RESEARCH ARTICLE

Time-Delay Twin Support Vector Regression-Based Adaptive Predistorter for LED Nonlinearity in Visible Light Communications

HAIPENG ZHA[1,2], KUN ZHANG[1], YUNJIAN JIA [ID][3], (Member, IEEE), AND WENQIANG LU [ID][1,2], (Member, IEEE)

[1]Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China
[2]School of Chongqing, University of Chinese Academy of Sciences, Chongqing 400714, China
[3]School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China

Corresponding author: Wenqiang Lu (wqlu@cigit.ac.cn)

ABSTRACT In visible light communication (VLC) systems based on light emitting diodes (LEDs), the nonlinearity of LEDs often leads to a high bit error rate (BER), which limits the system's performance. While artificial neural networks (ANNs) have been used as predistorters to mitigate LED nonlinearity, their effectiveness is hampered by overfitting. This paper proposes an adaptive predistorter based on amplitude time-delay twin support vector regression (ATD-TSVR) to address the nonlinearity of LEDs in orthogonal frequency division multiplexing (OFDM)-based VLC systems. The authors demonstrate through experiments that LED nonlinearity is the primary source of signal distortion in nonlinear VLC systems. Simulation results show that the proposed ATD-TSVR predistorter achieves superior BER, Inputs/Outputs curves, power spectral density (PSD), and constellation plots in an 80Mbit/s OFDM nonlinear VLC system. Meanwhile, as compared to the traditional SVR approach, the CPU training time of ATD-TSVR can be reduced by more than four times. The adaptive pre-distortion method herein is generally applicable to broadband VLC systems and also proves the application prospect and effectiveness of TSVR in VLC system.

INDEX TERMS ATD-TSVR, adaptive predistorter, nonlinearity, VLC, indirect learning architecture.

I. INTRODUCTION

Wradio frequency (RF) spectrum resources becoming increasingly limited, Visible Light Communication (VLC) technology has gained attention as an alternative communication method. However, VLC systems contain many nonlinear components that can lead to signal distortion and degrade system performance [1], [2]. The light-emitting diode (LED), which serves as the transmitter in VLC systems, is a major source of nonlinearity [3], [4]. To address the issue of nonlinearity in VLC systems, there are generally two approaches: designing waveforms that are insensitive to nonlinearities and mitigating nonlinear distortion. As LED

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Gao [ID].

chip design continues to improve and LED devices become more widely deployed, the research into VLC technology is gaining extensive attention. Mitigating nonlinear distortion is crucial for optimizing VLC system performance, and PAPR reduction is a practical approach for addressing LED nonlinearity [5], [6], [7].

Another approach is to use distortion compensation techniques to linearize the nonlinearity, which can be divided into pre-distortion and post-distortion compensation techniques. [1]. Digital pre-distortion (DPD) is a popular approach for modeling and estimating predistorters directly from sampled data by inverting the data stream. However, most DPD schemes rely on behavioral models of the LED, including memoryless and memory models. The passage also mentions specific models that have been used, such as

a fifth-order polynomial model and a lookup table (LUT) method. Inan et al. used a fifth-order polynomial to describe the behavior of a LED [8], [9]. Elgala et al. proposed a pre-distorter based on the lookup table (LUT) method to overcome the nonlinear effects of LED in [10] and [11]. To overcome the limitations of LUT-based predistortion, adaptive nonlinear pre-distortion schemes have been used to mitigate nonlinear impairments [12], [13], [14]. Additionally, equalizers based on the Volterra series have been used to compensate for dynamic nonlinear distortion in VLC systems [15], [16], [17], [18]. However, the computing complexity of models from the Volterra family increases exponentially with nonlinear order. To address this issue, models from the Volterra family have been clipped, and other modeling techniques have been explored. For instance, W. Zhao et al. proposed an orthogonal polynomial model to solve the problem of instability of the coefficients of traditional polynomial modeling(MP) by using a Hammerstein model [4]. Wiener model is another subset of the Volterra model,which has the opposite structure of the Hammerstein model [3], [19], and Qian and Haas et al. used this to model the nonlinearity of LEDs.

Recently, researchers have proposed using nonlinear adaptive algorithms based on neural networks to improve the performance of visible light communication (VLC) systems [12], [20], [21], [22]. Additionally, support vector machine (SVM) based classification has already been utilized to mitigate nonlinearity in both fiber communication and VLC systems [23], [24], [25], [26]. Another machine learning algorithm, the K-means algorithm, has also been presented as a technique for mitigating memoryless nonlinear phase noise in optical communication [27], [28]. In summary, machine learning (ML) approaches have proven to be effective in enhancing the performance of highly nonlinear VLC systems.

Furthermore, Support Vector Regression (SVR) is a method that is commonly used for estimation, fitting, and regression of data. It is based on Support Vector Machines (SVM) and has unique advantages over traditional Artificial Neural Network (ANN) models. According to reports in [29], [30], and [31],SVR can be used to improve optical communication. Additionally, there are studies that have used SVR to model power amplifiers [32], [33]. However, the computational complexity of training the SVR model can be a limiting factor. We learn that the training computational complexity of the SVR model is $8O(n^3)$, while that of the TSVR is $2O(n^3)$. As the number of samples increases, the learning speed of SVR decreases significantly due to the minimization of convex quadratic functions that are bound by a pair of linear inequalities in all training samples [34]. To address this issue, we propose a dynamic behavior model of Light Emitting Diodes (LED) based on amplitude time-delay twin support vector regression (ATD-TSVR) to accelerate the training process and improve model precision. This model takes into account the LED memory effects of nonlinearities and provides a detailed derivation of the TSVR theory. We perform numerical simulations to validate the feasibility

of the proposed ATD-TSVR modeling and compensation scheme in an 80-Mbit/s DCO-OFDM-based nonlinear VLC system. We also provide a performance comparison between the proposed scheme and the conventional SVR, MP, and Generalized Memory Polynomial (GMP) schemes.

The remainder of this paper is organized as follows. In Section II, we provide a review of the basic theory of SVR, and introduce TSVR in detail. In Section III, we present the ATD-TSVR model. Then, we provide simulation and experimental results for the proposed model. Finally, in Section IV, we conclude this paper.

II. THEORETICAL DERIVATION OF MODEL
A. SUPPORT VECTOR REGRESSION THEORY

SVR is a variation of the support vector machine technique created by Vapnik and Cortes [32]. Given a training sample $D = \{x_i, y_i\}_{i=1}^n \in (R^n \times y)^l$,where $x_i \in R^n$ denotes input variables with output value $y_i \in y = R$, we want to obtain a regression model as in (1) so that the value of $f(x)$ is as close as possible to y. ω and b are the model parameters waiting to be determined, which can be expressed as

$$f(x) = <\omega \cdot x> +b. \tag{1}$$

In Fig.1a, the variable ε is introduced to construct an isolation band of width 2ε to address the regression problem, and if the training sample falls into this interval band, it is considered to be correctly predicted. Since there are only a finite number of training points in the training set, the isolation band always exists when ε is sufficiently large. And the value of ε that can make the isolation band exist is not too small; it should be greater than the optimal value ε_{inf} of the following optimization problem:

$$\min_{\omega,b,\varepsilon} \varepsilon$$
$$s.t. -\varepsilon \leq y_i - (<\omega \cdot x> +b) \leq \varepsilon, i = 1,\ldots,l. \tag{2}$$

Consequently, the SVR problem can be described as

$$\min_{\omega,b} \left\{ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^n \ell_\varepsilon(f(x_i) - y_i) \right\} \tag{3}$$

where C is the regularization parameter,ℓ_ε is the ε-insensitive loss function.

$$[\ell_\varepsilon(z) = \begin{cases} 0, & if\ |z| \leq \varepsilon \\ |z| - \varepsilon, & otherwise. \end{cases} \tag{4}$$

Introducing the slack variable ξ_i^* and ξ_i, formula (3) was rewritten as

$$\min_{\omega,b,\xi_i^*,\xi_i} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^n (\xi_i + \xi_i^*)$$
$$s.t. \begin{cases} f(x_i) - y_i \leq \varepsilon + \xi_i^*, \\ y_i - f(x_i) \leq \varepsilon + \xi_i, \\ \xi_i^* \geq 0, \xi_i \geq 0, i = 1, 2\ldots n. \end{cases} \tag{5}$$

By introducing Lagrange multipliers $\alpha^* \geq 0$, $\alpha \geq 0$, $\mu^* \geq 0$, $\mu \geq 0$, the Lagrangian function of formula (5) can be obtained by the Lagrangian multiplier method

$$L(\omega, b, \alpha^*, \alpha, \mu^*, \mu)$$
$$= \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i^* + \xi_i) - \sum_{i=1}^{n}\mu_i\xi_i - \sum_{i=1}^{n}\mu_i^*\xi_i^*$$
$$+ \sum_{i=1}^{n}\alpha_i(f(x_i) - \xi_i - y_i - \varepsilon) + \sum_{i=1}^{n}\alpha_i^*(f(x_i) - \xi_i^* - y_i - \varepsilon).$$
$$(6)$$

Substituting formula (1) into (6), and making $\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi_i} = \frac{\partial L}{\partial \xi_i^*} = 0$, we obtain

$$\begin{cases} \omega = \sum_i^n (\alpha_i^* - \alpha_i)x_i \\ \sum_i^n (\alpha_i^* - \alpha_i) = 0 \\ C - \alpha_i - \mu_i = 0 \\ C - \alpha_i^* - \mu_i^* = 0. \end{cases} \quad (7)$$

Substituting (7) into (6), we obtain the dual form of the original problem

$$\max_{\alpha_i^*, \alpha_i} \frac{1}{2}\sum_{i}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)x_i^T x_j$$
$$+ \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)y_i - \varepsilon(\alpha_i^* + \alpha_i)$$
$$s.t. \begin{cases} \sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i)x_i = 0 \\ 0 \leq \alpha_i^*, \alpha_i \leq C. \end{cases} \quad (8)$$

The SVR fitting equation can be obtained once the dual problem has been solved

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)x_i^T x_i + b. \quad (9)$$

Regarding the feature mapping, the matching linear function in (8) is substituted with the kernel function for the LED modeling application. Thus, the nonlinear SVR function can be characterized as
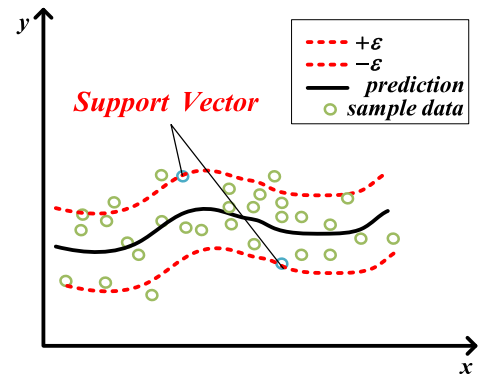
$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\kappa(x_i, x_j) + b \quad (10)$$

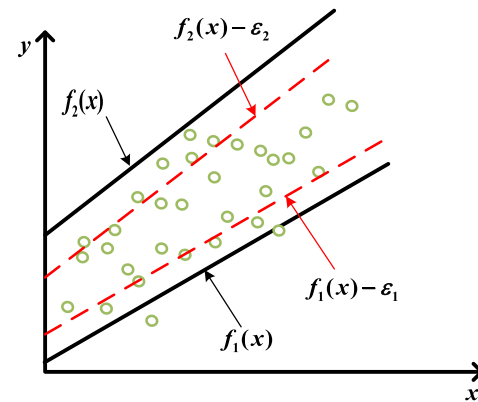where $\kappa(x_i, x_j)$ Gaussian kernel, described as

$$\kappa(x_i, x_j) = \phi(x_i)^T\phi(x_j) = e^{(-\gamma|x_i - x_j|^2)}. \quad (11)$$

B. TWIN SUPPORT VECTOR REGRESSION (TSVR) THEORY
By introducing the idea of twin support vector machines into the regression problem [35], [36], Peng proposed εtwin support vector regression(ε-TSVR) [37], and in the case of

FIGURE 1. Geometric interpretation of SVR and TSVR. (a) SVR. (b) TSVR.

linear problems, ε-TSVR considers finding a pair of linear functions like in(12)

$$\begin{cases} f_1(x) = \omega_1 x + b_1 \\ f_2(x) = \omega_2 x + b_2. \end{cases} \quad (12)$$

where $\omega_1 \in R^n$, $\omega_2 \in R^n$, $b_1 \in R^n$, $b_2 \in R^n$, the pair of linear functions is reduced and increased by ε as the control lower and upper bound of the prediction function $f(x) = 1/2(f_1(x) + f_2(x))$, respectively. Therefore, ε-TSVR tries to keep the training samples between the upper and lower bounds, and compresses the width between the upper and lower bounds as much as possible. As shown in Fig.1b, this width is controlled by the parameter ε in actual operation, and ε-TSVR solves the following two optimization problems:

$$\min \frac{1}{2}\sum_{i=1}^{l}\left(y_i - \varepsilon_1 - (\omega_1^T x_i + b_1)\right)^2 + C_1\sum_{i=1}^{l}\xi_i^*$$
$$s.t. \begin{cases} \omega_1^T x_i + b_1 + \varepsilon_1 - y_i \leq \xi_i^* \\ \xi_i^* \geq 0, \end{cases} \quad (13)$$

$$\min \frac{1}{2}\sum_{i=1}^{l}\left(y_i + \varepsilon_2 - (\omega_2^T x_i + b_2)\right)^2 + C_2\sum_{i=1}^{l}\xi_i$$
$$s.t. \begin{cases} y_i - (\omega_2^T x_i + b_2 - \varepsilon_1) \leq \xi_i \\ \xi_i \geq 0. \end{cases} \quad (14)$$

Here $\varepsilon_1 \geq 0, \varepsilon_2 \geq 0$ are insensitive parameters. ξ_i^* and ξ_i are relaxation variables to introduce for the purpose of measurement error, given by the user, which may vary depending on the application of the regression model. This constraint requires that the distance from the training sample point to the estimation function be at least ε_1 and ε_2. By introducing Lagrangian multipliers $\mu_i \geq 0, \alpha_i \geq 0, \lambda_i \geq 0, \beta_i \geq 0$, the function (12) can be defined as

$$
\begin{cases}
L\left(\omega_1, b_1, \alpha, \mu, \xi^*\right) = \frac{1}{2}\sum_{i=1}^{l}\left(y_i - \varepsilon_1 - \left(\omega_1^T x_i + b_1\right)\right)^2 \\
\quad + C_1 \sum_{i=1}^{l} \xi_i^* - \sum_{i=1}^{l} \mu_i \xi_i^* \\
\quad + \sum_{i=1}^{l} \alpha_i\left(\left(\omega_1^T x_i + b_1\right) - y_i + \varepsilon_1 - \xi_i^*\right) \\
L\left(\omega_2, b_2, \beta, \lambda, \xi_i\right) = \frac{1}{2}\sum_{i=1}^{l}\left(y_i - \varepsilon_2 - \left(\omega_2^T x_i + b_2\right)\right)^2 \\
\quad + C_2 \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \mu_i \xi_i \\
\quad + \sum_{i=1}^{l} \alpha_i\left(\left(\omega_2^T x_i + b_2\right) - y_i + \varepsilon_1 - \xi_i\right).
\end{cases}
\tag{15}
$$

To facilitate the implementation of the algorithm, the above equation is rewritten in matrix form, then the pair of Lagrangian functions can be written in the following form:

$$
\begin{aligned}
&L\left(\omega_1, b_1, \alpha, \mu, \xi^*\right) \\
&= \frac{1}{2}(Y - e\varepsilon_1 - (A\omega_1 + eb_1))^T (Y - e\varepsilon_1 - (A\omega_1 + eb_1)) \\
&\quad + C_1 e^T \xi^* + \alpha^T\left((A\omega_1 + eb_1 + e\varepsilon_1) - Y - \xi^*\right) - \mu^T \xi^*,
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
&L\left(\omega_2, b_2, \beta, \lambda, \xi\right) \\
&= \frac{1}{2}(Y + e\varepsilon_2 - (A\omega_2 + eb_2))^T (Y + e\varepsilon_2 - (A\omega_2 + eb_2)) \\
&\quad + C_2 e^T \xi + \beta^T\left(Y - (A\omega_2 + eb_2 - e\varepsilon_2) - \xi\right) - \lambda^T \xi.
\end{aligned}
\tag{17}
$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$, $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$, $\beta = (\beta_1, \beta_2, \ldots, \beta_n)^T$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$ are Lagrangian multiplier vectors. The training data are represented by the matrix A. Each row of the matrix A represents a set of training data, which is described as $A_i = (A_{i,1}, A_{i,2}, \ldots, A_{i,m})$ and the output signal is denoted as $Y(y_1, y_2, \ldots, y_n)$. In order to transform a nonlinear quadratic programming problem into a linear one, the KKT (Karush-Kuhn-Tucker) principle requirement needs to be satisfied, and the KKT can be described in the following form

$$
\begin{cases}
-A^T\left(Y - A\omega_1 - eb_1 - e\varepsilon_1\right) + A^T\alpha = 0 \\
-e^T\left(Y - A\omega_1 - eb_1 - e\varepsilon_1\right) + e^T\alpha = 0 \\
e^T C_1 - \alpha^T e - \mu^T e = 0 \\
Y - (A\omega_1 + eb_1) \geq e\varepsilon_1 - \xi^*, \quad \xi^* \geq 0 \\
\alpha^T\left(Y - (A\omega_1 + eb_1) - e\varepsilon_1 + \xi^*\right) = 0, \quad \alpha \geq 0 \\
\mu^T \xi = 0, \quad \mu \geq 0.
\end{cases}
\tag{18}
$$

From the KKT condition (18), it follows that

$$
-\begin{bmatrix} A^T \\ e^T \end{bmatrix}\left((Y - e\varepsilon_1) - [A\ e]\begin{bmatrix} \omega_1 \\ b_1 \end{bmatrix}\right) + \begin{bmatrix} A^T \\ e^T \end{bmatrix}\alpha = 0.
\tag{19}
$$

For descriptive convenience, we define the following matrix

$$
H = [A\ e], \ f = Y - e\varepsilon_1, \ p_1 = \begin{bmatrix} \omega_1 \\ b_1 \end{bmatrix}.
\tag{20}
$$

So equation (19) can be described as

$$
-H^T f + H^T H p_1 + H^T \alpha = 0.
\tag{21}
$$

Solving equation (19), we can obtain the parameters p_1

$$
p_1 = \left(H^T H\right)^{-1} H^T (f - \alpha).
\tag{22}
$$

From equation (22), we can see that p_1 does not necessarily have a solution because it is possible that $\left(H^T H\right)$ cannot exist as an inverse. To solve this problem, a very tiny parameter δ is introduced, so that equation (22) can be rewritten as

$$
p_1 = \left(H^T H + \delta I\right)^{-1} H^T (f - \alpha).
\tag{23}
$$

by the same token

$$
p_2 = \left(H^T H + \delta I\right)^{-1} H^T (\beta + h).
\tag{24}
$$

Here $h = Y + e\varepsilon_2, p_2 = \begin{bmatrix} \omega_2 \\ b_2 \end{bmatrix}$ combining (15), (18) and (22), we obtain the following dual problem of (13) and (14)

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\alpha^T H\left(H^T H + \delta I\right)^{-1} H^T \alpha \\
& + f^T H\left(H^T H + \delta I\right)^{-1} H^T \alpha - f^T \alpha \\
s.t. \quad & 0 \leq \alpha \leq C_3 e,
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\beta^T H\left(H^T H + \delta I\right)^{-1} H^T \beta \\
& - h^T H\left(H^T H + \delta I\right)^{-1} H^T \beta + h^T \beta \\
s.t. \quad & 0 \leq \beta \leq C_4 e.
\end{aligned}
\tag{26}
$$

After obtaining the vectors p_1, p_2 from equations (23) and (24), the up- and down-bound functions can be determined. Following is a construction of the calculated regression functions

$$
\begin{aligned}
f(x) &= \frac{1}{2}(f_1(x) + f_2(x)) \\
&= \frac{1}{2}(\omega_1 + \omega_2)^T x + \frac{1}{2}(b_1 + b_2).
\end{aligned}
\tag{27}
$$

C. KERNEL TWIN SUPPORT VECTOR REGRESSION THEORY

Using kernel approaches, we may achieve nonlinear twin support vector regression [34]. The entire model's regression function can be illustrated as

$$
\begin{cases}
f_1(x) = \kappa(x^T, A^T)\omega_1 + b_1, \\
f_2(x) = \kappa(x^T, A^T)\omega_2 + b_2.
\end{cases}
\tag{28}
$$

The kernel function uses the same Gaussian kernel in(13). Therefore, the quadratic programming problem in (13) and (14) can be expressed as follows:

$$\min \frac{1}{2}\left(Y - e\varepsilon_1 - \left(\kappa(A, A^T)\omega_1 + eb_1\right)\right)^T$$
$$* \left(Y - e\varepsilon_1 - \left(\kappa(A, A^T)\omega_1 + eb_1\right)\right) + C_1 e^T \xi^*$$
$$s.t. \quad Y - \left(\kappa(A, A^T)\omega_1 + eb_1\right) \geq e\varepsilon_1 - \xi^*, \quad \xi^* \geq 0,$$
$$(29)$$

$$\min \frac{1}{2}\left(Y + e\varepsilon_2 - \left(\kappa(A, A^T)\omega_2 + eb_2\right)\right)^T$$
$$* \left(Y + e\varepsilon_2 - \left(\kappa(A, A^T)\omega_2 + eb_2\right)\right) + C_2 e^T \xi$$
$$s.t. \quad \left(\kappa(A, A^T)\omega_2 + eb_2\right) - Y \geq e\varepsilon_2 - \xi, \quad \xi \geq 0. \quad (30)$$

As before, the Lagrangian multiplier is introduced

$$L\left(\omega_1, b_1, \alpha, \mu, \xi^*\right)$$
$$= \frac{1}{2}\left(Y - e\varepsilon_1 - \left(\kappa(A, A^T)\omega_1 + eb_1\right)\right)^T$$
$$* \left(Y - e\varepsilon_1 - \left(\kappa(A, A^T)\omega_1 + eb_1\right)\right) + C_1 e^T \xi^*$$
$$+ \alpha^T \left(\left(\kappa(A, A^T)\omega_1 + eb_1 + e\varepsilon_1\right) - Y - \xi^*\right) - \mu^T \xi^*.$$
$$(31)$$

The above process requires that the KKT condition be satisfied

$$- \kappa(A, A^T)^T \left(Y - \kappa(A, A^T)\omega_1 - eb_1 - e\varepsilon_1\right)$$
$$+ \kappa(A, A^T)^T \alpha = 0$$
$$- e^T \left(Y - \kappa(A, A^T)\omega_1 - eb_1 - e\varepsilon_1\right) + e^T \alpha = 0$$
$$e^T C_1 - \alpha^T e - \mu^T e = 0$$
$$Y - \left(\kappa(A, A^T)\omega_1 + eb_1\right) \geq e\varepsilon_1 - \xi^*, \quad \xi^* \geq 0,$$
$$\alpha^T \left(Y - \left(\kappa(A, A^T)\omega_1 + eb_1\right) - e\varepsilon_1 + \xi^*\right) = 0, \quad \alpha \geq 0,$$
$$\mu^T \xi = 0, \quad \mu \geq 0. \quad (32)$$

Define the following expressions

$$H = \left[\kappa(A, A^T) \; e\right], \; f = Y - e\varepsilon_1, \; p_1 = \begin{bmatrix} \omega_1 \\ b_1 \end{bmatrix}. \quad (33)$$

Thus the dual of (29) and (30) is described as follows

$$\max \; -\frac{1}{2}\alpha^T H \left(H^T H + \delta I\right)^{-1} H^T \alpha$$
$$+ f^T H \left(H^T H + \delta I\right)^{-1} H^T \alpha - f^T \alpha$$
$$s.t. \; 0 \leq \alpha \leq C_3 e, \quad (34)$$
$$\max \; -\frac{1}{2}\beta^T H \left(H^T H + \sigma I\right)^{-1} H^T \beta$$
$$- h^T H \left(H^T H + \sigma I\right)^{-1} H^T \beta + h^T \beta$$
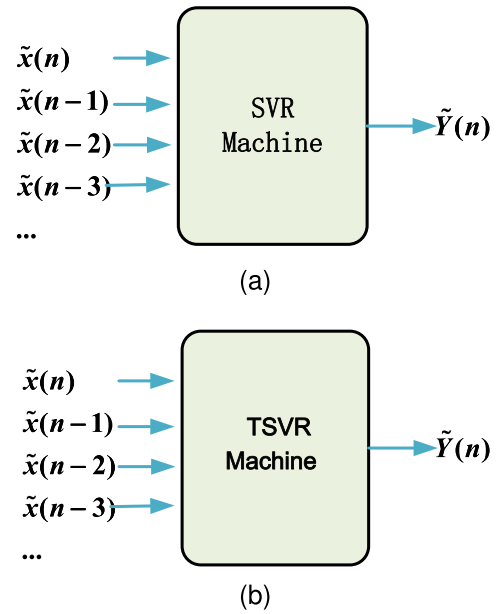$$s.t. \; 0 \leq \beta \leq C_4 e. \quad (35)$$

FIGURE 2. Behavioral model based on time-delay SVR and TSVR. (a) SVR. (b) TSVR.

Hence, the final intended regression function can be defined as the mean of the up- and down-bound functions using the following functional expression:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x))$$
$$= \frac{1}{2}(\omega_1 + \omega_2)^T \kappa(A, x) + \frac{1}{2}(b_1 + b_2). \quad (36)$$

D. EXPERIMENT SETUP AND LED MODELING DESCRIPTION

As mentioned in the introduction, the goal of LED modeling is to find an explicit function that accurately describes the nonlinear behavior of the LED.

For high-rate, wide-bandwidth transmissions, memory effects cannot be ignored. To accurately model the behavior of LEDs, we must consider both the current value and part of the historical values of the input signal. Therefore, the time delay terms used to characterize LED memory characteristics are incorporated into the TSVR modeling method proposed in this work. Fig.2 shows the behavioral model diagram illustrating the relationship between the input and output signals.

$$\tilde{Y}(n) = f(\tilde{x}(n), \tilde{x}(n-1), \tilde{x}(n-2), \ldots) \quad (37)$$

In VLC systems, LEDs must strike a balance between illumination and communication. Typically, LEDs have a turn-on voltage, which means that they require positive real-valued signals to operate.3 shows a diagram of a DCO-OFDM VLC system with an adaptive indirect learning architecture predistorter. As in the traditional method, the training data are two sets of input and output data, expressed as $D = \{x_i, y_i\}_{i=1}^n \in (R^n \times y)^l$ and we will be able to predict LED behavior using the model after training is complete.
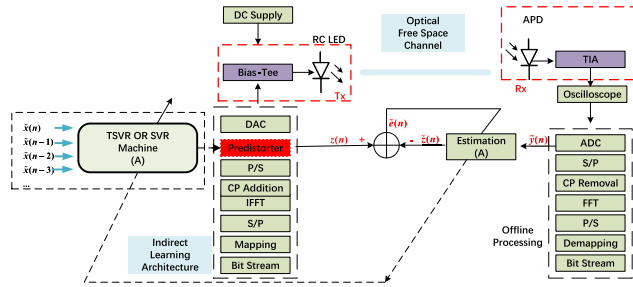
FIGURE 3. Diagram of a DCO-OFDM VLC system with an adaptive indirect learning architec.
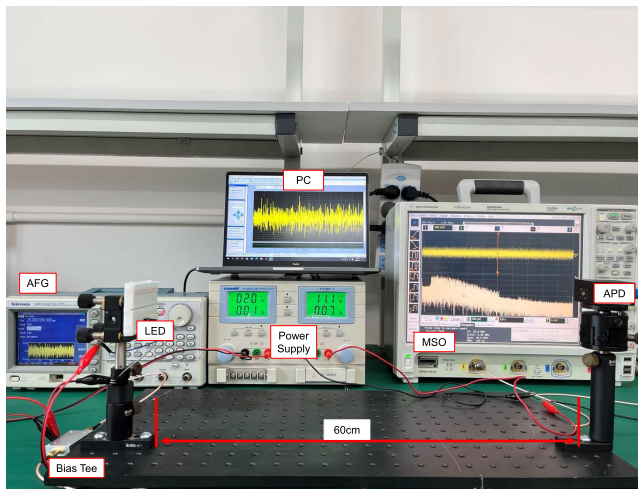

FIGURE 4. Experimental setup of the DCO-OFDM VLC platform.

In the following research, we implement TSVR algorithms and compare them to traditional modeling methods such as memory polynomials and generalized memory polynomials. Fig.4 shows the experimental setup of the VLC system. First, a randomly generated bit sequence is mapped to 16QAM symbols on the transmitter side. Then, the QAM signals are transmitted to the OFDM modulation encoder in MATLAB. In the optical communication experiment, an arbitrary function generator (AFG, Tektronix AFG3102) is used to feed the OFDM data. Subsequently, a 660nm wavelength LED (Hamamatsu L10762) is modulated via a Bias Tee to combine the signal and DC-bias voltage. After a short transit in free space, the optical signal is converted to an electrical signal in an APD photodiode and recorded by an MSO (Agilent MSO9254A) for further offline signal processing.

E. MODEL EXTRACTION

The four different kernel functions considered for the TSVR algorithm are listed below: 1) A linear function. 2) An RBF. 3) A sigmoidal function. 4) A polynomial function. The specific details of these kernel functions are listed in Table 1. Once the appropriate hyperparameters, which include the regularization parameters C_1, C_2, C_3, C_4, and the boundary sensitive parameters $\varepsilon_1, \varepsilon_2$. Once the appropriate

hyperparameters and their kernel functions are determined, the best model can be obtained simultaneously.

TABLE 1. Four different kernals.

Kernel	Formulation	Parameter
Linear	$K(x_i, x_j) = \gamma(x_i, x_j)$	$\gamma = 1$
Polynomial	$K(x_i, x_j) = (\gamma(x_i, x_j) + l)^d$	$\gamma = 1, d = 0.9851$
RBF	$K(x_i, x_j) = e^{(-\gamma|x_i - x_j|^2)}$	$\gamma = 0.9851$
Sigmoid	$K(x_i, x_j) = \tanh(\gamma(x_i, x_j) + l)$	$l = 1, \gamma = 0.9851$

Thus, based on information in Table 1, the Sigmoid function and the RBF are good choices. However, using Sigmoid in the model extraction process would lead to a model close to the singular value performance. Meanwhile, because the model is used for nonlinear behavioral prediction, the linear function is not suitable. Thus, the RBF is used as the kernel function in this paper. A detailed comparison of these kernels is displayed later in Table 2, as a way of providing a quantitative motivation for choosing the above kernel function.

TABLE 2. Extraction time and accuracy with different kernels.

Kernel	Extraction Time(sec)	NMSE(dB)
Linear	0.0421	-19.5216
Polynomial	0.1844	-28.2892
RBF	0.1508	-38.9673
Sigmoid	0.1834	-38.9115

Furthermore, to avoid causing errors in kernel function selection due to the number of modeled samples, the relationship between the kernel function and the number of modeled samples is further investigated, as shown in Fig.5. The figure clearly and intuitively shows the advantages of the RBF kernel function.
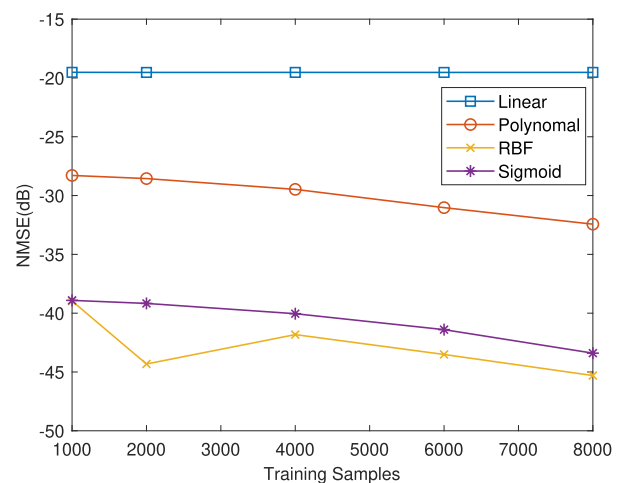

FIGURE 5. Comparison of modeling accuracy of different kernel functions.

C_1, C_2, C_3, C_4 are the regularization parameters and $\varepsilon_1, \varepsilon_2$ are the sensitive function control parameters. TSVR wants the found the upper ε_1-band of the lower bound function and the lower ε_2-band of the upper bound function both contain

TABLE 3. Characteristics comparison for different regularization parameters.

TSVR Parameters	Kernel	NMSE(dB)
$C_1 = C_2 = 10^{-5}, C_3 =$ $C_4 = 2.8, \varepsilon_1 = \varepsilon_2 = 1$	$K(x_i, x_j) = e^{(-\gamma|x_i - x_j|^2)}$	-45.3030
$C_1 = C_2 = 10^{-7}, C_3 =$ $C_4 = 2.8, \varepsilon_1 = \varepsilon_2 = 1$	$K(x_i, x_j) = e^{(-\gamma|x_i - x_j|^2)}$	-51.0632
$C_1 = C_2 = 10^{-7}, C_3 =$ $C_4 = 3, \varepsilon_1 = \varepsilon_2 = 1$	$K(x_i, x_j) = e^{(-\gamma|x_i - x_j|^2)}$	-50.1438
$C_1 = C_2 = 10^{2}, C_3 =$ $C_4 = 2.8, \varepsilon_1 = \varepsilon_2 = 1$	$K(x_i, x_j) = e^{(-\gamma|x_i - x_j|^2)}$	-21.3128

as many sample. This causes TSVR to lose sparsity, and the different choices of bandwidths $\varepsilon_1, \varepsilon_2 > 0$, also affects the fitting accuracy. From Table 3, it can be seen that the highest fitting accuracy is achieved when $C_1 = C_2 = 10^{-7}, C_3 = C_4 = 2.8, \varepsilon_1 = \varepsilon_2 = 1$. However, the problem of large parameter settings of $\varepsilon_1, \varepsilon_2$ also tends to make the model performance close to the singular value performance and reduces the modeling accuracy. Thus, we use the $C_1 = C_2 = 10^{-7}, C_3 = C_4 = 2.8, \varepsilon_1 = \varepsilon_2 = 1$.

III. NUMERICAL SIMULATION AND VLC EXPERIMENT RESULT

In this section, we demonstrate the signal distortion caused by LEDs as the primary nonlinear source. We use numerical simulations performed in MATLAB to investigate the performance of the proposed TD-TSVR-based LED nonlinear modeling and compensation scheme in the case of DCO-OFDM. In Fig.2, the delay term is used as the training term. Additionally, since the VLC system uses intensity modulation, the amplitude of the input signal is also used as the training term. The simulation results show that adding the amplitude term for training significantly improves the modeling precision. This means that the relationship between the final input and output signals is better represented by the model, resulting in improved compensation of the signal distortion caused by LEDs. Specifically, the modeling precision is represented by $\tilde{Y}(n) = f(abs(\tilde{x}(n)), \tilde{x}(n), \tilde{x}(n-1), \tilde{x}(n-2), \ldots)$. We refer to it as amplitude time-delay twin support vector regression (ATD- TSVR). The main simulation parameters are listed in Table 4.

TABLE 4. Simulation parameters.

Parameter	Value
Modulation bandwidth	20 MHz
Number of data subcarriers	512
QAM constellation order	16
Raw data rate	80 Mbit/s
Short Guard Interval	128
Size of FFT/IFFT	512

Figure 6 shows the signal acquired from our experiment, which clearly demonstrates the low-pass effect and distortion caused by the LED. These nonlinear effects significantly decrease the efficiency of the VLC system. Therefore,
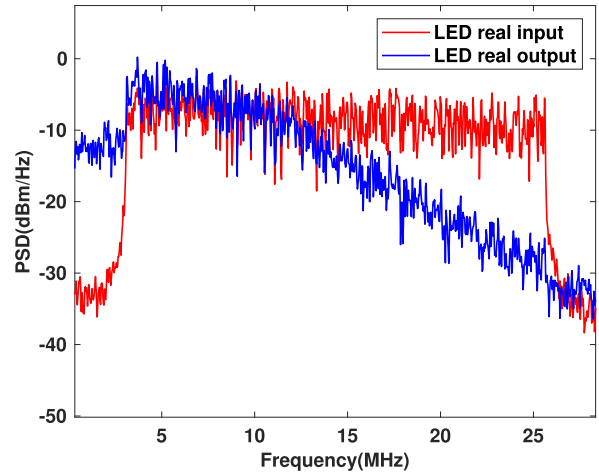

FIGURE 6. LED input and output power spectrum density.

we assume that the LED's nonlinearity with memory effects can be modeled using the Memory Rapps model, which consists of a memoryless nonlinear (NL) block and a linear time-invariant (LTI) block to increase the memory effect [38]. The NL block can be modeled using the Rapps model with some modifications to account for the specific characteristics of the LED [3].

$$z(n) = \begin{cases} (r(n) - V_{TOV})/(1 + (\frac{r(n)-V_{TOV}}{I_{\max}})^{2k})^{1/2k} \\ , r(n) \geq V_{TOV} \\ 0, r(n) < V_{TOV}, \end{cases} \quad (38)$$

the LTI block be presented as:

$$q(n) = \sum_{l=0}^{L-1} b_l z(n - l). \quad (39)$$

where L is the memory depth, b_l is the delayed tap weight, V_{TOV} is the led turn-on voltage, $I_{\max}$ is the maximum current, k controls the transition between non-linear and linear. The parameters of the Memory Rapps model are shown in Table 5. Fig.7a shows the transfer characteristics of the NL block and Fig.7b shows the frequency response of the Wiener LED model. The frequency response of the Wiener LED model is not flat, which indicates that the LED exhibits memory effects.

TABLE 5. Details of the wiener LED model.

LTI			NL		
b_0	b_1	b_2	$I_{\max}$	k	V_{TOV}
1	0.15	0.1	0.5	2	0.2

Normalized Mean Square Error is a regularly used method to evaluate modeling efficacy, define the accuracy of behavioral modeling, as well as measure the degree of distortion of the model, which is expressed by the following equation

$$NMSE(dB) = 10 \log 10 \left[\frac{\sum_{n=1}^{N} |y(n) - \hat{y}(n)|^2}{\sum_{n=1}^{N} |y(n)|^2} \right] \quad (40)$$
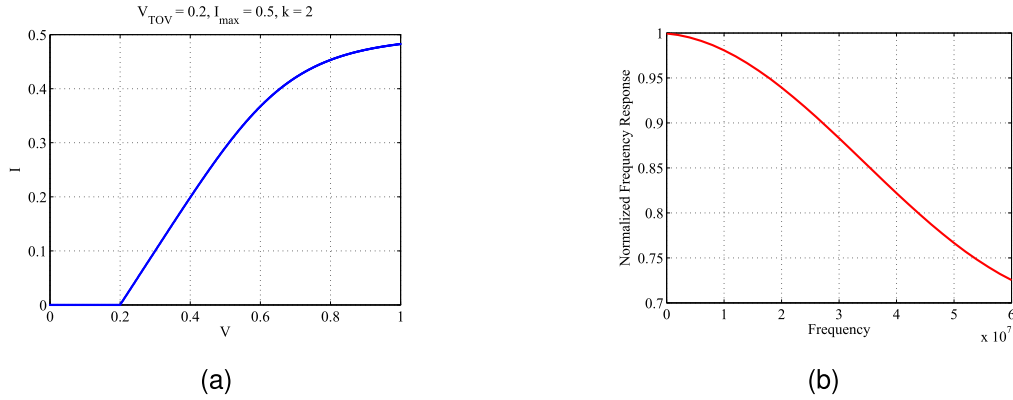
FIGURE 7. (a) shows the transfer characteristics of the Rapps model. (b) shows the frequency response of the Wiener LED model.

where, n is the total number of samples, and $y(n)$ and $\hat{y}(n)$ represent the measured and estimated data, respectively.

A. PSD PERFORMANCE

Initially, the prediction performance of the ATD-TSVR model on the test dataset is validated. We compare the measured spectrum with the spectrum obtained using the ATD-SVR and ATD-TSVR models. Fig.8 gives a comparison of the measured spectrum and the spectrum obtained using MP,GMP,ATD-SVR, and ATD-TSVR. And we can observe that the ATD-TSVR in-band channel and the alternating channel achieve an excellent match.
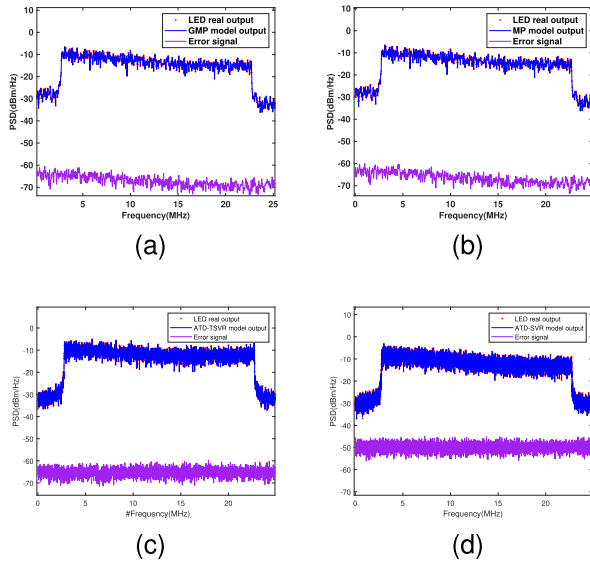

FIGURE 8. Output PSD performance comparision of different model schemes. (a) is Measured and GMP modeled PSD. (b) is Measured and MP modeled PSD. (c) is Measured and ATD-TSVR modeled PSD. (d) is Measured and ATD-SVR modeled PSD.

Fig.9 illustrates the comparison between the real signal of time-domain baseband data and the model output data. It is evident from the figure that the ATD-SVR and

ATD-TSVR models predict the measured data with high accuracy. The low-pass effect in the output power spectral density also confirms the accuracy of the LED nonlinear model used in our study. Both Fig.8 and Fig.9 demonstrate that the ATD-TSVR model performs well in both frequency and time domains.

B. IMPACT OF THE NUMBER OF MEASUREMENTS

In this section, we use the ATD-TSVR approach to characterize the behavior of LEDs. Additionally, we compare the performance of three other approaches: ATD-SVR, MP [39], and GMP [40]. The comparison results, using 10000 samples for modeling with different Memory depth (M) and Nonlinear order (P), are presented in Table 6. The results show that TSVR can reduce CPU training time by more than four times compared to the conventional SVR method. This is because TSVR is designed to solve two small-scale quadratic programming problems (QPPs) to learn the upper and lower bound functions, and then obtain the fitting functions. The computational complexity of TSVR is $2O(n^3)$, which is significantly faster than that of SVR. Fig.10 visually demonstrates the relationship between modeling time and samples for different models. The figure highlights the potential for further improvement in the modeling efficiency of machine learning methods.

TABLE 6. The characteristics comparison for different model.

Model Name	Memory depth(M) and Nonlinear order(P)	Modeling time(s)	NMSE(dB)
GMP	M=2,P=9	0.2741	-50.21
	M=3,P=9	0.1709	-45.85
	M=3,P=11	0.1654	-44.94
MP	M=2,P=9	0.0325	-46.28
	M=3,P=11	0.0424	-45.24
ATD-SVR	M=2	1106.8	-49.90
	M=3	1682.5	-49.48
ATD-TSVR	M=2	215.0061	-50.50
	M=3	283.1879	-49.56

Meanwhile, the different memory depths have an impact on the precision of ATD-TSVR modeling. Table 6 shows
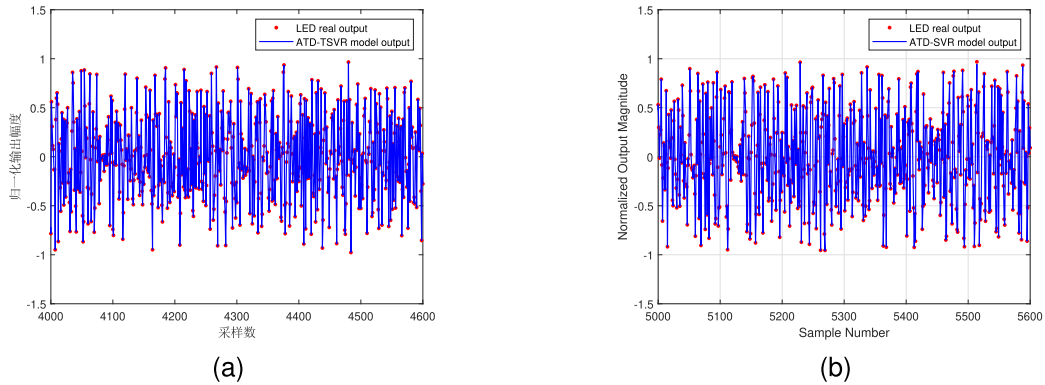
FIGURE 9. Measured and different modeled waveforms from LED. (a) is Measured and ATD-TSVR modeled time domain waveforms of the LED;(b) is Measured and ATD-SVR modeled time domain waveforms of the LED.
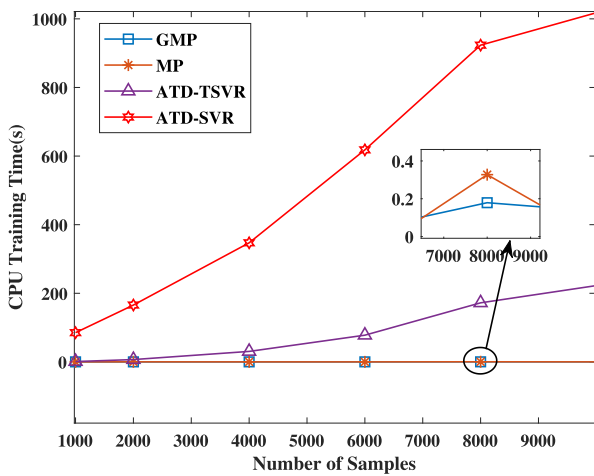

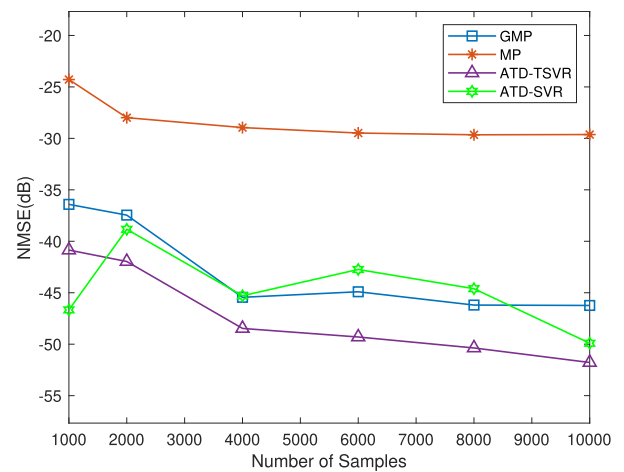**FIGURE 10.** Comparison of modeling time for different models.

FIGURE 11. NMSE performance of different number of samples.

that the precision of ATD-TSVR is better when M=2 than when M=3. Fig. 11 illustrates the relationship between the modeled sample points and NMSE, where M=3, P=9 for MP, M=2, P=9 for GMP, M=2 for ATD-TSVR, and M=2 for ATD-SVR. The figure shows that ATD-TSVR performs exceptionally well in terms of modeling performance. Overall, the precision of ATD-TSVR is higher than the other three methods.

The NMSE of the MP model is strongly influenced by the modeling sample, as the memory polynomial coefficients are unstable. In contrast, GMP introduces more terms to describe the behavioral model of LEDs and improves model precision. Meanwhile, both ATD-TSVR and ATD-SVR show great NMSE performance, and the variation of NMSE does not fluctuate much, indicating good stability and generalization of these two methods. However, the computational efficiency of ATD-SVR is much lower than that of ATD-TSVR. Further analysis revealed that the limit of ATD-TSVR was not reached at 10,000 modeled samples, indicating that TSVR has the capability to model LED nonlinearities for large samples.

C. LINEARIZATION CAPABILITY AND PREDISTORTION

The goal of predistortion is to make the LED and predistorter behave linearly when combined. To achieve this, this paper proposes an adaptive indirect learning predistorter based on the TSVR approach. Fig. 3 shows a more detailed block diagram of the predistorter. To compensate for the LED's nonlinear properties, the predistorter first constructs an inverse model using the input and output signals of the LED to estimate the back-inverse parameters. Then, the predistorter replicates the back-inverse parameters directly into the predistorter as a pre-inverse. The figure also illustrates the Input/Output distortion correction effect of the LED predistortion system based on TSVR. in Fig.12.

The Input/Output curve of the LED without pre-distortion clearly demonstrates nonlinear distortion and memory effect, whereas the Input/Output curve after pre-distortion is nearly linear, indicating that the proposed TSVR-based pre-distorter with adaptive indirect learning architecture effectively compensates the LED's nonlinear distortion and memory effect. To investigate whether the proposed predistorter mitigates the low-pass effect of LEDs, the compensation performance of
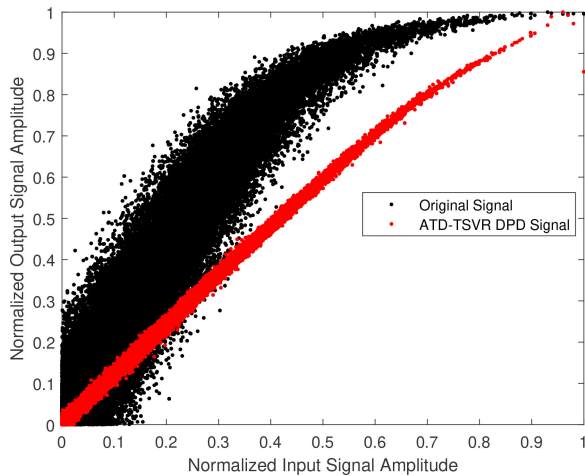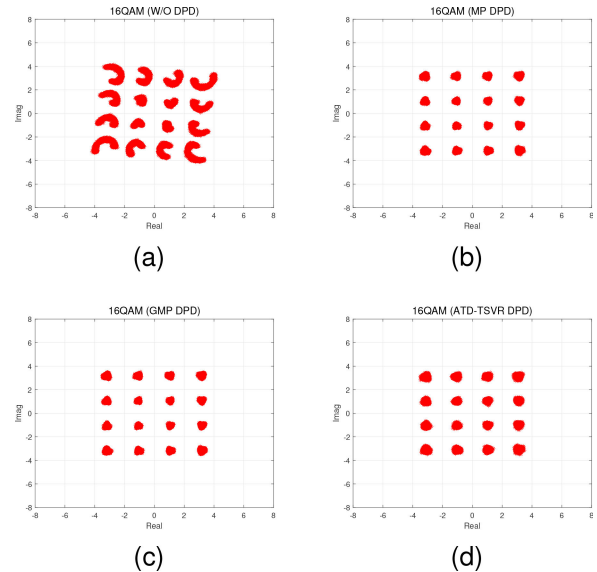
FIGURE 12. Comparison of AM/AM correction.

FIGURE 13. PSD performance comparison of the output signal with different pre-distortion schemes.

FIGURE 14. Constellation diagrams with 40 dB SNR.(a) W/O DPD. (b) MP DPD. (c) GMP DPD. (d) TSVR DPD.

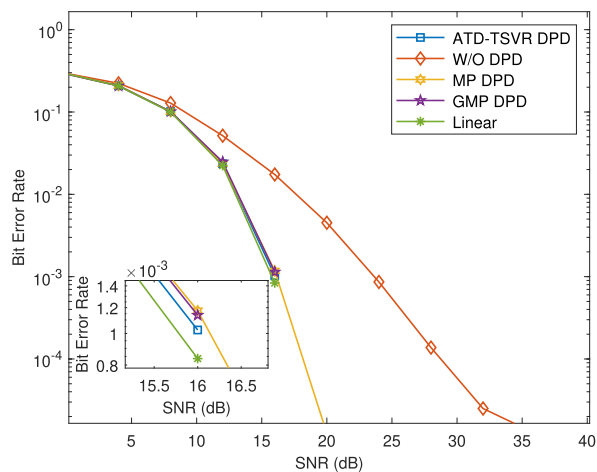**FIGURE 15. BER performance of DCO-OFDM VLC system with and without predistorter for different models.**

the predistortion system is evaluated, and Fig.13 depicts the power spectral density of the output signal through various predistorters compared to the original output signal. By comparing, we determined that the output power spectral density of the model has an effective high-frequency compensation effect. Moreover, the suppression of adjacent channel power demonstrates that the approach efficiently suppresses the out-of-band spectrum expansion.

In order to quantify the compensation impact, the adjacent channel power ratio(ACPR) measurements under each model compensation are presented in Table 7. Compared to the original output signal, the model has a great ability

to suppress out-of-band spectrum regeneration, suppressing 7.16dB at 2.7 MHz and 2.03 dB at 22.7 MHz, respectively. The high-frequency compensation and out-of-band spectrum suppression demonstrate that the proposed ATD-TSVR DPD model also has excellent frequency domain performance.

D. BER PERFORMANCE
The constellation diagram was drawn for the 16QAM symbols at 40dB SNR in Fig.14. In Fig.14(a),(b),(c),(d), the constellation map with pre-distortion are clearer than the constellation map without pre-distortion. By contrast, the constellation diagram is very blurry because static nonlinearity and memory effects produce phase and amplitude distortion without pre-distortion, which is shown in Fig.12(a). The BER performance degrades significantly due to the interference between different constellation points. The channel is assumed to be a line-of-sight channel, and additive white

TABLE 7. ACPR comparison of the different predistorters.

DPD MODEL	ACPR (dB)	
	2.7MHz	22.7MHz
W/O DPD	-22.42	-19.46
MP	-25.93	-16.36
GMP	-29.63	-21.65
ATD-TSVR	-29.58	-21.49

Gaussian noise (AWGN) is introduced at the receiver. Fig.15 shows the BER performance of the VLC system with the different pre-distorters. The SNR is defined as $SNR = E(z(n)^2)/\sigma_n^2$, where $z(n)$ is output signal, and σ_n^2 is the variance of the measurement noise. In general, the GMP model has higher complexity and better performance than the MP model. However, when linear memory effects are predominantly present in the LED, the performance of both models is comparable. Different pre-distortion were compared by design, and it was found that in-band compensation is generally about the same. In pre-distortion design, the accuracy of the forward model and the inverse model of the LED is not identical. When using the same modeling parameters as the forward model, the inverse model of the LED obtained by ATD-TSVR is more similar to that obtained by the polynomial model, which is the reason for the close BER performance. However, the forward modeling accuracy of ATD-TSVR is far superior to that of the traditional polynomial model. Additionally, the complexity of ATD-TSVR is lower than that of support vector regression, and the efficiency of extracting the model is higher.

IV. CONCLUSION

Nonlinearity in VLC systems is primarily caused by the LED, and we propose a pre-distortion technique based on ATD-TSVR with an adaptive indirect learning architecture to compensate for this nonlinearity. Unlike SVR, TSVR generates two small-scale quadratic programming problems (QPPs) to learn the regression function before attempting to solve a larger-scale QPP. This leads to a significant reduction in CPU training time, by more than four times in comparison to conventional SVR techniques. Additionally, the proposed method outperforms certain current traditional methods, as shown by the NMSE, BER, ACPR, PSD, and Constellation Plots results.

Experiment and numerical simulation results demonstrate that the proposed scheme efficiently compensates for nonlinear distortion, including nonlinear memory effects and static nonlinearity of LEDs. The adaptive nature of ATD-TSVR-based pre-distortion enables its usage with a variety of nonlinear light sources and broadband VLC systems to reduce nonlinear distortion in VLC systems. Furthermore, TSVR has a variety of variant algorithms, which makes it a versatile solution to LED nonlinearity.

REFERENCES

[1] K. Ying, Z. Yu, R. J. Baxley, H. Qian, G.-K. Chang, and G. T. Zhou, "Nonlinear distortion mitigation in visible light communications," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 36–45, Apr. 2015.

[2] H. Haas, E. Sarbazi, H. Marshoud, and J. Fakidis, "Visible-light communications and light fidelity," in *Optical Fiber Telecommunications VII*. Amsterdam, The Netherlands: Academic, Jan. 2020, pp. 443–493.

[3] 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, Apr. 2014.

[4] W. Zhao, Q. Guo, J. Tong, J. Xi, Y. Yu, P. Niu, and X. Sun, "Orthogonal polynomial-based nonlinearity modeling and mitigation for LED communications," *IEEE Photon. J.*, vol. 8, no. 4, pp. 1–12, Aug. 2016.

[5] J. Tan, Z. Wang, Q. Wang, and L. Dai, "Near-optimal low-complexity sequence detection for clipped DCO-OFDM," *IEEE Photon. Technol. Lett.*, vol. 28, no. 3, pp. 233–236, Feb. 1, 2016.

[6] H. Zhang, L.-L. Yang, and L. Hanzo, "Piecewise companding transform assisted optical-OFDM systems for indoor visible light communications," *IEEE Access*, vol. 5, pp. 295–311, 2017.

[7] K. Jia, B. Yang, M. Cao, Y. Lin, S. Li, and L. Hao, "An adaptive symbol decomposition with serial transmission for O-OFDM-based VLC system," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 916–920, Mar. 2021.

[8] B. Inan, S. C. J. Lee, S. Randel, I. Neokosmidis, A. M. J. Koonen, and J. W. Walewski, "Impact of LED nonlinearity on discrete multitone modulation," *J. Opt. Commun. Netw.*, vol. 1, no. 5, pp. 439–451, Oct. 2009.

[9] I. Neokosmidis, T. Kamalakis, J. W. Walewski, B. Inan, and T. Sphicopoulos, "Impact of nonlinear LED transfer function on discrete multitone modulation: Analytical approach," *J. Lightw. Technol.*, vol. 27, no. 22, pp. 4970–4978, Nov. 15, 2009.

[10] H. Elgala, R. Mesleh, and H. Haas, "Non-linearity effects and predistortion in optical OFDM wireless transmission using LEDs," *Int. J. Ultra Wideband Commun. Syst.*, vol. 1, no. 2, pp. 143–150, Jan. 2009.

[11] H. Chen, "Low-complexity multi-symbol multi-modulus weighted lookup table predistortion in UWOC system," *J. Lightw. Technol.*, vol. 40, no. 13, pp. 4224–4236, Jul. 2022.

[12] J.-K. Lain and Y.-H. Chen, "An ANN-based adaptive predistorter for LED nonlinearity in indoor visible light communications," *Electron.*, vol. 10, no. 8, pp. 948–961, Apr. 2021.

[13] Y. Zhou, J. Zhang, and N. Chi, "A novel memoryless power series based adaptive nonlinear pre-distortion scheme in high speed visible light communication," in *Proc. Opt. Fiber Commun. Conf.*, 2017, pp. 1–3.

[14] 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, May 15, 2016.

[15] G. Stepniak, M. Marzecki, and J. Bojarczuk, "Volterra predistorter for the dynamic nonlinearity of LED," *Opt. Lett.*, vol. 47, no. 5, pp. 1161–1164, Mar. 2022.

[16] P. Miao, G. Chen, X. Wang, Y. Yao, and J. A. Chambers, "Adaptive nonlinear equalization combining sparse Bayesian learning and Kalman filtering for visible light communications," *J. Lightw. Technol.*, vol. 38, no. 24, pp. 6732–6745, Dec. 1, 2020.

[17] 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, pp. 1–7, Jun. 2015.

[18] 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. 15, 2013.

[19] 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. 1, 2010.

[20] H. Chen, W. Niu, G. Li, Z. He, J. Zhang, N. Chi, and Z. Li, "Computationally efficient pre-distortion based on adaptive partitioning neural network in underwater visible light communication," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2022, pp. 1–3.

[21] W. Niu, H. Chen, F. Hu, J. Shi, Y. Ha, G. Li, Z. He, S. Yu, and N. Chi, "Neural-network-based nonlinear tomlinson-harashima precoding for bandwidth-limited underwater visible light communication," *J. Lightw. Technol.*, vol. 40, no. 8, pp. 2296–2306, Apr. 15, 2022.

[22] 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. 26700–26711, Oct. 2018.

[23] W. Niu, Y. Ha, and N. Chi, "Support vector machine based machine learning method for GS 8QAM constellation classification in seamless integrated fiber and visible light communication system," *Sci. China Inf. Sci.*, vol. 63, no. 10, pp. 1–12, Oct. 2020.

[24] T. Nguyen, S. Mhatli, E. Giacoumidis, L. Van Compernolle, M. Wuilpart, and P. Megret, "Fiber nonlinearity equalizer based on support vector classification for coherent optical OFDM," *IEEE Photon. J.*, vol. 8, no. 2, pp. 1–9, Apr. 2016.

[25] H. Sun, Y. Zhang, F. Wang, J. Zhang, and S. Shi, "SVM aided signal detection in generalized spatial modulation VLC system," *IEEE Access*, vol. 9, pp. 80360–80372, 2021.

[26] Y. Yuan, M. Zhang, P. Luo, Z. Ghassemlooy, L. Lang, D. Wang, B. Zhang, and D. Han, "SVM-based detection in visible light communications," *Optik*, vol. 151, pp. 55–64, Dec. 2017.

[27] X. Lu, K. Wang, L. Qiao, W. Zhou, Y. Wang, and N. Chi, "Nonlinear compensation of multi-CAP VLC system employing clustering algorithm based perception decision," *IEEE Photon. J.*, vol. 9, no. 5, pp. 1–9, Oct. 2017.

[28] X. Lu, M. Zhao, L. Qiao, and N. Chi, "Non-linear compensation of multi-CAP VLC system employing pre-distortion base on clustering of machine learning," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2018, pp. 1–3.

[29] E. Giacoumidis, S. Mhatli, T. Nguyen, S. T. Le, I. Aldaya, M. A. McCarthy, and B. J. Eggleton, "Kerr-induced nonlinearity reduction in coherent optical OFDM by low complexity support vector machine regression-based equalization," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2016, pp. 1–3.

[30] L. Han, Y. Wang, C. Li, Z. Gao, and X. Xin, "Nonlinear damage compensation using support vector regression," in *Proc. 19th Int. Conf. Opt. Commun. Netw. (ICOCN)*, Aug. 2021, pp. 1–3.

[31] Q. Sun, X. Chen, and Y. Zhou, "Performance monitoring of PAM4 optical communication system based on principal component analysis and support vector regression," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2019, pp. 535–539.

[32] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Oct. 1995.

[33] J. Cai, C. Yu, L. Sun, S. Chen, and J. B. King, "Dynamic behavioral modeling of RF power amplifier based on time-delay support vector regression," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 2, pp. 533–543, Feb. 2019.

[34] J. Xu, W. Jiang, L. Ma, M. Li, Z. Yu, and Z. Geng, "Augmented time-delay twin support vector regression-based behavioral modeling for digital predistortion of RF power amplifier," *IEEE Access*, vol. 7, pp. 59832–59843, 2019.

[35] R. Khemchandani and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905–910, May 2007.

[36] S. Zhang, C. Liu, T. Zhou, and L. Sun, "Twin least squares support vector regression of heteroscedastic Gaussian noise model," *IEEE Access*, vol. 8, pp. 94076–94088, 2020.

[37] X. Peng, "TSVR: An efficient twin support vector machine for regression," *Neural Netw.*, vol. 23, no. 3, pp. 365–372, Apr. 2010.

[38] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proc. IEEE*, vol. 69, no. 12, pp. 1557–1573, Dec. 1981.

[39] L. Ding, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[40] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 4, pp. 1166–1174, Apr. 2012.

KUN ZHANG received the Ph.D. degree from Harbin Engineering University, in 2022. He joined the Institute of Green Intelligent Technology, Chinese Academy of Sciences, Chongqing, China, as a special Research Assistant, in 2022. His research interests include infrared sensing and semiconductor detector technology.

YUNJIAN JIA (Member, IEEE) received the Ph.D. degree in engineering from Osaka University, Japan, in 2006. He worked as a tenured Researcher with the Central Research Institute of Hitachi, Japan, from 2006 to 2012, where he was the Leader of the Mobile Network Advance Research Group and the Next Generation Mobile Communication International Standard Group. Since 2007, he has been leading a team to participate in the work of 3GPP, an authoritative international standards organization, and researched and developed 4G communication standards with representatives from major enterprises and research institutions around the world. After returning to China, he systematically carried out research on the new architecture and technology of 5G communication networks, and as a member of the National IMT-2020 (5G) Working Group and one of the main authors of the white paper "5G Network Technology Architecture" in China. He actively promoted to the China's 5G research and international standardization work, especially the key technology of network intelligence.

HAIPENG ZHA received the B.Sc. degree from the School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China, in 2016. He is currently pursuing the master's degree with the University of Chinese Academy of Sciences. His research interests include communication signal processing, machine learning, and digital predistortion implementation in FPGA.

WENQIANG LU (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Physics, Nankai University, in 1998 and 2005, respectively. He was with the Computer Hard Disk Research and Development Center, Showa Denko Corporation, Ichihara, Chiba, Japan, from October 1998 to August 2000. In 2012, he was a Visiting Scholar with the Department of Materials, College of Engineering Center for Information Materials Technology (MINT), The University of Alabama, USA. His research interests include the preparation of gallium oxide third generation (broadband) semiconductor nanomaterials and their photoelectric sensing performance research (to explore its application research in photoelectricity, photocatalysis, and mechanical sensing). He is the Chairperson of the IEEE Solid State Circuits Society Chongqing Chapter.

• • •