Kernel Recursive Maximum Versoria Criterion Based Post-Distorter for VLC Using Kernel-Width Sampling

Sandesh Jain, Rangeet Mitra, *Member, IEEE*, Ondrej Krejcar, Jamel Nebhen, and Vimal Bhatia, *Senior Member, IEEE*

***Abstract*—Visible light communication (VLC) has emerged as a potential candidate for next generation wireless communication systems. However, nonlinear characteristics of light emitting diode (LED), user-mobility, and DC-bias fluctuations are the major factors that limit the throughput of a VLC link, and makes the overall additive distortion as non-Gaussian distributed. To mitigate this non-Gaussian noise processes encountered in VLC systems due to LED nonlinearity, and user-mobility, recently a random Fourier features (RFF) based kernel recursive maximum Versoria criterion (KRMVC) based post-distortion algorithm is proposed, which delivers better performance as compared to the classical polynomial series, and kernel recursive least squares (KRLS) algorithms due to the incorporation of higher order statistics of error. However, the performance of RFF-KRMVC algorithm is sensitive to the choice of kernel-width, and results in approximation errors due to imperfect choice of kernel-width. This paper proposes a novel RFF-KRMVC algorithm using a kernel-width sampling (KWS) technique called as RFF-KWS-KRMVC, which implements the post-distortion under a hyperparameter-free finite memory budget. Furthermore, analytical expressions for mean square error, and error rate are quantified for the proposed RFF-KWS-KRMVC post-distorter, and corroborated by Monte-Carlo simulations performed over standard VLC channel models.**

Manuscript received January 6, 2022; revised March 7, 2022; accepted March 25, 2022. Date of publication March 31, 2022; date of current version April 28, 2022. This work was supported in part by the Project of Operational Programme Integrated Infrastructure: Independent research and development of technological kits based on wearable electronics products, as tools for raising hygienic standards in a society exposed to the virus causing the COVID-19 disease, ITMS2014+ code 313011ASK8, in part by the European Regional Development Fund, and in part by the Grant Agency of Excellence, University of Hradec Kralove, Faculty of Informatics and Management, Czech Republic under Project ID 2204/2022. *(Corresponding author: Sandesh Jain.)*

Sandesh Jain is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: sandeshjain@iisc.ac.in).

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

Ondrej Krejcar is with the Faculty of Informatics and Management, University of Hradec Kralove, 50003 Hradec Kralove, Czech Republic, with the Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia, and also with the University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia (e-mail: ondrej.krejcar@uhk.cz).

Vimal Bhatia is with the Faculty of Informatics and Management, University of Hradec Kralove, 50003 Hradec Kralove, Czech Republic, and also with the Department of Electrical Engineering, Center for Advanced Electronics, Indian Institute of Technology Indore, 453552 Indore, India (e-mail: vbhatia@iiti.ac.in).

Jamel Nebhen is with the College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Alkharj 11942, Saudi Arabia (e-mail: j.nebhen@psau.edu.sa).

Digital Object Identifier 10.1109/JPHOT.2022.3163714

***Index Terms*—VLC, user-mobility, random Fourier features, maximum Versoria criterion, kernel-width sampling.**

I. Introduction

VISIBLE light communication [1], [2] has evolved as an attractive technology for supporting high-speed data communication for the next generation wireless communication systems. VLC relies on intensity modulation of light emitting diode (LED) lamps according to the input message signal at a speed imperceptible to the human eye with photodiodes (PD) used at the receiver for conversion of optical signal to electrical current. The advantages offered by VLC over the classical radio frequency (RF) communication are low cost, license-free wide spectrum, better security due to opaque boundaries, immunity to electromagnetic interference, better spatial reusability, and high signal-to-noise ratio (SNR) [3]. Due to the aforementioned advantages, VLCs are used in wide variety of applications such as light fidelity (Li-Fi) systems, smart lighting, intelligent transportation systems (like vehicle-to-vehicle (V2V) communication, and vehicle-to-infrastructure communication), underwater communication, Internet of Things (IoTs) based eco-systems, and wearable devices.

Despite the aforesaid desirable features, performance of a typical VLC link is degraded by the following channel impairments: (a) LED nonlinearity [4] that results in an uncorrelated additive distortion at the receiver (stems from the Bussgang's theorem [5] of decomposition of nonlinear function), (b) intersymbol interference due to limited modulation bandwidth of LED, (c) user-mobility [6], [7] that leads to an effective multiplicative distortion, which degrades the VLC link due to outages and non-Gaussian additive distortion, and (d) fluctuations in DC-bias. The aforementioned VLC channel impairments lowers the overall SNR, and degrades the error rate performance significantly.

A. Related Works

Existing works are based on classical polynomial filtering [4], [8], reproducing kernel Hilbert space (RKHS) [9], [10], and artificial/deep neural networks (ANN/DNN) [11], [12] based equalizers/post-distorters for mitigating LED nonlinearity encountered in VLC systems. However, performance of polynomial filtering based approaches is limited by approximation errors caused by finite order truncation of polynomial

series [13]. Furthermore, the existing ANN/DNN based equalizers (like multilayer perceptron, radial basis functions) are non-convex [14, Table 1.1], computationally complex, and are sensitive to choice of various hyperparameters like number of hidden layers, number of nodes in each layer, choice of activation function [15]. Alternatively, Alternatively, RKHS based approaches are convex, computationally efficient, and are found to outperform the conventional polynomial filters since the parameterization of the inverse of LED nonlinearity (or any nonlinear transformation in general) is guaranteed to exist in RKHS, and to learn this parameter, the apriori knowledge of LED characteristics at the receiver is not required [14].

However, the RKHS based approaches [9], [10] are subjected to erroneous observations whilst learning a dictionary during the early training phase. Furthermore, computational complexity of dictionary-based RKHS algorithms [9], [10] cannot be estimated apriori since cardinality of dictionary is data-dependent, which limits its practical implementation under a finite memory budget. To circumvent this limitation, further research is directed towards developing random Fourier features (RFF) [16]–[18] based post-distortion algorithms, which are basically Monte-Carlo approximations of an RKHS, and have lower computational budget whilst ensuring similar convergence performance as compared to the classical RKHS based methods.

The aforementioned RKHS/RFF based post-distorters (namely kernel least mean square (KLMS) [19], kernel minimum symbol error rate (KMSER) [9], KLMS with decision feedback equalizer (KLMS-DFE) [20], and RFF-KMSER-DFE [18]) are based on stochastic gradient descent (SGD) paradigms that have lower computational complexity, however higher steady-state mean square error (MSE). Further, the aforesaid SGD based post-distortion algorithms considers only instantaneous error for formulation of loss function, and hence delivers suboptimal performance under high date rate regime [21]. Furthermore, to achieve a trade-off between the convergence rate, and the MSE, the existing KLMS-DFE, and RFF-KMSER-DFE based post-distorters consider adaptive step-size, which itself are sensitive to the choice of various hyperparameters as given in [18, eq. (17)–(18)].

To further improve the convergence performance of RKHS based algorithms, least-squares based kernel recursive least squares (KRLS) [14], least KMSER [21], and kernel recursive maximum Versoria criterion (KRMVC) [22] algorithms have been proposed in the literature for a VLC post-distortion algorithm. These algorithms consider all the previous error terms for objective function formulation, and hence outperforms the existing SGD based approaches. Recently, Versoria criterion [23]–[25] based RFF-KRMVC algorithm is found to outperform the conventional minimum mean square error (MMSE) based RFF-KRLS post-distorter due to the incorporation of order statistics, and lower steady-state misadjustment as has been proved in the literature [22], [25].

However, the performance of RKHS/RFF based post-distorters [18], [21], and the Gaussian kernel based DNN (GK-DNN) equalizers [12] are sensitive to the choice of appropriate kernel-width, and results in degraded performance upon improper choice of kernel-width. There are several methods in the signal processing literature to select proper kernel-width like: (a) Silverman's and Scott's rule of thumb [14], (b) adaptive techniques [26], [27] for learning kernel-width based on various criteria such as MMSE, minimum symbol error rate (MSER) etc., and (c) Bayesian leaning methods [28]. However, Silverman's rule is simple to implement but prone to estimation errors. Further, adaptive techniques require additional computational steps, and large number of iterations to converge upon poor initialization of kernel-width. Bayesian learning techniques rely on computation of complex likelihood functions for optimization of kernel-width. Recently, RFF-KRLS based post-distorter is proposed using kernel-width sampling (KWS) technique [29], that alleviates the need for hyperparameter initialization. However, RFF-KWS-KRLS [29] algorithm rely on MMSE criterion, which is not suitable to deliver optimal performance over non-Gaussian distortions.

B. Contributions

In this paper, for the first time, the paradigm of KWS is explored for Versoria criterion based algorithm. Following are the major contributions of this paper:

1) In this paper, a novel KWS technique is proposed for RFF-KRMVC to yield RFF-KWS-KRMVC algorithm, which mitigates the non-Gaussian additive distortion due to LED nonlinearity and user-mobility, and facilitates hyperparameter-free solution to a VLC post-distortion problem. Furthermore, the proposed RFF-KWS-KRMVC post-distorter is derived from the second order adaptive filters, which updates both the mean and the covariance matrix, and hence are found to outperform the existing first order RKHS (KLMS [10], KMSER [9], KLMS-DFE [20], and RFF-KMSER-DFE [18]) based algorithms which updates mean alone.
2) Analytical MSE for the proposed RFF-KWS-KRMVC post-distorter is quantified, which is found to be smaller than the existing RFF-KWS-KRLS algorithm.
3) The probability density function (PDF) of the effective additive distortion for a VLC link degraded by user-mobility, LED nonlinearity, and unstable DC-bias is quantified.
4) Using the derived PDF of the effective additive distortion, an analytical upper bound on bit error rate (BER) is derived for the proposed post-distortion algorithm for a VLC link that considers user-mobility, LED nonlinearity, and DC-bias error together as opposed to user-mobility, and LED nonlinearity alone in the existing works [18], [22].
5) Monte-Carlo simulations are performed for random waypoint (RWP) mobility model for both memoryless, and memory LED nonlinearity as opposed to memoryless nonlinearilty alone in the existing works [9], [18], [22]. Simulations indicate that the proposed RFF-KWS-KRMVC post-distorter outperforms the existing RFF-KRLS, RFF-KWS-KRLS, and Volterra-RLS approaches.

Furthermore, the contributions of the proposed post-distorter for VLC as compared to the existing post-distortion algorithms are enlisted in Table I.

TABLE I
COMPARISON OF THE PROPOSED ALGORITHM WITH THE EXISTING APPROACHES

Algorithm	VLC Channel impairments				Dictionary or RFF	SGD or least-squares	Learning Criterion	Whether suitable for non-Gaussian distortions	Hyper-parameter-free
	LED non-linearity	ISI	user-mobility	DC-bias error					
KLMS [10]	✓	✗	✗	✗	Dictionary	SGD	MMSE	✗	✗
KMSER [9], [19]	✓	✓	✗	✗	Dictionary	SGD	MSER	✓	✗
KLMS-DFE [20]	✓	✓	✗	✗	Dictionary	SGD	MMSE	✗	✗
RFF-KMSER-DFE [18]	✓	✓	✓	✗	RFF	SGD	MSER	✓	✗
RFF-KRLS [21]	✓	✓	✗	✗	RFF	least-squares	MMSE	✗	✗
RFF-KRMVC [22]	✓	✓	✓	✗	RFF	least-squares	MVC	✓	✗
RFF-KWS-KRLS [29]	✓	✓	✗	✗	RFF	least-squares	MMSE	✗	✓
Proposed RFF-KWS-KRMVC	✓	✓	✓	✓	RFF	least-squares	MVC	✓	✓

C. Organization

Rest of the paper is structured as follows: System model for VLC is presented in Section II. An overview of RFF based learning in RKHS is given in Section III. Proposed RFF-KWS-KRMVC post-distortion algorithm is described in Section IV. Analytical results for the proposed algorithm are derived in Section V. Section VI presents the simulation results, and finally Section VII concludes the paper.

D. Notations

Capital, and small boldface letters represent matrices, and vectors, respectively while scalars are denoted by simple lower-case letters. $(\cdot)_k$ or $(\cdot)(k)$ denotes $(\cdot)$ at k^{th} time instance. Inner product in RKHS $\mathscr{H}$ is denoted by $\langle\cdot,\cdot\rangle_{\mathscr{H}}$. Hermitian transpose is represented by $(\cdot)^H$, $\mathrm{Tr}(\cdot)$ denotes trace of a matrix, and $\mathbf{I}_{(\cdot)}$ represents identity matrix of order $(\cdot)$. A Gaussian distribution is denoted by $\mathcal{N}(\cdot,\cdot)$, uniform distribution by $\mathcal{U}(\cdot,\cdot)$, and gamma distribution by $\mathrm{Gamma}(\cdot,\cdot)$. The quantity $[g(h)]_{h=h_{\min}}^{h=h_{\max}}$ is defined as $g(h_{\max}) - g(h_{\min})$, and $\Gamma(\cdot,\cdot)$ denotes upper incomplete gamma function.

II. SYSTEM MODEL

In this section, the considered VLC system model is described, where the transmitted signal model is described in Section II(A) followed by VLC channel model, and received signal model in Section II(B), and Section II(C), respectively.

A. Transmitted Signal Model

In this subsection, the transmitted carrierless amplitude phase QAM (CAP-QAM) signal model, and the LED nonlinear model is described. First, the transmitted data bits are modulated using QAM technique, and decomposed into in-phase, and quadrature-phase symbols. Next, the real and the imaginary parts of QAM symbols are given as an input to orthogonal in-phase $h_I(t)$, and quadrature-phase $h_Q(t)$ filter, respectively, which are related as Hilbert transform pair, and given as [30], [31]

$$h_I(t) = p(t)\cos(2\pi f_0 t), \text{ and } h_Q(t) = p(t)\sin(2\pi f_0 t), \quad (1)$$

where $p(t)$ denotes the impulse response of baseband filter. The overall real CAP-QAM transmitted signal is written as [30], [31]

$$x_k(t) = \sum_{k=-\infty}^{\infty} s_I(k)h_I(t-kT) - s_Q(k)h_Q(t-kT), \quad (2)$$

where $s_I(k)$, and $s_Q(k)$ represents real, and imaginary parts of QAM symbols, respectively. Next, sufficient DC offset is added to the transmitted symbol to operate the LEDs in forward biased mode. LED response exhibit nonlinearity since current-voltage, and current-optical power conversion is nonlinear. In particular, LEDs characteristics are nonlinear for signals with large dynamic range, and at high switching frequencies. This work considers Wiener model for modeling LEDs nonlinearity with memory effects [32], which consists of a cascade of linear time-invariant (LTI) system followed by a memoryless Rapp's nonlinear block. The output of the LTI block is written as [32]

$$x'_k = \sum_{m=0}^{M-1} b_m(x_{k-m} + b) \quad (3)$$

where M, and $\{b_m\}_{m=0}^{N}$ represents the memory, and weights for LTI system, and b is an appropriate DC bias. The memoryless Rapp's model for LED, denoted by $A_{\mathrm{LED}}(\cdot)$ is given as [33], [34]

$$A_{\mathrm{LED}}(x'_k) = \begin{cases} \dfrac{(x'_k - V_{\mathrm{th}})}{\left(1+\left(\frac{(x'_k - V_{\mathrm{th}})}{V_{\max}}\right)^{2p}\right)^{\frac{1}{2p}}} & x'_k \geq V_{\mathrm{th}} \\ 0 & x'_k < V_{\mathrm{th}} \end{cases} \quad (4)$$

where $V_{\max}$ is the maximum saturation voltage, V_{th} is the cut-in voltage of LED, and p controls the level of nonlinearity called as knee factor. Finally, the composite optical signal is transmitted by VLC link. The considered VLC channel model is described in the next subsection.

B. VLC Channel Model

In this subsection, the VLC channel model considering user-mobility is described. The standard IEEE 802.15 PAN VLC channel models [35] are deterministic in nature where the users are assumed to be fixed at their locations, and user-mobility is neglected. However, in a practical scenario, users can move while reading, talking on a cell phone, shopping in a mall etc.,

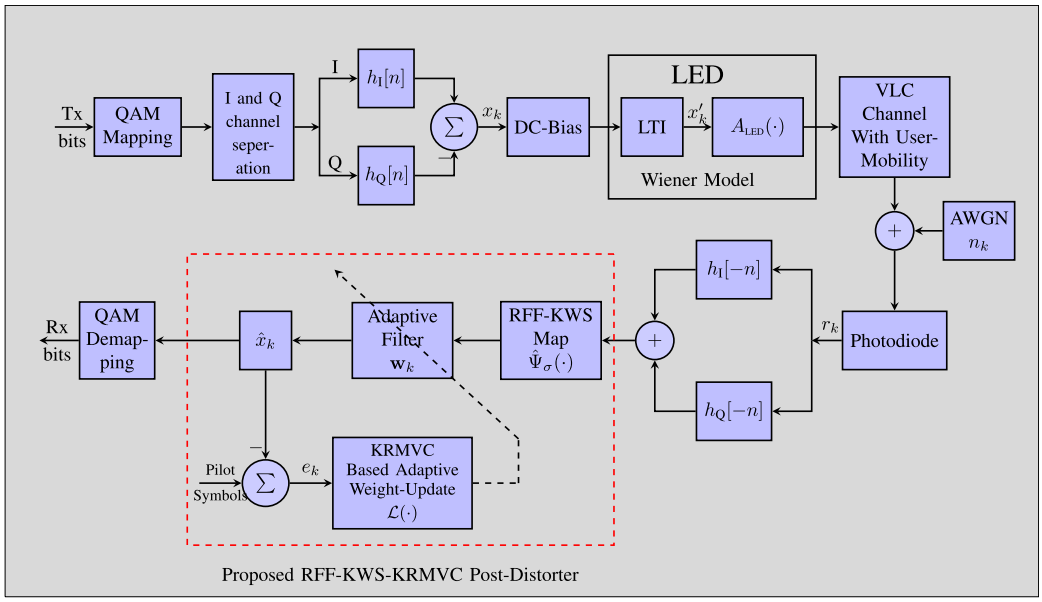


Fig. 1. Block diagram of the proposed system model.

which makes the overall VLC channel probabilistic in nature. In this work, a RWP mobility model is considered [7], which considers non-uniform distribution of users, and is more realistic as compared to the uniform distribution of users considered in [36]. The RWP probabilistic model for VLC (denoted by $f_H(h)$) is given as [7], [37]

$$f_H(h) = \begin{cases} \sum_{i=1}^{4} K_i h^{-c_i} & h \in [h_{\min},\, h_{\max}] \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $c_1 = (\frac{2}{m+3}) + 1$, $c_2 = c_4 = (\frac{4}{m+3}) + 1$, $c_3 = (\frac{6}{m+3}) + 1$, $K_1 = \tilde{K}\mathcal{K}$, $K_2 = \frac{-35\tilde{K}}{r_e^2}\kappa^{2\beta}$, $K_3 = \frac{8\tilde{K}}{r_e^4}\kappa^{4\beta}$, $K_4 = \frac{-16\tilde{K}L^2}{r_e^4}\kappa^{2\beta}$, $h_{\min} = \frac{\kappa}{(r_e^2+L^2)^{\frac{m+3}{2}}}$, and $h_{\max} = \frac{\kappa}{L^{m+3}}$, where $\tilde{K} = \frac{12\kappa^{2\beta}}{73(m+3)r_e^2}$, $\mathcal{K} = 27 + \frac{35L^2}{r_e^2} + \frac{8L^4}{r_e^4}$, $\beta = \frac{1}{m+3}$, r_e denotes the user's maximum coverage radius, $\kappa = \frac{A_d}{2\pi}(m+1)L^{m+1}$, A_d represents geometric area of receiver, and $m = -\frac{\log(2)}{\log(\cos(\phi_{\frac{1}{2}}))}$, where $\phi_{\frac{1}{2}}$ denotes half power semi-angle of LED.

C. Received Signal Model

The composite received signal by the PD is written as

$$r_k = \begin{cases} hA_{\text{LED}}(x'_k) + n_k & \text{for memory nonlinearity} \\ hA_{\text{LED}}(x_k + b) + n_k & \text{for memoryless nonlinearity,} \end{cases} \tag{6}$$

where $n_k \sim \mathcal{N}(0, \sigma_n^2)$ denotes VLC noise, which consists of ambient light, and thermal noise [38], [39]. The received symbol vector $\mathbf{r}_k = [r_k,\, r_{k-1}, \ldots, r_{k-n+1}]^T$ is sampled and passed through in-phase $h_I(-n)$, and quadrature-phase $h_Q(-n)$ filters to extract real, and imaginary parts of the received symbol, and applied as an input to the proposed RFF-KWS-KRMVC post-distorter detailed in Section IV. Lastly, symbols are passed through a QAM demodulator to generate estimate of transmitted bits $\hat{b}_k$.

III. Overview of RFF Based Learning in RKHS

In this section, mathematical insights for RFF is provided in order to motivate the proposed RFF-KWS-KRMVC post-distorter for VLC. From representer theorem, any arbitrary function $g(\cdot)$ can be represented in RKHS as a weighted combination of Mercer kernel functions (denoted by $\mathcal{K}(\cdot,\cdot)$) given as [14]

$$g(\mathbf{r}_k) = \sum_{i=1}^{N} a_i \mathcal{K}(\mathbf{r}_i, \mathbf{r}_k) \tag{7}$$

However, (7) relies on growing network of kernel evaluation functions with newly added input at each iteration. Hence, a Mercer kernel function $\mathcal{K}(\cdot,\cdot)$ can be approximated by using RFF based explicit feature maps, which is directly motivated from Bochner's theorem restated as follows [40], [41]:

Theorem 1: A continuous function $\psi(\cdot) \sim \mathbb{R}^n$ is positive definite iff there exists a finite non-negative Borel measure $P(\cdot)$ on $\mathbb{R}^d$ such that

$$\psi(\mathbf{r}) = \int_{\mathbb{R}^n} \exp(j\boldsymbol{\omega}^T \mathbf{r}) P(\boldsymbol{\omega}) d(\boldsymbol{\omega}) \tag{8}$$

In other words, Fourier transform $P(\boldsymbol{\omega})$ (where ω is the angular frequency) of a shift-invariant positive definite function is a probability distribution.

Using Bochner's theorem, shift-invariant positive definite kernel function $\mathcal{K}(\cdot,\cdot)$ can be written as

$$\mathcal{K}(\mathbf{r}, \mathbf{r}') = \mathcal{K}(\mathbf{r} - \mathbf{r}') = \int_{\mathbb{R}^n} \exp(j\boldsymbol{\omega}^T(\mathbf{r} - \mathbf{r}')) P(\boldsymbol{\omega}) d(\boldsymbol{\omega}), \tag{9}$$

where $\mathbf{r}$, and $\mathbf{r}'$ are arbitrary dummy variables. Defining $\hat{\Psi}_{\boldsymbol{\omega}}(\mathbf{y}) = \exp(j\boldsymbol{\omega}^T\mathbf{r})$, an unbiased estimate of (9) can be expressed as [17]

$$\mathcal{K}(\mathbf{r}, \mathbf{r}') = \mathbb{E}\left\{\hat{\Psi}^T_{\boldsymbol{\omega}}(\mathbf{r})\hat{\Psi}_{\boldsymbol{\omega}}(\mathbf{r}')\right\} \tag{10}$$

Using Monte-Carlo sampling technique, (10) is approximated using D independent random features as [40], [41]

$$\mathcal{K}(\mathbf{r}, \mathbf{r}') = \frac{1}{D}\sum_{i=1}^{D}\hat{\Psi}^T_{\boldsymbol{\omega}_i}(\mathbf{r})\hat{\Psi}_{\boldsymbol{\omega}_i}(\mathbf{r}'), \tag{11}$$

where $\{\boldsymbol{\omega}_i\}_{i=1}^D \sim P(\boldsymbol{\omega})$, and explicit feature map $\hat{\Psi} : \mathbb{R}^n \longrightarrow \mathbb{R}^D$ defined as [42]

$$\hat{\Psi}(\mathbf{r}) = \frac{1}{\sqrt{D}}[\exp(j\boldsymbol{\omega}_1^T\mathbf{r}), \exp(j\boldsymbol{\omega}_2^T\mathbf{r}), \ldots, \exp(j\boldsymbol{\omega}_D^T\mathbf{r})]^T$$

For real Gaussian kernel, the explicit feature map $\hat{\Psi}(\cdot)$ can be further approximated by using cosine functions $\cos(\cdot)$ given as follows

$$\hat{\Psi}(\mathbf{r}) = \sqrt{\frac{2}{D}}\begin{bmatrix}\cos(\boldsymbol{\omega}_1^T\mathbf{r} + \alpha_1)\\ \cos(\boldsymbol{\omega}_2^T\mathbf{r} + \alpha_2)\\ \vdots \\ \cos(\boldsymbol{\omega}_D^T\mathbf{r} + \gamma_D)\end{bmatrix}, \tag{12}$$

where $\{\boldsymbol{\omega}_i\}_{i=1}^D \sim \mathcal{N}(\mathbf{0}_D, \frac{1}{\sigma^2}\mathbf{I}_D)$ (σ denotes the kernel-width), and $\{\alpha_i\}_{i=1}^D \sim \mathcal{U}(0, 2\pi)$ is drawn from a uniform distribution [17]. However, the conventional RFF's depends upon the appropriate choice of kernel-width, and results in degraded performance upon improper choice of kernel-width. In the next section, we propose a probabilistic technique for assigning kernel-widths for Versoria criterion based approaches.

IV. Proposed RFF-KWS-KRMVC Based Post-Distorter for VLC

This section describes the proposed RFF-KWS-KRMVC post-distortion algorithm for VLC. For conventional RFFs, the frequencies $\{\omega_i\}_{i=1}^D$ corresponding to Gaussian kernel is drawn from a Gaussian distribution $\mathcal{N}(0, \frac{1}{\sigma^2}\mathbf{I}_D)$, which stems from Bochner's theorem. Similarly, the ensemble of kernel-widths $\{\sigma_i^2\}$ can be drawn from a inverse-Gamma distribution, which is found to be a viable prior distribution for kernel-width from the perspective of Bayesian learning [28]. A RFF map using KWS technique (RFF-KWS) for complex symbols, denoted by $\hat{\Psi}_{\tilde{\sigma}}(\cdot) : \mathbb{R}^{2n} \to \mathbb{R}^D$, is written as [29]

$$\hat{\Psi}_\sigma(\mathbf{r}_k) = \sqrt{\frac{2}{D}}\begin{bmatrix}\cos(\sigma_1\boldsymbol{\omega}_1^T[\Re(\mathbf{r}_k); \Im(\mathbf{r}_k)] + \alpha_1)\\ \cos(\sigma_2\boldsymbol{\omega}_2^T[\Re(\mathbf{r}_k); \Im(\mathbf{r}_k)] + \alpha_2)\\ \vdots \\ \cos(\sigma_D\boldsymbol{\omega}_D^T[\Re(\mathbf{r}_k); \Im(\mathbf{r}_k)] + \alpha_D)\end{bmatrix}, \tag{13}$$

where $\{\boldsymbol{\omega}_i\}_{i=1}^D$ is drawn from standard Gaussian distribution $\mathcal{N}(\mathbf{0}_D, \mathbf{I}_D)$, and $\{\sigma_i\}_{i=1}^D \sim \sqrt{\text{Gamma}(\gamma_1, \gamma_2)}$, where γ_1, and γ_2 are shape, and inverse scale parameter of gamma distribution, respectively. Using the periodicity property of discrete Fourier spectra, and from [29], [43], the suitable values of γ_1, and γ_2 are given as

$$\gamma_1 = \begin{cases}\frac{M}{4} & \text{M-PAM}\\ \frac{\sqrt{M}}{2} & \text{M-CAP-QAM}\end{cases} \tag{14}$$

$$\gamma_2 = \frac{\sigma_r^2}{4\pi^2}, \tag{15}$$

where σ_r^2 denotes the variance of regressors. Based on Versoria function, and training regressors $\{x_i, \hat{\Psi}_\sigma(\mathbf{r}_i)\}_{i=1}^k$, the cost function for the proposed post-distorter is given as

$$\mathcal{L}(\mathbf{w}) = \min_{\mathbf{w}} \sum_{i=1}^{k}\lambda^{k-i}\frac{1}{1 + \tau|x_i - \langle\mathbf{w}_i, \hat{\Psi}_\sigma(\mathbf{r}_i)\rangle_{\mathscr{H}}|^2}, \tag{16}$$

where λ is the forgetting factor close to unity, and τ is the Versoria shape parameter. To solve (16), the gradient $\frac{\partial\mathcal{L}(\mathbf{w})}{\partial\mathbf{w}}$ is set to 0 which yields

$$\mathbf{w}(k) = \mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k)\mathbf{p}^{\text{RFF-KWS}}_{x\hat{\Psi}_\sigma}(k), \tag{17}$$

where

$$\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k) = \left(\sum_{i=1}^{k}\lambda^{k-i}f^{\text{MVC}}(e_i)\hat{\Psi}_\sigma(\mathbf{r}_i)\hat{\Psi}^H_\sigma(\mathbf{r}_i)\right)^{-1}$$
$$= \left(\lambda\mathbf{R}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k-1) + f^{\text{MVC}}(e_k)\hat{\Psi}_\sigma(\mathbf{r}_k)\hat{\Psi}^T_\sigma(\mathbf{r}_k)\right)^{-1} \tag{18}$$

is the inverse of auto-correlation matrix $\mathbf{R}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k)$, and $\mathbf{p}^{\text{RFF-KWS}}_{x\hat{\Psi}_\sigma}(k)$ is the cross-correlation vector given as

$$\mathbf{p}^{\text{RFF-KWS}}_{x\hat{\Psi}_\sigma}(k) = \sum_{i=1}^{k}\lambda^{k-i}f^{\text{MVC}}(e_i)x_i^*\hat{\Psi}_\sigma(\mathbf{r}_i)$$
$$= \lambda\mathbf{p}^{\text{RFF-KWS}}_{x\hat{\Psi}_\sigma}(k-1) + f^{\text{MVC}}(e_k)x_k^*\hat{\Psi}_\sigma(\mathbf{r}_k), \tag{19}$$

where $f^{\text{MVC}}(e_k) = \frac{1}{(1+\tau|e_k|^2)^2}$, and $e_k = x_i - \langle\mathbf{w}_i, \hat{\Psi}_\sigma(\mathbf{r}_i)\rangle_{\mathscr{H}}$ is the prediction error. For $\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^H$, matrix inversion lemma can be written as $\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{C}^T\mathbf{B}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{B}$ [44], and assuming $A = \mathbf{R}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k)$, $\mathbf{B} = \lambda^{-1}(\mathbf{R}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k-1))^{-1}$, $\mathbf{C} = \sqrt{f^{\text{MVC}}(e_k)}\hat{\Psi}_\sigma(\mathbf{r}_k)$, and $\mathbf{D} = \mathbf{I}$, the recursive update equation for $\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k)$ is written as

$$\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k)$$
$$= \frac{1}{\lambda}\left(\mathbf{I} - f^{\text{MVC}}(e_k)\mathbf{g}(k)\hat{\Psi}^T_\sigma(\mathbf{r}_k)\right)\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k-1), \tag{20}$$

where

$$\mathbf{g}(k) = \frac{\lambda^{-1}\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k-1)\hat{\Psi}_\sigma(\mathbf{r}_k)}{1 + \lambda^{-1}f^{\text{MVC}}(e_k)\hat{\Psi}^T_\sigma(\mathbf{r}_k)\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma\hat{\Psi}_\sigma}(k-1)\hat{\Psi}_\sigma(\mathbf{r}_k)} \tag{21}$$

is the gain vector. Substituting (19)-(21) in (17), the weight update equation for the proposed RFF-KWS-KRMVC post-distorter is written as

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \mathbf{g}(k)f^{\text{MVC}}(e_k)e_k. \tag{22}$$

As observed from (22), the term $f^{\text{MVC}}(e_k) \to 0$ for large values of error $e_k \to \infty$, and hence the parameter learning for the proposed RFF-KWS-KRMVC post-distorter is unaffected by transients due to non-Gaussian distortions that arises due to residual nonlinearity, and user-mobility. Furthermore, there is no requirement of tuning hyperparameters (like kernel-width, γ_1, and γ_2) for the proposed RFF-KWS-KRMVC post-distorter, which makes the proposed algorithm attractive for post-distortion over VLC links.

Furthermore, for the existing fixed kernel-width based RFF-KRMVC algorithm, and GK-DNN equalizer, kernel-width needs to be tuned/estimated separately for different deployment scenarios as in [12], [22]. On the other hand, the kernel-width for the proposed RFF-KWS-KRMVC post-distorter is probabilistic in nature, and hence there is no requirement for hyperparameter tuning for the proposed RFF-KWS-KRMVC, which makes the proposed approach generic for different deployment scenarios/noise distributions. The pseudo code for the proposed algorithm is given in **Algorithm** 1.

Algorithm 1: Proposed RFF-KWS-KRMVC Post-Distorter for VLC.

Input: Received observations $\mathbf{r}_k$

Initialization:

- $k = 1$, Number of pilots: NUMPILOTS, λ, ϵ, τ, γ_1, γ_2, $\mathbf{w}_1 = \mathbf{0}$, $\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma \hat{\Psi}_\sigma}(1) = \epsilon \mathbf{I}_{D\times D}$, and RFF Dimension D
- Generate iid $\{\boldsymbol{\omega}_i\}_{i=1}^{D} \sim \mathcal{N}(0, \mathbf{I}_D)$ from normal distribution.
- Generate iid $\{\gamma_i\}_{i=1}^{D} \sim \mathcal{U}(0,\, 2\pi)$ from uniform distribution.
- Generate iid $\{\sigma_i\}_{i=1}^{D} \sim \sqrt{\text{Gamma}(\gamma_1, \gamma_2)}$

Computation:

while $k \leq$ NUMPILOTS **do**

Compute RFF-KWS vector: $\hat{\Psi}_\sigma(\mathbf{r}_k)$ via (13).

Compute post-distorter output: $\hat{x}_k = \mathbf{w}^T(k)\hat{\Psi}_\sigma(\mathbf{r}_k)$

Compute prediction error: $e_k = x_k - \hat{x}_k$

Update gain vector $\mathbf{g}(k)$ via (21)

Update weight vector:

$\mathbf{w}(k) = \mathbf{w}(k-1) + \mathbf{g}(k) f^{\text{MVC}}(e_k) e_k$

Update covariance matrix:

$\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma \hat{\Psi}_\sigma}(k) =$

$\frac{1}{\lambda}(\mathbf{I} - f^{\text{MVC}}(e_k)\mathbf{g}(k)\hat{\Psi}^T_\sigma(\mathbf{r}_k))\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_\sigma \hat{\Psi}_\sigma}(k-1)$

end while

Output: Estimated symbols $\hat{\mathbf{x}}_k$

A. Computational Complexity

The step-wise computational complexity for the proposed RFF-KWS-KRMVC algorithm is given as follows: (1) Computation of RFF map $\hat{\Psi}_\sigma(\mathbf{r}_k)$ involves D independent inner products of two $2n \times 1$ vectors $\mathbf{r}_k$, and $\{\boldsymbol{\omega}_i\}_{i=1}^{D}$ (which requires $2nD$ number of multiplications, and $(2n-1)D$ additions), and D additions for addition of $\{\alpha_i\}_{i=1}^{D}$ to $\{\sigma_i \boldsymbol{\omega}_i^T \mathbf{r}_k\}_{i=1}^{D}$. (2) Computation of output $\hat{x}_k$ and error e_k involves inner product of two $D \times 1$ vectors, which require D multiplications and D additions. (3) Computation of $\mathbf{g}(k)$ requires a matrix-vector product in the numerator, which requires D^2 multiplications, and $D(D-1)$ additions, and a vector-vector product, which requires D multiplications, and $(D-1)$ additions. (4) The weight update in (22) involves multiplication of a scalar $f^{\text{MVC}}(e_k)e_k$ with each element of $D \times 1$ vector $\mathbf{g}(k)$, which requires D multiplications, and addition of two $D \times 1$ vectors ($\mathbf{w}_{k-1}$, and $\mathbf{g}(k)f^{\text{MVC}}(e_k)e_k$), which requires D additions. (5) Lastly, the covariance update in (20) involves a matrix-vector product, outer product of two vectors, and subtraction of two $D \times D$ matrices, which requires $D(3D+1)$ multiplications, and $D(2D-1)$ additions. Hence, the proposed RFF-KWS-KRMVC algorithm requires $(4D^2 + D(n+4)) \approx O(D^2)$ multiplications per iteration, $3D^2 + D(n+1) \approx O(D^2)$ additions per iteration, and memory requirement of order $O(D^2)$ which is similar to the existing RFF-KRMVC [22], and RFF-KRLS algorithms [21].

V. Analytical Results for the Proposed RFF-KWS-KRMVC Post-Distorter

In this section, analytical results dictating MSE, and BER for the proposed algorithm is quantified in Section V(A), and Section V(B), respectively.

A. Convergence Analysis

In this subsection, convergence analysis for the proposed RFF-KWS-KRMVC algorithm is performed. Let $\widetilde{\mathbf{w}}(k) = \mathbf{w}_o - \mathbf{w}(k)$ be the deviation of weight vector $\mathbf{w}(k)$ from the optimal weight vector $\mathbf{w}_0$. The desired input can be written as: $\mathbf{x}_k = \mathbf{w}_0^T \hat{\Psi}_\sigma(\mathbf{r}_k) + n_k$, where n_k is the measurement noise, and $\widetilde{\hat{\Psi}}_{\sigma_o}(\cdot) = \hat{\Psi}_\sigma(\cdot) - \hat{\Psi}_{\sigma_o}(\cdot)$, where σ_o is the optimal kernel-width. Using the standard assumptions in adaptive filters theory [22], [44], the MSE for the proposed algorithm is given by the following theorem:

Theorem 2: The overall MSE at the k^{th} iteration for the proposed RFF-KWS-KRMVC post-distorter under the high SNR regime is given as

$$S_k^{\text{RFF-KWS-KRMVC}} \approx \underbrace{\frac{(1-\lambda^2)}{3(k-n-1)}\left(\sum_{\forall r}\left(\frac{1}{\rho_r}\right) + \frac{2\sigma_o^2 \mathcal{A}||\mathbf{w}||_2^2}{D}\right)\sigma_n^2}_{\sigma^2_{\text{RFF-KWS-KRMVC}}} + \sigma_n^2, \quad k > n+1, \tag{23}$$

where $\{\rho_i\}_{i=1}^{i=r}$ are the Eigen values of Gram matrix, and $\mathcal{A}$ is the arbitrary constant given in [29].

Proof: See Appendix A.

Following conclusions are observed from **Theorem 2**, which are given as follows:

- *Corollary 1:* The excess MSE (EMSE) for the proposed RFF-KWS-KRMVC post-distorter (denoted by $\sigma^2_{\text{RFF-KWS-KRMVC}}$) is similar to the EMSE of the existing

RFF-KRMVC algorithm [22] except for an additional offset error which arises due to the imperfection in kernel-width, and decays with the increase in number of RFF dimensions as $O(\frac{1}{3D})$.

- *Corollary 2:* The offset error for the proposed RFF-KWS-KRMVC algorithm is $\frac{2\sigma_o^2\mathcal{A}\|\mathbf{w}\|_2^2}{3D}$ that is three times smaller as compared to the offset error of existing RFF-KWS-KRLS algorithm [29, eq. (22)], which is the desirable property of the proposed approach.

B. BER Analysis

In this subsection, we derive the analytical BER for the proposed algorithm. After subtracting DC bias at the receiver, using Bussgang's theorem for decomposition of memoryless nonlinear functions, and at convergence of the proposed algorithm, (6) is rewritten as [18]

$$r_k = \alpha x_k + \underbrace{h^{-1}n_k + \delta + \Delta b}_{\text{additive distortion}\tilde{n}}, \tag{24}$$

where $\alpha = \frac{\mathbb{E}\{A(x_k)x_k\}}{\mathbb{E}\{|x_k|_2^2\}}$ is the scaling correlation factor, and $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$ is the additive distortion component due to LED nonlinearity, which is modeled as Gaussian distribution with variance σ_δ^2, and $\Delta b \sim \mathcal{U}(-\epsilon, \epsilon)$ is the DC-bias error term which arises due to unstable DC-bias. Hence, the probabilistic model for the overall additive distortion $\tilde{n}$ in (24) is given by the following theorem:

Theorem 3: The PDF of the overall additive distortion for a VLC link impaired by user-mobility, LED nonlinearity, and DC-bias fluctuations is given in (25) at the bottom of this page, where $d_i = (2 - c_i)$, u_p, and Ω_p denote zeros, and weights of the P^{th} order Gauss Hermite function, respectively, and $\{l_j, m_j\}_{j=1}^{N_e}$ are set of coefficients (N_e is the number of terms) used for approximation of erfc$(\cdot)$ function.

Proof: See Appendix B.

Capitalizing on the derived PDF in (25), next we proceed to derive analytical BER for the proposed algorithm. Using $\alpha \to 1$, and error-variance $\sigma_\delta^2 \approx \sigma^2_{\text{RFF-KWS-KRMVC}}$ [27] at convergence of the proposed RFF-KWS-KRMVC post-distorter, an analytical BER is derived, which is given by the following theorem.

Theorem 4: The analytical upper bound on BER for the proposed algorithm (denoted by $P_e^{\text{RFF-KWS-KRMVC}}$) for a VLC link degraded by LED nonlinearity, user-mobility, and DC-bias fluctuations is defined in (28), shown at the bottom of the next page, for generalized M-PAM, and M-QAM technique.

In (28), the constants $z(M)$, and A are defined as

$$z(M) = \begin{cases} \left(1 - \frac{1}{M}\right) & \text{for M-PAM} \\ 2\left(1 - \frac{1}{\sqrt{M}}\right) & \text{for M-QAM,} \end{cases} \tag{26}$$

$$A = \begin{cases} \sqrt{\frac{6E_s}{M^2-1}} & \text{for M-PAM} \\ \sqrt{\frac{3E_s}{M-1}} & \text{for M-QAM,} \end{cases} \tag{27}$$

where E_s is the average energy of the transmitted constellation.

Proof: See Appendix C.

VI. Simulations

This section presents simulation results to demonstrate the performance of the proposed RFF-KWS-KRMVC post-distorter, which is compared with the following existing post-distortion algorithms: (1) Volterra-RLS [4], (2) SGD based RFF-KMSER-DFE [18], (3) RFF-KRLS [21], (4) RFF-KWS-KRLS [29], and (5) RFF-KRMVC [22] with fixed kernel-width estimated by (a) Silverman's rule, and (b) brute force trial and error [29]. Kernel-width σ for RFF-KRLS, and RFF-KMSER-DFE based post-distorters is set by brute force trial and error method to achieve best performance [29], and Versoria-shape parameter $\tau = 2$ is chosen for RFF-KRMVC, and the proposed RFF-KWS-KRMVC algorithms [22]. Simulations are performed for 16-CAP-QAM, and 64-CAP-QAM with an ensemble of 10^5 symbols over 100 Monte-Carlo iterations. In all the simulations, number of RFFs $D = 100$ is chosen for RFF based post-distorters (unless and until specified explicitly in the legends), and data rate is set to 0.8 Gbps. Simulation parameters for RWP channel model [37], and Wiener LED model [32] are enlisted in Table II.

SNR vs BER performance are depicted in Figs. 2(a), and 3(a) for the proposed RFF-KWS-KRMVC algorithm for memoryless nonlinearity for 16-CAP-QAM, and 64-CAP-QAM technique, respectively. It can be observed from Figs. 2(a), and 3(a) that the proposed RFF-KWS-KRMVC algorithm delivers 4 dB gain at BER of 10^{-5} for 16-CAP-QAM, and 6 dB gain at BER of 10^{-4} for 64-CAP-QAM over the existing MMSE based RFF-KRLS, and RFF-KWS-KRLS post-distorters due to the incorporation of higher order statistics of error. Next, Figs. 2(b), and 3(b) shows evolution of MSE with the number of iterations for memoryless nonlinearity at SNR of 25 dB for 16-CAP-QAM, and 64-CAP-QAM technique, respectively. It is observed from Figs. 2(b), and 3(b) that the proposed RFF-KWS-KRMVC post-distorter delivers fast convergence, and lower MSE as compared to the existing RFF-KRLS, and RFF-KWS-KRLS post-distorters.

$$f_{\tilde{N}}(\tilde{n}) \approx \sum_{j=1}^{N_e}\sum_{i=1}^{4}\sum_{p=1}^{P}\Omega_p\frac{l_j}{\sqrt{m_j}}\frac{K_i 2^{\frac{d_i}{2}-1}\sigma_n^{d_i-1}\sigma_\delta}{4\sqrt{\pi}\epsilon}\left(|\tilde{n} - \sqrt{\frac{2}{m_j}}\sigma_\delta u_p - \epsilon|^{-d_i}\left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n} - \sqrt{\frac{2}{m_j}}\sigma_\delta u_p - \epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}}\right.$$
$$\left. - |\tilde{n} - \sqrt{\frac{2}{m_j}}\sigma_\delta u_p + \epsilon|^{-d_i}\left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n} - \sqrt{\frac{2}{m_j}}\sigma_\delta u_p + \epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}}\right) \tag{25}$$

TABLE II
CHANNEL PARAMETERS

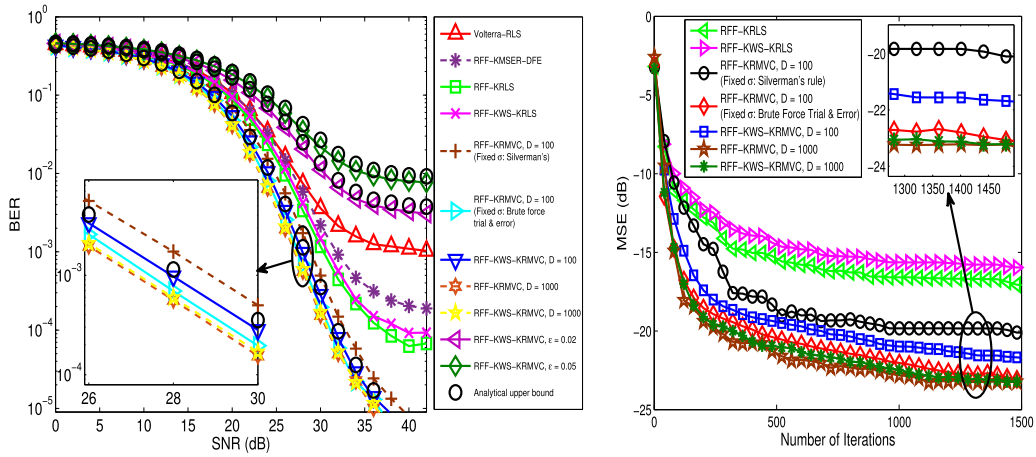
PARAMETERS	SPECIFICATIONS
Vertical distance between LED and PD L	2.25 m
Maximum coverage radius r_e	2.5 m
Effective aperture area of PD A_d	10^{-4} m^2
Field of view (FOV) Ψ_c	60°
LED half angle $\phi_{\frac{1}{2}}$	50°
Maximum saturation voltage of LED $V_{\max}$	0.5 V
Knee factor p	0.5
Cut-in voltage of LED V_{th}	0.2 V
Coefficients of LTI block b_0, b_1, and b_2	1, 0.5, and 0.1

Further, It is observed from Figs. 2 and 3 that the proposed probabilistic kernel-width based RFF-KWS-KRMVC algorithm delivers better BER, and MSE performance over the existing fixed kernel-width (estimated by the Silverman's rule) based RFF-KRMVC algorithm. This is attributed to the presence of statistical approximation error that arises due to erroneous kernel-width estimation by Silverman's rule of thumb whilst computing standard deviation with insufficient regressors/datasets.[1] Furthermore, it is observed from Figs. 2, and 3 that there is a marginal BER performance gap, and higher steady-state MSE floor for the proposed RFF-KWS-KRMVC algorithm over the existing fixed best kernel-width (estimated by brute force trail & error) based RFF-KRMVC algorithm for smaller number of RFFs D $=$ 100. However, this BER, and MSE performance gap reduces with the increase in the number of RFFs since offset error due to sampled kernel-width is negligible under the assumption of sufficiently large number of RFF dimensions as observed from the theoretical derived expression of MSE in (23). It is inferred from Figs. 2 and 3 that the BER, and the MSE of the proposed RFF-KWS-KRMVC algorithm overlaps with the existing fixed best kernel-width based RFF-KRMVC algorithm for higher number of RFFs D $=$ 1000, which highlights the viability of proposed algorithm for hyperparameter-free post-distortion under finite memory budget. The analytical BER curves for the proposed RFF-KWS-KRMVC are shown in Figs. 2(a), and 3(a) for $N_e = 2$ since approximation in (34) holds good for $N_e = 2$ [45], where $l_1 = \frac{1}{6}$, $l_2 = \frac{1}{2}$, $m_1 = 1$, and $m_2 = \frac{4}{3}$. Lastly, as inferred from Figs. 2(a) and 3(a), the analytical upper bound on BER for the proposed RFF-KWS-KRMVC post-distorter agrees well with the simulated BER for different values of ϵ, which validates the derived BER in (28), shown at the bottom of previous page.

Next, Fig. 4 shows SNR vs BER performance for the proposed RFF-KWS-KRMVC post-distorter for memory nonlinearity. As inferred from Fig. 4, the proposed RFF-KWS-KRMVC algorithm delivers gain of 4 dB at BER of 10^{-4}, and 6 dB gain at BER of 10^{-3} (which is below the forward error correction limit) for memory nonlinearity for 16-CAP-QAM, and 64-CAP-QAM,

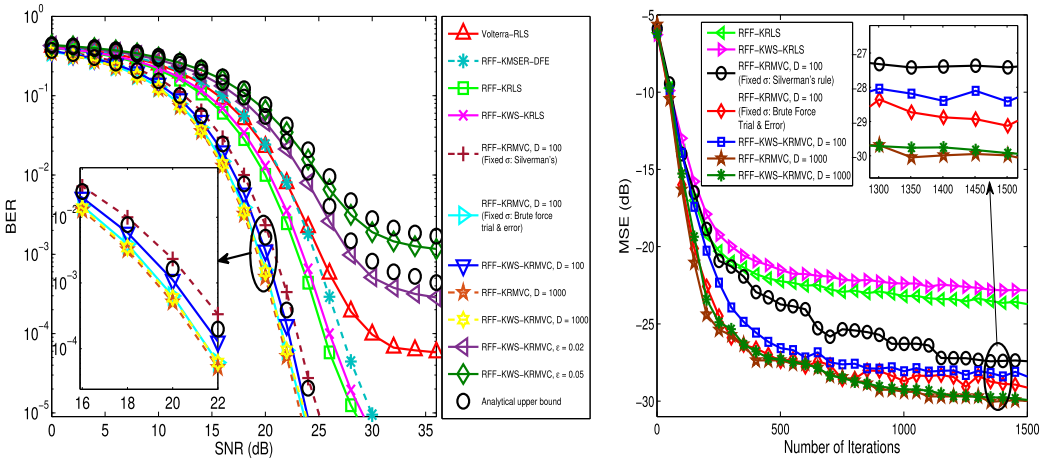
[1]The approximation error due to erroneous kernel-width estimation by Silverman's rule is unbounded for the generic fixed kernel-width based RFF-KRMVC algorithm. On the other hand, stochastic sampling of kernel-width makes the approximation error for the proposed RFF-KWS-KRMVC post-distorter bounded since averaging is done for the deviation due to improper kernel-width over the Gamma distribution, and this approximation error decays as $O(\frac{1}{3D})$, and becomes negligible under the assumption of large number of RFF dimensions D.

$$
\begin{aligned}
P_e^{\text{RFF-KWS-KRMVC}} \leq\ & z(M)\sum_{j=1}^{N_e}\sum_{i=1}^{4}\sum_{p=1}^{P}\Omega_p\frac{l_j}{\sqrt{m_j}}\frac{K_i 2^{\left(\frac{d_i}{2}-1\right)}\sigma_n^{d_i-1}\sigma_\delta}{4\sqrt{\pi}\epsilon(d_i-1)} \\
& \times\Bigg(\Bigg\{\left(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon\right)^{(1-d_i)}\left[\Gamma\left(\frac{d_i}{2},\frac{h^2(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}} \\
& -\left(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon\right)^{(1-d_i)}\left[\Gamma\left(\frac{d_i}{2},\frac{h^2(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}} \\
& +\left(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon\right)^{(1-d_i)}\left[\Gamma\left(\frac{d_i}{2},\frac{h^2(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}} \\
& -\left(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon\right)^{(1-d_i)}\left[\Gamma\left(\frac{d_i}{2},\frac{h^2(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}}\Bigg\} \\
& +\Bigg[2\sqrt{\pi}\left(\frac{h}{\sqrt{2}\sigma_n}\right)^{d_i-1}\Bigg\{Q\left(\frac{h(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)}{\sigma_n}\right)-Q\left(\frac{h(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon)}{\sigma_n}\right) \\
& +Q\left(\frac{h(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)}{\sigma_n}\right)-Q\left(\frac{h(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon)}{\sigma_n}\right)\Bigg\}\Bigg]_{h=h_{\max}}^{h=h_{\min}}\Bigg)
\end{aligned}
\tag{28}
$$



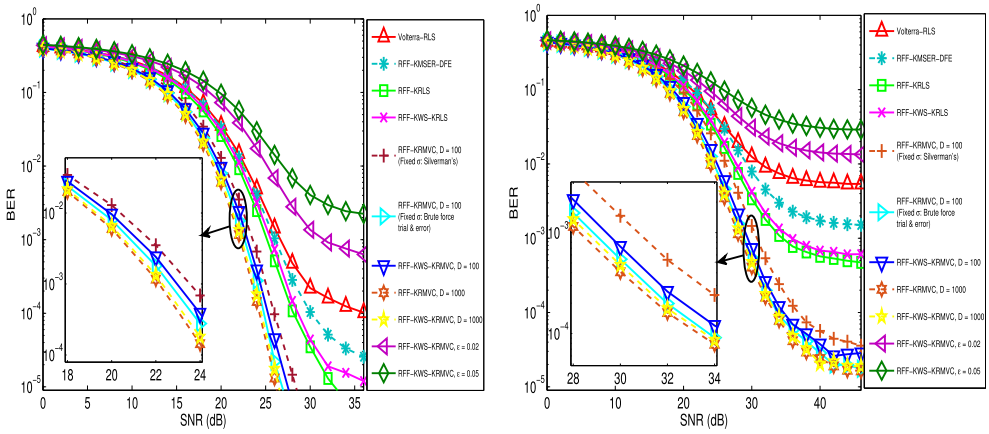
(a) SNR vs BER (b) MSE vs Number of iterations

Fig. 3. BER and convergence performance for the proposed RFF-KWS-KRMVC post-distorter, and existing Volterra-RLS, KMSER-DFE, RFF-KRLS, RFF-KWS-KRLS and RFF-KRMVC algorithms for memoryless nonlinearity using 64-CAP-QAM.



(a) SNR vs BER (b) MSE vs Number of iterations

Fig. 2. BER and convergence performance for the proposed RFF-KWS-KRMVC post-distorter, and existing Volterra-RLS, KMSER-DFE, RFF-KRLS, RFF-KWS-KRLS and RFF-KRMVC algorithms for memoryless nonlinearity using 16-CAP-QAM.



(a) 16-CAP-QAM (b) 64-CAP-QAM

Fig. 4. SNR vs BER performance for the proposed RFF-KWS-KRMVC post-distorter, and existing Volterra-RLS, RFF-KMSER-DFE, RFF-KRLS, RFF-KWS-KRLS and RFF-KRMVC algorithms for memory nonlinearity.

respectively over the existing RFF-KRLS, and RFF-KWS-KRLS post-distorters. Further, gains in BER performance are more visible for 64-CAP-QAM as compared to 16-CAP-QAM since LED nonlinearity is more dominant for higher order constellation having high peak-to-average power ratio, which is mitigated by the proposed RFF-KWS-KRMVC algorithm. Furthermore, it is observed from Figs. 2(a), 3(a), and 4 that the system performance degrades with the increase in DC-bias error ϵ, and results in error rate floor at high SNR regime.

VII. Conclusion

A novel RFF-KWS-KRMVC based post-distortion is proposed in this paper, which assigns random kernel-widths using a pre-designed probability density function, which alleviates kernel-width dependence. Convergence analysis of the proposed RFF-KWS-KRMVC is performed, and analytical MSE for the proposed RFF-KWS-KRMVC algorithm is found to be smaller as compared to the existing RFF-KWS-KRLS algorithm, which highlights the viability of the proposed algorithm for post-distortion over VLC links. Analytical bound for BER is derived for the proposed post-distortion algorithm for a nonlinear, and mobility impaired VLC link in the presence of unstable DC-offset. Furthermore, the computer simulations performed over a RWP mobility model indicate that the proposed RFF-KWS-KRMVC algorithm delivers better BER performance as compared to the existing Volterra-RLS, and RFF-KWS-KRLS based post-distorters under the same computational budget. Hence, the proposed RFF-KWS-KRMVC post-distorter is viable for practical VLC systems deployments for the future beyond 5G/6G communication technologies.

Appendix A
Proof of Theorem 2

Substituting (19) in (17), the evolution of weight error vector is written as

$$\widetilde{\mathbf{w}}(k) = \left(\mathbf{G}^{\text{RFF-KWS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_0}}(k) + \widetilde{\mathbf{G}}^{\text{RFF-KWS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_o}}(k)\right) \times \sum_{i=1}^{k} \lambda^{k-i} f_{\text{MVC}}(e_i)\left(\hat{\Psi}_{\sigma_o}(\mathbf{r}_i) + \widetilde{\hat{\Psi}}_{\sigma_o}(\mathbf{r}_i)\right) n_i. \quad (29)$$

Under the assumption of $\widetilde{\mathbf{G}}^{\text{RFF-KWS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_o}}(k)$, and $\widetilde{\hat{\Psi}}_{\sigma_o}(\mathbf{r}_i)$ to be statistically independent [29], the weight error correlation matrix is written as [22, eq. 11]

$$\begin{aligned}\boldsymbol{\Xi}(k) &= \mathbb{E}(\widetilde{\mathbf{w}}(k)\widetilde{\mathbf{w}}^H(k)) \\ &= \sigma_n^2 \mathbb{E}\left[\mathbf{G}^{\text{RFF-KWS-KRLS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_0}}(k) + \widetilde{\mathbf{G}}^{\text{RFF-KWS-KRLS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_o}}(k)\right] \\ &\quad \times \left(\frac{\mathbb{E}(f^2_{\text{MVC}}(e_k))}{[\mathbb{E}(f_{\text{MVC}}(e_k))]^2}\right), \end{aligned} \quad (30)$$

where $\mathbf{G}^{\text{RFF-KWS-KRLS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_0}}(k) = (\sum_{i=1}^{k}\lambda^{(k-i)}\hat{\Psi}_{\sigma_o}(\mathbf{r}_k)\hat{\Psi}^H_{\sigma_o}(\mathbf{r}_k))^{-1}$, and $\widetilde{\mathbf{G}}^{\text{RFF-KWS-KRLS}}_{\hat{\Psi}_{\sigma_o}\hat{\Psi}_{\sigma_0}}(k) = (\sum_{i=1}^{k}\lambda^{(k-i)}\widetilde{\hat{\Psi}}_{\sigma_o}(\mathbf{r}_k)\widetilde{\hat{\Psi}}^H_{\sigma_o}(\mathbf{r}_k))^{-1}$.

Hence, the MSE at k^{th} iteration for the proposed RFF-KWS-KRMVC based post-distorter is written as

$$\begin{aligned}S_k^{\text{RFF-KWS-KRMVC}} &= \sigma_n^2 + \text{Tr}(\boldsymbol{\Xi}(k)) = \sigma_n^2 + \frac{\sigma_n^2(1-\lambda^2)}{(k-n-1)} \\ &\times \left(\sum_{\forall g}\left(\frac{1}{\rho_g}\right) + \mathbb{E}\left(\widetilde{\hat{\Psi}}_{\sigma_o}(\mathbf{r}_k)\widetilde{\hat{\Psi}}^H_{\sigma_o}(\mathbf{r}_k)\right)\right) \\ &\times \left(\frac{\mathbb{E}(f^2_{\text{MVC}}(e_k))}{[\mathbb{E}(f_{\text{MVC}}(e_k))]^2}\right)\end{aligned}$$

Using the integrals given in [22], the ratio $\left(\frac{\mathbb{E}(f^2_{\text{MVC}}(e_k))}{[\mathbb{E}(f_{\text{MVC}}(e_k))]^2}\right) \to \frac{1}{3}$ [22, eq. 26] at high SNR, and from [29, eq. (15)], $\mathbb{E}(\widetilde{\hat{\Psi}}_{\sigma_o}(\mathbf{r}_k)\widetilde{\hat{\Psi}}^H_{\sigma_o}(\mathbf{r}_k)) = \frac{2\sigma_o^2\mathcal{A}||\mathbf{w}||_2^2}{D}$, the overall MSE can be derived as given in (23).

Appendix B
Proof of Theorem 3

Using function of random variables, and from [46], the overall PDF of $z = \hat{h}^{-1}n_k$ (denoted by $q_Z(z)$) is given as

$$q_Z(z) = \sum_{i=1}^{4}\frac{K_i}{\sqrt{\pi}}2^{\frac{d_i-3}{2}}\sigma_n^{d_i-1}|z|^{-d_i}\left[\Gamma\left(\frac{d_i}{2}, \frac{z^2(h)^2}{2\sigma_n^2}\right)\right]_{h=h_{\text{max}}}^{h=h_{\text{min}}}.$$

Using central limit theorem, the PDF of sum of Gaussian, and uniform random variable $n' = \delta + \Delta b$ in (24) is given as [27, 11]

$$g_{N'}(n') \approx \frac{1}{4\epsilon}\left(\text{erfc}\left(\frac{n'-\epsilon}{\sqrt{2\sigma_\delta^2}}\right) - \text{erfc}\left(\frac{n'+\epsilon}{\sqrt{2\sigma_\delta^2}}\right)\right), \quad (31)$$

Under the assumption of z, and n' to be independent, the PDF of $\tilde{n} = z + n'$ is written as

$$\begin{aligned}f_{\tilde{N}}(\tilde{n}) &\approx \sum_{i=1}^{4}\frac{K_i 2^{\frac{d_i-3}{2}}\sigma_n^{d_i-1}}{4\sqrt{\pi}\epsilon} \\ &\times \underbrace{\left(\int_{-\infty}^{\infty}|\tilde{n}-n'|^{-d_i}\left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n}-n')^2}{2\sigma_n^2}\right)\right]_{h=h_{\text{max}}}^{h=h_{\text{min}}}\right.}_{} \\ &\times \underbrace{\left.\left\{\text{erfc}\left(\frac{n'-\epsilon}{\sqrt{2\sigma_\delta^2}}\right) - \text{erfc}\left(\frac{n'+\epsilon}{\sqrt{2\sigma_\delta^2}}\right)\right\}\right)dn'}_{\mathscr{I}_1}\end{aligned} \quad (32)$$

Upon change of variables $x = \frac{n'-\epsilon}{\sqrt{2\sigma_\delta^2}} \to dn' = \sqrt{2}\sigma_\delta dx$, and $y = \frac{n'+\epsilon}{\sqrt{2\sigma_\delta^2}}$, $\mathscr{I}_1$ reduces to

$$\begin{aligned}\mathscr{I}_1 &= \sqrt{2}\sigma_\delta\int_{-\infty}^{\infty}\left\{|\tilde{n}-\sqrt{2}\sigma_\delta x-\epsilon|^{-d_i}\right. \\ &\times \left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n}-\sqrt{2}\sigma_\delta x-\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\text{max}}}^{h=h_{\text{min}}} \\ &\left. - |\tilde{n}-\sqrt{2}\sigma_\delta x+\epsilon|^{-d_i}\right.\end{aligned}$$

$$\times \left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n}-\sqrt{2}\sigma_\delta x+\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}}\Bigg\}\text{erfc}(x)dx \tag{33}$$

From [45, eq. (8)], a complementary error function $\text{erfc}(x)$ can be approximated as sum of exponentials as follows:

$$\text{erfc}(x) \approx \sum_{j=1}^{N_e} l_j \exp(-m_j x^2). \tag{34}$$

Invoking (34) in (33), and using change of variables $t = \sqrt{m_j}x \rightarrow dx = \frac{1}{\sqrt{m_j}}dt$, (33) reduces to

$$f_{\tilde{N}}(\tilde{n}) \approx \sum_{j=1}^{N_e}\sum_{i=1}^{4} \frac{l_j}{\sqrt{m_j}} \frac{K_i 2^{\frac{d_i}{2}-1}\sigma_n^{d_i-1}\sigma_\delta}{4\sqrt{\pi}\epsilon} \int_{-\infty}^{\infty}$$

$$\Bigg\{|\tilde{n}-\sqrt{\frac{2}{m_j}}\sigma_\delta t-\epsilon|^{-d_i}$$

$$\times \left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n}-\sqrt{\frac{2}{m_j}}\sigma_\delta t-\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}}$$

$$-|\tilde{n}-\sqrt{\frac{2}{m_j}}\sigma_\delta t+\epsilon|^{-d_i}$$

$$\times \left[\Gamma\left(\frac{d_i}{2}, \frac{(h)^2(\tilde{n}-\sqrt{\frac{2}{m_j}}\sigma_\delta t+\epsilon)^2}{2\sigma_n^2}\right)\right]_{h=h_{\max}}^{h=h_{\min}}\Bigg\}\exp(-t^2)dt. \tag{35}$$

Using Gauss Hermite numerical integration $\int_{-\infty}^{\infty} f(t)e^{-t^2}dt = \sum_{p=1}^{P}\Omega_p f(u_p)$, the PDF of the overall additive distortion $\tilde{n}$ is obtained as given in (25).

APPENDIX C
PROOF OF THEOREM 4

Let $a_1 \in \{\pm(2m+1)A \pm j(2n+1)A\}_{m,n=1}^{m,n=(\frac{M}{2}-1)}$ denotes M-QAM symbols. The analytical upper bound on error rate for M-PAM is given as

$$P_e^{\text{M-PAM}} \leq \left(1-\frac{1}{M}\right)(\Pr(\tilde{n}<-A)+\Pr(\tilde{n}>A))$$

$$\leq \left(1-\frac{1}{M}\right)\Bigg(\underbrace{\int_{-\infty}^{A} f_{\tilde{N}}(\tilde{n})d\tilde{n}}_{\mathscr{I}_2} + \underbrace{\int_{A}^{\infty} f_{\tilde{N}}(\tilde{n})d\tilde{n}}_{\mathscr{I}_3}\Bigg). \tag{36}$$

Invoking (25) in (36), and using change of variables $x = (\tilde{n}-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p-\epsilon)^2 \rightarrow d\tilde{n} = \frac{1}{2}x^{-\frac{1}{2}}dx$, and $y = (\tilde{n}-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2 \rightarrow d\tilde{n} = \frac{1}{2}y^{-\frac{1}{2}}dy$, and then replacing dummy variable $y = x$, integrals $\mathscr{I}_2$, and $\mathscr{I}_3$ can be written as

$$\mathscr{I}_2 \approx \sum_{j=1}^{N_e}\sum_{i=1}^{4}\sum_{p=1}^{P}\Omega_p \frac{l_j}{\sqrt{m_j}} \frac{K_i 2^{\left(\frac{d_i}{2}-1\right)}\sigma_n^{d_i-1}\sigma_\delta}{4\sqrt{\pi}\epsilon}$$

$$\times \int_{(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2}^{(A+\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2} \frac{x^{-d_i'}}{2}\left[\Gamma\left(\frac{d_i}{2},\tilde{h}x\right)\right]_{h=\frac{h_{\max}^2}{2\sigma_n^2}}^{h=\frac{h_{\min}^2}{2\sigma_n^2}} dx \tag{37}$$

$$\mathscr{I}_3 \approx \sum_{j=1}^{N_e}\sum_{i=1}^{4}\sum_{p=1}^{P}\Omega_p \frac{l_j}{\sqrt{m_j}} \frac{K_i 2^{\left(\frac{d_i}{2}-1\right)}\sigma_n^{d_i-1}\sigma_\delta}{4\sqrt{\pi}\epsilon}$$

$$\times \int_{(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2}^{(A-\sqrt{\frac{2}{m_j}}\sigma_\delta u_p+\epsilon)^2} \frac{x^{-d_i'}}{2}\left[\Gamma\left(\frac{d_i}{2},\tilde{h}x\right)\right]_{h=\frac{h_{\max}^2}{2\sigma_n^2}}^{h=\frac{h_{\min}^2}{2\sigma_n^2}} dx, \tag{38}$$

where $l_i = \frac{d_i+1}{2}$. Using the integral

$$\int_A^B x^{-l}\Gamma(a,kx)dx = \frac{A^{(1-l)}\Gamma(a,kA)-B^{(1-l)}\Gamma(a,kB)}{(l-1)}$$

$$+\frac{k^{(l-1)}}{l-1}\left(\Gamma(1+a-l,kB)-\Gamma(1+a-l,kA)\right), \tag{39}$$

$\Gamma(0.5,x) = 2\sqrt{\pi}Q(\sqrt{2x})$, and $P_e^{\text{M-QAM}} = 2P_e^{\sqrt{\text{M}}-\text{PAM}}$, the final bound on BER is determined as given in (28).

REFERENCES

[1] A. Jovicic, J. Li, and T. Richardson, "Visible light communication: Opportunities, challenges and the path to market," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 26–32, Dec. 2013.

[2] D. Karunatilaka, F. Zafar, V. Kalavally, and R. Parthiban, "LED based indoor visible light communications: State of the art," *IEEE Commun. Surv. Tut.*, vol. 17, no. 3, pp. 1649–1678, Jul.–Sep. 2015.

[3] Y. Wang, L. Tao, X. Huang, J. Shi, and N. Chi, "8-Gb/s RGBY LED-based WDM VLC system employing high-order CAP modulation and hybrid post equalizer," *IEEE Photon. J.*, vol. 7, no. 6, Dec. 2015, Art. no. 7904507.

[4] 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, Aug. 2013.

[5] W. Zhang, "A general framework for transmission with transceiver distortion and some applications," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 384–399, Feb. 2012.

[6] F. Miramirkhani, O. Narmanlioglu, M. Uysal, and E. Panayirci, "A mobile channel model for VLC and application to adaptive system design," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1035–1038, May 2017.

[7] A. Gupta and P. Garg, "Statistics of SNR for an indoor VLC system and its applications in system performance," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1898–1901, Sep. 2018.

[8] Y. Wang, L. Tao, X. Huang, J. Shi, and N. Chi, "Enhanced performance of a high-speed WDM CAP64 VLC system employing Volterra series-based nonlinear equalizer," *IEEE Photon. J.*, vol. 7, no. 3, Jun. 2015, Art. no. 7901907.

[9] R. Mitra and V. Bhatia, "Adaptive sparse dictionary-based kernel minimum symbol error rate post-distortion for nonlinear LEDs in visible light communications," *IEEE Photon. J.*, vol. 8, no. 4, Aug. 2016, Art. no. 7905413.

[10] S. Jain, R. Mitra, and V. Bhatia, "Multivariate-KLMS based post-distorter for nonlinear RGB-LEDs for color-shift keying VLC," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2019, pp. 1–6.

[11] P. A. Haigh, Z. Ghassemlooy, S. Rajbhandari, I. Papakonstantinou, and W. Popoola, "Visible light communications: 170 Mb/s using an artificial neural network equalizer in a low bandwidth white light configuration," *J. Lightw. Technol.*, vol. 32, no. 9, pp. 1807–1813, 2014.

[12] N. Chi, Y. Zhao, M. Shi, P. Zou, and X. Lu, "Gaussian kernel-aided deep neural network equalizer utilized in underwater PAM8 visible light communication system," *Opt. Exp.*, vol. 26, no. 20, pp. 26 700–26712, 2018.

[13] 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, Feb. 2019.

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

[15] R. Mitra and G. Kaddoum, "Random Fourier feature based deep learning for wireless communications," *IEEE Trans. Cogn. Commun. Netw.*, 2022, doi: 10.1109/TCCN.2022.3164898.

[16] P. Bouboulis, S. Pougkakiotis, and S. Theodoridis, "Efficient KLMS and KRLS algorithms: A random Fourier feature perspective," in *Proc. IEEE Stat. Signal Process. Workshop*, 2016, pp. 1–5.

[17] S. Wang, L. Dang, B. Chen, S. Duan, L. Wang, and K. T. Chi, "Random fourier filters under maximum correntropy criterion," *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 65, no. 10, pp. 3390–3403, Oct. 2018.

[18] S. Jain, R. Mitra, and V. Bhatia, "Kernel MSER-DFE based post-distorter for VLC using random Fourier features," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16241–16246, Dec. 2020.

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

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

[21] 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, Apr. 2020.

[22] S. Jain, R. Mitra, and V. Bhatia, "Kernel recursive maximum versoria criterion algorithm using random Fourier features," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 68, no. 7, pp. 2725–2729, Jul. 2021.

[23] F. Huang, J. Zhang, and S. Zhang, "Maximum versoria criterion-based robust adaptive filtering algorithm," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 64, no. 10, pp. 1252–1256, Oct. 2017.

[24] S. Jain, R. Mitra, and V. Bhatia, "Kernel adaptive filtering based on maximum versoria criterion," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Sys.*, 2018, pp. 1–6.

[25] S. S. Bhattacharjee, D. Ray, and N. V. George, "Adaptive modified versoria zero attraction least mean square algorithms," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 67, no. 12, pp. 3602–3606, Dec. 2020.

[26] 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.

[27] 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, Jan. 2020.

[28] 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, Apr. 2015.

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

[30] 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. 27002–27013, 2017.

[31] V. Bhatia, S. Jain, K. Garg, and R. Mitra, "Performance analysis of RKHS based detectors for nonlinear NLOS ultraviolet communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3625–3639, Apr. 2021.

[32] 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, Aug. 2014, Art. no. 7901508.

[33] 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, Jun. 2010.

[34] R. Mitra and V. Bhatia, "Unsupervised multistage-clustering-based Hammerstein postdistortion for VLC," *IEEE Photon. J.*, vol. 9, no. 1, Feb. 2017, Art. no. 7900310.

[35] F. Miramirkhani and M. Uysal, "Channel modeling and characterization for visible light communications," *IEEE Photon. J.*, vol. 7, no. 6, Dec. 2015, Art. no. 7905616.

[36] L. Yin, W. O. Popoola, X. Wu, and H. Haas, "Performance evaluation of non-orthogonal multiple access in visible light communication," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5162–5175, Dec. 2016.

[37] K. R. Sekhar and R. Mitra, "MBER combining for MIMO VLC with user mobility and imperfect CSI," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 376–380, Feb. 2020.

[38] H. Ma, L. Lampe, and S. Hranilovic, "Hybrid visible light and power line communication for indoor multiuser downlink," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 8, pp. 635–647, Aug. 2017.

[39] S. Jain, R. Mitra, and V. Bhatia, "On BER analysis of nonlinear VLC systems under ambient light and imperfect/outdated CSI," *OSA Continuum*, vol. 3, no. 11, pp. 3125–3140, 2020.

[40] S. Bochner, *Harmonic Analysis and the Theory of Probability*. Univ. California Press, 2020.

[41] P. Bouboulis, S. Chouvardas, and S. Theodoridis, "Online distributed learning over networks in RKH spaces using random Fourier features," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1920–1932, Apr. 2018.

[42] F. Liu, X. Huang, Y. Chen, and J. A. Suykens, "Random features for kernel approximation: A survey on algorithms, theory, and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, doi: 10.1109/TPAMI.2021.3097011.

[43] R. Mitra, V. Bhatia, S. Jain, and K. Choi, "Performance analysis of random Fourier features based unsupervised multistage-clustering for VLC," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2659–2663, Aug. 2021.

[44] S. S. Haykin, *Adaptive Filter Theory*. Delhi, India: Pearson Education, 2005.

[45] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.

[46] R. Mitra and V. Bhatia, "Minimum error entropy criterion based channel estimation for Massive-MIMO in VLC," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1014–1018, Jan. 2019.