Anchored Beam-Convolutional Prediction for Delayed Beam Selection

Hao Hoang Tran[a], Thanh Thien-An Dang[a], Viet Anh-Huy Le[a], Minh Dinh Nguyen[a], Kiet Nguyen-Tuan Tran[a] and Sungrae Cho[a,*]

[a]Department of Computer Science and Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, 06974, South Korea

ARTICLE INFO

Keywords:
Beam prediction
Convolutional networks
Millimeter wave
Power measurements

ABSTRACT

Measurement and processing delays can misalign narrow beams before transmission, motivating prediction from recent noisy power sweeps. We propose anchored beam-convolutional prediction (ABCP), which combines a causal fitted present-power profile with a shallow convolutional correction along the beam axis. Tests on disjoint simulated routes show that ABCP uses substantially fewer learned parameters than the evaluated dense and recurrent baselines, while primary gain comparisons remain inconclusive. Ablations support ordered history and local filtering relative to restricted controls, and an equivalent NumPy implementation achieves lower processor latency than the measured PyTorch multilayer perceptron. Additional route partitions and external-source spatial evaluations characterize distribution sensitivity. The findings cover simulated full-sweep observations with noise and implementation limitations.

1. Introduction

Mobility and channel variation can misalign narrow beams during measurement and processing delays, while sounding incurs additional overhead [1, 2]. For example, a controller retaining eight full sweeps sampled every 10 ms may need to select a beam for transmission 40 or 80 ms ahead. Camera-based vehicular prediction uses additional sensing [3], whereas we predict future power from noisy sweeps without positioning, phase-coherent channels or online clean labels.

Related work uses phase-less received signal strength (RSS) for long short-term memory (LSTM) beam prediction [4], while a convolutional neural network (CNN) infers beam pairs from reference signal received power (RSRP) subsets [5] (Table 1). The residual variant of CSI-PNet predicts complex channel state information (CSI) using small convolutional kernels with circular padding and temporal input channels [6]. Related approaches include R2-TBP, which combines a convolutional residual LSTM (CRLSTM) with attention to predict future RSRP [7], and mobility-aware deep beam tracking [8]. Beyond beam prediction, MambaX uses state predictive control for image super-resolution [9], while SpectralGPT learns ordered spatial and spectral representations [10]. Their image objectives differ from power-sweep forecasting.

We ask how a compact power-only predictor can preserve ordered history and neighboring-beam structure from full sweeps. ABCP combines a causal fitted present-power anchor with a shallow shared beam-axis correction, using established convolution, temporal channels, residual learning and circular padding. Controlled comparisons examine history, locality, anchoring and padding, alongside route partitions, external transfer and cost. Sections 2–4 describe the model and experiments, followed by results and limitations.

Table 1
Related predictors use different observations and prediction targets. The entries describe architectures without ranking their performance, and † denotes an abstract-based summary.

Method	Input	Target	Core model
Domae et al. [4]	Phase-less RSS	Future beams	LSTM
Jalali et al. [5]	RSRP subset	Beam pair	CNN
CSI-PNet [6]	Complex CSI	Future CSI	Circular CNN
R2-TBP [7]†	Beam history	Future RSRP	CRLSTM
ABCP (ours)	Power sweeps	Future power	Anchor + CNN

2. Observation and decision model

Figure 1(a) illustrates delayed beam selection between a base station (BS) and user equipment (UE). A half-wavelength uniform linear array (ULA) uses $N = B = 32$ elements and discrete Fourier transform (DFT) beams [11]

$$[\boldsymbol{w}_b]_n = N^{-1/2}\exp(-j\pi n u_b), \qquad u_b = -1+2b/B, \quad (1)$$

where $n, b \in \{0, \dots, B-1\}$ index elements and beams. Let $v_{t,b,f}$ be the beamformed channel voltage scaled by the square root of the transmit-to-noise power ratio at tone f. Under additive Gaussian receiver noise [11], define clean and measured powers as

$$g_{t,b} = F^{-1}\sum_{f=1}^{F}|v_{t,b,f}|^2, \quad (2)$$

$$y_{t,b} = F^{-1}\sum_{f=1}^{F}|v_{t,b,f}+\epsilon_{t,b,f}|^2, \quad (3)$$

with nominal noise $\epsilon_{t,b,f} \sim \mathcal{CN}(0,1)$. During online inference, predictors observe only noisy sweeps $\boldsymbol{y}_{0:t}$, while clean

*Corresponding author.
hhtran@uclab.re.kr (H.H. Tran); attdang@uclab.re.kr (T.T. Dang); lvahuy@uclab.re.kr (V.A. Le); mdnguyen@uclab.re.kr (M.D. Nguyen); knttran@uclab.re.kr (K.N. Tran); srcho@cau.ac.kr (S. Cho)
ORCID(s): 0009-0006-8180-9486 (H.H. Tran); 0009-0008-9595-9527 (T.T. Dang); 0009-0008-6134-046X (V.A. Le); 0000-0003-1879-688X (S. Cho)

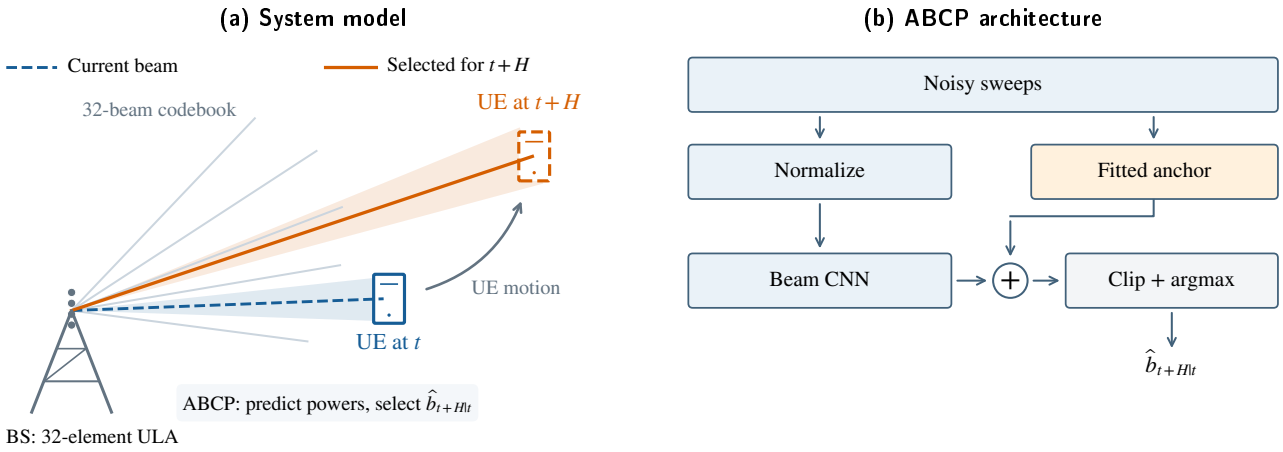


Figure 1: ABCP corrects a fitted present-power profile for delayed beam selection. Panel (a) selects for $t+H$ from noisy powers through t. Positions indicate motion rather than predictor inputs, and beam directions depict neither propagation paths nor alignment guarantees. Panel (b) combines a CNN correction from eight normalized sweeps with a clipped, normalized anchor fitted to four raw sweeps. The sum is clipped before beam selection.

powers $g_{t,b}$ serve as offline training targets and evaluation ground truth.

Given the predicted profile, the system selects

$$\hat{b}_{t+H|t} = \arg\max_b \hat{g}_{t+H|t,b}, \tag{4}$$

with ties resolved at the lowest index. With 10 ms sounding intervals, horizons $H \in \{4, 8\}$ give 40 ms and 80 ms lookahead delays. We define normalized beam gain and use the Gaussian-channel rate expression [11] for gross spectral efficiency,

$$G_{t,H} = g_{t+H,\hat{b}} / \max_b g_{t+H,b}, \tag{5}$$

$$R_{t,H} = F^{-1} \sum_f \log_2(1 + |v_{t+H,\hat{b},f}|^2). \tag{6}$$

Gain outage is defined as $G_{t,H} < 0.5$, rate outage as $R_{t,H} <$ 1 bit/s/Hz, and hit rate denotes the empirical frequency of selecting a strongest-power beam.

3. Anchored beam-convolutional prediction (ABCP)

ABCP combines the anchor and learned correction shown in Fig. 1(b).

3.1. Causal anchor and history

For nonnegative beam powers $\boldsymbol{z} \in \mathbb{R}^B$, define

$$\mathcal{N}(\boldsymbol{z}) = \frac{B\boldsymbol{z}}{\max(\mathbf{1}^\mathsf{T}\boldsymbol{z}, 10^{-12})}. \tag{7}$$

With $\mathbf{1}$ denoting the all-ones vector, $\mathcal{N}$ rescales the beam powers while preserving their ratios, and the denominator floor prevents division by zero. The input $X_t \in \mathbb{R}^{8\times B}$ stacks $\mathcal{N}(\boldsymbol{y}_{t-7}), \ldots, \mathcal{N}(\boldsymbol{y}_t)$ chronologically to distinguish temporal trends, repeating the earliest available sweep at startup.

For each beam, a least-squares line [12] fitted to the four latest observations and evaluated at t gives the clipped, normalized anchor

$$\boldsymbol{a}_t = \mathcal{N}\Big(\big[-0.2\boldsymbol{y}_{t-3} + 0.1\boldsymbol{y}_{t-2} + 0.4\boldsymbol{y}_{t-1} + 0.7\boldsymbol{y}_t\big]_+\Big), \tag{8}$$

while earlier steps fit all available samples. The fixed weights follow from equally spaced sampling, sum to one and recover an affine history's present endpoint before clipping. The negative oldest weight distinguishes the fit from a moving average, and zero CNN correction recovers the anchor.

3.2. Local learned correction

The correction uses standard CNN convolution [13] with circular beam indexing,

$$[\mathcal{C}_K(X)]_{o,b} = \beta_o + \sum_i \sum_{s=-r}^{r} K_{o,i,s} X_{i,(b+s) \bmod B}. \tag{9}$$

Two width-three, 16-channel layers with rectified linear unit (ReLU) σ [13] precede a width-one linear output, yielding

$$Z_1 = \sigma(\mathcal{C}_{K_1}(X_t)), \quad Z_2 = \sigma(\mathcal{C}_{K_2}(Z_1)), \tag{10}$$

$$\boldsymbol{d}_{t,H} = \mathcal{C}_{K_3}(Z_2), \tag{11}$$

$$\hat{\boldsymbol{g}}_{t+H|t} = [\boldsymbol{a}_t + \boldsymbol{d}_{t,H}]_+. \tag{12}$$

Training at each horizon minimizes mean squared error (MSE) against $\mathcal{N}(\boldsymbol{g}_{t+H})$, without confidence gating, angular decomposition or online adaptation. The width-three layers cover $1+2+2 = 5$ normalized bins across eight sweeps without recurrence, favoring local beam patterns while limiting direct interactions despite normalization's global coupling. The width-one head supplies a signed residual correction [14] for noisy fits or abrupt changes, while circular padding connects endpoint beams.

Normalization, anchoring, shared convolutions and clipping commute with circular beam-index shifts P_s, giving

$$F_H(P_s\boldsymbol{y}_{0:t}) = P_s F_H(\boldsymbol{y}_{0:t}). \tag{13}$$

Table 2
Simulation and training parameters.

Setting	Value
Carrier / bandwidth / tones	28 GHz / 100 MHz / 16
Route / sampling / speed	320 samples / 100 Hz / 3, 10 m/s
BS / UE height, elements	10 / 1.5 m, isotropic, $\lambda/2$
Transmit power / receiver noise	20 / −87 dBm
Noise density / noise figure	−174 dBm/Hz / 7 dB
Train / validation / test	208 / 64 / 120 disjoint routes
Optimizer / loss / batch	AdamW / MSE / 1024
Epoch cap / patience / decay	30 / 6 / 0

SI equivalents are 0.1 W for transmit power, 2.00×10^{-12} W for receiver noise and 3.98×10^{-21} W/Hz for noise density, with noise values rounded.

Thus, in exact arithmetic a unique maximum shifts accordingly without requiring isotropic propagation.

4. Experimental setup

QuaDRiGa 2.8.1 [15] in Octave 9.4.0 generates 3GPP TR 38.901 urban-microcell line-of-sight (LOS) and non-line-of-sight (NLOS) channels with spherical-wave drifting (Table 2). Straight tracks retain native pathloss, shadowing and cluster normalization, with equal tone powers and no temporal interpolation. Predictors share observations and targets across four propagation/speed strata, with trajectory details in Supplementary Section S7. Test routes are balanced across propagation/speed strata and disjoint from development routes (Table 2), with model selection completed before testing. Stress tests add 32 routes at 6 m/s and paired observation variance scaling from $\sigma_0^2 = 1$ to 4 on the main test channels.

Learned baselines share development data and five seeds. The multilayer perceptron (MLP) and gated recurrent unit (GRU) use ABCP's anchor and clipping, with MSE fitting and checkpoint selection by validation gain. Supplementary Sections S3 and S7 give architectures and fitting protocols. Gain, gross spectral efficiency (SE) and outages are averaged over routes, strata and five fits at $t = 40, \ldots, 319 - H$. Primary 40 ms comparisons use 5000 paired bootstrap replicates that resample routes within strata and the five fits, with Bonferroni adjustment for the three primary comparisons [16]. Other comparisons use individual 95% confidence intervals without adjustment for multiple comparisons. These intervals reflect variation across test routes and training runs for the fixed training and validation split.

Five-fold cross-validation (CV) partitions 272 development routes by physical route, stratified by propagation/speed, with checkpoint and learning-rate selection on inner-validation routes. Original test and stress routes are excluded, with partition and fitting details in Supplementary Section S1. External transfer uses DeepMIMO static ray tracing in Seoul, New York and Boston [17], with 30 routes per city and all five unchanged QuaDRiGa fits. Horizons four/eight count spatial samples, with no test-label fitting (Supplementary Section S2).

5. Numerical results

5.1. Performance comparison

Table 3 reports performance at 40/80 ms, with bold values marking descriptive extrema and neural counts excluding the earlier gated predictor's fitted decision tree. At 40 ms, ABCP's estimated gain difference from MLP is +0.252 percentage points (pp), with a Bonferroni-adjusted interval from −1.379 to 1.632 pp. Against GRU, the difference is +0.394 pp, with limits −1.269 and 1.860 pp, and against no-anchor it is +0.321 pp, with limits −0.195 and 1.389 pp. Since all primary intervals include zero, neither superiority nor equivalence is established. At 80 ms, secondary differences of +0.740 and +1.091 pp also have intervals spanning zero.

At 40 and 80 ms, secondary ABCP-minus-MLP gross-SE differences are 0.00825 and 0.01537 bit/s/Hz. Because gross SE excludes full-sweep sounding overhead, these differences do not imply net throughput gains. Supplementary Sections S4–S5 provide profiles, distributions, additional intervals and rate-outage results.

5.2. Ablation analysis

Ablations remove the anchor, apply zero padding, use a *pointwise* variant with width-one kernels and 433 parameters, repeat the latest sweep across learned channels in the *current-repeat* variant, or use an anchored *tiny MLP* with 256–4–32 units and 1188 parameters. Current-repeat preserves anchor history, while pointwise filtering reduces capacity. Current-repeat reduces gain by 4.972/5.708 pp at 40/80 ms, with both secondary intervals excluding zero, as shown in Fig. 2(e,f). Pointwise losses of 0.764/1.170 pp have positive intervals, but reduced capacity confounds the spatial interpretation. The approximately parameter-matched tiny MLP loses 3.354/4.330 pp. These results support useful temporal and local structure relative to the tested controls.

The anchor's primary 40 ms interval is inconclusive. Its secondary 80 ms advantage is 1.938 pp ([0.558, 3.413]), with no all-zero no-anchor forecasts to indicate clipping collapse. Zero padding yields mean gain 0.8253/0.8215, with ABCP-minus-zero-padding intervals containing zero at both horizons. Thus, circular padding remains a pre-specified choice without a demonstrated gain benefit.

For the exploratory delay sweep, the 20–100 ms sweep in Fig. 2(a–d) reuses the 120 test routes and 40/80 ms checkpoints, with additional fits at other delays. Simultaneous gain intervals include zero at all nine delays, while gross-SE bands favor ABCP over GRU throughout and over MLP at 40–100 ms (Supplementary Section S6).

5.3. Partition stability and spatial transfer

ABCP has higher pooled route-CV gain than MLP and GRU, and Fig. 3(a,b) shows how gain varies across folds. These results describe partition sensitivity of the developed architecture and complement the original test intervals (Supplementary Section S1).

Across the external tests in Fig. 3(c–f), descriptive gains over the stronger anchor/persistence control are generally

Table 3
Prediction performance at 40/80 ms.

Method	Neural params	Normalized gain 40 ms	Normalized gain 80 ms	Gross SE (bit/s/Hz) 40 ms	Gross SE (bit/s/Hz) 80 ms	Gain outage (%) 40 ms	Gain outage (%) 80 ms
ABCP	**1201**	**0.8215**	**0.8192**	**4.2387**	**4.2346**	16.67	**16.87**
MLP	51424	0.8190	0.8118	4.2304	4.2192	16.69	17.46
GRU	51072	0.8176	0.8083	4.2284	4.2097	16.72	17.84
Physics residual	51489	0.8209	0.8136	4.2326	4.2163	**16.55**	17.35
Small MLP	11328	0.8145	0.8065	4.2273	4.2146	17.28	18.21
Anchor only	0	0.7701	0.7599	4.2039	4.1805	22.87	23.98
Persistence	0	0.7562	0.7463	4.1921	4.1702	24.69	25.78
Adaptive three	0	0.7628	0.7532	4.2013	4.1903	24.05	25.21
Earlier gated predictor	0	0.7678	0.7611	4.2037	4.1939	23.47	24.28

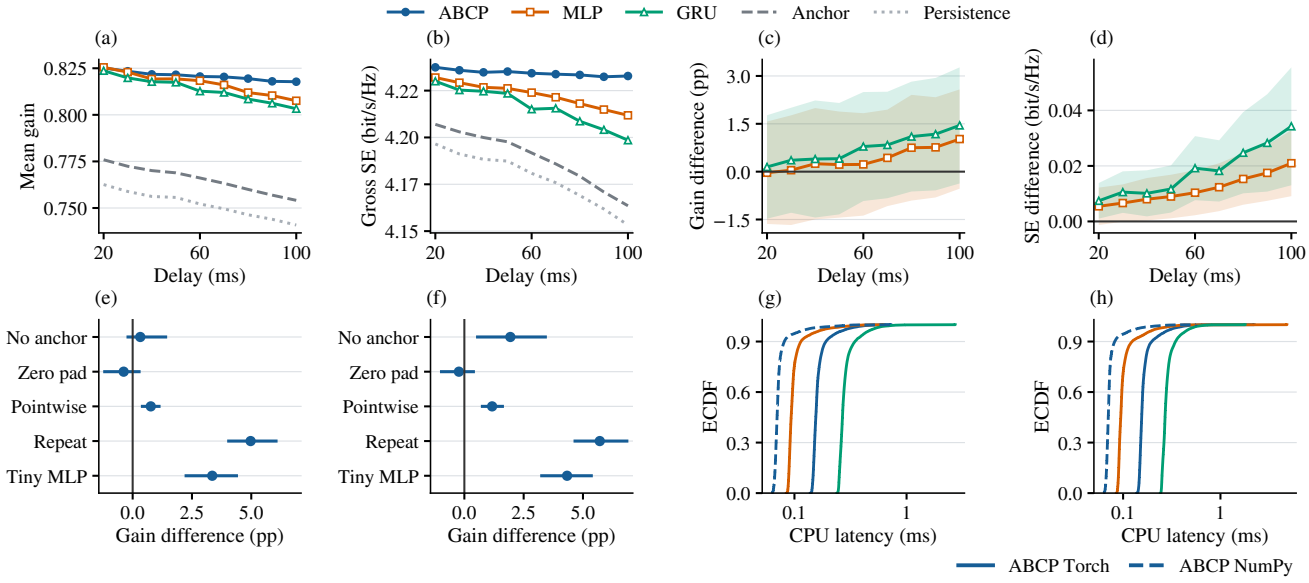


Figure 2: Delay sensitivity (a–d), ablation effects (e,f) and processor latency (g,h), with approximate joint 95% bands in (c,d) and pointwise 95% intervals in (e,f) except the adjusted 40 ms no-anchor comparison.

Table 4
Model complexity and processor latency at 40/80 ms.

Model	Params	MACs	Median (ms)	p95 (ms)
ABCP (T)	1201	37376	0.1565/0.1564	0.2373/0.2570
ABCP (N)	1201	37376	0.0705/0.0709	0.1019/0.1057
MLP (T)	51424	51200	0.0943/0.0954	0.1474/0.1735
GRU (T)	51072	378400	0.2702/0.2720	0.4254/0.4126

small. At four samples, Boston persistence scores 0.8335 versus ABCP's 0.8325 (Supplementary Table S2). These tests combine source, band, bandwidth, spacing and noise changes without isolating geographic effects or validating temporal performance.

5.4. Complexity and robustness

Shared filters allow ABCP to process more beams without increasing its learned parameter count.

For B beams, history length L, hidden-channel width C and kernel width k, the two hidden convolutions and width-one output require $B(kLC + kC^2 + C)$ multiply-accumulate (MAC) operations per prediction. History normalization adds $O(LB)$ work, while four-sweep anchoring, clipping and beam selection each require $O(B)$, giving

$$\mathcal{T}_{\mathrm{ABCP}} = O\big(B(kLC + kC^2 + C + L)\big). \tag{14}$$

ABCP has $kLC + kC^2 + 3C + 1$ learned parameters including biases, and $O(B(L+C))$ logical storage for history and activations. At fixed L, C, k, all three networks have $O(B)$ inference work and working storage, while ABCP's learned-weight storage is $O(1)$ in B, compared with $O(B)$ for MLP and GRU. Supplementary Section S3 gives detailed baseline formulas, memory counts and measurement conditions.

ABCP uses over 97% fewer parameters than MLP and GRU, but the timings in Table 4 and Fig. 2(g,h) vary with implementation. PyTorch ABCP is slower than PyTorch

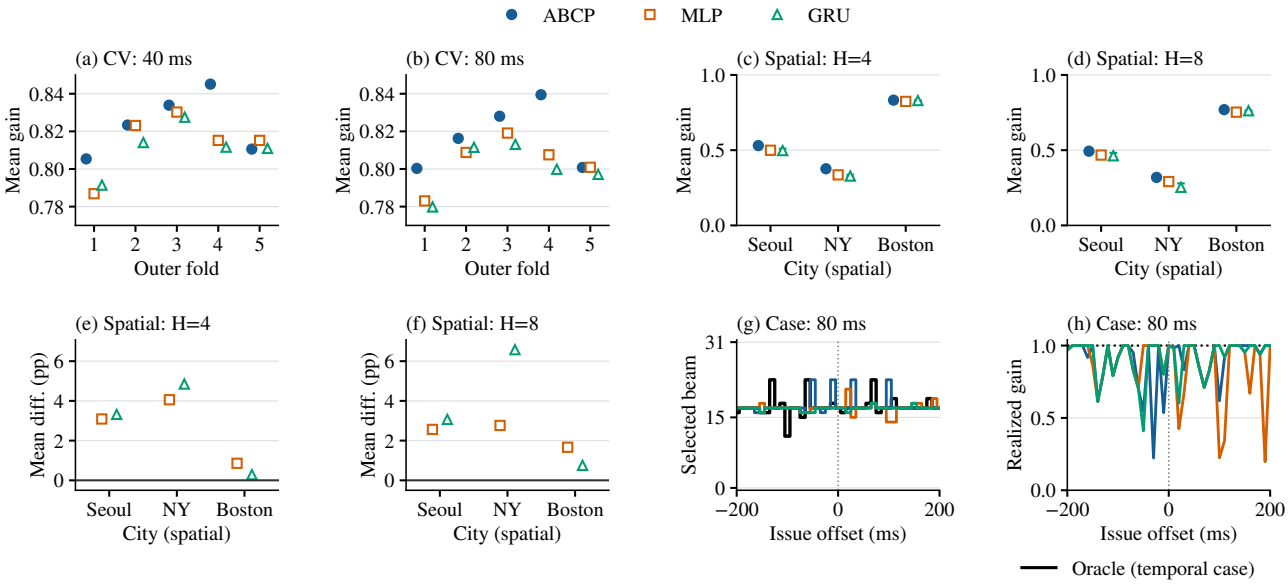


Figure 3: Route cross-validation (a,b), spatial transfer at four/eight samples (c–f) and a fixed-rule QuaDRiGa example at 80 ms (g,h), with external whiskers spanning fit ranges rather than confidence intervals.

MLP, whereas equivalent NumPy inference is approximately 25% faster than the measured PyTorch MLP. Parameter count therefore does not determine execution cost, and an equally optimized MLP was not evaluated. T/N denotes PyTorch/NumPy, and MAC counts exclude preprocessing, bias additions and nonlinearities. Central processing unit (CPU) latency includes feature construction, prediction and argmax but excludes radio acquisition and communication.

Under higher noise, ABCP's mean gain is below MLP by 0.736/0.254 pp, with both intervals spanning zero. On separate 6 m/s routes, secondary advantages of 2.908 pp ([1.385, 4.649]) and 3.676 pp ([2.299, 5.200]) support speed interpolation. Main-test means are below MLP at 3 m/s NLOS and higher at 10 m/s NLOS.

6. Limitations and future work

ABCP requires full sweeps and delayed beam decisions from finite history without recurrence. System-level assessment must therefore include sounding, communication and scheduling costs. Circular padding assumes endpoint adjacency without guaranteeing generalization, and residual learning guarantees neither easier optimization nor smaller errors. Straight tracks omit turns, cluster transitions and within-sweep motion. Route CV conditions on the developed architecture, and the delay sweep reuses the original test routes. Static ray tracing evaluates combined source/protocol transfer, including Seoul's different frequency, without measured temporal validation or broad geographic coverage.

Future work should first evaluate ABCP on measured temporal channels with repeated trajectories, turns and blockage events. Such data would allow prediction errors to be related to observed changes in beam power over the actual decision delay. Training and evaluation should separate physical routes and acquisition sessions, while additional environments and antenna configurations would test whether the same architecture remains useful beyond the present simulations. Comparing a fixed model with one retrained on limited target data would also help distinguish direct transfer from the benefit of local adaptation.

Partial-sweep operation is a second direction, since the current input assumes every beam is sounded at each observation. A predictor could combine the available powers with an observation mask and timestamps, with evaluation accounting for information loss as sounding is reduced. On-device measurements should then include acquisition, feature construction, inference and beam-switching delays under the same resource budget for all methods. Testing optimized ABCP, MLP and GRU implementations on a common platform would clarify whether the smaller parameter count translates into lower decision latency or energy consumption. Conditioning predictions on the expected transmission time could support several delay horizons within one model, with tests assessing sensitivity to errors in processing and scheduling estimates.

Predictive network control [18], reconfigurable transmission [19], drone mobility [20] and satellite coordination [21] require separate observation and timing assessments before establishing net rate or energy benefits. A further step is to couple the forecasts to a decision rule that accounts for uncertainty and the cost of additional sounding, then assess useful data rate over complete communication cycles. Evaluation of this decision rule should track additional sounding requests, outage frequency and useful data rate under a fixed sounding budget, so that reliability is assessed together with communication cost.

7. Conclusion

ABCP combines a fitted present-power anchor with a shallow beam-axis correction. Its 1201 parameters are over 97% fewer than the evaluated dense and recurrent baselines, while primary gain comparisons remain inconclusive. Ablations support ordered history and local filtering, although the pointwise control reduces capacity. NumPy ABCP achieves lower measured latency than PyTorch MLP in these simulated full-sweep tests, subject to noise, padding and implementation limitations.

CRediT authorship contribution statement

Hao Hoang Tran: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing. **Thanh Thien-An Dang:** Validation, Formal analysis, Investigation, Writing – Review & Editing. **Viet Anh-Huy Le:** Software, Validation, Investigation, Writing – Review & Editing. **Minh Dinh Nguyen:** Investigation, Writing – Review & Editing, Visualization. **Kiet Nguyen-Tuan Tran:** Validation, Writing – Review & Editing, Visualization. **Sungrae Cho:** Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Chung-Ang University Young Scientist Scholarship in 2025.

References

[1] Q. Xue, C. Ji, S. Ma, J. Guo, Y. Xu, Q. Chen, W. Zhang, A survey of beam management for mmWave and THz communications towards 6G, IEEE Commun. Surveys Tuts. 26 (3) (2024) 1520–1559. doi:10.1109/COMST.2024.3361991.

[2] M. Ahangar, Q. Ahmed, M. Hafeez, S. Bashir, Artificial intelligence-aided beam tracking in autonomous vehicles: State of the art and future directions, IEEE Trans. Intell. Transp. Syst. 26 (11) (2025) 18385–18403. doi:10.1109/TITS.2025.3597346.

[3] G. Charan, M. Alrabeiah, T. Osman, A. Alkhateeb, Camera based mmWave beam prediction: Towards multi-candidate real-world scenarios, IEEE Trans. Veh. Technol. 74 (4) (2025) 5897–5913. doi:10.1109/TVT.2024.3506948.

[4] B. W. Domae, V. Boljanovic, R. Li, D. Cabric, Machine learning prediction for phase-less millimeter-wave beam tracking, in: 2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC), 2022, pp. 1–5. doi:10.1109/SPAWC51304.2022.9833935.

[5] J. Jalali, J. Roa, Y. Song, R. Zhao, B. Sheen, Fast best beam prediction and overhead reduction for 6G networks: A deep learning approach, in: 2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring), 2024, pp. 1–6. doi:10.1109/VTC2024-Spring62846.2024.10683207.

[6] D. Burghal, Y. Li, P. Madadi, Y. Hu, J. Jeon, J. Cho, A. F. Molisch, J. Zhang, Enhanced AI-based CSI prediction solutions for massive MIMO in 5G and 6G systems, IEEE Access 11 (2023) 117810–117825. doi:10.1109/ACCESS.2023.3324399.

[7] Y. Zhao, L. Zhao, J. Xu, H. Shen, Temporal beam prediction for mmWave MIMO systems based on deep learning, in: Communications and Networking. ChinaCom 2024, Vol. 645 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, 2026, pp. 152–165. doi:10.1007/978-3-032-03131-0_12.

[8] S. H. Lim, S. Kim, B. Shim, J. W. Choi, Deep learning-based beam tracking for millimeter-wave communications under mobility, IEEE Trans. Commun. 69 (11) (2021) 7458–7469. doi:10.1109/TCOMM.2021.3107526.

[9] C. Li, D. Hong, B. Zhang, Z. Pan, N. Yokoya, J. Chanussot, MambaX: Image super-resolution with state predictive control, IEEE Trans. Pattern Anal. Mach. Intell. (2026). doi:10.1109/TPAMI.2026.3721958.

[10] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, P. Ghamisi, X. Jia, A. Plaza, P. Gamba, J. A. Benediktsson, J. Chanussot, SpectralGPT: Spectral remote sensing foundation model, IEEE Trans. Pattern Anal. Mach. Intell. 46 (8) (2024) 5227–5244. doi:10.1109/TPAMI.2024.3362475.

[11] D. Tse, P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, 2005. doi:10.1017/CBO9780511807213.

[12] S. Boyd, L. Vandenberghe, Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares, Cambridge University Press, 2018. doi:10.1017/9781108583664.

[13] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[15] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials, IEEE Trans. Antennas Propag. 62 (6) (2014) 3242–3256. doi:10.1109/TAP.2014.2310220.

[16] O. J. Dunn, Multiple comparisons among means, J. Amer. Statist. Assoc. 56 (293) (1961) 52–64. doi:10.1080/01621459.1961.10482090.

[17] DeepMIMO, [dataset] DeepMIMO scenario database, official scenario catalog. City37 Seoul and archived New York and Boston scene products, accessed 19 October 2026. https://www.deepmimo.net/.

[18] T. T.-A. Dang, D. Won, J. Kim, V. A.-H. Le, S. Cho, Forecasting before acting: Recent advances in learning-based traffic prediction and intelligent control in Open RAN, ICT Express 12 (4) (2026) 1041–1061. doi:10.1016/j.icte.2026.07.007.

[19] C. Song, D. Lee, Y. Lee, W. Noh, S. Cho, Deep learning based energy-efficient transmission control for STAR-RIS aided cell-free massive MIMO networks, ICT Express 11 (2) (2025) 341–347. doi:10.1016/j.icte.2025.02.001.

[20] S. Park, C. Park, J. Kim, Learning-based cooperative mobility control for autonomous drone-delivery, IEEE Trans. Veh. Technol. 73 (4) (2024) 4870–4885. doi:10.1109/TVT.2023.3330460.

[21] S. Park, G. S. Kim, Z. Han, J. Kim, Quantum multi-agent reinforcement learning is all you need: Coordinated global access in integrated TN/NTN cube-satellite networks, IEEE Commun. Mag. 62 (10) (2024) 86–92. doi:10.1109/MCOM.010.2400001.