Journal of Computers Vol. 37 No. 5, October 2026, pp. 183-201
https://doi.org/10.63367/19911599202610370501 2

Transformer-KF Aided Sparse mmWave Channel Estimation for Complex Dynamic Scenarios

Jiancheng Chen*

Xiamen Intretech Inc., Xiamen 361000, P.R. China

xmcjc@intretech.com

Received 1 May 2026; Revised 1 August 2026; Accepted 15 August 2026

Abstract. Millimeter-wave (mmWave) MIMO systems provide large bandwidth and high array gain, but reliable channel estimation is challenged by high-dimensional arrays, severe path loss, blockage, and dynamic dominant-path switching. Sub-6 GHz channels are easier to observe and can provide out-of-band angular priors for mmWave sparse recovery. However, Root-MUSIC angle observations may contain outliers under low SNR, blockage, and path switching. This paper proposes a Sub-6 GHz aided sparse mmWave channel estimation method. Sub-6 GHz pilots are first processed by least-squares channel estimation and Root-MUSIC angle observation. A gated residual Transformer-Kalman filter (TF-KF) then tracks the angular prior, and the resulting angle and confidence are transferred to confidence-aided AWP-AMP for mmWave sparse channel recovery. The proposed TF-KF does not replace the Kalman filter; instead, it learns Root-MUSIC observation confidence, bounded residual compensation, and measurement-covariance scaling while preserving the recursive KF structure. In a representative fixed-seed test trajectory, TF-KF reduces the average AoA/AoD RMSE from 14.903 deg for conventional KF to 2.937 deg. In sparse recovery, the TF-KF prior with a strong fixed threshold coefficient lambda = 0.5 obtains an average NMSE of -1.5647 dB, almost identical to the learned threshold-calibration result of -1.5646 dB. These results indicate that the dominant performance gain comes from reliable confidence-aware angular priors, while AMPParamNet mainly acts as a lightweight threshold-calibration module in the current setting.

Keywords: mmWave, Sub-6 GHz assistance, Root-MUSIC, Transformer-KF, AWP-AMP, sparse channel estimation

1 Introduction

mmWave communication provides large bandwidth and narrow beams, but also introduces high-dimensional channel estimation and severe path-loss challenges [1]. Since mmWave channels are often sparse in the angular domain, compressed sensing and AMP-type algorithms have been widely used to reduce pilot overhead [2], [3]. In dynamic scenarios, however, angular support is not static. Blockage, competing reflected paths, and user motion may cause dominant-path switching, making fixed angular priors or overly smooth filters unreliable.

Sub-6 GHz and mmWave links may share geometric scatterer information; therefore, low-frequency channels can provide out-of-band assistance for mmWave estimation [4, 5]. Measurement-based mmWave modeling also emphasizes spatial consistency and blockage, which motivates dynamic path evolution in simulation [6]. Nevertheless, scatterer visibility and dominant-path selection are not perfectly aligned across bands, and Root-MUSIC may produce abnormal observations at low SNR. Thus, angular priors should be used with confidence rather than as deterministic truth.

Existing mmWave channel-estimation methods can be roughly divided into three categories. Sparse-recovery methods exploit angular sparsity but are sensitive to dictionary resolution, thresholds, and support priors; VAMP provides a representative message-passing reference for structured linear inverse problems [7]. Deep-learning methods directly map observations or low-frequency channels to mmWave channels, but their interpretability and cross-scenario generalization remain concerns; learning-based beamspace and hybrid-transceiver channel estimation also highlights the importance of exploiting angular structure [8, 9]. Out-of-band assisted methods reduce overhead by using low-frequency geometry. This paper follows the third direction, but places angle observation, observation confidence, and sparse-recovery thresholding into one interpretable pipeline.

* Corresponding Author

1.1 Literature Review and Research Gap

Research on mmWave MIMO has established that sparse angular structure is a key resource for reducing pilot overhead. Early hybrid-precoding and channel-estimation studies exploit the limited scattering property of mmWave propagation and formulate channel acquisition as a sparse reconstruction problem [2, 10]. The virtual channel representation further explains why large antenna arrays can transform geometric propagation into a sparse angle-domain model [11]. These studies provide the signal-processing basis for using compressed sensing and AMP-type recovery in the mmWave stage.

Another line of work focuses on signal processing for practical mmWave transceivers. Hardware constraints, hybrid analog-digital architectures, beam training overhead, and frequency-dependent propagation all make direct full-dimensional channel acquisition difficult [12, 13]. As a result, side information, structural priors, and learning-based modules are often introduced to reduce the effective search space. However, when the prior is inaccurate, a narrow angular window may suppress the correct path and degrade sparse recovery.

Out-of-band assistance uses the fact that Sub-6 GHz and mmWave links may share scatterers and large-scale geometry [4, 5]. This idea is especially attractive because Sub-6 GHz pilots are easier to estimate robustly. Nevertheless, the cross-band relation is not exact: the dominant path at mmWave may disappear under blockage, new reflected paths may become dominant, and Root-MUSIC observations obtained from noisy LS estimates may contain local jumps. Therefore, this paper treats Sub-6G-derived angles as uncertain observations rather than fixed labels.

Deep-learning channel estimation has shown strong performance in beamspace and hybrid-transceiver scenarios [8, 9]. Yet a fully data-driven channel estimator may hide the physical meaning of the intermediate variables and can be sensitive to distribution shift. For dynamic angle tracking, a learning-aided Kalman filter is more interpretable because the KF still handles recursive prediction and covariance propagation. The remaining challenge is how to let the network decide when to trust Root-MUSIC and when to rely on the prediction.

Compared with the above methods, the proposed framework is designed around confidence-aware prior transfer. The Transformer-KF module produces not only an angle estimate but also observation confidence. This confidence is then used by AWP-AMP to adjust the angular prior width and threshold behavior. Hence, the method connects angle tracking and sparse recovery through a physically meaningful confidence variable instead of using the tracked angle as a deterministic input.

1.2 Motivation of the Proposed Framework

The main difficulty in the considered scenario is the coexistence of true path switching and false observation jumps. Both appear as large angle changes in raw Root-MUSIC outputs, but their meanings are opposite. A true dominant-path switch should be followed rapidly, whereas a noise-induced Root-MUSIC outlier should be rejected. A conventional KF tends to smooth both cases and therefore suffers from tracking lag. A direct Root-MUSIC prior has no memory and may follow false jumps. A generic Deep-KF can learn temporal smoothing, but without an explicit confidence output it is difficult to transfer its reliability to the sparse recovery stage.

This observation motivates the gated residual structure. The gate controls the information flow from Root-MUSIC to the fused angle prior. A high gate indicates that the observation should be trusted, which helps the filter follow a real path switch. A low gate weakens the influence of an abnormal observation, preventing the sparse recovery window from being dragged to a wrong angular region. The residual branch compensates for the remaining mismatch of the constant-velocity model. In this way, the proposed TF-KF is not merely a smoother; it is a confidence-aware fusion module.

The main contributions are: 1) a complete LS-Root-MUSIC-TF-KF-AWP-AMP framework; 2) a physically interpretable gated residual TF-KF, where the gate denotes Root-MUSIC observation confidence and the residual compensates KF model mismatch; 3) confidence transfer from TF-KF to AWP-AMP, where the gate controls the Gaussian angular-prior width instead of treating the tracked angle as a hard constraint; 4) threshold-calibration and strong fixed-threshold ablations showing that the final sparse-recovery gain is mainly caused by reliable angular priors and an appropriate shrinkage range, rather than by an unconstrained black-box channel estimator.

The complete processing flow is illustrated in Fig. 1, where the Sub-6G front-end angle observation, TF-KF tracking, and confidence-aided AWP-AMP recovery are connected in one pipeline.

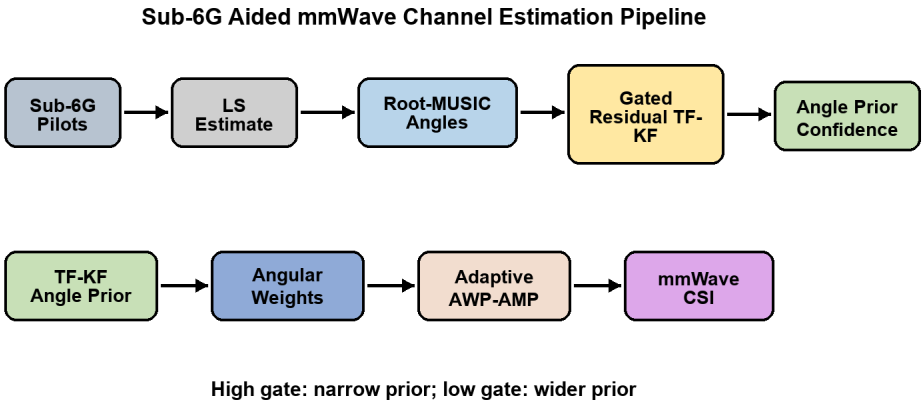


Fig. 1. Sub-6G aided mmWave channel-estimation pipeline

1.3 Organization of the Paper

The rest of this paper is organized as follows. Section 2 describes the Sub-6G and mmWave signal models and clarifies how LS estimation and Root-MUSIC observation are connected. Section 3 presents the gated residual Transformer-KF and explains the roles of observation confidence and bounded residual compensation. Section 4 introduces the confidence-aided AWP-AMP recovery procedure and the adaptive threshold module. Section 5 reports the dynamic channel simulation settings. Section 6 discusses tracking, ablation, NMSE, and learning-curve results. Sections 7 and 8 analyze complexity, runtime overhead, practical implications, and limitations. Section 9 concludes the paper.

2 System Model and Front-End Angle Observation

2.1 Array Response and mmWave Sparse Channel Model

For a uniform linear array, the steering vector for angle θ is

$$\mathbf{a}_N(\theta) = \frac{1}{\sqrt{N}}\left[1, e^{j\pi\sin\theta}, \ldots, e^{j\pi(N-1)\sin\theta}\right]^T \tag{1}$$

The geometric mmWave channel at frame (k) is

$$\mathbf{H}_{\mathrm{mm},k} = \sum_{l=1}^{L_k} \alpha_{l,k}\mathbf{a}_r(\theta_{l,k})\mathbf{a}_t^H(\phi_{l,k}) \tag{2}$$

After angular dictionary representation and compressed observation, the mmWave pilot model is

$$y = D_{s+n} \tag{3}$$

2.2 Sub-6 GHz Pilot Observation and LS Estimation

The Sub-6 GHz pilot observation follows the standard linear pilot model shown in the displayed equation. The channel matrix denotes the observable Sub-6 GHz MIMO channel, the pilot matrix denotes the transmitted training block, the received pilot block is contaminated by receiver noise, and the LS estimate is obtained from the normal-equation form. When orthogonal pilots are used and the pilot Gram matrix is well conditioned, the estimator reduces to the pseudo-inverse implementation.

For the mmWave recovery stage, the compressed observation is written in an angular sparse form. The variables represent the mmWave channel, the pilot-combining sensing matrix, the angular dictionary, the equivalent sensing matrix, the sparse angular-domain channel vector, measurement noise, the number of compressed measurements, and the angular-grid size. These definitions are used consistently in the AWP-AMP recursion.

The symbols and dimensions used in the following derivations are summarized in Table 1.

Table 1. Main notation used in this paper

Symbol	Meaning
theta, phi	AoA and AoD angles, reported in degrees for plots and RMSE.
z_{raw}, z_{pred}, z_{prior}	Root-MUSIC observation, KF predicted angle observation, and TF-KF fused angular prior.
x_k	Four-dimensional KF state [theta, phi, angular velocity of theta, angular velocity of phi].
s	Sparse angular-domain mmWave channel vector; it is different from the KF state x_k.
D	Equivalent compressed sensing matrix formed by the pilot-combining matrix and angular dictionary.
g	Observation-confidence gate; a larger value means Root-MUSIC is trusted more strongly.
delta	Bounded residual compensation added to the predicted angular observation.
λ_t, beta	AMP threshold coefficient and fixed damping factor.
τ_i^t	Weighted threshold applied to the i-th angular coefficient at AMP iteration t.
R_{eff}	Confidence-scaled KF measurement covariance used during the update step.

$$H_{LS} = YX^{\dagger} \tag{4}$$

2.3 Root-MUSIC AoA/AoD Extraction from LS Channel

Root-MUSIC is a classical high-resolution subspace angle-estimation method [14]. In this paper, the AoA covariance is constructed from the LS channel estimate and its Hermitian transpose, while the AoD covariance uses the corresponding transmit-side product. After eigendecomposition, the dominant signal subspace represents the selected main propagation component, while the noise subspace is used to form the Root-MUSIC polynomial or pseudo-spectrum. The number of tracked dominant paths is one in the present experiments, so the strongest physically consistent root or peak is selected as the raw AoA/AoD observation.

It is important that Root-MUSIC uses the LS-estimated channel rather than the true channel. Therefore, low SNR, finite pilot length, blockage-induced attenuation, and local subspace failures can produce abnormal observations. A Root-MUSIC outlier is marked when the observation deviates strongly from the local prediction while the true path does not show a corresponding physical switch. True angles are used only to build supervised labels and evaluation metrics in simulation; they are unavailable during online testing.

$$\begin{aligned} R_{AoA} &= H_{LS} H_{LS}^{H}, \\ R_{AoD} &= H_{LS}^{H} H_{LS}. \end{aligned} \tag{5}$$

The true angle is defined as the instantaneous dominant mmWave path. True dominant-path switching is retained as valid channel dynamics, whereas Root-MUSIC outliers are treated as observation noise and are not used as testing labels.

The LS estimate is used only as the observable Sub-6G channel. Root-MUSIC is then applied to the covariance matrices derived from the LS estimate, not from the true channel. This setting is important because it preserves the effect of pilot noise, finite pilot length, and occasional subspace estimation failure. The true AoA

and AoD are used only for supervised training and evaluation, not as available online information.

The simulated dynamic channel follows two principles. First, most frames obey continuous motion with moderate nonlinear acceleration, so temporal tracking should be useful. Second, a small number of frames contain dominant-path switching or blockage, so the true trajectory is not artificially smoothed. This distinction makes the task harder than fitting a smooth curve: the estimator must suppress false Root-MUSIC outliers while preserving true channel discontinuities.

The steering-vector convention is kept consistent between channel generation and Root-MUSIC estimation. Angles are internally handled in radians for array-response computation and converted to degrees only for plotting, thresholding, and RMSE reporting. This avoids unit inconsistency, which is a common reason for unstable angle observations in simulation studies.

2.4 Dynamic Channel and Front-End Observation Errors

The Sub-6G front end introduces several error sources before the angle prior is transferred to the mmWave stage. Pilot noise affects the LS channel estimate directly, while finite pilot length limits the rank and reliability of the estimated covariance matrices. When the received signal is weak, the signal and noise subspaces may become poorly separated, and Root-MUSIC may produce a spurious peak. These errors are not removed from the simulation because they are exactly the observations that a real tracking module must handle.

A second source of uncertainty comes from cross-band mismatch. Although Sub-6G and mmWave links may share some scatterers, their dominant components are not guaranteed to be identical. A path that is strong at Sub-6G may be attenuated at mmWave, and a reflected path that is weak at Sub-6G may become important at mmWave because of directional beamforming. The proposed framework therefore uses the Sub-6G angle as a soft prior rather than as a deterministic support constraint.

The third source of difficulty is temporal nonstationarity. In a static or slowly varying channel, a conventional KF can perform well because the constant-velocity assumption is approximately valid. In the considered dynamic scenario, however, user motion, blockage, and path competition can change the dominant direction abruptly. This is why the learning module must distinguish true path evolution from isolated observation errors.

3 Gated Residual TF-KF Angle Tracking

Kalman filtering provides recursive prediction and measurement update [15]. The angular state is

$$x_k = \left[\theta_k, \phi_k, \dot{\theta}_k, \dot{\phi}_k\right]^T, \quad z_k = \left[\theta_k, \phi_k\right]^T \tag{6}$$

3.1 KF Prediction Baseline and Observation Innovation

Conventional KF uses a constant-velocity transition matrix. It is stable but may lag under nonlinear motion or path switching. Directly trusting Root-MUSIC, on the other hand, can be harmful when outliers occur. Therefore, the proposed Transformer assists the information fusion process rather than replacing recursive filtering.

3.2 Transformer Input and Output Design

Each frame feature includes Root-MUSIC observation, KF prediction, angular velocity, innovation, and covariance:

$$s_k = [\theta^{raw}, \phi^{raw}, \theta^{pred}, \phi^{pred}, \dot{\theta}, \dot{\phi}, v_\theta, v_\phi, P_\theta, P_\phi]^T \tag{7}$$

The most recent L frames form S_k. Transformer self-attention is suitable for extracting short-term temporal dependency [16], and the network outputs gate, residual, and confidence terms:

$$[g_{\theta,k}, g_{\phi,k}, \Delta\theta_k, \Delta\phi_k] = f_{TF}(S_k). \tag{8}$$

3.3 Gate-Residual Fusion and Confidence-Scaled KF Update

Here g denotes Root-MUSIC observation confidence and lies between 0 and 1. The implemented TF-KF update has two coupled effects. First, the Transformer constructs a confidence-corrected angular prediction according to the display equation below. Second, the same confidence information scales the measurement covariance used by the subsequent KF update.

$$z_{prior,k} = z_{pred,k} + g_k \odot (z_{raw,k} - z_{pred,k}) + \delta_k,$$
$$\delta_k = \delta_{\max} \tanh(o_{\delta,k}),$$
$$x_k^- = Ax_{k-1}, \quad P_k^- = AP_{k-1}A^T + Q, \quad z_{pred,k} = Hx_k^-,$$
$$\tilde{x}_k^-(1:2) = z_{prior,k}, \quad \tilde{x}_{\bar{k}}(3:4) = x_{\bar{k}}(3:4),$$
$$R_{eff,k} = diag(r_{\theta,k}, r_{\phi,k})R, \quad r_{\theta,k}, \quad r_{\phi,k} \geq 1,$$
$$K_k = P_k^- H^T (HP_k^- H^T + R_{eff,k})^{-1},$$
$$x_k = \tilde{x}_k^- + K_k(z_{raw,k} - H\tilde{x}_{\bar{k}}), \quad P_k = (I - K_k H)P_k^-.$$

Thus, the Transformer does not directly output the final tracked angle and does not create a second independent observation. It modifies the predicted angular state and the measurement covariance, after which the standard Kalman gain, measurement update, and covariance update are applied using the Root-MUSIC observation. The final reported TF-KF angle is the first two entries of the updated KF state.

The training objective follows the implementation logs and is a weighted multi-task loss. Its dominant term is the fused-angle Smooth-L1 loss. Auxiliary terms supervise the gate, jump indicator, residual branches, and measurement-covariance scaling, while residual regularization prevents excessively large corrections:

The loss weights follow the final implementation: the dominant fused-angle term is combined with gate, jump, residual-branch, measurement-scale, and residual-regularization terms.

In the manuscript-level algorithm, internal residual-strength variables are absorbed into the effective residual delta. They are not treated as separate physical channel parameters.

The internal structure of the gated residual TF-KF is illustrated in Fig. 2.

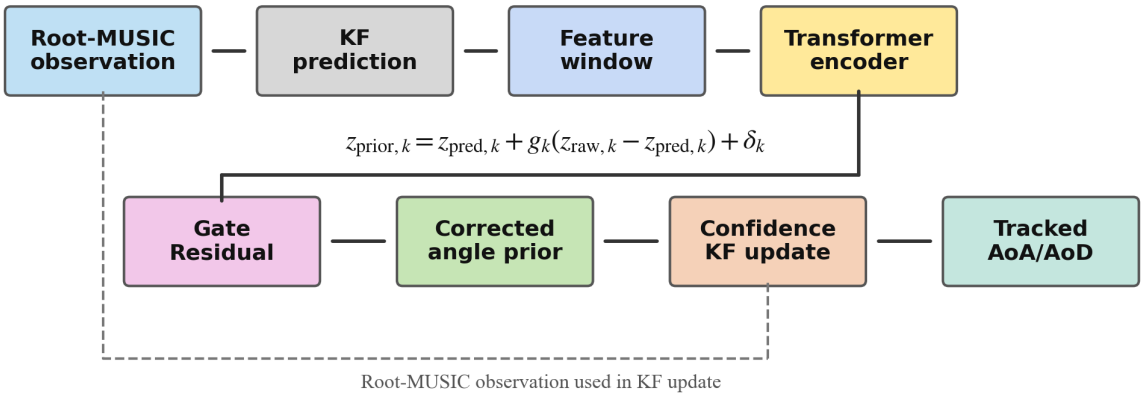


Fig. 2. Gated residual TF-KF module

The gate and residual branches play different roles. The gate branch is responsible for observation selection, whereas the residual branch is responsible for prediction correction. If only the gate is used, the fused result is restricted to a weighted combination of Root-MUSIC and KF prediction. When both sources are biased, such a structure cannot generate a new correction direction. The residual branch resolves this limitation by adding a bounded correction learned from the recent temporal context.

The Transformer input window also carries uncertainty information through the covariance terms. A large covariance indicates that the recursive tracker is less confident in the current prediction, while a large innovation indicates inconsistency between prediction and Root-MUSIC. Self-attention can compare these quantities over multiple frames and distinguish isolated observation outliers from persistent changes. This is the reason for using a short temporal window rather than a single-frame multilayer perceptron.

To avoid turning the neural network into an unconstrained post-processor, the final angle is obtained only after the KF update. The network modifies the predicted observation through gate and bounded residual compensation and also scales the measurement covariance. The KF keeps the recursive state, covariance propagation, Kalman gain, and measurement update. This hybrid design improves interpretability and reduces the risk of unstable extrapolation outside the training distribution.

3.4 Training Objective and Confidence Interpretation

The TF-KF module is trained by supervising the final fused angle rather than only supervising an intermediate residual. This choice is important because the final objective of angle tracking is not to reproduce a particular residual value, but to provide an accurate and reliable angular prior. The gate is encouraged to approximate observation reliability: it should increase when Root-MUSIC is closer to the true angle than the prediction and decrease when Root-MUSIC behaves like an outlier.

The residual branch is deliberately bounded. Without a bound, the neural network may overcompensate the KF prediction and follow noise-induced patterns in the training set. With a bound, the residual mainly corrects model mismatch and short-term nonlinear motion, while large jumps are primarily handled by the observation gate. This division of labor keeps the model interpretable and prevents the Transformer from becoming an unconstrained black-box angle estimator.

The temporal window helps the network exploit persistence. A single abnormal Root-MUSIC point usually appears as a sudden innovation spike that is not supported by neighboring frames. A true path switch, in contrast, tends to be followed by a persistent change in subsequent observations. Self-attention can compare current innovation with previous velocity, covariance, and observation patterns, which makes the gate more robust than a single-frame threshold rule.

Table 2 clarifies the difference among conventional KF, the Deep-KF baseline, and the proposed TF-KF. The Deep-KF baseline is implemented as a learning-assisted KF smoother using the same temporal observation features but without an explicitly supervised observation-confidence gate and without transferring confidence to AWP-AMP. It therefore represents a generic neural correction baseline, whereas the proposed TF-KF decomposes the problem into observation selection, bounded model-error compensation, and confidence-aware prior transfer.

Unlike generic gating or deep Kalman filtering, the proposed gate is not an internal attention weight. It has a clear physical meaning: Root-MUSIC observation confidence. The residual does not replace the observation; it compensates the bounded mismatch of the fixed motion model. More importantly, the gate is transferred to mmWave AWP-AMP to control angular-prior width, forming confidence coupling from angle tracking to sparse recovery.

Table 2. Structural comparison of angle-tracking methods

Method	Deep network	Explicit gate	Residual compensation	Confidence to AMP
Conventional KF	No	No	No	No
Deep-KF baseline	Yes	No	Learnable correction	No
Proposed TF-KF	Yes	Yes	Yes	Yes

4 TF-KF Aided Threshold-Calibrated AWP-AMP

4.1 Confidence-Controlled Angular Prior Weight

Using the TF-KF angle and gate, a two-dimensional Gaussian angular weight is constructed:

$$w(\theta_i,\phi_j)=\exp\left[-\frac{(\theta_i-\hat{\theta})^2}{2\sigma_\theta^2}-\frac{(\phi_j-\hat{\phi})^2}{2\sigma_\phi^2}\right] \tag{11}$$

where

$$\sigma_\theta=\sigma_{\min}+(1-g_\theta)\sigma_{extra},\quad \sigma_\phi=\sigma_{\min}+(1-g_\phi)\sigma_{extra}. \tag{12}$$

High gate values concentrate the prior, whereas low gate values widen it to avoid suppressing the true path. The AMP threshold is

$$T_t=\lambda_t\frac{\left\|r^t\right\|_2}{\sqrt{M}},\quad T_i^t=\frac{T_t}{w_i+\varepsilon}. \tag{13}$$

AMPParamNet outputs a bounded threshold coefficient from residual and gate features. In the final analysis it is interpreted as a lightweight threshold-calibration module, and a fixed lambda = 0.5 strong baseline is reported explicitly:

$$\lambda_t=\lambda_{\min}+(\lambda_{\max}-\lambda_{\min})\sigma(o_t) \tag{14}$$

Damping is fixed:

$$x^{t+1}=\beta\eta(u^t;\tau^t)+(1-\beta)x^t. \tag{15}$$

4.2 Weighted AMP Recursion with Fixed Damping

Because the sensing matrix is formed by compressed measurements and angular dictionaries, it does not strictly satisfy the i.i.d. Gaussian condition required by classical AMP state evolution. Therefore, this paper does not claim a global convergence proof. Instead, stability is encouraged by bounded threshold coefficients, lower-bounded angular weights, fixed damping, finite iterations, and residual monitoring.

The full AWP-AMP recursion used in the sparse recovery stage is summarized below. The AMP sparse estimate is denoted separately from the KF state to avoid symbol confusion.

$$\begin{aligned}
u^t&=x_{amp}^t+D^Hr^t\\
\tau^t&=\lambda_t\frac{\left\|r^t\right\|_2}{\sqrt{M}},\quad \tau_i^t=\frac{\tau^t}{w_i+\varepsilon}\\
x_{new}^t(i)&=\eta(u^t(i),\tau_i^t)=sign(u^t(i))\max(\left|u^t(i)\right|-\tau_i^t,0)\\
x_{amp}^{t+1}&=\beta x_{new}^t+(1-\beta)x_{amp}^t\\
r^{t+1}&=y-Dx_{amp}^{t+1}+b_tr^t
\end{aligned} \tag{16}$$

$$stop \ \ if \ \frac{\left|\left\|r^{t+1}\right\|_2-\left\|r^{t}\right\|_2\right|}{\left\|r^{t}\right\|_2+\varepsilon}<\varepsilon_{stop} \ \ or \ \ t=T_{\max}.$$

The correction coefficient b_t follows the AMP Onsager correction form and is estimated from the active fraction or shrinkage derivative. Fixed damping beta = 0.55 is used in the final experiment because it produced more stable behavior than learning both threshold and damping simultaneously.

At each AMP iteration, AWP-AMP first forms a matched-filtered pseudo observation from the current sparse estimate and residual. It then applies weighted soft thresholding, where the threshold of each angular bin is inversely related to the TF-KF prior weight. The estimate is updated by fixed damping, and the residual is updated with the standard AMP correction term. The iteration stops when the maximum iteration number is reached or the normalized residual change becomes sufficiently small. In the reported implementation, fixed damping is used for stability, while AMPParamNet only adapts the threshold coefficient.

The residual and estimate variation metrics are

$$\boldsymbol{\epsilon}_r^t=\frac{\left\|r^t-r^{t-1}\right\|_2}{\left\|r^{t-1}\right\|_2+\varepsilon}, \ \ \boldsymbol{\epsilon}_x^t=\frac{\left\|x^t-x^{t-1}\right\|_2}{\left\|x^{t-1}\right\|_2+\varepsilon}. \tag{16}$$

The overall algorithm is summarized as follows. First, Sub-6G LS channel estimation is computed from pilot observations. Second, Root-MUSIC extracts raw AoA/AoD observations from the LS-derived covariance matrices. Third, KF prediction and Transformer-based gate/residual inference produce the TF-KF angle prior. Fourth, the gate controls Gaussian angular weights and the measurement covariance used by tracking. Finally, AMPParamNet outputs the threshold coefficient, and AWP-AMP performs weighted soft thresholding with fixed damping.

The role of AWP-AMP is to use the TF-KF angle prior without making the recovery overly dependent on a single angle. A narrow prior is beneficial when TF-KF is confident, because it reduces dictionary ambiguity and suppresses noise-like coefficients. When the TF-KF gate is low, the prior width is enlarged so that the sparse recovery process can still search neighboring angular bins. This prevents the angle tracker from becoming a hard constraint. This weighted recovery stage is related to the broader AMP/GAMP family, where iterative residual correction and thresholding are used to solve high-dimensional linear inverse problems [17].

The threshold-calibration module is intentionally lightweight. AMPParamNet does not estimate the channel directly; it only outputs a threshold coefficient according to iteration status, residual norm, residual variation, estimate variation, and TF-KF confidence. In the final experiment, the learned coefficient is very close to the upper bound 0.5, and its NMSE is almost identical to a fixed lambda = 0.5 baseline. Therefore, the paper treats AMPParamNet as threshold calibration and diagnostic support, while the primary source of improvement is the confidence-aware TF-KF angular prior.

Algorithm 1. Transformer-KF aided threshold-calibrated AWP-AMP

Input: Sub-6G pilots, mmWave pilot observation, pilot matrix, angular dictionary, trained TF-KF, and optional threshold-calibration module.
Step 1: Estimate the Sub-6G channel by LS and construct AoA/AoD covariance matrices.
Step 2: Apply Root-MUSIC to obtain raw AoA and AoD observations for each frame.
Step 3: Run KF prediction and form the temporal Transformer input window from observation, prediction, innovation, velocity, and covariance features.
Step 4: Use TF-KF to infer observation gate, residual compensation, and a confidence-corrected pseudo measurement.
Step 5: Perform one KF measurement update using the pseudo measurement and confidence-scaled measurement covariance, then convert the resulting angle and gate into a Gaussian angular-prior weight map.
Step 6: In each AMP iteration, use either a fixed threshold coefficient or AMPParamNet to determine the shrinkage threshold; fixed damping is used for stability.
Step 7: Form weighted thresholds, apply soft thresholding, update the residual with the AMP correction term, and use fixed damping for stability.
Output: Estimated sparse angular-domain channel and reconstructed mmWave MIMO channel.

Note: The Transformer and AMPParamNet assist prior construction and threshold calibration; they do not replace LS, Root-MUSIC, KF recursion, or AMP recovery.

4.3 Threshold Calibration and Prior Reliability

In AWP-AMP, the angular prior influences recovery through the weight map and the threshold. If the prior is too narrow when the angle estimate is wrong, the true sparse coefficient may be excessively penalized. If the prior is too wide when the angle estimate is reliable, the algorithm loses the benefit of out-of-band assistance. The TF-KF gate provides a natural way to adjust this tradeoff by controlling the width of the Gaussian angular weight.

The threshold coefficient controls shrinkage strength. At low SNR, a larger threshold suppresses noise components, whereas at high SNR a smaller threshold can preserve weak valid paths. The final experiment shows that a strong fixed threshold coefficient lambda = 0.5 is already a competitive baseline, and the learned AMPParamNet output converges to nearly the same value. This observation is reported transparently to avoid overstating the independent adaptivity of the threshold network.

Fixed damping is retained for stability. Earlier experiments showed that learning both threshold and damping can introduce unnecessary variation when the training set is limited. By keeping damping fixed and comparing fixed lambda = 0.08, fixed lambda = 0.5, and AMPParamNet, the final version separates the influence of angular prior quality from threshold-range selection.

5 Simulation Setup

5.1 Dynamic Channel Generation

The simulated channel includes continuous motion, nonlinear angular variation, dominant-path switching, blockage, distance-dependent path loss, low-SNR observations, and local Root-MUSIC outliers. The setting follows the path-loss, spatial-consistency, blockage, and dynamic propagation ideas in 3GPP and NYUSIM-type mmWave channel modeling [6, 18]. Therefore, this paper preserves true dominant-path switching rather than artificially smoothing the true angle. Training, validation, and testing are split by complete trajectories to avoid temporal leakage.

The training, validation, and testing sets are generated from the same channel simulator but are separated by complete trajectories. This avoids temporal leakage caused by randomly mixing frames from the same trajectory. Each split contains normal continuous frames, path-switching frames, and Root-MUSIC outlier frames. This split strategy is more demanding than random frame sampling because the model must generalize to unseen motion sequences.

The simulation also separates true channel dynamics from observation errors. Dominant-path switching and blockage are treated as valid physical events in the true channel. Root-MUSIC outliers are generated through noisy LS observations, low-SNR frames, and local subspace estimation failures. Therefore, a good estimator should not simply smooth all sudden changes; it should identify whether the sudden change is supported by the temporal context and observation confidence.

5.2 SNR, Distance, and Cross-Band Mismatch Settings

Distance and SNR are varied to reflect mmWave propagation sensitivity. At long distance or low SNR, the LS and Root-MUSIC front end becomes less reliable, and the confidence-aware prior becomes more important. At high SNR, most prior-aided AMP variants tend to perform similarly because the angular support is already clear. The proposed method is therefore mainly evaluated under difficult regimes where prior reliability changes over time.

5.3 Dataset Split and Dynamic Scenario Coverage

The dataset is designed to cover both easy and difficult tracking conditions. Normal continuous frames dominate the data distribution, but each split also contains path-switching, blockage, low-SNR, and Root-MUSIC outlier frames. These event frames are essential because they determine whether the proposed method is only a smoother or can actually improve the reliability of the angle prior under dynamic channel changes.

Training, validation, and testing trajectories are separated at the trajectory level. This prevents the model from

seeing neighboring frames of the same motion pattern during training and testing. The validation set is used to select the best checkpoint, which reduces the risk that the Transformer or AMPParamNet only memorizes a small group of simulated motions.

All reported main experiments are obtained with a fixed random seed to make the complete LS-Root-MUSIC-TF-KF-AWP-AMP chain exactly reproducible. The angle-tracking RMSE values in the main tables are measured on one representative test trajectory after the Transformer window warm-up, not on a full multi-seed statistical sweep. The sparse-recovery NMSE is averaged over the specified SNR and distance grids for that same representative setting.

The tracking curves are therefore interpreted as representative trajectory-level behavior rather than isolated frame-level samples. A method that performs well on this setting must maintain smooth tracking during continuous motion, respond quickly after true dominant-path switching, and avoid being dragged by isolated Root-MUSIC outliers.

The main simulation parameters are listed in Table 3.

Table 3. Main simulation parameters

Parameter	Value
Train / validation / test trajectories	240 / 48 / 16
Frames per test trajectory	120
Sub-6G / mmWave array size	16 / 64
Sub-6G pilot length	48
Angle-tracking training SNR	-15 to 20 dB
mmWave sparse-recovery SNR	-20 to 15 dB
Distance set	5, 20, 50, 100, 200, 300 m
Observation gate threshold	8 degrees
Transformer window / layers / heads	10 / 2 / 4
AMP iterations / measurements	12 / 192
Threshold coefficient range / fixed damping	[0.005, 0.5] / 0.55

5.4 Evaluation Metrics and Reproducibility

The NMSE is defined as

$$NMSE_{dB} = 10\log_{10}\frac{\left\|H_{mm}^{est} - H_{mm}\right\|_F^2}{\left\|H_{mm}\right\|_F^2}. \tag{17}$$

For each channel sample, NMSE is computed in the linear domain as the squared channel-estimation error normalized by the squared true-channel norm. For each SNR or distance point, the linear NMSE values are averaged first and then converted to dB using 10log10(.). The average NMSE reported in Table 4 is the arithmetic average of the resulting SNR-wise dB values. Lower NMSE in dB indicates better estimation. AoA RMSE and AoD RMSE are computed separately in degrees, and the reported average angle RMSE is the arithmetic mean of the two.

6 Results and Analysis

6.1 Angle Tracking Performance

Fig. 3 shows that Root-MUSIC responds quickly in many frames but contains obvious outliers. Conventional KF and Deep-KF are smoother but lag near dominant-path switching. The proposed TF-KF suppresses abnormal observations through the gate and compensates prediction bias through the residual branch, therefore staying closer to the true path

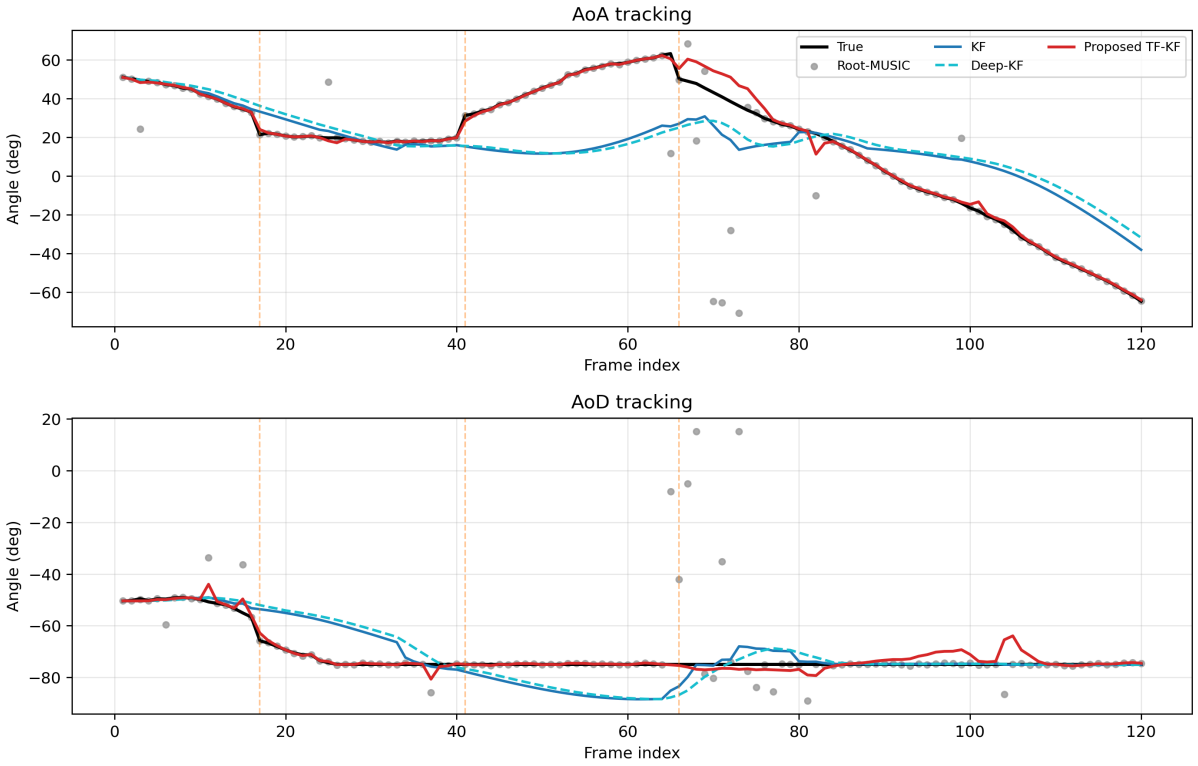


Fig. 3. AoA/AoD tracking under dynamic path switching

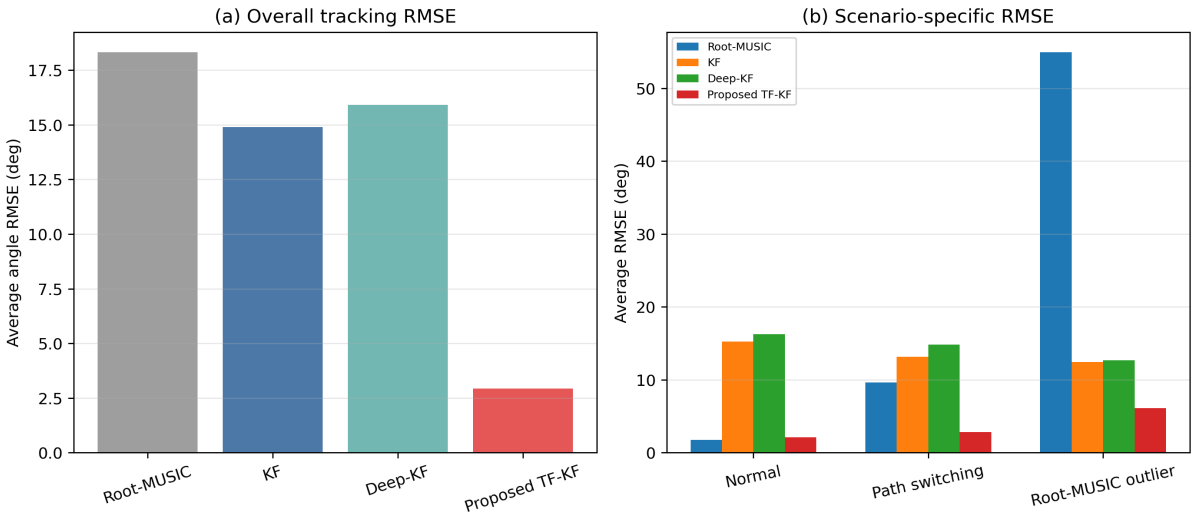


Fig. 4. Angle-tracking error and TF-KF ablation

Fig. 4 shows that, in the representative valid-frame tracking case, the average angle RMSE of TF-KF is 2.937 degrees, lower than Root-MUSIC at 18.320 degrees, conventional KF at 14.903 degrees, and Deep-KF at 15.909 degrees. In the ablation study, Gate-only can become worse than KF because it can only interpolate between two biased sources and cannot actively correct motion-model mismatch. The complete TF-KF benefits from both observation-confidence selection and bounded residual compensation.

6.2 Sparse Channel Estimation NMSE

Fig. 5 indicates that the TF-KF angular prior clearly improves sparse recovery when combined with an appropriate threshold range. The fixed small-threshold setting lambda = 0.08 is weaker, whereas fixed lambda = 0.5 and AMPParamNet produce almost identical NMSE. This means that the present sparse-recovery gain should be attributed mainly to reliable TF-KF angular priors and a sufficiently strong shrinkage coefficient, not to large per-iteration threshold fluctuations.

The tracking results should be interpreted together with the Root-MUSIC scatter points. In frames where Root-MUSIC is close to the true path, a high gate enables fast response and prevents KF lag. In frames with isolated abnormal observations, the gate should decrease so that the fused result remains close to the predicted trajectory. The residual term is expected to remain moderate in normal frames and become more active only when the prediction model cannot explain the current motion.

The ablation results reveal why both components are necessary. Gate-only may become worse than KF because it can only interpolate between two imperfect sources. Residual-only improves model flexibility but may still be affected by unreliable observations. The complete TF-KF uses the gate to control observation reliability and the residual to compensate prediction mismatch, which explains the lower overall angle RMSE.

For sparse channel recovery, a lower angle RMSE does not always produce a proportional NMSE improvement. NMSE also depends on angular-grid resolution, path-gain distribution, sensing-matrix coherence, threshold selection, and noise realization. When all prior-aided methods already place high weight near the correct angular support, their NMSE curves can become close. The advantage of TF-KF is more visible when Root-MUSIC outliers or path switching would otherwise misplace the angular prior.

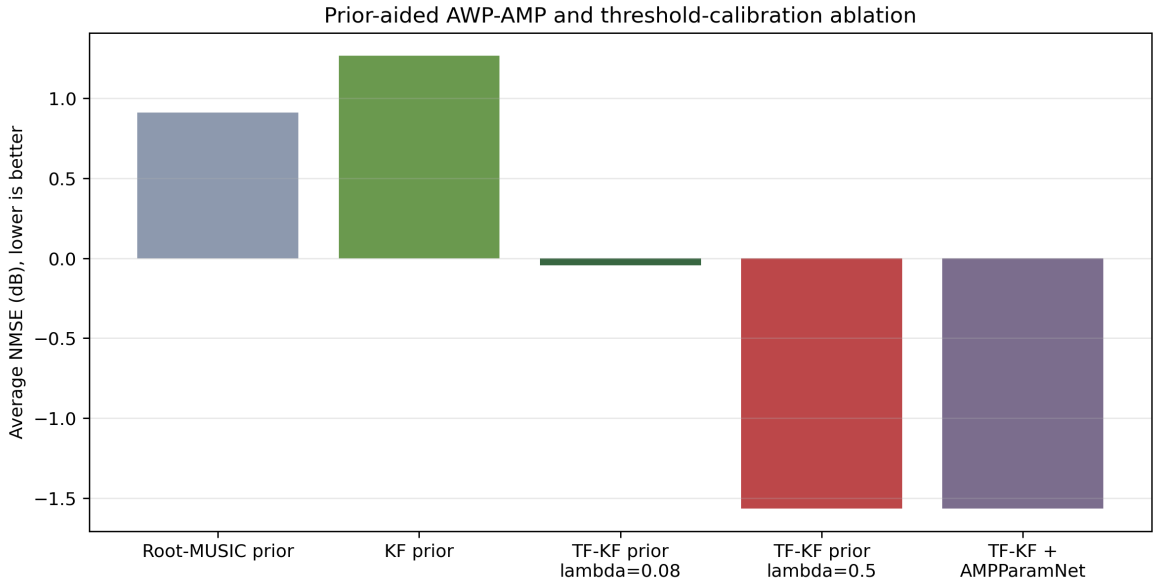


Fig. 5. NMSE comparison of prior-aided AWP-AMP

6.3 Training Diagnostics

Fig. 6 reports the training behavior of the two learning modules. The TF-KF objective includes fused-angle loss, gate supervision, jump supervision, residual supervision, and regularization. The AMPParamNet objective is a compressed NMSE-related diagnostic loss. Since the learned threshold stays close to lambda = 0.5, the AMPParamNet curve is interpreted as threshold-calibration behavior rather than evidence of a large adaptive threshold gain. The best validation model is used for testing, rather than the last epoch.

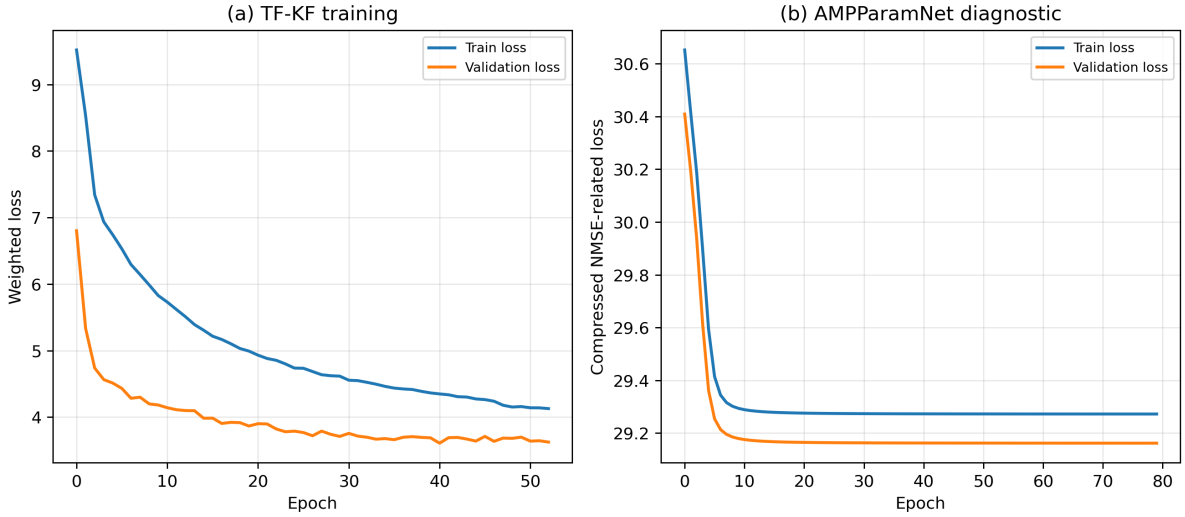


Fig. 6. Training and validation losses of TF-KF and AMPParamNet

6.4 Main Results, Scenario Analysis, and Ablation

Table 4 summarizes the representative valid-frame tracking results and AMP stability diagnostics from the same final experiment. The residual monotonic ratio is about 0.92 and the average oscillation count is low, indicating stable behavior under bounded thresholds and fixed damping. Since the average iteration count equals the maximum iteration budget of 12, it should be interpreted as fixed-budget termination rather than proof of earlier convergence. The no-prior fixed AMP baseline produces extremely large NMSE in this setting and is therefore interpreted as numerical divergence or severe instability, not as a meaningful ordinary average.

Table 4. Main representative results and convergence diagnostics

Metric	Value
Root-MUSIC representative angle RMSE	18.3197 deg
Conventional KF representative angle RMSE	14.9030 deg
Deep-KF representative angle RMSE	15.9094 deg
TF-KF representative angle RMSE	2.9374 deg
Fixed AMP without angular prior	Unstable; raw mean 170.2389 dB
TF-KF prior AWP-AMP, fixed lambda = 0.08	-0.0445 dB
TF-KF prior AWP-AMP, fixed lambda = 0.5	-1.5647 dB
TF-KF + AMPParamNet threshold calibration	-1.5646 dB
Threshold coefficient mean / min / max	0.4999 / 0.4965 / 0.5000
Fixed damping beta	0.55
Residual monotonic ratio / oscillation count	0.9208 / 0.1625
Average iteration count	12.0
Angle RMSE reporting scope	One representative fixed-seed test trajectory
Fixed lambda = 0.5 no-retraining seed check	-1.2069 +/- 0.0597 dB over five seeds

Table 5 shows scenario-specific behavior in the representative valid-frame tracking case. Root-MUSIC is already very accurate in normal continuous frames, so TF-KF is not claimed to be best in every individual scenario. The main advantage of TF-KF appears in the overall result, path-switching frames, and Root-MUSIC outlier frames, where conventional KF or Deep-KF either lag behind the new dominant path or are affected by abnormal observations. This is the most relevant behavior for constructing a robust angular prior for subsequent sparse recovery.

Table 5. Scenario-specific average AoA/AoD RMSE on representative valid frames (deg)

Scenario	Frames	Root-MUSIC	KF	Deep-KF	TF-KF
Representative all valid frames	111	18.3197	14.9030	15.9094	2.9374
Normal continuous	96	1.7680	15.2249	16.2780	2.1293
Path switching	3	9.6394	13.1723	14.8278	2.8159
Root-MUSIC outlier	12	54.9465	12.4553	12.6971	6.0821

Table 6 separates the effect of angular-prior source, confidence-controlled prior width, and threshold calibration. The fixed lambda = 0.5 row is included as a strong baseline. Its result is almost the same as AMPParamNet, showing that the threshold network mainly learns a high-threshold calibration strategy in the current experiment. A separate no-retraining five-seed sanity check using the archived TF-KF angle outputs gives -1.2069 +/- 0.0597 dB for ATKF prior with fixed lambda = 0.5, compared with 0.2884 +/- 0.1281 dB for ATKF prior with fixed lambda = 0.08. This seed check supports the threshold choice, but it is not a replacement for a full multi-seed retraining study.

Table 6. Confidence-transfer and threshold-calibration ablation

Variant	Angular prior	Prior width	Threshold	Average NMSE (dB)
Fixed AMP without prior	None	None	Fixed	Unstable; raw mean 170.2389
Root-MUSIC prior AWP-AMP	Root-MUSIC	Fixed prior rule	Fixed lambda = 0.08	0.9105
KF prior AWP-AMP	KF	Fixed prior rule	Fixed lambda = 0.08	1.2644
TF-KF prior AWP-AMP	TF-KF	Gate-controlled	Fixed lambda = 0.08	-0.0445
TF-KF prior AWP-AMP	TF-KF	Gate-controlled	Fixed lambda = 0.5	-1.5647
TF-KF + threshold calibration	TF-KF	Gate-controlled	AMPParamNet	-1.5646
Adaptive threshold only	None	None	AMPParamNet	Not isolated in final log

The loss curves are used primarily as training diagnostics rather than direct performance metrics. For TF-KF, the objective combines fused-angle error, gate supervision, jump supervision, residual supervision, and regularization, so its scale is not directly comparable with angle RMSE. For AMPParamNet, the loss is compressed to reduce domination by very large early-stage NMSE values. The validation-best checkpoint is used for testing to reduce overfitting.

In addition to average RMSE and NMSE, scenario-specific behavior is important. Normal frames test whether the method avoids unnecessary corrections. Path-switching frames test whether the method can follow a real change quickly. Root-MUSIC outlier frames test whether the method can reject unreliable observations. These scenario-level diagnostics are more informative than a single averaged metric when evaluating dynamic channel estimation.

6.5 Interpretation of Tracking and NMSE Results

The improvement in angle tracking should be understood as an improvement in prior reliability. Root-MUSIC can be very accurate in many frames, but its errors are often impulsive. KF reduces impulsive noise but may create long tracking lag after path switching. Deep-KF improves the smoothness of the trajectory but does not explicitly expose observation confidence. TF-KF reduces the most harmful errors by combining memory, gate control, and bounded residual compensation.

The NMSE results depend on whether the sparse recovery problem is prior-limited, threshold-limited, or noise-limited. When the angular prior is the dominant bottleneck, a more reliable TF-KF prior leads to a clear NMSE gain. When the threshold is too small, as in the fixed lambda = 0.08 setting, noise-like angular coefficients are insufficiently suppressed. When lambda is increased to 0.5, performance becomes close to AMPParamNet, indicating that the learned threshold mostly selects a strong shrinkage regime. This does not weaken the TF-KF contribution; rather, it clarifies that the main novelty lies in confidence-aware angular prior transfer.

The training loss curves should also be interpreted carefully. Their numerical scale is determined by the weighted objective, not by raw RMSE. A smooth decrease in training loss indicates that the network is learning the intended regression and confidence targets, while the validation-best checkpoint avoids using an overfitted

final epoch. The reported testing metrics are therefore based on generalization performance rather than the lowest training error.

7 Complexity and Runtime Overhead

The additional overhead of the proposed method mainly comes from TF-KF and AMPParamNet, but neither replaces the original signal-processing pipeline. Root-MUSIC is dominated by covariance eigendecomposition. TF-KF operates on a short window with L = 10, so its inference cost is much smaller than large angular-grid sparse recovery. AWP-AMP is dominated by repeated multiplications with the equivalent sensing dictionary and its Hermitian transpose. AMPParamNet is a lightweight MLP and is negligible compared with AMP matrix-vector products.

The dominant operations of each module are summarized in Table 7.

Table 7. Main computational complexity

Module	Dominant operation	Complexity order	Comment
LS estimation	Matrix inverse/multiplication	Pilot and array dependent	Front-end estimation
Root-MUSIC	Covariance EVD	$O(Nant^2K + Nant^3)$	Nant is the array size
KF	Low-dimensional recursion	$O(nx^3)$	nx = 4, negligible
TF-KF	Short-window attention	$O(L^2d + Ldf)$	Lightweight inference
AWP-AMP	Dictionary matrix-vector products	O(T M Ng)	Main sparse-recovery cost
AMPParamNet	Small MLP forward pass	Much smaller than AMP	Threshold output only

The computational overhead of the proposed learning modules is limited because the networks do not process the full channel matrix. TF-KF operates on a short sequence of low-dimensional angle features, and AMPParamNet is a small MLP evaluated once per AMP iteration. The dominant operations remain Root-MUSIC eigendecomposition and AMP dictionary matrix-vector products.

In practical deployment, the learning modules can be executed after the Sub-6G pilot processing and before mmWave sparse recovery. Since the angle state dimension is fixed and the Transformer window is short, TF-KF inference can be batched across frames. AMPParamNet adds only a few fully connected operations per iteration, so its cost is negligible compared with multiplying by the sensing dictionary and its Hermitian transpose.

Since strict timing was not performed on a fixed hardware platform, this paper reports complexity and iteration diagnostics rather than absolute runtime. In the formal experiment, AMP uses the fixed iteration budget of 12 rounds on average, with residual monotonic ratio 0.90 and average oscillation count 0.24. Therefore, the iteration statistic should be interpreted as a stability and budget indicator rather than a strict convergence-time measurement.

7.1 Runtime Considerations

The proposed framework is suitable for offline training and online inference. During online operation, LS and Root-MUSIC process the Sub-6G pilot observations, TF-KF performs a lightweight forward pass on low-dimensional temporal features, and AWP-AMP performs sparse recovery on the mmWave observation. The neural modules do not require retraining for each frame.

The dominant runtime cost remains the matrix-vector products in AWP-AMP and the subspace decomposition in Root-MUSIC. TF-KF and AMPParamNet contribute only small additional overhead because their input dimensions are much lower than the full channel dimension. This makes the architecture attractive for systems where a modest learning component can be tolerated but a fully neural channel estimator is undesirable.

Memory consumption is also moderate. The tracker stores only a short window of angle features, covariance entries, and innovations. AMP stores the current sparse estimate, residual, dictionary products, and weight map. These quantities are already needed in conventional sparse recovery, so the additional memory caused by confidence-aware adaptation is limited.

8 Discussion

The results show that the gated residual structure is not merely a smoother. Root-MUSIC responds fast and can be the most accurate method in clean normal frames, but its errors are impulsive under outliers. Conventional KF is stable but may lag; Deep-KF can learn some nonlinear correction but lacks explicit observation-confidence modeling. TF-KF improves the reliability of the angular prior by combining memory, gate control, bounded residual compensation, and confidence-scaled KF updating.

The inferior Gate-only result is also informative. Gate-only cannot create a new correction direction when both Root-MUSIC and KF prediction are biased, and it may over-trust abnormal observations near path switching. Residual-only is slightly better than KF but lacks outlier rejection. The full TF-KF combines both mechanisms and therefore performs best.

A key practical implication is that the proposed framework does not require the learning model to perfectly detect every path switch. Even when jump sensitivity is imperfect, the gate and residual can still improve the prior by reducing large Root-MUSIC errors and limiting KF lag. For the subsequent sparse recovery stage, the most important requirement is not a perfectly labeled jump probability, but a reliable angular region and a confidence-controlled prior width.

The method also avoids using true angles during online inference. True angles are needed only for supervised training and evaluation in simulation. During testing, the system uses LS-estimated Sub-6G channels, Root-MUSIC observations, KF states, and learned confidence outputs. This separation is important for making the simulation closer to a realistic signal-processing pipeline.

Limitations remain. The channel is generated by statistical simulation and should be further validated using ray-tracing or measured data. The main angle-tracking RMSE is obtained from a representative fixed-seed trajectory, while the additional five-seed test only checks the fixed-threshold AWP-AMP behavior without retraining TF-KF or AMPParamNet. Future work should include repeated full-chain multi-seed trials and report mean plus standard deviation for both angle tracking and channel recovery. AMPParamNet currently learns the threshold while damping is fixed for stability, but in the final experiment its output is close to the fixed upper-bound baseline. The current model is narrowband and two-dimensional, leaving wideband beam squint, planar arrays, and multi-user interference for future work.

8.1 Practical Limitations and Extension Directions

The present work is simulation-based, and the channel generator approximates dynamic mmWave behavior through path switching, blockage, distance-dependent path loss, angular spread, NLoS loss, shadowing, and cross-band angular mismatch. Although these factors make the simulation more realistic than a purely smooth trajectory, measured data or ray-tracing data would still be valuable for validating cross-band geometric consistency. A wider cross-band mismatch sweep is also useful for quantifying how much Sub-6G geometric information can be trusted under different propagation environments.

A dedicated cross-band angular mismatch sweep with standard deviations of 0, 2, 5, 8, and 10 degrees is recommended as the next reproducibility experiment. The present final log uses a fixed mismatch standard deviation of 1.5 degrees; therefore, this paper does not report multi-point mismatch results that were not actually generated.

Another limitation is that the current architecture considers a dominant angular prior. In rich scattering or multi-user scenarios, several angular clusters may need to be tracked simultaneously. The gate-residual idea can be extended to multiple candidate paths by allowing the Transformer to output a confidence vector or a small set of angular hypotheses. The AWP-AMP weight map could then be constructed as a mixture of Gaussian priors rather than a single peak.

Finally, the present AMPParamNet is trained to adjust the threshold coefficient while the damping factor is fixed. This choice improves stability and keeps the ablation study clear. Since the final learned threshold is close to lambda = 0.5, future work should either design a more diverse training set where adaptive threshold variation is necessary, or simplify the recovery stage to a fixed strong-threshold AWP-AMP when the same behavior is observed. Such extensions should preserve the current interpretability principle: learned modules should regulate model-based steps rather than obscure the signal-processing structure.

9 Conclusion

This paper presented a Sub-6G aided sparse mmWave channel estimation method for complex dynamic scenarios. The gated residual TF-KF learns Root-MUSIC observation confidence and bounded prediction correction, while the confidence-aided AWP-AMP uses the resulting angle prior to control angular weights and threshold calibration. Under the reported representative valid-frame tracking case, TF-KF reduces the average angle RMSE from 14.903 degrees for conventional KF to 2.937 degrees. In sparse recovery, TF-KF aided AWP-AMP with fixed lambda = 0.5 achieves -1.5647 dB average NMSE, while AMPParamNet obtains -1.5646 dB, showing that the current threshold network behaves similarly to a strong fixed-threshold baseline. The final conclusion is therefore that reliable confidence-aware angular priors are the dominant source of gain, and threshold calibration should be interpreted conservatively. Future work will validate the method under measured or ray-tracing channels, full-chain multi-seed experiments, wider cross-band mismatch, pilot-overhead sweeps, and multi-path prior tracking.

References

[1] T.S. Rappaport, R.W. Heath Jr., R.C. Daniels, J.N. Murdock, Millimeter Wave Wireless Communications, Prentice Hall, Upper Saddle River, NJ, 2014.

[2] A. Alkhateeb, O. El Ayach, G. Leus, R.W. Heath Jr., Channel estimation and hybrid precoding for millimeter wave cellular systems, IEEE Journal of Selected Topics in Signal Processing 8(5)(2014) 831-846. https://doi.org/10.1109/JSTSP.2014.2334278

[3] D.L. Donoho, A. Maleki, A. Montanari, Message passing algorithms for compressed sensing: I. motivation and construction, in: Proc. IEEE Information Theory Workshop, 2010. https://doi.org/10.1109/ITWKSPS.2010.5503193

[4] F. Pasic, M. Hofer, M. Mussbah, S. Caban, S. Schwarz, T. Zemen, C.F. Mecklenbraeuker, Channel estimation for mmWave MIMO using sub-6 GHz out-of-band information, in: Proc. International Conference on Smart Applications, Communications and Networking, 2024, pp. 1-6. https://doi.org/10.1109/SmartNets61466.2024.10577648

[5] F. Pasic, L. Eller, S. Schwarz, M. Rupp, C.F. Mecklenbraeuker, Deep learning-based mmWave MIMO channel estimation using sub-6 GHz channel information: CNN and UNet approaches, in: Proc. IEEE Conference on Computer Communications Workshops (INFOCOM Workshops), 2025. https://doi.org/10.1109/INFOCOMWKSHPS65812.2025.11152870

[6] S. Ju, O. Kanhere, Y. Xing, T.S. Rappaport, A millimeter-wave channel simulator NYUSIM with spatial consistency and human blockage, in: Proc. IEEE Global Communications Conference, 2019. https://doi.org/10.1109/GLOBECOM38437.2019.9013273

[7] S. Rangan, P. Schniter, A.K. Fletcher, Vector approximate message passing, IEEE Transactions on Information Theory 65(10)(2019) 6664-6684. https://doi.org/10.1109/TIT.2019.2916359

[8] H. He, C.-K. Wen, S. Jin, G.Y. Li, Deep learning-based channel estimation for beamspace mmWave massive MIMO systems, IEEE Wireless Communications Letters 7(5)(2018) 852-855. https://doi.org/10.1109/LWC.2018.2832128

[9] A.M. Elbir, K.V. Mishra, M.R.B. Shankar, B. Ottersten, A family of deep learning architectures for channel estimation and hybrid beamforming in multi-carrier mm-wave massive MIMO, IEEE Transactions on Cognitive Communications and Networking 8(2)(2022) 642-656. https://doi.org/10.1109/TCCN.2021.3132609

[10] O.E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, R.W. Heath Jr., Spatially sparse precoding in millimeter wave MIMO systems, IEEE Transactions on Wireless Communications 13(3)(2014) 1499-1513. https://doi.org/10.1109/TWC.2014.011714.130846

[11] A.M. Sayeed, Deconstructing multiantenna fading channels, IEEE Transactions on Signal Processing 50(10)(2002) 2563-2579. https://doi.org/10.1109/TSP.2002.803324

[12] R.W. Heath Jr., N. Gonzalez-Prelcic, S. Rangan, W. Roh, A.M. Sayeed, An overview of signal processing techniques for millimeter wave MIMO systems, IEEE Journal of Selected Topics in Signal Processing 10(3)(2016) 436-453. https://doi.org/10.1109/JSTSP.2016.2523924

[13] J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, K. Venugopal, R.W. Heath Jr., Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems, IEEE Transactions on Wireless Communications 17(5)(2018) 2946-2960.

https://doi.org/10.1109/TWC.2018.2804943
[14] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Transactions on Antennas and Propagation 34(3)(1986) 276-280.
https://doi.org/10.1109/TAP.1986.1143830
[15] R.E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering 82(1)(1960) 35-45.
https://doi.org/10.1115/1.3662552
[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proc. Advances in Neural Information Processing Systems, 2017.
https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
[17] S. Rangan, Generalized approximate message passing for estimation with random linear mixing, in: Proc. IEEE International Symposium on Information Theory, 2011.
https://doi.org/10.1109/ISIT.2011.6033942
[18] 3GPP, Study on channel model for frequencies from 0.5 to 100 GHz, 3GPP Technical Report TR 38.901. <https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/>, (accessed 25.10.2026).