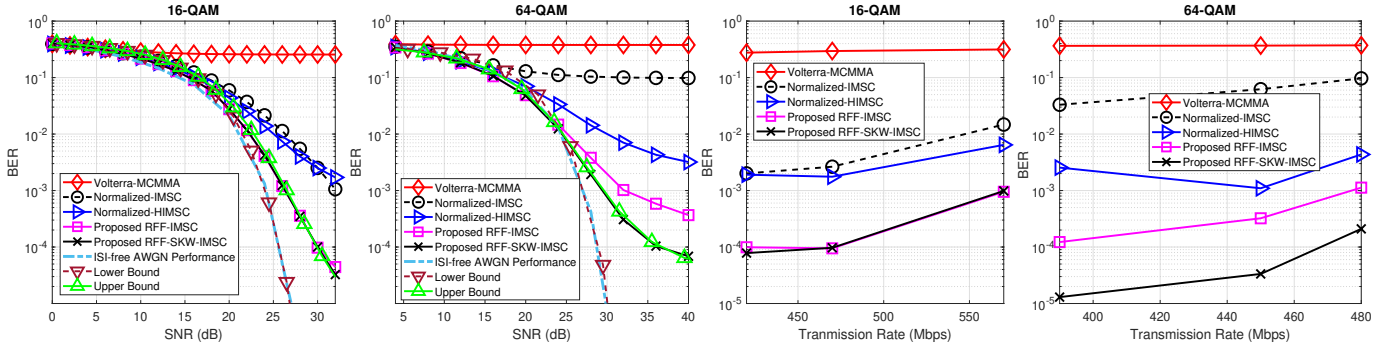


Fig. 2. BER vs SNR and BER vs Sum-Rate for the existing Volterra-MCMMA, IMSC, HIMSC and the proposed RFF-IMSC post-distorter for 16-QAM and 64-QAM over Manufacturing Cell scenario.

algorithm. However in Figs. 1 and 2, the gains for the RFF-SKW-IMSC are more visible for the 64-QAM scenario due the inexactness of the MMA-type post-distorters[1]. Additionally, it is observed that the choice of same hyperparameter-values for the RFF-IMSC as for the case of 16-QAM delivers a degraded BER performance. Besides, the BER performance is further evaluated by varying the transmit data rate at an SNR of 25 dB for 16-QAM and 38 dB for 64-QAM considering relevant CIRs for for "residential" scenario in Fig. 1, and for the "manufacturing cell" scenario in Fig. 2 [3]. Lastly, in the presented simulations for the BER vs SNR performance for the RFF-IMSC and the RFF-SKW-IMSC, the analytical bounds in (8) are validated for both 16-QAM and 64-QAM in Fig. 1 and Fig. 2. From the simulations, it is found that among other unsupervised post-distorters, the proposed RFF-IMSC, and RFF-SKW-IMSC post-distorter support higher data rates and deliver improved BER performance.

## V. Conclusion

In this paper, RFF-IMSC based unsupervised post-distortion is proposed to enhance VLC links impaired by LED nonlinearity and ISI. To mitigate these impairments, this paper proposes an RFF-IMSC based unsupervised post-distorter for VLC. Furthermore, a hyperparameter free RFF-IMSC post-distorter, namely the RFF-SKW-IMSC, is proposed using a kernel width sampling technique, and generic optimality guarantees are presented for these RFF based approaches. Computer-simulations performed over standardized IEEE 802.15 PAN VLC channels indicate that the proposed unsupervised RFF-IMSC post-distorter delivers improved BER-performance as compared to the existing polynomial series based Volterra-MCMMA and Hammerstein-IMSC based approaches.

[1]From classical works, this deviation is found to increase for higher modulation-orders [35, Fig. 7].

## Appendix A
## Proof of Theorem I

From [14], it is noted that the unsupervised RFF-IMSC cost-function is an approximation of the MCC loss-function. The MCC cost-function, $J_{\text{MCC}}$, is denoted as:

$$J_{\text{MCC}} = \exp(-\upsilon|s_{n-\tau} - \boldsymbol{\Omega}^T\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)|^2), \tag{11}$$

$\upsilon$ denotes the spread-factor for MCC. At convergence, the following holds [14]:

$$|\mathcal{J}^R_{\text{RFF-IMSC}} + \mathcal{J}^I_{\text{RFF-IMSC}} - \mathcal{J}^R_{\text{MCC}} - \mathcal{J}^I_{\text{MCC}}| < \epsilon(\boldsymbol{\Omega}), \tag{12}$$

for a small and bounded $\epsilon(\boldsymbol{\Omega})$, with $\|\epsilon\| < C$, where $C$ is a constant. Noting that $\mathcal{J}^R_{\text{RFF-IMSC}} + \mathcal{J}^I_{\text{RFF-IMSC}}$ is Lipschitz continuous in the argument $e_{\text{RFF-IMSC}}$ and for a supervised Lipschitz continuous convex loss-function [36], $\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n), \boldsymbol{\Omega}))$:

$$|\mathcal{J}_{\text{RFF-IMSC}}| \leq \Upsilon|e_{\text{RFF-IMSC}}|, \tag{13}$$
$$|\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n), \boldsymbol{\Omega}))| \leq \Theta|s_{n-\tau} - \boldsymbol{\Omega}^T\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)|,$$

where $\Upsilon, \Theta$ are Lipschitz constants. and $\Theta$ is a Lipschitz constant. Additionally, for high SNR, we have $e_{\text{RFF-IMSC}} \to 0$ and $|s_{n-\tau} - \boldsymbol{\Omega}^T\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)| \to 0$. Invoking the asymptotic equivalence of IMSC and the MCC cost-function, we write:

$$|e_{\text{RFF-IMSC}} - s_{n-\tau} + \boldsymbol{\Omega}^T\hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)| < \Delta, \tag{14}$$

for an infinitesimal $\Delta$. This enables re-expression of (13) as:

$$|\mathcal{J}_{\text{RFF-IMSC}}| \leq \Upsilon(|s_{n-\tau} - \mathbf{\Omega}^T \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)| + \Delta), \quad (15)$$
$$|\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n), \mathbf{\Omega}))| \leq \Theta(|s_{n-\tau} - \mathbf{\Omega}^T \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)|)$$

Simplifying (15) gives the following value of $C$:

$$C = (\Upsilon - \Theta)|s_{n-\tau} - \mathbf{\Omega}^T \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)| + \Delta. \quad (16)$$

## APPENDIX B
## PROOF OF THEOREM II

Using the log-sum-exp inequality for the supervised squared-error objective, we have $(\Upsilon - \Theta) \approx \sqrt{2\rho \log(Q)}$. At convergence, the variance of the mis-adjustment of the proposed approach, $\sigma^2_{\text{RFF-IMSC}}$, is bounded as:

$$\sigma_n^2\Big(1 + \frac{\eta}{2D}\Big) \leq \sigma^2_{\text{RFF-IMSC}} \leq \sigma_n^2\Big(1 + \frac{\eta}{2D}\Big) + \frac{2\rho \log(Q)}{Q}. \quad (17)$$

Noting that both the underlying and the residual estimation-error are zero mean Gaussian random-variables [11], [37], the variance-range in (17) is mapped to the BER-range in (8).

## APPENDIX C
## PROOF OF THEOREM III

For the IMSC-RFF-SKW, [27] finds that $\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF-SKW}}_{\tilde{\sigma}}(\mathbf{x}_n), \mathbf{\Omega}))$ lies in an error-neighborhood of $\mathcal{L}(s_{n-\tau}; \hat{\Psi}^{\text{RFF}}_{\tilde{\sigma}}(\mathbf{x}_n), \mathbf{\Omega}))$. In details, this error-radius is bounded by $\mathcal{O}(\frac{1}{D})$. Using the SKW loss-function leads us to the following equivalent expression of $C$ in (12), $C_{\text{SKW}}$:

$$C_{\text{SKW}} = (\Upsilon - \Theta)|s_{n-\tau} - \mathbf{\Omega}^T \hat{\Psi}^{\text{RFF}}(\mathbf{x}_n)| + \Delta + \frac{\mathcal{A}}{D}, \quad (18)$$

where $\mathcal{A}$ is an arbitrary bounded constant. Notably the offset $\frac{\mathcal{A}}{D}$ decays with increase in the number of RFFs.

## REFERENCES


[1] H. Haas *et al.*, "Introduction to indoor networking concepts and challenges in LiFi," *J. Opt. Commun. Netw.*, vol. 12, no. 2, pp. A190–A203, 2020.

[2] S. Dimitrov and H. Haas, *Principles of LED light communications: towards networked Li-Fi*. Cambridge University Press, 2015.

[3] M. Uysal, F. Miramirkhani, T. Baykas, and K. Qaraqe, "IEEE 802.11bb reference channel models for indoor environments," doc.: IEEE 11-18-1582-02-00bb, 2018.

[4] R. Mitra and V. Bhatia, "Adaptive sparse dictionary-based kernel minimum symbol error rate post-distortion for nonlinear LEDs in visible light communications," *IEEE Photonics J.*, vol. 8, no. 4, pp. 1–13, 2016.

[5] Z. Ghassemlooy, L. N. Alves, S. Zvanovec, and M.-A. Khalighi, *Visible Light Communications: Theory and Applications*. CRC press, 2017.

[6] J. K. Kim, K. Hyun, and S. K. Park, "Adaptive predistorter using NLMS algorithm for nonlinear compensation in visible-light communication system," *Electronics Letters*, vol. 50, no. 20, pp. 1457–1459, 2014.

[7] R. Mitra and V. Bhatia, "Chebyshev polynomial-based adaptive predistorter for nonlinear LED compensation in VLC," *IEEE Photon. Technol. Lett.*, vol. 28, no. 10, pp. 1053–1056, 2016.

[8] G. Stepniak, J. Siuzdak, and P. Zwierko, "Compensation of a VLC phosphorescent white LED nonlinearity by means of Volterra DFE," *IEEE Photon. Technol. Lett.*, vol. 25, no. 16, pp. 1597–1600, 2013.

[9] R. Mitra, F. Miramirkhani, V. Bhatia, and M. Uysal, "Mixture-kernel based post-distortion in RKHS for time-varying VLC channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1564–1577, 2018.

[10] S. Jain, R. Mitra, and V. Bhatia, "KLMS-DFE based adaptive post-distorter for visible light communication," *Optics Communications*, vol. 451, pp. 353–360, 2019.

[11] R. Mitra, F. Miramirkhani, V. Bhatia, and M. Uysal, "Low Complexity Least Minimum Symbol Error Rate Based Post-Distortion for Vehicular VLC," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11 800–11 810, 2020.

[12] H. Qian, S. Yao, S. Cai, and T. Zhou, "Adaptive postdistortion for nonlinear LEDs in visible light communications," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–8, 2014.

[13] Y. Wang, L. Tao, Y. Wang, and N. Chi, "High speed WDM VLC system based on multi-band CAP64 with weighted pre-equalization and modified CMMA based post-equalization," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1719–1722, 2014.

[14] R. Mitra and V. Bhatia, "Unsupervised multistage-clustering-based Hammerstein postdistortion for VLC," *IEEE Photon. J.*, vol. 9, no. 1, pp. 1–10, 2017.

[15] C. Chen, X. Deng, Y. Yang, P. Du, H. Yang, and L. Zhao, "LED nonlinearity estimation and compensation in VLC systems using probabilistic bayesian learning," *Applied Sciences*, vol. 9, no. 13, p. 2711, 2019.

[16] L. Tao, Y. Wang, Y. Gao, A. P. T. Lau, N. Chi, and C. Lu, "Experimental demonstration of 10 Gb/s multi-level carrier-less amplitude and phase modulation for short range optical communication systems," *Optics express*, vol. 21, no. 5, pp. 6459–6465, 2013.

[17] L. Tao, Y. Ji, J. Liu, A. P. T. Lau, N. Chi, and C. Lu, "Advanced modulation formats for short reach optical communication systems," *IEEE Network*, vol. 27, no. 6, pp. 6–13, 2013.

[18] S. Chen, S. McLaughlin, P. M. Grant, and B. Mulgrew, "Multi-stage blind clustering equaliser," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 701–705, 1995.

[19] R. Mitra, S. Singh, and A. Mishra, "Improved multi-stage clustering-based blind equalisation," *IET communications*, vol. 5, no. 9, pp. 1255–1261, 2011.

[20] B. Chen, Y. Zhu, J. Hu, and J. C. Principe, *System parameter identification: information criteria and algorithms*. Newnes, 2013.

[21] P. Bouboulis, S. Pougkakiotis, and S. Theodoridis, "Efficient KLMS and KRLS algorithms: A random Fourier feature perspective," *2016 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 1–5, June 2016.

[22] R. Mitra, S. Jain, and V. Bhatia, "Least Minimum Symbol Error Rate Based Post-Distortion for VLC Using Random Fourier Features," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 830–834, 2020.

[23] R. Mitra and V. Bhatia, "Low complexity post-distorter for visible light communications," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 1977–1980, 2017.

[24] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*. John Wiley & Sons, 2011, vol. 57.

[25] B. Chen, J. Liang, N. Zheng, and J. C. Principe, "Kernel least mean square with adaptive kernel size," *Neurocomputing*, vol. 191, pp. 95–106, 2016.

[26] Y. Mohsenzadeh and H. Sheikhzadeh, "Gaussian kernel width optimization for sparse Bayesian learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 4, pp. 709–719, 2014.

[27] R. Mitra, G. Kaddoum, and V. Bhatia, "Hyperparameter-free transmit-nonlinearity mitigation using a kernel-width sampling technique," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2613–2627, 2021.

[28] M.-A. Khalighi, S. Long, S. Bourennane, and Z. Ghassemlooy, "PAM- and CAP-Based Transmission Schemes for Visible-Light Communications," *IEEE Access*, vol. 5, pp. 27 002–27 013, 2017.

[29] S. Jain, R. Mitra, and V. Bhatia, "Adaptive precoding-based detection algorithm for massive MIMO visible light communication," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1842–1845, 2018.

[30] H. Elgala, R. Mesleh, and H. Haas, "An LED model for intensity-modulated optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 22, no. 11, pp. 835–837, 2010.

[31] J. Grubor, S. Randel, K.-D. Langer, and J. W. Walewski, "Broadband information broadcasting using LED-based interior lighting," *J. Lightwave Technol.*, vol. 26, no. 24, pp. 3883–3892, 2008.

[32] M. Uysal, T. Baykas, F. Miramirkhani, N. Serafimovski, and V. Jungnickel, "TG7r1 channel model document for high rate PD communications," Tech. Rep., 2015.

[33] M. Uysal, F. Miramirkhani, O. Narmanlioglu, T. Baykas, and E. Panayirci, "IEEE 802.15. 7r1 reference channel models for visible light communications," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 212–217, 2017.

[34] S. Wang *et al.*, "Random Fourier filters under maximum correntropy criterion," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 65, no. 10, pp. 3390–3403, 2018.

[35] S. Abrar and R. A. Axford Jr, "Sliced multi-modulus blind equalization algorithm," *ETRI journal*, vol. 27, no. 3, pp. 257–266, 2005.

[36] J. Heinonen, *Lectures on Lipschitz analysis*. University of Jyväskylä, 2005, no. 100.

[37] S. Jain, R. Mitra, and V. Bhatia, "Hybrid Adaptive Precoder and Post-Distorter for Massive-MIMO VLC," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 150–154, 2019.