Scientific Reports

https://doi.org/10.1038/s41598-026-44747-3

Article in Press

OTFS channel estimation method based on IBO-dynamic gated Bi-GRU

Received: 10 September 2025

Accepted: 13 March 2026

Published online: 26 March 2026

Cite this article as: Hou J., Wei Z., Ji Y. *et al.* OTFS channel estimation method based on IBO-dynamic gated Bi-GRU. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-44747-3

Jie Hou, Zhaochuan Wei, Yuanfa Ji & Xiaofang Deng

OTFS Channel Estimation Method Based on IBO-Dynamic Gated Bi-GRU

Jie Hou1, Zhaochuan Wei1, *,□, Yuanfa Ji1,2,3, *,□, and Xiaofang Deng1

[1]Information and Communication School, Guilin University of Electronic Technology, Guilin, 541004, China
[2]Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin, 541004, China
[3]International Joint Research Laboratory of Spatio-temporal Information and Intelligent Location Services, Guilin University of Electronic Technology, Guilin 541004, China
*E-mail: wei510102@sina.com
*E-mail: jiyuanfa@163.com
□these authors contributed equally to this work

Abstract

This paper introduces a novel channel estimation method for Orthogonal Time Frequency Space (OTFS) systems affected by nonlinear distortion from High-Power Amplifiers (HPA). The method integrates a Bidirectional Gated Recurrent Unit (Bi-GRU) with a dynamic gating mechanism driven by the Input Back-Off (IBO) parameter of the HPA, combined with a multi-head attention network. The dynamic gating mechanism adaptively adjusts the update gate of the Gated Recurrent Unit (GRU) based on real-time IBO values, optimizing the trade-off between historical memory and current input during training. The multi-head attention module further captures long-range dependencies in the channel response. Theoretical analysis indicates that the proposed IBO-driven dynamically gated Bi-GRU achieves a computational complexity reduction of 20–46.7% compared to a Bi-GRU architecture. Simulation results demonstrate the superior performance of the proposed method across both bit error rate (BER) and normalized mean square error (NMSE) metrics under high mobility and nonlinear distortion. It achieves up to 22.6 quantified in decibels (dB) lower NMSE and, at a signal-to-noise ratio (SNR) of 30 dB, a 15.2 dB reduction in logarithmic BER compared to conventional methods, along with a 3–4 dB improvement over deep learning baselines at the same SNR. It also provides over 7 dB peak-to-average power ratio (PAPR) reduction over traditional methods, confirming strong robustness and accuracy in challenging communication scenarios.

Keywords

Channel estimation, Dynamic Gating, Orthogonal Time Frequency Space, Bidirectional Gated Recurrent Unit

Introduction

With the rapid development of modern transportation, the mobility of existing transportation tools has increased significantly, leading to a growing demand for wireless communication in high-speed mobile environments[1]. Among these, vehicular communication, as one of the typical scenarios of high-speed mobile wireless communication, has increasingly demanding requirements for high data rates, high reliability, and high connectivity. This development trend poses significant challenges to existing wireless communication technologies.

Traditional Orthogonal Frequency Division Multiplexing (OFDM) technology, renowned for its high spectral efficiency, has been widely adopted in 4G/5G wireless communication and has become the dominant modulation scheme in current wireless communications[2]. However, in high-speed mobile environments, this technology is susceptible to Inter-Carrier Interference (ICI) and is highly sensitive to time-varying channels and frequency selectivity. These factors collectively cause significant degradation in communication performance, making it difficult to meet the high reliability requirements of future vehicular networks[3]. To address this challenge, the Orthogonal Time Frequency Space (OTFS) modulation proposed by Hadani et al. provides a new approach to wireless communication[4]. Unlike traditional modulation schemes, OTFS modulation operates in the Delay-Doppler (DD) domain, transforming the time-varying channel into a sparse, time-invariant channel. This effectively separates the delay and Doppler paths, obtaining full diversity in the channel and enabling reliable information transmission in high-speed mobile environments[5].

In this context, channel estimation in OTFS systems becomes particularly important. Accurate channel estimation is not only a key factor in realizing the advantages of OTFS technology, but its performance also directly impacts the overall effectiveness of the communication system. In [6], author proposed a classical channel estimation method that uses embedded pilot estimation with guard intervals, utilizing thresholds for channel estimation at the receiver. This method effectively separates pilot data from user data but sacrifices spectral efficiency, as the dedicated guard intervals consume time-frequency resources that could otherwise carry data symbols. Furthermore, the pilot pulse power influences the estimation performance because the accuracy of the least-squares estimate is directly proportional to the signal-to-noise ratio (SNR) of the received pilots. To achieve better estimation results, relatively high power is required. This, however, creates a fundamental trade-off: increasing pilot power improves estimation accuracy but simultaneously elevates the peak-to-average power ratio (PAPR) of the transmitted frame. The high PAPR can drive the high-power amplifier (HPA) into its nonlinear region, causing in-band distortion and spectral regrowth that ultimately degrade the overall system performance. In article[7], author introduced a channel estimation algorithm based on cross-correlation, assuming that the channel is stationary for a period of time and uses a single frame to transmit pilot data. While this method also effectively separates pilot data and user data, it significantly reduces spectral efficiency. In[8], introduced a low-overhead OTFS channel estimation method that reduces overhead

and complexity through one-dimensional frequency or time-domain estimation. However, this method faces performance limitations in fractional Doppler scenarios in the frequency domain, and the time-domain training sequence introduces a high PAPR. To address the PAPR problem, In article[9], author presented a joint channel estimation method for the zero-padded OTFS (ZP-OTFS) system with low PAPR. By inserting Zadoff-Chu sequences into the zero-frequency band, the overhead and PAPR are reduced, while a two-step sparse recovery algorithm improves estimation accuracy. In[10], author proposed a hybrid SLM and PTS algorithm to reduce the PAPR of OTFS systems effectively. Furthermore, In article[11], author introduces a method that combines Deep Neural Networks (DNN) and Least Squares (LS) algorithms for OTFS channel estimation, significantly improving estimation accuracy in high-mobility environments while reducing complexity. However, although the Long Short-Term Memory (LSTM)-based estimator proposed by dos Reis et al[12]. incorporates HPA nonlinear distortion into its modeling, it—like other learning-based approaches—does not explicitly account for the dynamic influence of the HPA's operating point, such as the Input Back-Off, on the estimation process. The impact of this crucial operating parameter on estimation performance remains largely unexamined. From a physical perspective, the IBO directly determines the severity of amplitude and phase distortion, yet this dependence is not leveraged to guide the estimation model. Motivated by this gap, our work investigates the integration of the IBO parameter as a dynamic gating signal into a deep learning framework, thereby enabling adaptive compensation for HPA nonlinear distortion that varies with transmit power.

It is worth noting that while existing studies have addressed the PAPR problem to some extent, they have not systematically examined the impact of HPA nonlinear distortion on channel estimation performance. To address this issue, this paper proposes a channel estimation method for compensating for HPA nonlinear distortion. This method is based on Bidirectional Gated Recurrent Unit (Bi-GRU) and attention mechanisms. By introducing Input Backoff (IBO) as a gating signal for the Bi-GRU, the model adaptively adjusts its focus on channel features under different power conditions. The proposed method not only exhibits low peak-to-average power ratio (PAPR) characteristics but also significantly improves system bit error rate (BER) and mean square error (NMSE) performance. It effectively enhances channel estimation accuracy and system robustness in environments with HPA nonlinear distortion.

The main contributions of this work are summarized as follows:

1) We propose a novel IBO-driven dynamic gating Bi-GRU architecture. By embedding the HPA's Input Back-Off parameter as a gating signal into the GRU update gate, the model adaptively characterizes the severity of nonlinear distortion. Compared with the conventional Bi-GRU, this architecture reduces computational complexity by approximately 33% through a dynamic sparsification strategy while preserving high estimation accuracy.

2) We develop a lightweight estimation framework that integrates the dynamic gating Bi-GRU with a multi-head attention mechanism. This framework jointly

models long-range channel dependencies and nonlinear distortion characteristics in the time-frequency domain. Under high-mobility and low-IBO conditions, the proposed framework achieves an NMSE improvement of up to 22.6 decibels (dB) and a BER gain of 15.2 dB compared to conventional threshold-based estimation methods.

3) We demonstrate the comprehensive advantages of the proposed method through systematic simulations. Compared to both the Threshold-based Channel Estimation (TCE) method[6] and the Cross-Correlation-based Channel Estimation (CCE) method[7], our method achieves a PAPR reduction exceeding 7 dB. Furthermore, it maintains stable estimation performance under extreme mobility (500 km/h) and severe nonlinear distortion (IBO=2dB), showcasing its strong robustness. This performance is achieved while also providing a superior trade-off between computational efficiency and estimation accuracy compared to other learning-based methods.

The remainder of the paper is organized as follows: Section 2 introduces the OTFS system model and the HPA model. Section 3 presents the proposed OTFS channel estimation algorithm based on dynamic IBO gating for Bi-GRU. Section 4 analyzes the simulation results, and the final section concludes the paper.

System Model

The OTFS channel estimation system model used in this study is depicted in Fig. 1. The most prevalent implementations of OTFS systems in the current literature are primarily based on two mathematical frameworks. One framework combines the Symplectic Finite Fourier Transform (SFFT)[13] with multi-carrier modulation, while the other employs a scheme based on the discrete Zak transform[3]. In this paper, we adopt a system architecture that utilizes the inverse SFFT (ISFFT)[13] along with the SFFT operation. This architecture enables the transformation of time-varying channels in the time-frequency domain into quasi-static channels in the DD domain.

First, a randomly generated bitstream is modulated using Quadrature Amplitude Modulation (QAM), producing the modulated signal $X_{DD}[l,k]$ in the DD domain, where the delay index is denoted as $l \in \{0,\ldots,M-1\}$ and the Doppler index as $k \in \{0,\ldots,N-1\}$. The corresponding time-frequency (TF) domain signal $X_{TF}[m,n]$ can subsequently be obtained via the inverse Symplectic Finite Fourier Transform (ISFFT). Next, the TF-domain signal is further processed using the Heisenberg transform, which maps it into the time domain signal $s(t)$. After undergoing nonlinear transformations by a high-power amplifier, the resulting signal $s_{NDL}(t)$ is transmitted through an antenna into the wireless channel.

Upon arrival at the receiver, the received time-domain signal $r(t)$ undergoes an inverse transformation. The signal is first processed using the Wigner transform to retrieve the TF-domain representation, followed by the application of the SFFT to

recover the DD-domain signal $Y_{DD}[k,l]$. Finally, the inverse QAM operation is performed to recover the original binary bitstream.

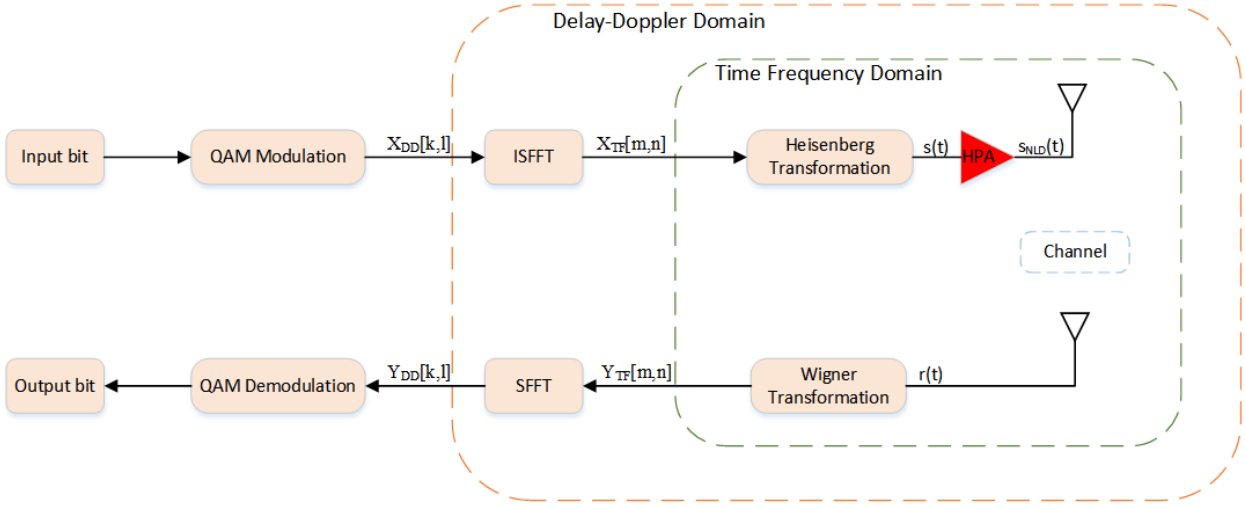

Fig.1 OTFS Systems

A. OTFS Modulation

As the core of OTFS modulation technology, the SFFT operation successfully facilitates the effective mapping of data between the DD domain and the TF domain. After the input bitstream undergoes QAM modulation, a two-dimensional DD-domain signal $X_{DD}[l,k]$ is obtained. This signal is then transformed into the TF-domain signal $X_{TF}[m,n]$ using the ISFFT, which can be expressed as:

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

where $m\in\{0,\ldots,M-1\}$ denotes the subcarrier index and $n\in\{0,\ldots,N-1\}$ represents the symbol index.

The TF-domain signal undergoes a Heisenberg transform to obtain the time-domain transmitted signal :

$$s(t)=\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}X_{\mathrm{TF}}[m,n]g_{\mathrm{tx}}(t-nT)e^{j2\pi m\Delta f(t-nT)} \tag{2}$$

where g_{tx} is the pulse function at the transmitter.

B. HPA Distortion Model

It is noteworthy that the adopted memoryless polynomial model effectively captures the static nonlinearity of the HPA, which is the dominant impairment in the considered context. While practical HPAs may exhibit memory effects, their impact is less pronounced compared to static nonlinearity for the subcarrier spacing and bandwidth under study. The proposed estimation framework, however, is not limited

to this specific model and can be extended to incorporate memory effects in future work by employing more complex models like the Volterra series.

As shown in the nonlinear distortion model of the HPA in Fig.1, the model follows the memoryless HPA description proposed in[14], which can describe both the Amplitude-to-Amplitude (AM/AM) and Amplitude-to-Phase (AM/PM) distortions.

In practice, to avoid or reduce the impact of nonlinear distortion, the HPA works at a given IBO starting from the 1dB compression point[15]. The 1dB compression point refers to the input power at which the amplifier's transfer characteristic deviates by 1 dB from the ideal linear characteristic. The IBO can be defined as:

$$IBO = 10\log_{10}\left(\frac{P_{1dB}}{P_s}\right) \tag{3}$$

Where P_{1dB} is the input power at the 1dB compression point, and P_s is the average power of the input signal.

Before entering the HPA, the signal $s(t)$ is scaled in amplitude by a gain a to achieve the desired IBO operating point. Given the average power of signal $s(t)$ as $P_s = \mathrm{E}[|x(t)|^2]$, the IBO in dB is defined as $IBO_{dB} = 10\log_{10}(P_{1dB}/P_s)$. To set the average input power to the HPA to $P_{1dB} \times 10^{-IBO_{dB}/10}$, the required amplitude gain is:

$$a = 10^{-\frac{IBO_{dB}}{20}} \tag{4}$$

The complex envelope of signal $s(t)$ can be written as:

$$s(t) = r(t)e^{jy(t)} \tag{5}$$

Where $r(t)$ is the amplitude of the input signal and $y(t)$ is the phase of the input signal.

The signal $s_{NDL}(t)$ can be written as:

$$s_{NDL}(t) = f_{\mathrm{AM-AM}}(r(t))e^{jf_{\mathrm{AM-PM}}(r(t)+y(t))} \tag{6}$$

Where $f_{\mathrm{AM-AM}}$ is the AM/AM characteristic and $f_{\mathrm{AM-PM}}$ is the AM/PM characteristic of the HPA.

This can be rewritten as:

$$s_{NDL}(t) = \mathrm{F}(r(t))e^{jy(t)} \tag{7}$$

Where $\mathrm{F}(r(t)) = f_{\mathrm{AM-AM}}e^{jf_{\mathrm{AM-PM}}(r(t))}$ is the complex envelope of the amplified signal.

According to the AM/AM and AM/PM characteristics of the Long Term Evolution (LTE) user equipment HPA presented in Reference[16] and 3GPP specifications, when

a memoryless HPA is modeled as a polynomial model of order $K=9$, the above expression can be simplified as:

$$s_{NDL}(t)=\sum_{l=1}^{K} a_l r(t)\,|r(t)|^{l-1} \tag{8}$$

Where a_l represents the complex coefficients of the polynomial approximation, computed using the classical LS method.

By combining Eq. (6) and (8), the complex envelope of the amplified signal can be written as:

$$s_{NDL}(t)=f_{\text{AM-AM}}\left[r(t)\right]e^{jf_{\text{AM-PM}}(r(t)+y(t))}=\sum_{l=1}^{K} a_l r(t)^{n} \tag{9}$$

C. Wireless Channel

After a signal $s_{NDL}(t)$ is transmitted over a time-frequency doubly-selective channel, the received signal can be expressed as:

$$r(t)=\int\int h(t,u)\,s_{NDL}(t-\tau)e^{2\pi u(t-\tau)}\,du\,d\tau \tag{10}$$

Where τ and u represent the delay and Doppler shift, respectively. In the Delay-Doppler domain, the channel is typically characterized by a limited number of reflectors, each associated with specific delay and Doppler shifts. Consequently, the channel impulse response can be represented by the following sparse model:

$$h_{DD}(t,u)=\sum_{i=1}^{P} h_i\,\delta(t-\tau_i)\,\delta(u-u_i) \tag{11}$$

Here, P denotes the number of propagation paths, h_i, τ_i and u_i represent the path gain, delay, and Doppler shift associated with the i-th path, respectively, and $\delta(\cdot)$ is the Dirac delta function. The delay and Doppler taps for the i-th path are given by:

$$\tau_i=\frac{l_{\tau_i}}{M\Delta f},\ u_i=\frac{k_{u_i}+\kappa_{u_i}}{NT} \tag{12}$$

Where $l_{\tau_i},k_{u_i}\in\mathbb{Z},-\frac{1}{2}<\kappa_{u_i}<\frac{1}{2}$. The indices l_{τ_i} and k_{u_i} correspond to the delay τ_i tap and Doppler frequency u_i tap, respectively. The term $\kappa_{u_i}\in[-0.5,0.5)$ denotes the fractional Doppler, representing the fractional offset from the nearest Doppler tap.

D. OTFS Demodulation

The time-domain received signal $r(t)$ is obtained from the time-domain transmitted signal in Eq.(2) after passing through the doubly-selective channel, as given by:

$$r(t)=\int\int h_{DD}(\tau,u)e^{j2\pi u(t-\tau)}s_{NLD}(t-\tau)\,d\tau\,du+w(t) \tag{13}$$

Where $w(t) \sim \mathrm{N}(0, \sigma_w^2)$ represents additive white Gaussian noise (AWGN) in the time domain.

Subsequently, the received signal undergoes a Wigner transform to yield the received signal in the TF domain:

$$Y(t,f) = \int g_{rx}^{*}(t' - t) r(t') e^{-j2\pi f(t'-t)} dt' \tag{14}$$

Here, g_{rx}^{*} denotes the conjugate of the receiver pulse shaping function. By sampling at intervals of $t = nT$ and $f = m\Delta f$ for time and frequency, respectively, we obtain the discretized received signal:

$$Y_{\mathrm{TF}}[n,m] = Y(t,f)|_{t=nT,\, f=m\Delta f} \tag{15}$$

The TF-domain signal is then transformed into the DD domain via the SFFT:

$$Y_{DD}[k,l] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y[n,m] e^{-j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} \tag{16}$$

Methods

This section presents a deep learning-based channel estimation method designed to address the nonlinear distortion introduced by HPA. The core innovation of this method lies in integrating the dynamic memory characteristics of a Bi-GRU[17] with an attention mechanism to achieve high-precision channel estimation in the TF domain.

The choice of a GRU-based architecture over a Transformer is primarily motivated by a trade-off between model efficiency and performance in complex channel environments. While Transformers excel in many sequence tasks, their quadratic-complexity attention mechanism and large parameter count introduce significant deployment overhead in resource-constrained scenarios requiring real-time channel prediction and continual learning[18]. In contrast, the gating mechanisms of GRUs effectively model the strong temporal correlations inherent in wireless channels. Crucially, GRUs maintain comparable prediction accuracy (typically within 0.5 dB NMSE difference compared to Transformers[18]) while offering substantially lower computational complexity and memory footprint. This makes them more suitable for HPA nonlinearity compensation tasks that demand rapid incremental updates and high energy efficiency.

First, an initial channel estimate is obtained and used as the input to the Bi-GRU network. Second, a dynamic gating mechanism[19] is designed, which incorporates the IBO parameter of the HPA as a gating signal into the GRU cells. This mechanism dynamically adjusts the update gate based on the IBO value, thereby optimizing the training process of the Bi-GRU. Subsequently, a multi-head attention mechanism[20] is employed to extract deeper features from the channel response. This module captures long-range dependencies among channel features from different representation subspaces by operating multiple attention heads in parallel. Finally,

an IBO weighted output layer performs a weighted correction on the prediction results, ensuring the output aligns more accurately with real-world requirements.

A. Frame Structure Design

The system's frame structure, illustrated in Fig.2, employs an efficient pilot insertion scheme. A known pilot sequence is inserted in the entire first column of the TF grid. Additionally, auxiliary pilot sequences are distributed across specific columns of the data matrix, collectively forming the complete pilot structure. The remaining positions are filled with data symbols. Crucially, the power level of all pilot sequences is maintained consistent with that of the data symbols; this design effectively reduces the system's PAPR. Inserting pilots in the TF domain efficiently minimizes the number of subcarriers allocated for pilot transmission in the DD domain, while preserving the same subcarrier spacing and bandwidth.

Let the set of active subcarriers be denoted by M_{on}, where $|M_{on}|$ represents the total number of active subcarriers. The set of pilot subcarriers is denoted by M_p, with $|M_p|$ being the number of pilot subcarriers. The set of data subcarriers is denoted by M_d, with $|M_d|$ being the number of data subcarriers, ensuring $M_p \cup M_d = M_{on}$.

Considering the data symbols in the TF domain, we have:

$$X_{TF}[m_d, n] = \frac{1}{\sqrt{NM_d}} \sum_{k=0}^{N-1} \sum_{l=0}^{M_d-1} X_{DD}[l,k] e^{j2\pi\left(\frac{nk}{N} - \frac{m_d l}{M_d}\right)} \tag{17}$$

After inserting the pilot subcarriers, the complete TF grid is obtained:

$$X_{TF}[m,n] = \begin{cases} pilots & if \quad m \in M_p \\ X_{TF}[m_d, n] & if \quad m \in M_d \end{cases} \tag{18}$$

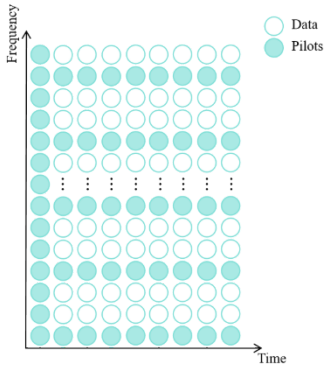

Fig.2 Frame structure

Crucially, the preamble sequence in our design is custom-optimized for low PAPR, distinguishing it from standard preambles (e.g., IEEE 802.11p) and contributing directly to the overall system power efficiency.

B. IBO-Dynamic Gated Bi-GRU Estimation

The received TF-domain signal undergoes initial channel estimation using the least squares (LS) algorithm to obtain a preliminary channel response. This process is expressed as:

$$H_{LS\text{-}TF}[m_{pilot}, n_{pilot}] = \frac{Y_{TF}[m_{pilot}, n_{pilot}]}{X_{TF}[m_{pilot}, n_{pilot}]} \tag{19}$$

Where $Y_{TF}[m_{pilot}, n_{pilot}]$ is the received data at pilot positions, and $X_{TF}[m_{pilot}, n_{pilot}]$ is the transmitted data at pilot positions.

It is feasible for the receiver to acquire the Input Back-Off parameter of the transmitter's High-Power Amplifier in practical systems, for example, through low-overhead signaling embedded in control information or via blind estimation techniques applied to the received signal. This work focuses on demonstrating how known IBO information can systematically enhance channel estimation performance under nonlinear distortion, rather than on the design of the IBO estimator itself. To establish a clear performance baseline, the following analyses assume that the receiver has accurate knowledge of the IBO value.

Following the initial LS estimation, this work utilizes a Bi-GRU integrated with a dynamic gating mechanism to perform interpolation and fitting for the data portion of the channel matrix. The Gated Recurrent Unit (GRU), a transformative improvement over LSTM networks, achieves more efficient temporal modeling through a streamlined gating structure, making it particularly suitable for feature extraction in time-varying channels. The core innovation of the GRU is its dual-gating mechanism, the structure of which is shown in Fig.3.

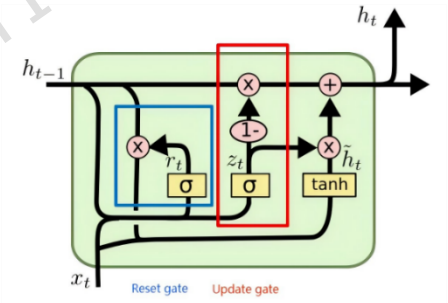

Fig.3 GRU structure

For the input $x_t \in \mathbb{R}^d$ at time step t and the previous hidden state $h_{t-1} \in \mathbb{R}^h$, the state transition of a GRU is formulated as follows:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \tag{20}$$

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r) \tag{21}$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \tag{22}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t \tag{23}$$

Here, $\sigma(\cdot)$ denotes the sigmoid function, $\odot$ represents the Hadamard product, and $\mathbf{W}_\cdot, \mathbf{U}_\cdot \in \mathbb{R}^{h \times d}$ are trainable parameter matrices.

The update gate z_t controls the retention ratio of historical channel state information (CSI), while the reset gate r_t performs selective memory filtering, demonstrating exceptional responsiveness to abrupt channel changes. The candidate state $\tilde{h}_t$ generates an innovation estimate based on the current input and the filtered historical information. The final state update h_t constitutes a convex combination optimization problem dynamically regulated by the update gate.

To further enhance model performance, a Bi-GRU architecture is adopted, as illustrated in Fig.4

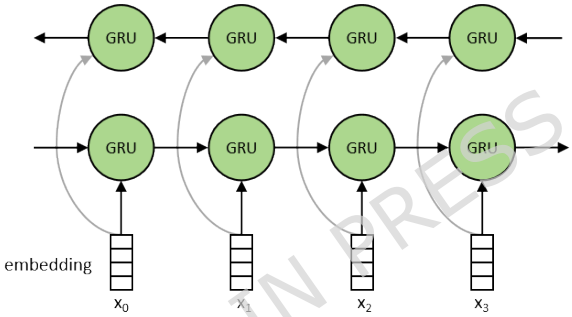

Fig.4 Bi-GRU structure

The state equations for the Bi-GRU can be expressed as:

$$\overrightarrow{\mathbf{h}}_t = \mathrm{GRU}_{\mathrm{forward}}(\mathbf{x}_t, \overrightarrow{\mathbf{h}}_{t-1}; \Theta_f) \tag{24}$$

$$\overleftarrow{\mathbf{h}}_t = \mathrm{GRU}_{\mathrm{backward}}(\mathbf{x}_t, \overleftarrow{\mathbf{h}}_{t+1}; \Theta_b) \tag{25}$$

$$\mathbf{H}_t = [\overrightarrow{\mathbf{h}}_t \| \overleftarrow{\mathbf{h}}_t] \times \mathbf{W}_o + \mathbf{b}_o \tag{26}$$

where $\Theta \in \{f, b\}$ represent the parameter sets of the forward and backward GRUs, respectively, and $\|$ denotes the vector concatenation operation.

To address HPA nonlinear distortion, this work innovatively introduces a dynamic gating mechanism. The choice of IBO as the gating signal is motivated by its direct correlation with the severity of HPA nonlinearity. A lower IBO value indicates the HPA is operating closer to saturation, introducing more severe nonlinear distortion. In this regime, the channel's characteristics become less reliable and more volatile. Therefore, the model should place less trust in the historical channel state and rely more on the current input to adapt to the rapid distortion changes. Conversely, a higher IBO implies a more linear operation, where the channel is more stable, and the model can confidently retain more historical information. By integrating the IBO

value into the update gate, the model physically interprets the operating state of the HPA and mathematically adjusts the memory retention ratio accordingly. By designing a dynamic gated GRU cell that adjusts the update ratio according to the input IBO value, the training process of the GRU cell is optimized, as illustrated in Fig.5.

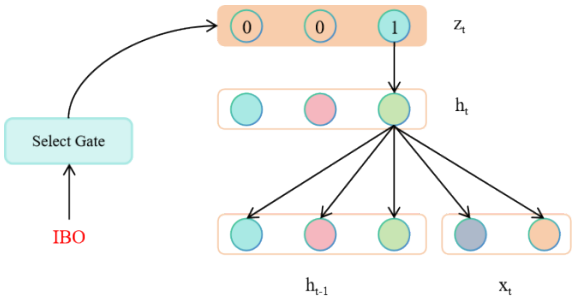

Fig.5 Dynamic gating mechanism

The traditional GRU update gate mechanism is reconfigured into a power-aware gating system:

$$\tilde{\mathbf{z}}_t = \Gamma\left(\mathbf{z}_t, \Phi(IBO)\right) \tag{27}$$

Where $\mathbf{z}_t$ is the conventional update gate from Eq.(20), $\tilde{\mathbf{z}}_t$ is the resulting dynamic gate, $\Gamma(\cdot)$ denotes the sparsification operation, and $\Phi(IBO)$ is a lightweight nonlinear mapping that outputs an update ratio $\rho \in [\rho_{\min}, \rho_{\max}]$.

The function $\Phi(\cdot)$ is implemented as follows:

$$\Phi(IBO) = \rho_{min} + (\rho_{max} - \rho_{min}) \times s(a \times IBO + b) \tag{28}$$

Here, $s(x)$ is the sigmoid function, $a = 0.1$ and $b = 5.0$ are fixed normalization parameters, and $\rho_{\min}$, $\rho_{\max}$ define the fixed output range of the ratio. The only trainable component in Φ is the nonlinear transformation inside the sigmoid.

The ratio $\rho = \Phi(IBO)$ directly governs the sparsification: it determines the number of active neurons

$$k = \lfloor \rho \cdot N_{GRU} \rfloor \tag{29}$$

Where N_{GRU} is the dimension of the GRU hidden state. The sparsification operator Γ then selects the k largest (in magnitude) components of $\mathbf{z}_t$:

$$\Gamma(\mathbf{z}_t; \rho) = \mathrm{Top}_k(\mathbf{z}_t) \tag{30}$$

Thus the dynamic gate $\tilde{\mathbf{z}}_t$ is a sparsified version of the original gate, with the sparsity level adapted instantaneously to the HPA's operating point via the IBO. Extensive experiments confirm that the model preserves its temporal modelling capability

while p varies within $[p_{min}, p_{max}]$. All parameters a, b, p_{min}, p_{max} are kept fixed in this work, leaving their joint optimisation as a future extension.

This design enables the update gate to adjust the trade-off between historical channel information and current input according to the severity of HPA nonlinear distortion, thereby delivering accurate compensation in realistic transmission scenarios.

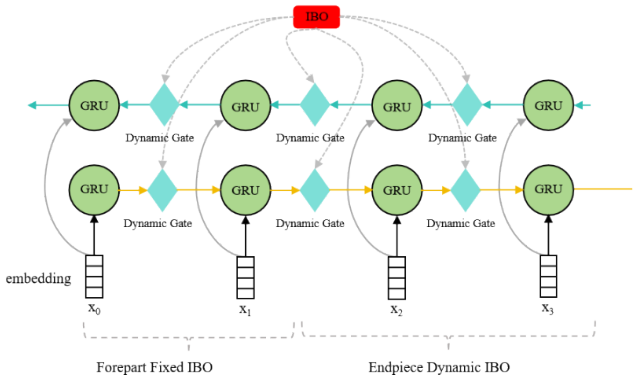

Fig.6 Dynamic gated Bi-GRU structure

The architecture of the proposed dynamic gated Bi-GRU network is shown in Fig.6. The network employs a layered processing structure: the input sequence is first mapped through an embedding layer, then processed by the dynamic gated Bi-GRU layer. During the initial training phase, the IBO value is fixed to stabilize the feature extraction process and establish a reliable base model. In the later training stages, dynamic IBO adjustment is introduced to enhance the model's capability for nonlinear distortion compensation.

To further strengthen the extraction of channel information, an attention mechanism is introduced to reinforce the extraction of critical channel features. The spatiotemporal features extracted by the dynamic gated Bi-GRU are projected via linear transformations into three subspaces: Query (Q), Key (K), and Value (V). The structure of the multi-head attention mechanism is shown in Fig.7.

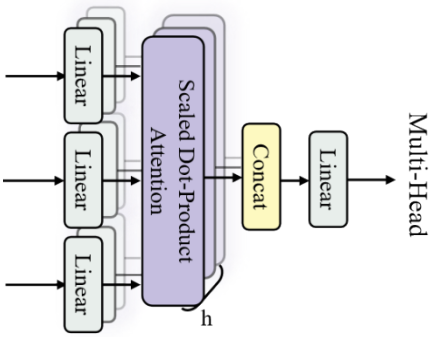

Fig.7 Multi-head attention mechanism

The overall proposed network structure for OTFS channel estimation is shown in Fig.8.

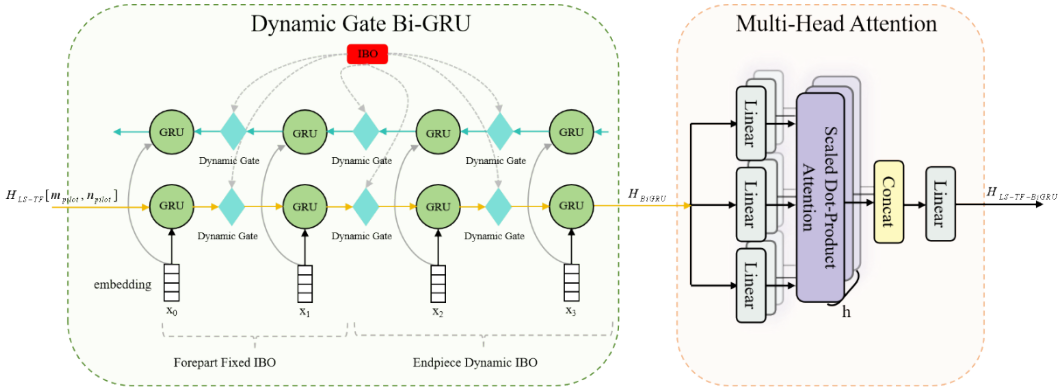

Fig.8 Channel estimation network structure

C. Complexity Analysis

The computational complexity of the proposed network is primarily influenced by the dynamic gated Bi-GRU and the multi-head attention network. The complexities of these two components are analyzed below.

For a traditional Bi-GRU, the computational complexity is:

$$O(2T(3d_{in}d_h+3d_h^2)) \tag{31}$$

Where T denotes the sequence length, d_{in} represents the input feature dimension, and d_h is the number of hidden units. The dynamic Bi-GRU incorporates an adaptive sparsification mechanism based on the IBO parameter, which selectively updates only a fraction p of neurons at each step, thereby compressing the candidate state computation from $3d_h^2$ to $(1+2p)d_h^2$. Specifically, when a neuron is not selected for update, both its reset gate and candidate state calculations are skipped, while its hidden state retains the previous value. Its final complexity is therefore:

$$O(2T(3d_{in}d_h+(1+2p)d_h^2)) \tag{32}$$

This leads to a theoretical computational reduction in the range of 20.0% to 46.7% compared to a Bi-GRU architecture, as derived from the formula:

$$h=1-\frac{(1+2p)}{3} \tag{33}$$

To validate the theoretical complexity reduction, comprehensive measurements were conducted on an Intel CPU platform (i7-9750H). The proposed dynamic gated Bi-GRU achieves approximately 26% reduction in both MACs and FLOPs compared to the standard Bi-GRU. This empirical result aligns well with the theoretical prediction of 20-30% computational saving derived from Eq. (33) with p=0.7. Furthermore, these computational savings translate to practical runtime

improvements, including reduced inference time and increased throughput, confirming the effectiveness of the adaptive sparsification mechanism in real-world deployment.

Method	MACs	FLOPs
Bi-GRU	279,900	559,800
Proposed	**207,000**	**414,000**

Table 1 Measured Computational Complexity Comparison

The computational complexity of the multi-head attention mechanism is primarily composed of attention score calculation and linear transformations. The complexity of linearly projecting the input sequence to generate the Q, K, and V matrices is $O(Td_m^2)$. The complexity of calculating the attention scores and performing the weighting which includes matrix multiplication of QK^T and multiplication with $V$is $O(T^2 d_m)$. The overall complexity of multi-head attention is therefore:

$$O(T^2 d_{in} + T d_{in}^2) \tag{34}$$

Consequently, the total complexity of the entire network is:

$$O(2T(3d_{in}d_h + (1+2\rho)d_h^2) + T^2 d_{in} + T d_{in}^2) \tag{35}$$

Combining Eqs. (32) and (35), the total complexity of the proposed network is:

$$O(2T(3d_{in}d_h + (1+2\rho)d_h^2) + T^2 d_{in} + T d_{in}^2) \tag{36}$$

To relate the abstract dimensions T, d_{in}, and d_h to the OTFS system parameters, we note that: T corresponds to the number of OTFS symbols N in a frame. d_{in} is proportional to the number of active subcarriers M_{on}, specifically $d_{in} = M_{on}$. d_h is a fixed hyperparameter independent of M_{on}, which contrasts with LSTM-based methods where the hidden state size scales with M_{on}.Therefore, Eq. (36) can be rewritten in terms of system parameters as:

$$O(NM_{on} + N^2 M_{on} + M_{on}^2) \tag{37}$$

Comparison with Existing Methods. Table 2 summarizes the computational complexity orders of the proposed method and several existing OTFS channel estimation techniques.

To place the computational efficiency of the proposed estimator in context, we compare its theoretical complexity order with established OTFS channel estimation methods, as summarized in Table 2. The conventional TCE[6] and CCE[7] methods are dominated by the cubic complexity $O(MN + M_d^3 N^3)$due to DD-domain equalization. The LSTM estimator shifts equalization to the TF domain, reducing that component to $O(M_d N)$, but its estimator core scales quadratically as $O(M_{on}^2)$ because its LSTM hidden state size χ grows with the number of subcarriers $(c = M_{on}/2)$. In contrast, the proposed architecture employs a fixed-size GRU hidden layer L, which changes

the scaling of the estimator core to linear in M_{on}. This represents a fundamental reduction in computational order when the number of subcarriers is large. Note that the multi-head attention contributes an $N^2 M_{on}$ term. While this may dominate for long sequences, the overall complexity remains favorable compared to conventional methods. Future work could explore optimized attention variants to mitigate this limitation.

Methods	Complexity
TCE[6]	$O\left(MN + M_d^3 N^3\right)$
CCE[7]	$O(kDMN\log N + M_d^3 N^3)$
LSTM12	$O(M_{on}^2 + M_p^2 + M_{on}M_p + M_d N)$
Proposed	$O(NM_{on} + N^2 M_{on} + M_{on}^2)$

Table 2 Comparison Complexity

D. Implementation Details and Hyperparameters

To ensure reproducible results, the complete architecture of the proposed network is shown in Fig.8, with its core hyperparameters summarized in Table 3 The model takes the time-frequency domain channel response obtained from an initial LS estimate as input, which is sequentially processed by an IBO-dynamic gated bidirectional GRU layer, a 4-head attention mechanism, and a fully connected output layer. Training is performed using the Adam optimizer with an initial learning rate of 0.0001, coupled with a learning rate reduction scheduler. Both input and label data are normalized using StandardScaler. All experiments are conducted based on this fixed configuration.

The hyperparameters listed in Table 3 remain unchanged across all experimental scenarios. This design leverages the adaptive capability of the proposed dynamic gating mechanism: the IBO parameter directly modulates the model's reliance on historical states through the gating function $f(\cdot)$ enabling the network to intrinsically adapt to varying degrees of nonlinear distortion, thereby reducing the need for environment-specific tuning. The performance evaluation in Section 4 validates the effectiveness of this design: under diverse conditions including speeds of 300 km/h and 500 km/h and IBO values of 2 dB and 4 dB, the model trained with the identical parameter set demonstrates stable and superior performance, confirming its strong generalization capability and engineering practicality.

Parameter	Values
GRU hidden units	30
Attention heads	4
MLP neurons	15
Number of neurons per hidden layer	44
Number of epochs	500
Batch size	32
Learning rate	0.0001
Number of training samples	8000

Number of testing samples	2000

Table 3 Parameter Configuration

Results

To validate the performance of the proposed method, the BER, NMSE, System Throughput Analysis and PAPR were selected as performance evaluation metrics. The validation platform used was Matlab 2022a. The simulation parameters are configured as shown in Table 4.

We compare our proposal with the benchmark methods TCE[6], CCE[7], and LSTM[12]. For TCE and CCE, which perform channel estimation using pilots in the delay-Doppler domain, the pilot SNR is set to $SNR_p = 40dB$ following their original designs. The LSTM baseline employs a time-frequency domain pilot structure with equal pilot and data power. Our proposed method uses a custom-designed, PAPR-optimized preamble sequence, distinct from the standard IEEE 802.11p preamble used in LSTM also maintaining equal pilot and data power.

The BER performance is evaluated at an SNR of 30 dB. The improvement of the proposed method is quantified in dB as the logarithmic ratio of the baseline BER to the proposed BER at this SNR.

Parameter	Values
Number of subcarriers	64
Number of symbols	14
Symbol Duration	8μs
Carrier Frequency	5.9GHz
Bandwidth	10MHz
Speed	{100,200,300,400,500}km/h
Channel Model	Vehicular A
Modulations	16-QAM
IBO	{2,4} dB

Table 4. OTFS Simulation Parameters

To assess the signal power distribution, we compared the impact of different channel estimation methods on the Complementary Cumulative Distribution Function (CCDF). The CCDF is a common metric for evaluating PAPR distortion; it statistically calculates the probability that the PAPR of a signal exceeds a threshold g:

$$CCDF = P\{PAPR > g\} \tag{38}$$

The PAPR is calculated as:

$$PAPR = 10\lg\left(\frac{\max(x[n]^2)}{mean(x[n]^2)}\right) \tag{39}$$

Where $x[n]$ represents the time-domain samples of the modulated signal.

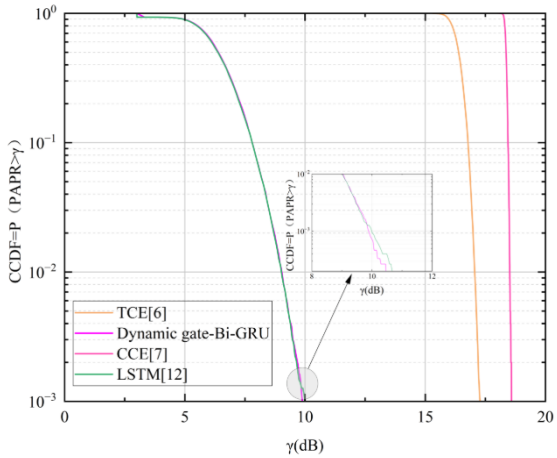

Fig.9 PAPR Comparison

Fig.9 PAPR Comparison shows that the transmission signal waveform of our proposed system enjoys a low PAPR property, achieving a gain of over 7 dB compared to both the TCE method[6] and the CCE method[7]. Notably, the PAPR characteristics of our proposed waveform, which employs a custom-designed preamble sequence, are comparable to those of the LSTM baseline that uses standard IEEE 802.11p preamble symbols. The subsequent channel estimation network is designed to effectively operate under this low-PAPR waveform.

To evaluate channel estimation performance, we also compared the NMSE of different channel estimation algorithms. The NMSE is defined as:

$$\mathrm{NMSE}=\frac{\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}\left|\hat{H}-H\right|^2}{\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}\left|H\right|^2} \tag{40}$$

Where $\hat{H}$ denotes the estimated channel matrix and H represents the true channel matrix.

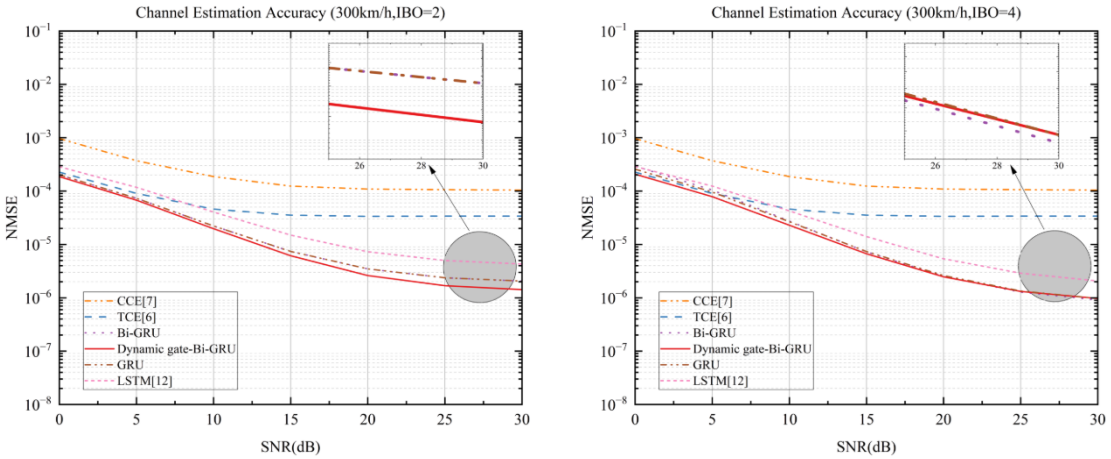

Fig. 10 NMSE analysis at 300 km/h

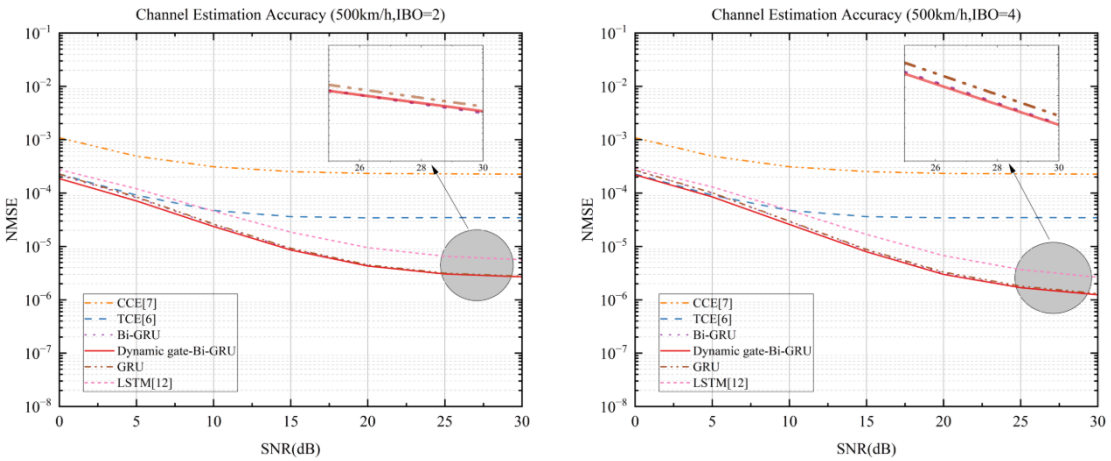

Fig. 11 NMSE analysis at 500 km/h

From Fig. 10, under high-mobility conditions of 300 km/h and with IBO levels of 2 dB and 4 dB, the proposed network-based estimation method outperforms conventional TCE, CCE, and the LSTM method from Ref. 12 across the entire SNR range. In particular, at an SNR of 30 dB, the proposed method achieves gains of 14.4 dB, 22.6 dB, and 3.3 dB over TCE, CCE, and LSTM, respectively. These values demonstrate that the proposed method significantly improves channel estimation accuracy in environments with HPA-induced nonlinear distortion compared to traditional methods, while also maintaining a clear advantage over other deep learning-based approaches. Notably, compared to traditional GRU and attention-enhanced Bi-GRU methods, the proposed method exhibits superior performance at low SNRs. This is benefitted from the IBO-dynamic gating mechanism, which adaptively adjusts information flow, suppresses noise interference in low-SNR environments, and simultaneously reinforces effective signal components. Fig. 11 shows that this advantage persists when the speed increases to 500 km/h with IBO levels of 2 dB and 4 dB.

To assess the detection capability of the proposed channel estimation algorithm in an OTFS system impaired by HPA-induced nonlinear distortion, we employed the BER as a comparative metric. Simulations were conducted under scenarios with speeds of 300 km/h and 500 km/h, and IBO values of 2 dB and 4 dB. An MMSE detector was used for signal detection, and signals underwent scrambling and convolutional encoding processing in the simulation.

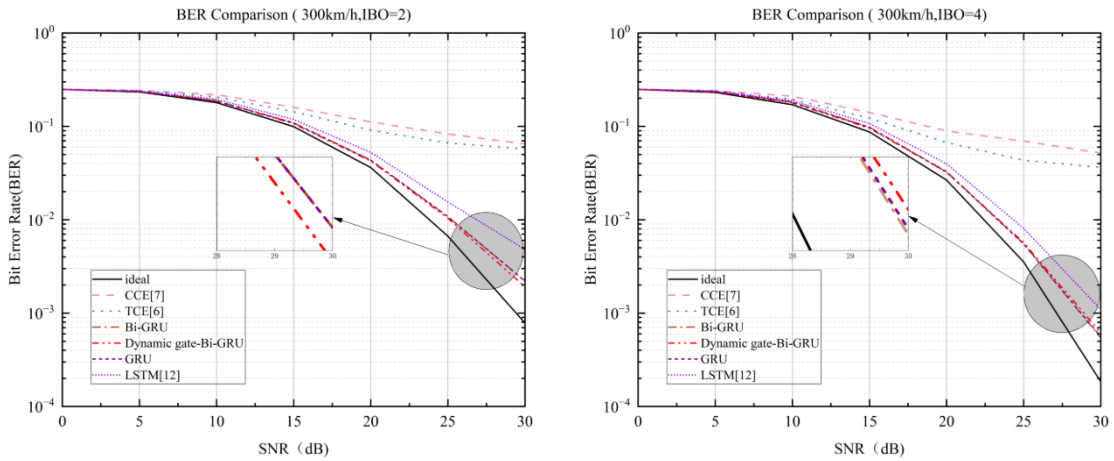

Fig. 12 BER analysis at 300 km/h

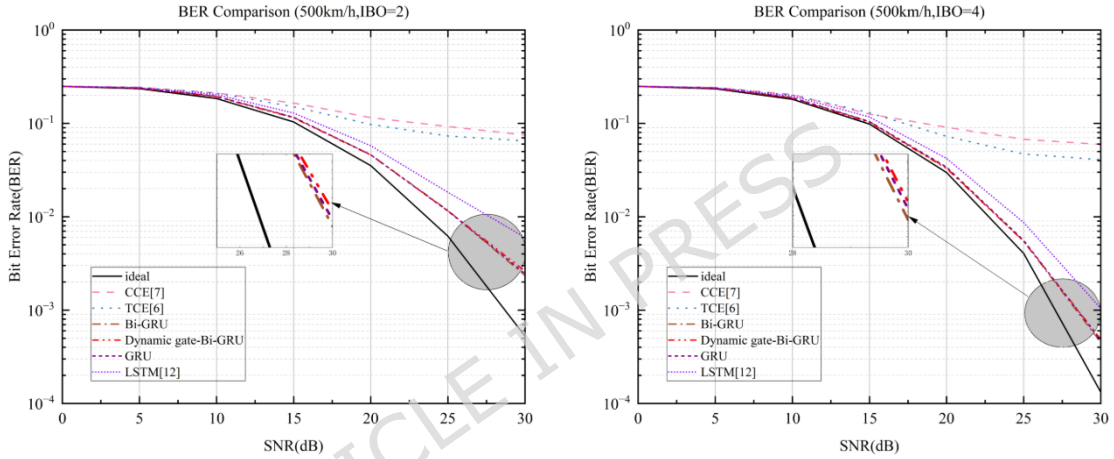

Fig. 13 BER analysis at 500 km/h

As observed in the left subfigure of Fig. 12 under the condition of IBO = 2 dB, the proposed method begins to outperform the comparative TCE, CCE, and LSTM methods in terms of BER once the SNR exceeds 10 dB. This performance advantage becomes more pronounced as the SNR increases beyond 25 dB. At the fixed SNR of 30 dB, the BER of the proposed method is 14.7 dB, 15.2 dB, and 4.0 dB lower than that of the TCE, CCE, and LSTM methods, respectively. These results highlight the exceptional estimation accuracy and strong robustness of the proposed approach under conditions involving nonlinear distortion and high-speed time-varying channels. The right subfigure of Fig. 12 and Fig. 13 further validate this conclusion, demonstrating consistent performance superiority even when the speed is increased to 500 km/h with IBO levels of 2 dB and 4 dB. Furthermore, while reducing the average computational complexity compared to the traditional attention-enhanced Bi-GRU estimation method, the proposed method maintains nearly identical BER performance, as evidenced in both Fig. 12 and Fig. 13.

Notably, the performance results presented in Fig. 12 and Fig. 13 are obtained using the identical hyperparameter set detailed in Table 1. Despite varying channel conditions—including speeds of 300 km/h and 500 km/h, and IBO levels of 2 dB and

4 dB—the proposed method consistently achieves low BER without any parameter re-tuning. As shown in the results, at an SNR of 30 dB the system maintains bit error rates below 3×10^{-3} under all considered scenarios, with specific performance figures visualized in Fig14 . This demonstrates the inherent adaptability and robustness of the IBO-driven dynamic gating mechanism, validating the design choice of fixed hyperparameters for diverse operating environments.

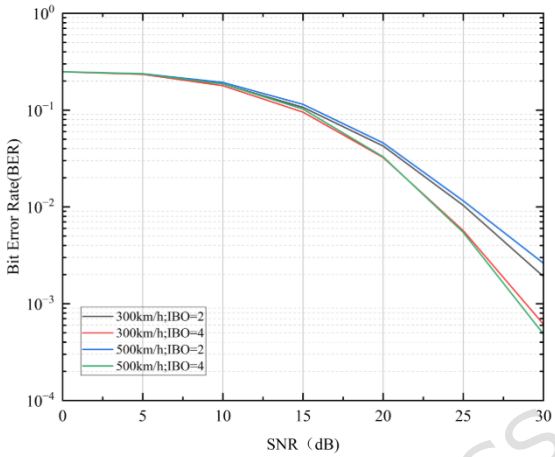

Fig. 14 BER Performance with Fixed Hyperparameters

System Throughput Performance. To directly quantify the impact on spectral efficiency, we evaluate the system throughput under varying SNR, as shown in Fig.14. Throughput is calculated based on successful data detection after channel estimation and MMSE equalization. The proposed Dynamic Gated Bi-GRU estimator achieves competitive throughput across all SNRs. It consistently outperforms the conventional CCE method[7] and shows comparable or slightly better performance than the LSTM baseline[12] across the evaluated range. However, the conventional TCE method[6] achieves the highest raw throughput under these linear detection conditions. This outcome is expected, as the TCE method employs high-power pilots that maximize the received SNR for data symbols in the absence of nonlinear compensation, albeit at the cost of high PAPR and susceptibility to HPA distortion. The key value of our method lies in delivering robust and efficient throughput under practical nonlinear HPA constraints, where traditional high-power pilot schemes like TCE would fail due to excessive distortion. The achievable throughput of our method under these realistic conditions directly validates its spectral efficiency improvement, stemming from accurate nonlinear channel estimation that enables reliable high-rate data transmission.

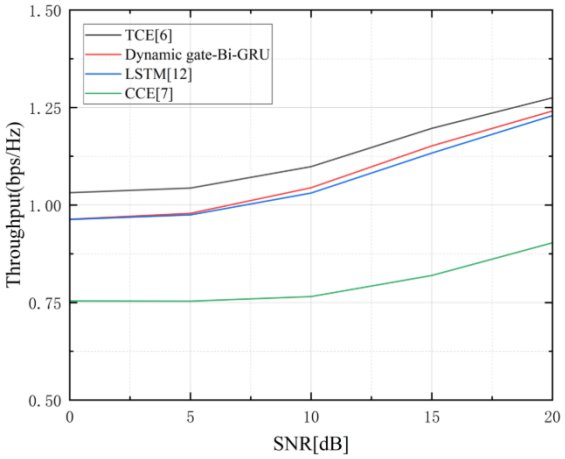

Fig. 15 Throughput Comparison (v = 500 km/h, 16-QAM modulation and IBO = 4dB)

To evaluate robustness against IBO inaccuracies, BER is tested at 500 km/h with true IBO = 4 dB and deviated values of 3, 5, 6, and 7 dB , with key results summarized in Fig. 16. The system shows strong tolerance to IBO overestimation (5–7 dB), with BER at SNR = 25 dB remaining within the same order of magnitude as the true IBO case. In contrast, underestimation to 3 dB leads to notable performance sensitivity, resulting in significant BER degradation. This asymmetry stems from the dynamic gating mechanism's conservative response to a low perceived IBO, which causes excessive reliance on historical channel state information and, in high-mobility scenarios, performance deterioration due to rapidly varying channel conditions.

This sensitivity analysis suggests that a conservative over-estimation strategy can be adopted in practice to mitigate the impact of IBO uncertainty. It is noted that the current model does not incorporate specific mechanisms to compensate for IBO mismatch; the observed asymmetry is an inherent characteristic of the proposed dynamic gating design under the evaluated training scheme.

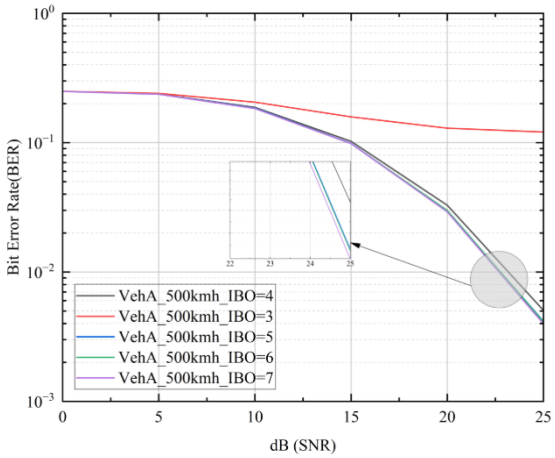

Fig. 16 BER performance of channel estimation under IBO mismatch

To verify the robustness of the proposed method against varying Doppler spreads, the system BER was tested under a fixed IBO of 4 dB across vehicle speeds ranging from 100 km/h to 500 km/h. Key results are summarized in Fig. 17. The results demonstrate that the proposed method maintains excellent and stable BER performance across the entire speed range from 100 km/h to 500 km/h. Even at the extreme speed of 500 km/h, the BER performance remains highly consistent with that at lower speeds across all SNR points. This conclusively proves that the proposed dynamic gated Bi-GRU based channel estimation method can effectively cope with severe Doppler shifts induced by high-speed mobility. Its performance is insensitive to variations in channel time selectivity, demonstrating strong environmental adaptability.

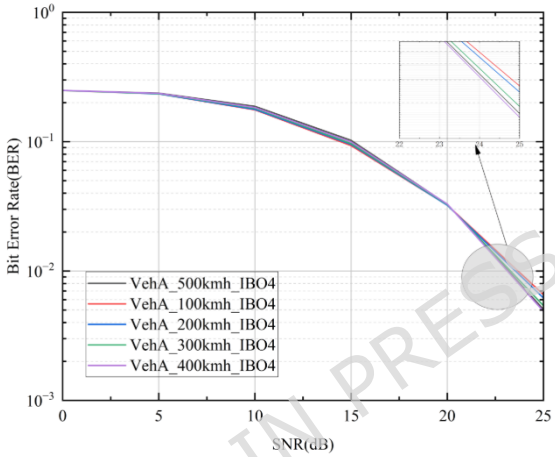

Fig. 17 System BER Performance under Different Vehicle Speeds

To assess the generalization capability of the proposed model across diverse channel conditions, its Bit Error Rate (BER) performance was evaluated under three representative channel models with a fixed IBO of 4 dB: Vehicular A (VehA, 500 km/h), Vehicular B (VehB, 300 km/h), and Pedestrian A (PedA, 3 km/h). The results are summarized in Fig. 18. The data indicates that the proposed method maintains superior performance across all tested channels. In the Vehicular B channel, which features severe delay spread, the performance is even slightly better than the baseline Vehicular A channel at most SNR points. Even in the Pedestrian A channel, which has vastly different multipath characteristics and minimal Doppler spread, the method still exhibits a rapidly declining BER curve for SNR ≥ 15 dB, demonstrating good learning and adaptation capabilities. This collectively verifies that the proposed algorithm is not overfitted to a specific channel scenario; the features learned by its dynamic gating and attention mechanisms exhibit strong generalization across different multipath structures and fading characteristics.

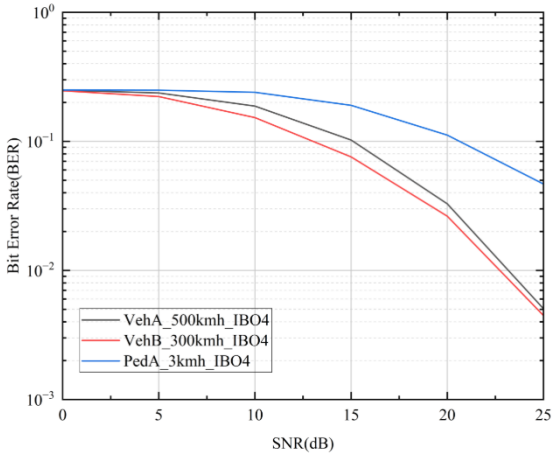

Fig. 18 System BER Performance under Different Channel Models

Discussion

This paper has addressed the challenge of channel estimation in OTFS communication systems impaired by nonlinear distortion introduced by HPAs. To tackle this problem, we proposed a novel channel estimation method that integrates an IBO-dynamic gated Bi-GRU network with a multi-head attention mechanism.

The observed spectral efficiency improvement is fundamentally enabled by the accurate nonlinear channel estimation of our IBO-aware estimator. Its superior estimation fidelity directly reduces the bit error rate at a given SNR. In a practical system, this translates into the ability to maintain reliable high-rate data transmission under the stringent constraints of HPA nonlinearity, a regime where traditional high-power pilot schemes falter due to excessive distortion. While simple methods may achieve higher throughput in linear channels, our method's key advantage is delivering robust and spectrally efficient performance when power amplifier nonlinearity is accounted for, thereby increasing the achievable data rate per unit bandwidth in real-world deployment scenarios.

Simulation results demonstrate that the proposed channel estimator exhibits superior estimation performance and strong robustness in environments with HPA-induced nonlinear distortion. Furthermore, through a dynamic sparsification strategy, the method maintains high estimation accuracy while ensuring relatively low computational complexity. This balance highlights its significant potential for practical deployment in future communication equipment.

The fixed-hyperparameter approach validated in this study offers notable advantages for practical deployment. By eliminating the need for environment-specific parameter tuning, it reduces calibration overhead and operational complexity. The consistent performance across varying speeds and IBO levels (Fig. 14) demonstrates strong generalization, suggesting that a single trained model can reliably operate across diverse vehicular scenarios. This robustness, combined with low computational complexity, facilitates the implementation of adaptive, learning-

based receivers in real-world systems.

The memoryless HPA model used in this study does not capture the memory effects of practical amplifiers. The algorithm's generalizability under complex channel conditions and its deployment cost on embedded devices also require further validation. The IBO-driven dynamic gating mechanism developed herein, however, establishes a generalizable framework for nonlinearity compensation that is not inherently limited to memoryless distortions. Future research will focus on HPA models with memory effects, investigate the transfer learning capabilities of the algorithm in complex scenarios. Future research will focus on HPA models with memory effects, investigate the transfer learning capabilities of the algorithm in complex scenarios.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request and with appropriate permission.

References

1. Fan, P., Zhao, J., & Chih-Lin, I. 5G high mobility wireless communications: Challenges and solutions. *China Communications*, *13*(2), 1-13 (2016).

2. Wen, M., Basar, E., Li, Q., Zheng, B., & Zhang, M. Multiple-mode orthogonal frequency division multiplexing with index modulation. *IEEE Transactions on Communications*, *65*(9), 3892-3906 (2017).

3. Wei, Z., Li, S., Yuan, W., Schober, R., & Caire, G. Orthogonal time frequency space modulation—Part I: Fundamentals and challenges ahead. *IEEE Communications Letters*, *27*(1), 4-8 (2022).

4. Hadani, R., et al. Orthogonal time frequency space modulation. In *2017 IEEE wireless communications and networking conference (WCNC),* (pp. 1-6) (2017).

5. Hadani, R., & Monk, A. OTFS: A new generation of modulation addressing the challenges of 5G. *arxiv preprint arxiv:1802.02623* (2018).

6. Raviteja, P., Phan, K. T., & Hong, Y. Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels. *IEEE transactions on vehicular technology*, *68*(5), 4906-4917 (2019).

7. Hashimoto, N., Osawa, N., Yamazaki, K., & Ibi, S. Channel estimation and equalization for CP-OFDM-based OTFS in fractional Doppler channels. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)* (pp. 1-7) (2021).

8. Zhang, H., Huang, X.& Zhang, J. A. Low-overhead OTFS transmission with frequency or time domain channel estimation. *IEEE Transactions on Vehicular*

Technology, *73*(1), 799-811 (2023).

9. Aghda, O. A., Omidi, M. J., & Saeedi-Sourck, H. Low-PAPR Joint Channel Estimation and Data Detection in ZP-OTFS System. *IEEE Transactions on Vehicular Technology* (2024).

10. Kumar, A., Gaur, N., Chakravarty, S., & Nanthaamornphong, A. Reducing the PAPR of OTFS modulation using hybrid PAPR algorithms. *Wireless Personal Communications*, *133*(4), 2503-2523 (2023).

11. Hu, J., et al. DNN and LS Based Channel Estimation in OTFS System. In *2023 IEEE 23rd International Conference on Communication Technology (ICCT)* (pp. 106-110) (2023).

12. dos Reis, A. F., Chang, B. S., Medjahdi, Y., Brante, G., & Bader, F. LSTM-Based Time-Frequency Domain Channel Estimation for OTFS Modulation. *IEEE Transactions on Vehicular Technology*, *73*(10), 15049-15060 (2024).

13. Surabhi, G. D., & Chockalingam, A. Low-complexity linear equalization for OTFS modulation. *IEEE communications letters*, *24*(2), 330-334 (2019).

14. Parkvall, S., Dahlman, E., Furuskar, A., & Frenne, M. NR: The new 5G radio access technology. *IEEE Communications Standards Magazine*, *1*(4), 24-30 (2018).

15. Colantonio, P., Giannini, F., & Limiti, E. *High efficiency RF and microwave solid state power amplifiers*. John Wiley & Sons (2009).

16. Shaiek, H., Zayani, R., Medjahdi, Y., & Roviras, D. Analytical analysis of SER for beyond 5G post-OFDM waveforms in presence of high power amplifiers. *IEEE Access*, *7*, 29441-29452 (2019).

17. Hou, J., Liu, H., Zhang, Y., Wang, W., & Wang, J. GRU-based deep learning channel estimation scheme for the IEEE 802.11 p standard. *IEEE Wireless Communications Letters*, *12*(5), 764-768 (2022).

18. Mohsin M A, Umer M, Bilal A, et al. Continual Learning for Wireless Channel Prediction[J]. *arXiv preprint arXiv:2506.22471* (2025).

19. Cheng, L., Pandey, A., Xu, B., Delbruck, T., & Liu, S. C. Dynamic gated recurrent neural network for compute-efficient speech enhancement. *arxiv preprint arxiv:2408.12425* (2024).

20. Fang, Y., Yang, S., Zhao, B., & Huang, C. Cyberbullying detection in social networks using bi-gru with self-attention mechanism. *Information*, *12*(4), 171 (2021).

Funding

This research was supported by the National Natural Science Foundation of China (U23A20280, 62161007, 62471153), the project of Guangxi Key Laboratory of

Precision Navigation Technology and Application. (No. DH202312) , and the Innovation Project of Guangxi Graduate Education(YCSW2025355).

Author contributions

J.H wrote the paper and did experiment, Z.W and Y.J provided supervision, X.D also provided supervision, All authors have reviewed the manuscript.