Data-driven Channel Estimation for OTFS Systems Affected by HPA

Xinlin Wang, Shuo Cui, Qiang Ai, *Member, IAENG*

***Abstract*—In modern wireless communications, efficient utilization of spectrum resources and enhancement of communication quality are key objectives. Orthogonal Time Frequency Space (OTFS) modulation technology has become a research hotspot due to its superior performance in the delay-Doppler domain. This paper focuses on the issue of nonlinear distortion in OTFS modulation systems caused by High Power Amplifier (HPA) effects. We propose a novel data-driven channel estimation algorithm based on Transformer and adaptive filters, which performs channel estimation in the time-frequency (TF) domain to effectively mitigate the nonlinear distortion introduced by HPA. Simulation results show that compared with traditional methods, our proposed method demonstrates higher estimation accuracy and better anti-interference capability in high dynamic environments, significantly improving the robustness of signal transmission, especially in high-speed mobile communication scenarios.**

***Index Terms*—OTFS, channel estimation, Transformer, simulation, wireless communications.**

I. INTRODUCTION

IN modern wireless communication systems, the efficient utilization of spectrum resources and the improvement of communication quality have always been focal points of research [1]. With the widespread adoption of 5G technology and the gradual development of 6G, the demand for communications in high-speed mobility environments is becoming increasingly urgent [2]–[4]. Traditional Orthogonal Frequency Division Multiplexing (OFDM) technology experiences significant performance degradation when faced with high Doppler shifts and multipath effects [5]. Consequently, Orthogonal OTFS modulation technology has become a hot topic in recent years due to its superior performance in the delay-Doppler domain [6], [7]. OTFS modulation, by operating in the delay-Doppler domain, transforms the signal transmission environment from a time-varying channel to a time-invariant channel, significantly enhancing signal robustness [8]. However, in practical applications, the acquisition of Channel State Information (CSI) for OTFS systems still poses numerous challenges. Particularly in high-speed mobility scenarios, the rapid changes in the channel make channel estimation complex and cost-intensive [9].

Due to the significant advantages and broad application prospects of OTFS technology, it has garnered widespread attention within the academic community. Particularly in OTFS systems, the channel estimation algorithm directly determines the performance of the transmission, making related research especially critical [10], [11]. Existing methods for channel estimation in OTFS primarily include techniques based on pilot symbols, compressive sensing, and machine learning approaches [12]–[14].

Pilot Symbol-Based Channel Estimation: This approach involves embedding pilot symbols within the transmitted data, which are then utilized at the receiver end to estimate the channel [15], [16]. While straightforward and intuitive, this method can be costly in terms of pilot overhead in scenarios with high Doppler shifts, and its accuracy is susceptible to noise and interference [17], [18].

Compressive Sensing Methods: Compressive sensing leverages the sparse characteristics of the channel to reduce the number of samples required for channel estimation. Typical compressive sensing algorithms include Orthogonal Matching Pursuit (OMP) and Least Absolute Shrinkage and Selection Operator (Lasso) [19]–[22]. These methods perform well in scenarios where the channel exhibits sparsity; however, they are computationally intensive and tend to underperform in non-sparse channel conditions.

Machine Learning Methods: In recent years, machine learning techniques have been extensively applied to channel estimation. These methods utilize large-scale data sets to train models, enabling effective estimation in complex channel environments. Common machine learning techniques include Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) [23]–[29]. However, these approaches generally require substantial training data and often exhibit limited generalizability in highly dynamic environments.

Although existing methods have improved the performance of OTFS channel estimation to some extent, numerous challenges persist. For instance, pilot-symbol-based methods are costly and unstable in high Doppler shift environments, compressive sensing techniques are computationally intensive, and machine learning approaches rely heavily on extensive data and have limited generalizability. Furthermore, nonlinear distortions caused by High Power Amplifiers (HPAs) and issues with high Peak-to-Average Power Ratio (PAPR) pose additional demands on channel estimation. High PAPR values can lead to nonlinear distortions in the presence of HPAs, reducing overall system performance. To address these issues, Reference [30] utilized mu-law compression to reduce the PAPR of OTFS signals by 2.5 dB, but this resulted in a slight increase in the Bit Error Rate (BER) compared to the original OTFS signals. Reference [31] confirmed the effectiveness of Selected Mapping (SLM) technique in suppressing the PAPR of OTFS modulation signals, with results indicating superior performance of OTFS over OFDM signals. Reference [32] combined Genetic Algorithms (GA) with SLM to handle the PAPR of OFDM signals, but GAs have high dimensions, significant complexity, and parameter

Manuscript received August 5, 2024; revised November 13, 2024.

Xinlin Wang is a lecturer at the Software College, Liaoning Technical University, Huludao 125105, China (wangxinlin1107@foxmail.com).

Shuo Cui is a postgraduate student at the Software College, Liaoning Technical University, Huludao 125105, China (Corresponding author to provide e-mail: 2727269416@qq.com).

Qiang Ai is a postgraduate student at the College of Computer, Qinghai Normal University, Xining 810008, China (e-mail: qiang.ai@outlook.com).

selection heavily dependent on experience. These studies have not yet integrated research on signal design and performance optimization for communication and sensing in OTFS-based systems. Therefore, developing a new method that can perform channel estimation efficiently and accurately in highly dynamic environments is particularly crucial. In response to these challenges, we propose a channel estimation method for OTFS modulation in the time-frequency domain, based on Transformer and Adaptive Filters. This method combines a preliminary coarse channel estimation using a preamble sequence with frequency domain pilots, utilizing the Transformer network to capture the temporal correlations of the channel, and employing adaptive filters to enhance noise reduction capabilities. Our approach not only significantly reduces pilot overhead but also effectively addresses nonlinear distortions caused by HPAs, thereby enhancing the accuracy and robustness of channel estimation.

II. PROBLEM EXPLANATION

A. OTFS System

OTFS is an emerging modulation technique designed to enhance the performance of wireless communication systems under high Doppler frequencies and complex multipath channel conditions. By representing signals in a two-dimensional time-frequency domain, OTFS effectively captures and processes the time-varying and frequency-varying characteristics of wireless channels, significantly boosting the robustness and reliability of communication systems. OTFS processes data in the Delay-Doppler (DD) domain, with its overall architecture illustrated in Fig. 1.

1) ISSFT: At the transmitter, the $M \times N$ information symbols $X_{DD} \in \mathbb{C}^{M\times N}$ from the modulation alphabet are assigned to a $M \times N$ grid in the DD domain, where M and N respectively represent the number of subcarriers and time slots used. The symbols in the DD domain are transformed into the TF domain through the Inverse Symplectic Finite Fourier Transform (ISFFT). Here, the TF domain is discretized into an $M \times N$ grid, and the element (m, n) of X_{TF} is represented as:

$$X_{TF}[m,n] = \frac{1}{\sqrt{MN}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} X_{DD}[k,l] e^{j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} \tag{1}$$

In the transformation process, $X_{DD}[k, l]$ represents the element at position (k, l) within the matrix X_{DD}, where k ranges from 0 to $M - 1$ and l ranges from 0 to $N - 1$.

2) Heisenberg transform: The discrete Heisenberg transform is applied to generate the time-domain transmission signal. The vector form of the transmission signal is expressed as:

$$s = \text{vec}(G_{\text{tx}} F_M^H X_{TF}) = (F_N^H \otimes G_{\text{tx}}) x_{DD} \tag{2}$$

The pulse shaping waveform, denoted by G_{tx}, is considered as a rectangular waveform lasting for duration T_s, which implies $G_{\text{tx}} = I_M$. The matrix X_{DD} is vectorized to form $x_{DD} = \text{vec}(X_{DD})$, expressed as $x_{DD} = [x_{DD}(0), \cdots, x_{DD}(MN - 1)]^T$. The vector $s \in \mathbb{C}^{MN\times 1}$ represents the form of the transmission signal, expressed as $s = [s(0), \cdots, s(n), \cdots, s(MN - 1)]^T$, where n ranges from 0 to $MN - 1$, and the expression for $s(n)$ is:

$$s(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{j2\pi \lfloor \frac{n}{M} \rfloor k/N} x_{DD}([n]_M + kM) \tag{3}$$

3) Channel: The OTFS wireless channel, characterized as a time-varying multipath channel, can be described through its impulse response in the Delay-Doppler domain.

$$h(\tau, v) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i)\delta(v - v_i) \tag{4}$$

The impulse response of the OTFS channel is characterized by the Dirac delta function, denoted by $\delta(\cdot)$. The path gain, h_i, for the i-th path, follows a normal distribution with mean zero and variance $1/P$, where P represents the total number of paths.

4) Wigner transform: At the receiver, the time-domain received signal is expressed by the following equation:

$$r(t) = \iint h(\tau, v) s(t - \tau) e^{j2\pi v(t-\tau)} d\tau dv + w(t) \tag{5}$$

The time-domain received signal, denoted as $s(t)$, and the channel characteristics in the DD domain, represented as $h(\tau, v)$, are considered. The received signal $r(t)$ is sampled at the time instances $t = \frac{n}{M\Delta f}$, where $n = 0, \cdots, MN - 1$. Consequently, the discretized received signal $r(n)$ is expressed as:

$$r(n) = \sum_{i=1}^{P} h_i e^{j2\pi \frac{k_i(n-l_i)}{MN}} s([n - l_i]_{MN}) + w(n) \tag{6}$$

The noise w is a complex-valued Gaussian white noise, independently and identically distributed following the complex normal distribution $N(0, \sigma_c^2 I)$, where σ_c^2 denotes the noise variance. $I_{MN}(l_i)$ is an $MN \times MN$ matrix obtained by cyclically shifting the columns of the identity matrix to the left. Δ is an $MN \times MN$ Doppler frequency shift diagonal matrix defined as $\Delta(k_i) = \text{diag}\left(e^{j2\pi k_i \frac{0}{MN}}, e^{j2\pi k_i \frac{1}{MN}}, \cdots, e^{j2\pi k_i \frac{MN-1}{MN}}\right)$, where $\text{diag}(\cdot)$ converts a vector into a diagonal matrix. Through the Wigner transform, the TF domain received signal $Y_{TF} \in \mathbb{C}^{M\times N}$ is expressed as:

$$Y_{TF} = F_M G_{\text{rx}} R \tag{7}$$

The matrix $R = \text{vec}^{-1}(r)$ is obtained by the inverse vectorization of vector r. The receive filter G_{rx} is a rectangular waveform with duration T_s, and is represented by $G_{\text{rx}} = I_M$.

5) SFFT: DD domain received signal $Y_{DD} \in \mathbb{C}^{M\times N}$ is obtained by performing the Symplectic Finite Fourier Transform (SFFT), expressed as:

$$Y_{DD} = F_M^H Y_{TF} F_N = F_M^H F_M G_{\text{rx}} R F_N = G_{\text{rx}} R F_N \tag{8}$$

Here, Y_{TF} represents the received signal in the TF domain as processed through the receiver's filter systems. The matrices F_M and F_N are Fourier transform matrices applying transformations across subcarriers and time slots, respectively, with F_M^H being the Hermitian of F_M, indicating the inverse Fourier transform. The receiver's filter G_{rx}, typically a rectangular waveform, shapes the waveform in signal processing. The vector r is transformed back into a matrix R through inverse vectorization. The signal

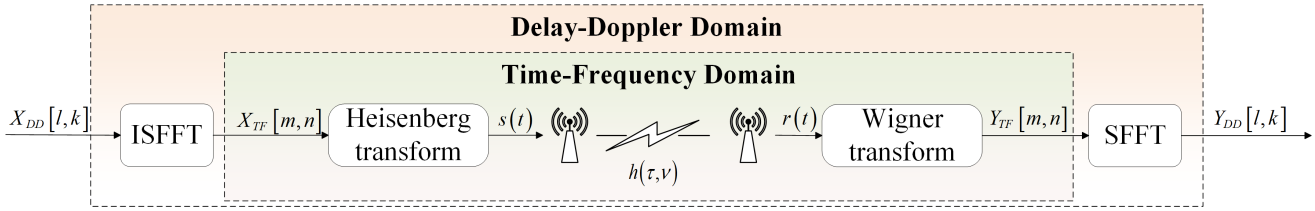

Fig. 1. Structure of OTFS System.

Y_{DD} in the DD domain is then obtained by applying the Symplectic Finite Fourier Transform (SFFT) through the operations described. The matrix Δ represents a Doppler frequency shift diagonal matrix, formulated as $\Delta(k_i) = \text{diag}(e^{j2\pi k_i \frac{0}{MN}}, e^{j2\pi k_i \frac{1}{MN}}, \ldots, e^{j2\pi k_i \frac{MN-1}{MN}})$ where $\text{diag}(\cdot)$ turns a vector into a diagonal matrix. The transformation simplifies to $G_{\text{rx}} R F_N$ because the product $F_M^H F_M$ simplifies to the identity matrix I, thus not altering the product but simplifying the expression.

B. HPA

HPAs are crucial components in wireless communication systems, responsible for amplifying modulated low-power signals to levels sufficient for transmission over long distances or in complex wireless environments. Ideally, an HPA's output is a simple amplification of the input signal; however, due to hardware limitations, the output often exhibits nonlinear distortions, especially when the input signal approaches or exceeds the amplifier's saturation point. This can be mathematically modeled as:

$$y = H(x) = ax + bx^2 + cx^3 + \cdots \tag{9}$$

where a, b, c, etc., are coefficients that describe the nonlinear characteristics of the HPA.

In OTFS modulation systems, the HPAs significantly affect system performance, particularly in handling high dynamic range signals. OTFS is designed to excel in environments with high Doppler shifts and multipath effects, which typically present signals with high PAPR. Such signals challenge HPAs to operate over a broad dynamic range without introducing significant nonlinear distortions.

Assuming the signal input to the HPA is s_{TF}, and the output from the HPA is s_{HPA}, the operation of the HPA can be expressed as:

$$s_{HPA}(t) = H(s_{TF}(t)) \tag{10}$$

Due to the high PAPR of OTFS signals, they are more susceptible to nonlinear distortions from the HPA, especially near the amplifier's saturation. These distortions not only increase interference in the signal but can also lead to increased BER.

C. Channel Estimation

Channel estimation is a crucial preprocessing step in communication systems, aimed at quantifying the various effects experienced by a signal as it passes through wireless or wired transmission media. The primary task of channel estimation is to reconstruct the channel characteristics based on the received signal so that the transmitted signal can be accurately recovered at the receiver, thus ensuring the accuracy and reliability of data transmission.

Channel estimation primarily involves determining the channel impulse response $h(t)$ or its discrete counterpart $h[n]$. This estimation process requires precise quantification of the channel's multipath effects, Doppler effects, path loss, and phase shifts. Specifically, the received signal $r(t)$ can be modeled as the convolution of the transmitted signal $x(t)$ with the channel impulse response $h(t)$, plus noise $n(t)$, expressed mathematically as:

$$r(t) = (h * x)(t) + n(t) \tag{11}$$

where $*$ denotes the convolution operation. In discrete-time systems, the model can be simplified to:

$$r[n] = \sum_{k=0}^{L-1} h[k]x[n-k] + n[n] \tag{12}$$

During the signal reception phase, the receiver first needs to estimate the channel impulse response, denoted as $\hat{h}$. This process is typically accomplished by transmitting a known pilot sequence and processing it at the receiver. The estimated channel matrix $\hat{h}$ can be obtained by minimizing the error between the received signals and the pilot sequence transmitted through the estimated channel model:

$$\hat{h} = \arg\min_{h} \|r - h * p\|^2 \tag{13}$$

Here, p represents the transmitted pilot sequence, $*$ denotes convolution, and r is the received signal.

Once channel estimation is obtained, it can be used to detect data symbols. At the receiver, the estimated channel matrix is used to equalize the received signal, compensating for the effects caused by the channel. Mathematically, this can be achieved through various methods, such as Zero Forcing (ZF) equalization or Minimum Mean Square Error (MMSE) equalization. For ZF equalization, the detected symbols $\hat{Y}_{DD}$ are given by:

$$\hat{Y}_{DD} = \hat{h}^{-1} \cdot Y_{DD} \tag{14}$$

Here, $\hat{h}^{-1}$ represents the inverse of the estimated channel matrix, and Y_{DD} is the received signal.

III. METHODOLOGY

A. Signal Generation

This paper proposes a novel channel estimation algorithm addressing the issue of HPA distortion in OTFS systems. As Fig 2 shows, the algorithm is trained and validated in a simulation environment. Firstly, we generate an OTFS communication dataset that includes the impact of HPA distortion using the simulation system. This simulation environment retains the true channel structure for each sample.

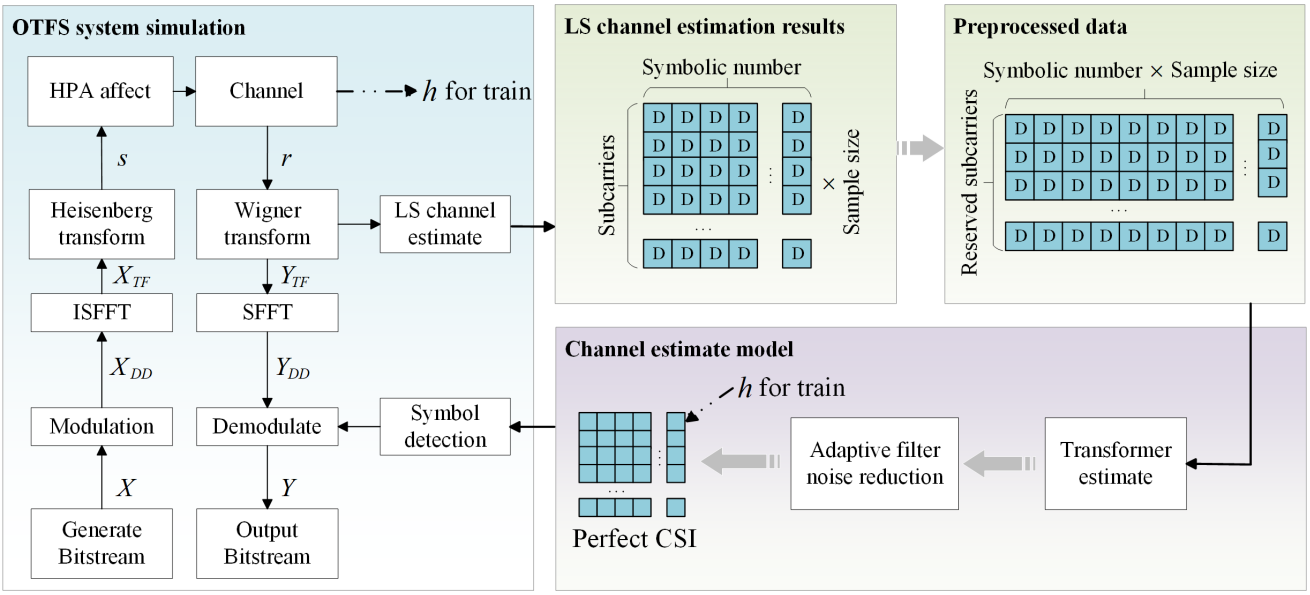

Fig. 2. Structure of the method.

During the simulation experiment, the program uses the Least Squares (LS) channel estimation method to obtain an initial channel estimate from the pilot sequences and frequency domain pilot signals, which, together with the true channel structure, forms a training dataset for offline training of the deep learning module. Subsequently, the accuracy of the channel structure estimated during online simulation is verified using the test dataset during the demodulation phase. This method efficiently handles channel estimation for nonlinear and highly selective channels while maintaining low pilot overhead.

This paper defines a basic transmission packet that includes a known deterministic preamble sequence (used for initial channel synchronization) and a data field. Additionally, a Cyclic Prefix (CP) is used to mitigate Inter-Symbol Interference (ISI) caused by multipath propagation. In the data field, each symbol utilizes M subcarriers, of which only M_{on} are active, with the remaining subcarriers being inactive. Here, $|M_{on}|$ represents the number of pilot subcarriers. Further, among the M_{on} subcarriers, M_p are allocated as pilots, and the remaining M_d subcarriers are used for data transmission.

In the signal, the pilots are embedded in the TF domain, resulting in a reduced number of subcarriers for data transmission in the DD domain while maintaining the subcarrier spacing and bandwidth. The signal is transformed to the TF domain, considering only the data part, resulting in an $M_d \times N$ matrix:

$$X_{TF_d}[m_d, n] = \frac{1}{\sqrt{NM_d}} \sum_{k=0}^{N-1} \sum_{l=0}^{M_d-1} X_{DD}[l,k] e^{j2\pi\left(\frac{nk}{N} - \frac{m_d l}{M_d}\right)} \tag{15}$$

Then pilots are then inserted into the subcarriers as follows:

$$X_{TF}[m,n] = \begin{cases} \text{pilots} & \text{if } m \in M_p \\ X_{TF_d}[m,n] & \text{if } m \in M_d \end{cases} \tag{16}$$

Fig. 3 illustrates the TF domain frame structure, where D represents data subcarriers, P represents pilot subcarriers, and PR represents the preamble sequence. This configuration ensures no overlap between pilot and data subcarriers, thereby optimizing the utilization of resources during channel estimation.

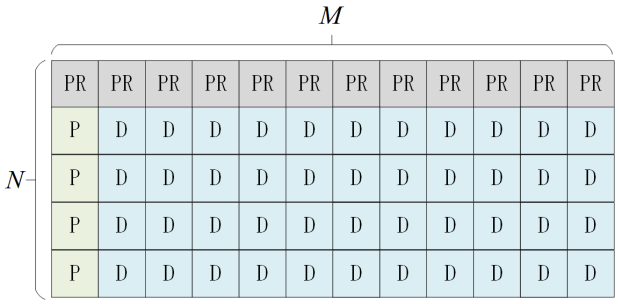

Fig. 3. TF domain frame structure.

B. Preliminary Estimate

LS channel estimation method in the frequency domain to process the preamble and pilot information, initiating the channel estimation process. The initial channel estimation process for the preamble is as follows:

$$\hat{H}_{TF}^{LS}[m, n_{\text{pr}}] = \frac{\mathcal{Y}[m, n_{\text{pr}}]}{\mathcal{P}[m]}, \forall m \tag{17}$$

where $\mathcal{Y}[m, n_{\text{pr}}]$ is the frequency domain signal of the m-th subcarrier obtained by demodulating the preamble at the n_{pr}-th symbol position. $\mathcal{P}[m]$ denotes the predefined frequency domain preamble sequence. For the pilot subcarriers, the initial channel estimation result is as follows:

$$\hat{H}_{TF}^{LS}[m_p, n] = \frac{\mathcal{Y}[m_p, n]}{\mathcal{S}[m_p, n]}, \forall n \tag{18}$$

Thus, for the n-th symbol of each frame, $\mathcal{Y}[m_p, n]$ and $\mathcal{S}[m_p, n]$ represent the received and transmitted frequency domain signals at the m_p pilot positions, respectively. In this experiment, the pilot and transmitted signals have the same power. The pilot information serves as the basis for

interpolating the data subcarrier information to obtain the final channel estimation.

C. Transformer Estimator

We utilize a Transformer-based receiver which demonstrates significant performance improvements in symbol-by-symbol estimation compared to traditional DNN receivers. The Transformer processes and learns data sequences through a self-attention mechanism, which effectively captures long-range dependencies within the sequence data. Unlike traditional feedforward DNN architectures, the Transformer can process data in parallel, thereby improving training and inference efficiency. Through the multi-head attention mechanism, the Transformer can focus on different parts of the input sequence simultaneously, considering more information when estimating the current output. The internal mechanism of the Transformer is as follows:

First, the Transformer generates the query (Q), key (K), and value (V) matrices from the input sequence through linear transformations using the self-attention mechanism, as shown in Equation (19):

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \tag{19}$$

Then, the attention weights are calculated using Equation (6):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{20}$$

where d_k is the dimension of the key vectors. The result of the self-attention calculation is processed by a feed forward neural network to enhance feature representation:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{21}$$

Here, $\max(0, x)$ denotes the activation function, and we chose the Rectified Linear Unit (ReLU) as the activation function. This choice is due to its fast computation, low parameter tuning requirements, and significant effectiveness in solving optimization problems. The weight matrices W_1 and W_2 correspond to the linear transformations of the input.

Through self-attention mechanisms, the Transformer can effectively learn and capture the temporal correlations of the channel, thereby excelling in symbol-by-symbol estimation. In the context of channel estimation, the Transformer's input must be related to the number of active and pilot subcarriers. Additionally, each Transformer network contains multiple encoder and decoder layers, which determine the steps of parallel operations. These processing characteristics enable the Transformer to learn the time-varying correlations of the channel and adjust the channel estimation accordingly. We use the initial channel estimation introduced in the previous section as the input to the Transformer interpolation algorithm to obtain the channel information for the entire frame.

D. Adaptive Filter Noise Compensation

To reduce noise, the output of the Transformer layer is further processed by Adaptive Filters (AF). The goal of the adaptive filters is to update the initial estimates and learn to correct the estimation errors compared to the perfect CSI channel.

The adaptive filter network is trained to determine the parameters $\theta^\star$ that minimize the loss function $\ell(\theta)$. This process can be expressed as:

$$\theta^\star = \arg\min_\theta \ell(\theta, \hat{H}_{TF}^{Transformer}[m,n], H_{TF}[m,n]) \tag{22}$$

where θ represents the weight vector of the adaptive filters, $H_{TF}[m,n]$ is the perfect channel response obtained from the training sample vectors used during the training phase of our proposed estimator, and $\hat{H}_{TF}^{Transformer}[m,n]$ is the output of the Transformer. The loss is calculated as:

$$\frac{1}{NM}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left|H_{TF}[m,n] - \hat{f}(\hat{H}_{TF}^{Transformer}[m,n], \theta)\right|^2 \tag{23}$$

During the training phase, this process is iterative, with the neural network adapting $\hat{f}(\hat{H}_{TF}^{Transformer}[m,n], \theta)$ based on each training sample. In summary, the final channel estimation is given by:

$$\hat{H}_{Transformer}[m,n] = \hat{f}(\hat{H}_{TF}^{Transformer}[m,n], \theta^\star) \tag{24}$$

IV. EXPERIMENTS

This chapter evaluates the performance of the proposed channel estimation method in a communication environment affected by nonlinear distortion through controlled experiments. The proposed method is compared with two baseline methods, TCE and CCE. In these methods, channel estimation is accomplished through the pilot responses in the DD domain. For these baseline methods, the pilot Signal-to-Noise Ratio (SNR) is set to $\text{SNR}_p = 40$ dB according to the design assumptions.

A. Dataset Simulation

To ensure compatibility, our system design adheres to the physical layer specifications of the IEEE 802.11p communication standard. Therefore, we consider a bandwidth of $B = 10$ MHz and a carrier frequency of $f_c = 5.9$ GHz, transmitting frames with M subcarriers and $N = 14$ symbols. Additionally, we use a Turbo LTE coding with a rate of 1/2 and perform equalization using the MMSE criterion. Unlike other estimators, our proposed method performs equalization within the Time-TF domain.

Furthermore, simulations are conducted under the ITU-T Vehicular-A channel model at a speed of $v = 300$ km/h, employing 16-QAM and QPSK modulation. Our evaluation also considers the effects of HPA non-linearity. For the highest modulation order, the Input Back-Off (IBO) is set to 4 dB, while for the non-linear effects of QPSK modulation, the IBO is set to 2 dB. Table I summarizes the parameters used in the simulations.

B. Evaluation Metrics

To comprehensively evaluate the performance of the proposed method, we utilize two key metrics: the BER and the PAPR. This article will detail each metric and its calculation methods.

The BER is defined as the ratio of the number of erroneous bits to the total number of transmitted bits, and it is one of the

TABLE I
SUMMARY OF PARAMETERS

Symbol	Description
Symbols Number	14
Subcarriers Number	52
Band-width	10 MHz
Carrier Frequency	5.9 GHz
Symbol Duration	8 μs
Subcarrier Spacing	156.25 kHz
Mobility Speed	300 km/h
Input Back-Off	2 dB, 4 dB
Modulation	QPSK, 16-QAM

important indicators of communication system performance. The definition of BER is as follows:

$$\text{BER} = \frac{N_e}{N_t} \tag{25}$$

This formula expresses the BER as a simple fraction or ratio, where the numerator is the number of bits that were incorrectly received, and the denominator is the total number of bits sent during the transmission. This metric is crucial for understanding the robustness and reliability of a communication link under various conditions, such as noise, interference, and signal distortion.

The PAPR is another critical metric used to assess the performance of communication systems, particularly in the context of OFDM transmissions, which are prone to high PAPR values. PAPR is defined as follows:

$$\text{PAPR} = \frac{\max_{0\le t\le T} |x(t)|^2}{\text{mean}_{0\le t\le T}|x(t)|^2} \tag{26}$$

The instantaneous value of the transmitted signal is denoted by $x(t)$. The maximum instantaneous power of the signal within the time interval $0 \le t \le T$ is represented by $\max_{0\le t\le T} |x(t)|^2$, and the average power over the same interval is given by $\text{mean}_{0\le t\le T}|x(t)|^2$.

High PAPR values can lead to inefficiencies in the power amplifiers used in communication systems, potentially causing nonlinear distortions if the amplifier enters saturation. Minimizing PAPR is essential for efficient amplifier operation and overall system performance.

C. Baseline Methods

1) TCE: This method has proposed an embedded pilot scheme that includes sufficiently large guard intervals around a unique pilot to improve the acquisition of responses in the DD domain. The allocation of pilot and data symbols within the OTFS frame is as follows:

$$X_{DD}[l,k] = \begin{cases} \text{pilots} & \text{if } l = l_p, k = k_p \\ 0 & \text{if } |l-l_p| \le G_l, |k-k_p| \le G_k \\ \text{data} & \text{otherwise} \end{cases} \tag{27}$$

In this setup, G_l and G_k represent the guard intervals along the delay and Doppler axes, respectively. The channel estimation method is based on the received symbols $Y_{TCE}[l,k]$ within the sub-grid defined by $(l_p - G_l \le l \le l_p + G_l, k_p - G_k \le k \le k_p + G_k)$.

Therefore, using a thresholding method, the channel estimate for this grid is expressed as:

$$\hat{h}_{DD_{TCE}}[l-l_p, k-k_p] = \begin{cases} \frac{Y_{TCE}[l,k]}{X_{DD}[l_p,k_p]} & \text{if } Y_{TCE}[l,k] \ge \vartheta \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Here, ϑ represents the detection threshold, which is set at $\vartheta = 3\sigma_w$. The σ_w denotes the standard deviation of the effective noise in the pilot signal.

2) CCE: This method introduces an estimator for OTFS systems that obtains the cross-correlation channel matrix through estimation in the DD domain. It assumes that the channel remains invariant over multiple symbol durations and that pilots and data are transmitted in separate frames. Consequently, in the DD domain, the pilot signals are considered as:

$$X_{DD}[l,k] = \begin{cases} 1 & \text{if } l = l_p, k = k_p \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

Consequently, the estimated channel response in the DD domain can be represented as:

$$\begin{aligned} \hat{h}_{DD_{CCE}}[l,k] = \sum_{\kappa=1}^{K} h_\kappa \delta((l-l_\kappa) - l_{\tau_\kappa}) e^{j\phi_\kappa} \cdot \psi_\kappa[l] \\ \times \Upsilon_N(k v_\kappa + \kappa v_\kappa - (k-k_\kappa)) \end{aligned} \tag{30}$$

In this representation, h_K denotes the path gain, while l_{τ_K} and k_{v_K} are the integer indices of delay and Doppler shifts, respectively, in the DD domain. The initial phase is represented by ϕ. The fractional Doppler shift, κ_{v_K}, uses a non-integer index to denote the Doppler value of the K-th path, where $(k_{v_K} + \kappa_{v_K})$ represents the DD domain representation of the path's fractional Doppler shift. The phase shift caused by the Doppler effect, ψ_K, and the function Υ_N are defined as shown in Equations (31) and (32):

$$\psi_K[l] e^{j2\pi(k_{v_K}+\kappa_{v_K})\frac{N_{CP}-l_{\tau_K}+l}{(M+N_{CP})N}} \tag{31}$$

$$\Upsilon_N(x) \sum_{n=1}^{N} e^{j2\pi(n-1)\frac{x}{N}} \tag{32}$$

In this context, N_{CP} represents the size of the cyclic prefix.

D. BER Analysis

Fig. 4 illustrates the BER performances of different channel estimation methods under QPSK modulation with an IBO set at 2 dB. As depicted, the performance of all estimators is similar at lower SNR. However, as the SNR increases beyond 25 dB, the error levels of the baseline estimators become more pronounced. In contrast, the method proposed in this paper exhibits superior performance, with its detection capability nearing ideal detection performance when considering perfect Channel State Information. Notably, at a high SNR of 30 dB, only the method proposed in this paper achieves a BER below the specified threshold. Compared to the baseline methods, the proposed method offers close to 10 dB gain at this specific BER threshold.

Fig. 5 displays the BER performance using 16-QAM modulation with an IBO of 4 dB. Similar to observations made under QPSK modulation, all estimation methods exhibit comparable performance at lower SNR. However, at higher SNR levels, the method proposed in this paper demonstrates significant performance gains. Furthermore, the two

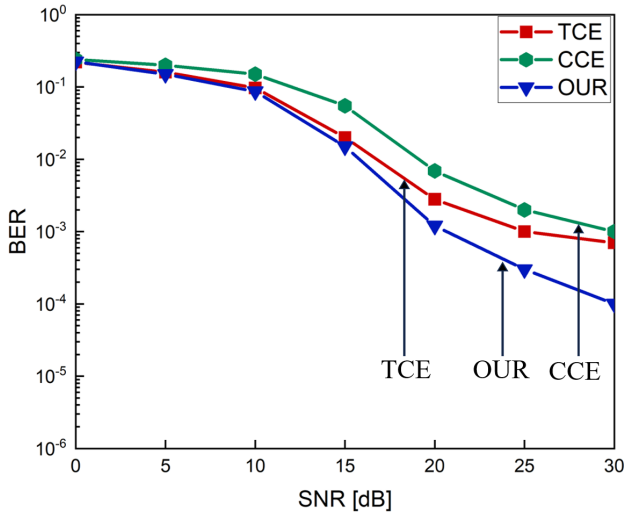

Fig. 4. BER Analysis under QPSK Modulation with IBO=2dB.

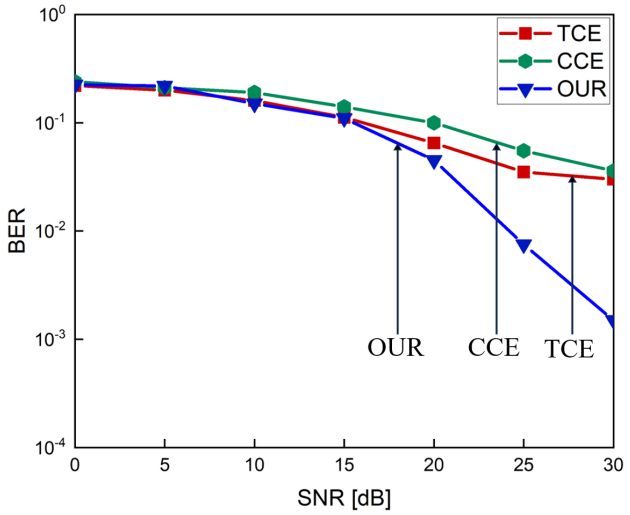

Fig. 5. BER Analysis under 16-QAM Modulation with IBO=4dB.

estimators from the baseline methods exhibit a pronounced error floor at high SNR values in scenarios involving higher modulation orders. This is primarily due to the fact that even as SNR continues to increase, the system's nonlinear distortion and other hardware limitations prevent further reduction in the BER. Consequently, these estimators are unable to reduce the BER below a specific threshold during the signal detection process.

In addition, we compared the BER accuracy of different deep learning estimators in data-driven channel estimation. The specific results are shown in Table II

TABLE II
ESTIMATION RESULTS BER ANALYSIS ON DIFFERENT MODEL

	SNR	LSTM	GRU	Transformer
QPSK & IBO=2dB	15	0.0159	0.0167	0.0150
	20	0.0019	0.0018	0.0012
	25	0.0004	0.0003	0.0003
	30	0.0001	0.0001	0.0001
16-QAM & IBO=4dB	15	0.1201	0.1302	0.1101
	20	0.0470	0.0480	0.0450
	25	0.0085	0.0080	0.0075
	30	0.0021	0.0018	0.0015

E. PAPR Analysis

We compared the impact of two benchmark channel estimation methods on the Complementary Cumulative Distribution Function (CCDF), a commonly used tool to assess signal PAPR distortion and provide intuitive information about the probability of the signal exceeding a given power level, λ. As shown in Fig. 6, we analyzed the CCDF results of different estimation methods. The figure reveals that, to facilitate channel estimation in the DD domain, the estimators allocate very high pilot power. It is evident that our method achieves at least a 10 dB PAPR threshold gain compared to the suboptimal benchmark method. Research indicates that higher symbol detection accuracy depends on higher pilot power, which leads to increased PAPR levels. However, the requirement for high pilot power can lead to inefficient resource utilization, potentially adversely affecting power efficiency. Compared to pilot configurations in the DD domain, pilot placements in the TF domain offer a more efficient use of resources.

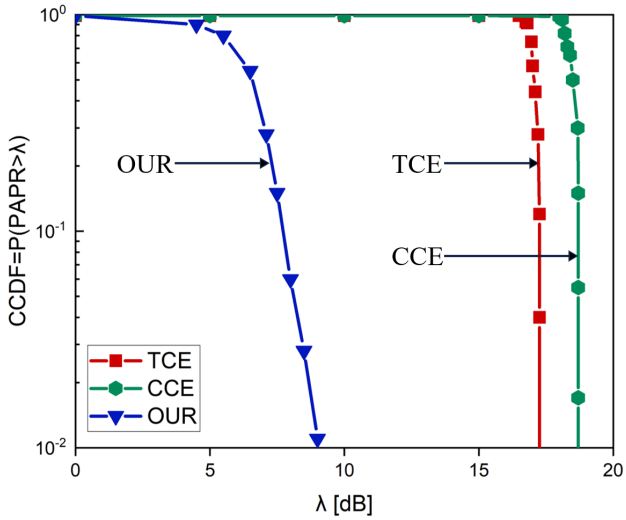

Fig. 6. PAPR Analysis.

The PAPR threshold is influenced by several factors, including the specific attributes of the communication system, the characteristics of the HPA, and the performance requirements of the application. The PAPR limitation caused by high power pilots in the DD domain is a significant challenge in existing research. Therefore, maintaining a lower PAPR is extremely important in mitigating distortion caused by HPAs, as the efficiency of HPAs decreases with the increase of the PAPR of the input signal. It is noteworthy that higher PAPR values can significantly impact the quality of communication and are generally challenging to achieve in practical applications. In contrast, our method offers a more practical approach to OTFS communication.

V. CONCLUSION

This paper addresses the issue of nonlinear distortion in OTFS systems caused by HPA and proposes a novel channel estimation algorithm that combines Transformer and adaptive filters. Simulation experiments demonstrate that the proposed algorithm significantly outperforms traditional pilot symbol estimation and compressed sensing methods

in highly dynamic channel environments. By conducting channel estimation in the TF domain, the proposed method successfully reduces the impact of HPA-induced nonlinear distortion and significantly lowers the BER at signal-to-noise ratios SNR up to 35 dB, showcasing its efficiency and reliability in complex channel conditions. This study provides a new approach for effectively obtaining and utilizing CSI in OTFS systems, offering important theoretical foundations and technical support for the development of future high-speed mobile communication systems.

REFERENCES

[1] S. Zhang, C. Xiang, and S. Xu, "6g: Connecting everything by 1000 times price reduction," *IEEE Open Journal of Vehicular Technology*, vol. 1, pp107–115, 2020.

[2] B. Mao, J. Liu, Y. Wu, and N. Kato, "Security and privacy on 6g network edge: A survey," *IEEE communications surveys & tutorials*, vol. 25, no. 2, pp1095–1127, 2023.

[3] G. Liu, Y. Huang, N. Li, J. Dong, J. Jin, Q. Wang, and N. Li, "Vision, requirements and network architecture of 6g mobile network beyond 2030," *China Communications*, vol. 17, no. 9, pp92–104, 2020.

[4] R. Sepúlveda, O. Montiel-Ross, J. C. Manzanarez, and E. E. Quiroz, "Fuzzy logic predictive algorithm for wireless-lan fast inter-cell handoff." *Engineering Letters*, vol. 20, no. 1, pp109–115, 2012.

[5] T. Xiaogang, Z. Yiqing *et al.*, "Survey of channel estimation method in delay-doppler domain for otfs," *Journal on Communications*, vol. 43, no. 12, pp188–201, 2022.

[6] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp1–6.

[7] Q. Li, Z. Xiang, P. Ren, and W. Li, "Variational autoencoder based receiver for orthogonal time frequency space modulation," *Digital Signal Processing*, vol. 117, p.103170, 2021.

[8] R. Hadani and A. Monk, "Otfs: A new generation of modulation addressing the challenges of 5g," *arXiv preprint arXiv:1802.02623*, 2018.

[9] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE transactions on wireless communications*, vol. 17, no. 10, pp6501–6515, 2018.

[10] I. A. Khan and S. K. Mohammed, "A low-complexity otfs channel estimation method for fractional delay-doppler scenarios," *IEEE Wireless Communications Letters*, vol. 12, no. 9, pp1484–1488, 2023.

[11] Z. Li, W. Yuan, B. Li, J. Wu, C. You, and F. Meng, "Reconfigurable-intelligent-surface-aided otfs: Transmission scheme and channel estimation," *IEEE Internet of Things Journal*, vol. 10, no. 22, pp19 518–19 532, 2023.

[12] S. Wang, J. Guo, X. Wang, W. Yuan, and Z. Fei, "Pilot design and optimization for otfs modulation," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp1742–1746, 2021.

[13] Y. Liang, Q. Wang, and P. Fan, "Pilot-aided channel estimation scheme based on frank array for otfs under rapidly time-varying channels," in *2022 IEEE 95th vehicular technology conference:(VTC2022-Spring)*. IEEE, 2022, pp1–6.

[14] H. Luo, X. Yu, Z. Zhang, and C. Gan, "Channel estimation for 5 g mmwave communications systems: A survey," *Telecommunication Engineering*, vol. 61, no. 2, pp254–262, 2021.

[15] T. Jiang, M. Song, X. Zhao, and X. Liu, "Channel estimation for millimeter wave massive mimo systems using separable compressive sensing," *IEEE access*, vol. 9, pp49 738–49 749, 2021.

[16] A. Munshi and S. Unnikrishnan, "Performance analysis of compressive sensing based ls and mmse channel estimation algorithm," *Journal of Communications Software and Systems*, vol. 17, no. 1, pp13–19, 2021.

[17] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for otfs in delay–doppler channels," *IEEE transactions on vehicular technology*, vol. 68, no. 5, pp4906–4917, 2019.

[18] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Performance analysis and window design for channel estimation of otfs modulation," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp1–7.

[19] Q. Bai, J. Wang, Y. Zhang, and J. Song, "Deep learning-based channel estimation algorithm over time selective fading channels," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp125–134, 2019.

[20] Q. Cheng, Z. Shi, J. Yuan, and M. Zhou, "Environment-robust signal detection for otfs systems using deep learning," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp5219–5224.

[21] X. Xu, S. Zhang, F. Gao, and J. Wang, "Sparse bayesian learning based channel extrapolation for ris assisted mimo-ofdm," *IEEE Transactions on Communications*, vol. 70, no. 8, pp5498–5513, 2022.

[22] D. Shi, W. Wang, L. You, X. Song, Y. Hong, X. Gao, and G. Fettweis, "Deterministic pilot design and channel estimation for downlink massive mimo-otfs systems in presence of the fractional doppler," *IEEE Transactions on Wireless Communications*, vol. 20, no. 11, pp7151–7165, 2021.

[23] X. Xu, M.-M. Zhao, M. Lei, and M.-J. Zhao, "A damped gamp detection algorithm for otfs system based on deep learning," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*. IEEE, 2020, pp1–5.

[24] X. Yi and C. Zhong, "Deep learning for joint channel estimation and signal detection in ofdm systems," *IEEE Communications Letters*, vol. 24, no. 12, pp2780–2784, 2020.

[25] Q. Peng, J. Li, and H. Shi, "Deep learning based channel estimation for ofdm systems with doubly selective channel," *IEEE Communications Letters*, vol. 26, no. 9, pp2067–2071, 2022.

[26] A. K. Gizzini, M. Chafii, A. Nimr, and G. Fettweis, "Deep learning based channel estimation schemes for ieee 802.11 p standard," *IEEE Access*, vol. 8, pp113 751–113 765, 2020.

[27] J. Fan, P. Liang, Z. Jiao, and X. Han, "A compressive sensing and deep learning-based time-varying channel estimation for fdd massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp8729–8738, 2022.

[28] T. Sun, J. Lv, and T. Zhou, "A transformer-based channel estimation method for otfs systems," *Entropy*, vol. 25, no. 10, p.1423, 2023.

[29] N. Shigei, I. Fukuyama, H. Miyajima, and Y. A. S. Yudo, "Battery-aware algorithms for mobile relay and route construction on wireless sensor network," *IAENG International Journal of Computer Science*, vol. 39, no. 3, pp321–328, 2012.

[30] C. Naveen and V. Sudha, "Peak-to-average power ratio reduction in otfs modulation using companding technique," in *2020 5th international conference on devices, circuits and systems (ICDCS)*. IEEE, 2020, pp140–143.

[31] S. Chennamsetty, S. Boddu, P. Chandhar, and K. C. Bulusu, "Analysis of papr in otfs modulation with classical selected mapping technique," in *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2023, pp319–322.

[32] T. H. Ali and A. Hamza, "A novel combined slm-pts technique based on genetic algorithms for papr reduction in ofdm systems," in *2020 second international conference on embedded & distributed systems (EDiS)*. IEEE, 2020, pp71–75.