

White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/247029/

Version: Accepted Version

Article:
Fatehi, Kavan, Rahmani, Mostafa, Grace, David et al. (2026) PhysDiff-Net: Physics-Guided Diffusion for Synthetic Wireless Traffic Generation. IEEE Transactions on Machine Learning in Communications and Networking. ISSN: 2831-316X

https://doi.org/10.1109/TMLCN.2026.3741046

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.





This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

IEEE ComSoc IEEE Communications Society | IEEE COMPUTER SOCIETY | IEEE Signal Processing Society | VTS Connecting the Mobile World | IEEE Transactions on Machine Learning in Communications and Networking

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/TMLCN.2022.1234567

PhysDiff-Net: Physics-Guided Diffusion for Synthetic Wireless Traffic Generation

Kavan Fatehi[1]*, Mostafa Rahmani Ghourtani[1]*, David Grace[1], Mahnaz Arvaneh[2], Kin K Leung[3], Hamed Ahmadi[1]

[1] School of Physics, Engineering and Technology, University of York, UK
[2] School of Electrical and Electronic Engineering, University of Sheffield, UK
[3] Department of Electrical and Electronic Engineering, Imperial College London, London, UK

Corresponding author: Hamed Ahmadi (email: Hamed.ahmadi@york.ac.uk).
*These authors contributed equally to this work.

ABSTRACT Synthetic network traffic for benchmarking wireless optimisation algorithms must satisfy statistical, temporal, and physical requirements that existing generative models do not meet simultaneously. To address this challenge, we propose **PhysDiff-Net**, a conditional denoising diffusion probabilistic model for synthetic WiFi network traffic generation. Employing a 6-layer Transformer encoder backbone (5.62M parameters), the model learns to iteratively denoise network traffic data conditioned on hour-of-day labels. We introduce three components: (1) Variational Hour Conditioning with an auxiliary classification loss that enforces strong temporal pattern encoding, enabling the generation of distinct traffic profiles corresponding to peak and off-peak periods; 2) Physics-Guided Training Losses that incorporate an SNR–user-density anti-correlation as an explicit wireless-domain prior, enabling the generator to represent interference-limited high-load conditions that are under-represented in the observational dataset; and (3) Variance-Guided Diffusion Sampling (VGDS), a sampling-time intervention that resolves the trade-off between the strong classifier-free guidance required for temporal steering and the preservation of realistic output variance. VGDS applies an adaptive post-hoc correction that constrains the output distribution to match real data statistics while preserving temporal patterns and physics correlations. PhysDiff-Net achieves a 100% Pattern Score, satisfying all three physics validation criteria, and preserves cross-feature correlation structure ($\rho_{\mathrm{FC}} = 0.856$), outperforming VAE, TimeGAN, and vanilla diffusion baselines on five of six evaluation metrics.

INDEX TERMS Network Digital Twins, Generative AI, Diffusion Models, Physics-Informed Machine Learning, Synthetic Data Generation

I. INTRODUCTION[1]

THE increasing densification and heterogeneity of wireless networks, driven by the deployment of a large number of access points (APs), diverse service requirements, and time-varying traffic characteristics, introduce significant challenges for efficient network operation. Consequently, data-driven network optimization techniques are regarded as a key enabler for beyond-5G and 6G systems [2]. Accurate evaluation and benchmarking of intelligent network control mechanisms, such as load balancing, interference management, and energy-efficient resource allocation, require large-scale, high-fidelity datasets that capture both statistical variability and physically meaningful network behaviour. However, obtaining such datasets from real-world deployments remains challenging due to limited measurement campaigns, privacy constraints, and the inherent rarity of extreme or stress scenarios [3], [4].

Operational networks record this behaviour as high-resolution cross-layer telemetry: traffic counters, radio measurements such as RSSI and SNR, retransmission statistics, and client-association states. These records support congestion diagnosis and the design of load-balancing, congestion-control, and resource-allocation algorithms [5], but privacy regulations, infrastructure-security policies, and commercial sensitivity keep long-duration, high-resolution traces within the operators that collect them [6]. The resulting scarcity

[1] The source code and dataset used in our experiments is available at [1].

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited a
content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

limits reproducibility and fair benchmarking, and restricts exposure to the diverse traffic regimes and rare congestion events against which optimisation algorithms must be validated [7]–[9]. Synthetic telemetry that preserves the statistical, temporal, and physical structure of real measurements offers an alternative that supports safe data sharing and stress testing without exposing operational information.

Generative Artificial Intelligence (GenAI) has recently emerged as a promising solution for synthetic network data generation, enabling the creation of scalable and controllable datasets for network planning, prediction, and optimization [10]. Early approaches based on Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and autoregressive models have demonstrated the feasibility of generating realistic traffic patterns [11]. Nevertheless, these models often struggle to simultaneously satisfy three key requirements essential for network digital twins: statistical fidelity, temporal controllability, and physical plausibility [12], [13]. In particular, unconstrained generative models are prone to learning spurious correlations present in limited datasets, leading to synthetic data that may appear statistically plausible but violates fundamental wireless principles [14].

Diffusion models have recently gained significant attention due to their strong mode coverage, stable training dynamics, and superior sample quality compared to GAN-based approaches [15]. Their application to time-series and network traffic generation has shown encouraging results in capturing complex spatio-temporal dependencies [16], [17]. However, enabling strong conditional controllability in diffusion models introduces a fundamental trade-off: guidance mechanisms such as classifier-free guidance often lead to variance inflation and degraded statistical fidelity in generated samples [18]. This challenge becomes particularly problematic in wireless networking applications, where preserving accurate statistical structure is essential for reliable network analysis and optimization. Moreover, most existing diffusion-based approaches remain physics-agnostic, limiting their reliability for wireless network optimization and digital twin applications [19]. In practical wireless systems, unrealistic correlations between network load and radio quality can lead to misleading insights when synthetic data is used for downstream tasks.

To address these challenges, we propose PhysDiff-Net, a physics-guided conditional diffusion framework for synthetic WiFi data generation in scenarios where real network measurements are limited or difficult to obtain. PhysDiff-Net integrates three complementary innovations: (i) hour-conditioned diffusion with auxiliary temporal supervision, enabling the generation of distinct traffic profiles corresponding to different times of day; (ii) physics-guided training losses, explicitly discouraging implausible correlations between user density and radio quality metrics; and (iii) Variance-Guided Diffusion Sampling (VGDS), a novel sampling-time intervention that resolves the variance inflation problem associated with strong conditional guidance without requiring model retraining. Using real-world WiFi hotspot measurements collected in a dense urban deployment, we demonstrate that PhysDiff-Net provides the strongest overall balance across the evaluated statistical, temporal, and domain-prior criteria.

The main contributions of this work are as follows:

1) A conditional diffusion framework (PhysDiff-Net) for synthetic WiFi network traffic generation that jointly satisfies statistical fidelity, temporal controllability, and physics compliance;
2) Physics-guided training losses that impose the SNR density anti-correlation of interference physics as an explicit domain prior, correcting a relationship that the observational data, shaped by provisioning and operating-regime effects, does not exhibit;
3) Variance-Guided Diffusion Sampling, a post-hoc correction mechanism that resolves the trade-off between strong classifier-free guidance and realistic output variance without model retraining.

The remainder of this paper is organised as follows. Section II reviews related work on generative and diffusion-based synthetic data generation, including physics-informed approaches. Section III describes the Blackpool WiFi deployment, characterises its statistical structure, and formalises the generation problem. Section IV presents the PhysDiff-Net architecture and its three components. Section V details the dataset processing, model configuration, training, baselines, and evaluation protocol. Section VI reports the experimental results and ablation studies. Section VII concludes and outlines future work.

II. Related Work

Our approach spans generative modelling, diffusion techniques, and physics-based constraints, and we organise the related work along these three lines. First, generative AI for network traffic synthesis and management, where prior studies employ GANs and VAEs to overcome data scarcity and privacy constraints in networking datasets. Second, diffusion models for network traffic generation, covering recent gains in fidelity and stability for complex spatio-temporal patterns. Third, physics awareness and constraint enforcement in generative models, where domain knowledge or wireless-domain priors are embedded to ensure feasibility in wireless scenarios.

A. Generative AI for Network Traffic and Management

Synthetic data generation for wireless networking has gained significant attention as it can alleviate the limited availability of large-scale datasets and enable privacy-preserving model development [20]. Early approaches relied on statistical traffic models (e.g., Poisson or Markov processes), but recent work increasingly adopts deep generative learning to capture nonlinear dynamics, mobility-driven variations, and bursty traffic [21]. GANs were among the first models applied

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

to network traffic and channel synthesis [22]. They often produce visually and statistically realistic samples, but suffer from training instability and mode collapse, which is particularly problematic for multi-modal traffic traces containing both normal and rare behaviors [23]. In parallel, VAEs have been used to generate traffic matrices and telemetry features with more stable training and probabilistic foundations [24]. However, VAEs frequently smooth high-frequency structures and may underrepresent burstiness and long-tail events due to posterior collapse or over-regularization. Overall, while GAN/VAE-based methods can reduce data scarcity, they are often limited in capturing long-range temporal dependencies and ensuring sample validity under networking rules and wireless-domain constraints.

B. Diffusion Models for Network Traffic Generation

Diffusion probabilistic models have recently shown strong performance in synthetic network data generation, largely due to stable training and improved coverage of complex distributions [25], [26]. Recent wireless/networking studies employ conditional and guided diffusion to incorporate contextual information such as service types, spatial semantics, or external data sources.

In addition, several studies have developed conditional diffusion frameworks that exploit auxiliary inputs (such as population statistics, points-of-interest (POI) signals, or legacy traffic traces) to synthesize realistic future demand patterns [27], [28]. These diffusion-based generators generally outperform GAN baselines in terms of statistical fidelity, temporal consistency, and robustness to multimodal traffic distributions. Nevertheless, diffusion models typically incur higher inference cost due to the iterative denoising process and often require careful conditioning and architectural design to remain controllable and scalable for large-scale wireless deployments.

Early structured diffusion-based traffic generators primarily focused on modeling temporal dynamics and statistical fidelity. For instance, NetDiff [29] proposes a service-guided hierarchical diffusion framework that explicitly models user service intents through a two-layer generation process, enabling improved controllability of network flow traces. Similarly, PacketDiff [30] introduces a flow-guided diffusion model that leverages an IP-graph representation together with classifier-free guidance to capture correlations between network flows and packet traces, demonstrating strong fidelity on IoT traffic datasets.

More recently, knowledge-enhanced diffusion approaches have been explored to incorporate richer contextual structure. Chai *et al.* proposed a series of knowledge-enhanced diffusion frameworks that incorporate structured urban context into mobile traffic generation. Their earliest effort, KG-Diff, introduces frequency-domain attention guided by urban knowledge graphs to capture periodic and small-scale patterns in traffic data [31]. This work was extended in STKD-Diff, which integrates mobility graphs and regional semantics via a dual-branch architecture for improved diversity and generalization [32]. Most recently, MCDiff jointly synthesizes user and traffic data conditioned on multimodal environmental features such as satellite imagery and population maps [33]. While these models improve fidelity and controllability through contextual conditioning, they remain fully data-driven and do not incorporate physics-based constraints to ensure wireless-domain consistency.

Complementary research has also explored hybrid paradigms. For example, FlowDiffGAN [34] incorporates diffusion processes into adversarial training to balance fidelity and efficiency, while GAN-based frameworks such as SPATGAN [35] focus on packet-level synthesis through multi-agent adversarial learning.

However, GAN-based and hybrid approaches may inherit training instability or lack the strong likelihood-based foundations of pure diffusion models. Despite these advances, most existing diffusion-based traffic generators remain largely physics-agnostic and lack explicit mechanisms to ensure wireless-domain consistency. This limitation motivates the growing interest in physics-aware generative modeling for reliable WiFi data synthesis.

C. Physics Awareness and Constraint Enforcement in Generative Models

A fundamental limitation of purely data-driven generators is the lack of intrinsic enforcement of network rules or wireless physics. Several studies report that unconstrained synthetic traces can violate protocol validity (e.g., inconsistent header/state combinations). To address this, an emerging line of work integrates domain knowledge into generative pipelines, either by embedding physical constraints directly into model training or by enforcing feasibility via guided sampling and validation [36]. In wireless-specific settings, conditional diffusion has been applied to generate channel state information and radio maps, with clear benefits over interpolation or generic neural baselines [37]–[39]. More recent efforts further integrate physical priors into the generation process. For example, physics-informed diffusion frameworks combine generative diffusion models with PINN-style constraints derived from the Helmholtz equation to enforce propagation consistency and improve generalization in sparse environments [40]. Similarly, WiFi-Diffusion adopts a generate-and-elect strategy that incorporates radio propagation models to select physically plausible candidates [41]. These developments highlight that physics-aware modeling is essential for trustworthy synthesis in wireless systems, particularly when extrapolating to rare events such as deep fades, coverage holes, or extreme traffic surges.

D. Research Gap

Diffusion models now match or exceed GANs and VAEs in fidelity and training stability for synthetic traffic generation, yet most treat physical feasibility and networking validity as secondary. Existing diffusion-based generators are largely

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

physics-agnostic, with no explicit mechanism to enforce wireless-domain consistency; the constraint-aware methods that do exist rely on soft penalties or post-processing and leave unaddressed the variance inflation that strong conditional guidance induces. PhysDiff-Net targets both gaps, pairing conditional diffusion with an explicit physics-guided objective and a variance-correction sampling stage so that generated data stays statistically realistic and physically consistent while remaining useful for downstream management and optimisation.

III. Problem Formulation and Dataset Characterisation

A. Problem Formulation

We consider a wireless network comprising N_{AP} APs, where the state of AP a at observation time τ is described by an F-dimensional cross-layer feature vector $\mathbf{x}_\tau^{(a)} \in \mathbb{R}^F$. The complete network-state snapshot is represented as

$$\mathbf{x}_\tau = \left[\mathbf{x}_\tau^{(1)}, \ldots, \mathbf{x}_\tau^{(N_{\mathrm{AP}})}\right]^{\mathsf{T}} \in \mathbb{R}^{N_{\mathrm{AP}} \times F}. \tag{1}$$

Each snapshot is associated with an hour-of-day condition $h_\tau \in \{0, \ldots, 23\}$. Given an observational dataset $\mathcal{D}_{\mathrm{real}} = \{(\mathbf{x}_\tau, h_\tau)\}_{\tau=1}^{N}$, our objective is to learn a conditional generative distribution $p_\theta(\mathbf{x} \mid h)$ from which synthetic network states $\hat{\mathbf{x}} \sim_\theta (\mathbf{x} \mid h)$ can be generated for a specified hour of day. The formulation is architecture-independent: at this stage, p_θ denotes a generic conditional generator rather than a particular Transformer or diffusion implementation. The specific PhysDiff-Net architecture used to realise this distribution is introduced in Section IV. For clarity, the principal notation used throughout the problem formulation, diffusion framework, training objectives, and evaluation is summarised in Table 1.

For network digital-twin and what-if applications, the generated distribution is required to satisfy three complementary properties. First, *statistical fidelity* requires the generated network states to reproduce the distributional characteristics of the observational telemetry, including feature marginals and cross-feature dependencies. At the distribution level, this objective can be expressed as

$$p_\theta(\mathbf{x}) \approx p_{\mathrm{data}}(\mathbf{x}). \tag{2}$$

Second, *temporal controllability* requires the conditional distribution $p_\theta(\mathbf{x} \mid h)$ to change meaningfully with the specified hour h, such that generated states reproduce the characteristic traffic and network-usage regimes associated with different periods of the day.

Third, *domain compliance* requires the framework to permit wireless-domain knowledge to regularise generation in operating regimes that are weakly represented in the observational record. In this work, we focus on an interference-limited high-load regime. Increasing the number of simultaneously active users can increase contention and co-channel interference and thereby degrade radio quality. We encode this behaviour through the domain prior

TABLE 1. Summary of notation.

Symbol	Definition
Dataset and dimensions	
$\mathbf{r}^{(c,a)} \in \mathbb{R}^{50}$	Raw telemetry record (client c, AP a)
$\mathbf{x}_\tau^{(a)} \in \mathbb{R}^F$	Engineered AP feature vector at time τ
$F = 29$	Engineered features per AP
$N_{\mathrm{AP}} = 11$	Number of access points
$T = 360$	Observations per 1-h sliding window
$N = 5{,}017$	Number of sliding windows
$N_s = 5{,}000$	Samples for the feature-correlation metric
$h \in \{0, \ldots, 23\}$	Hour-of-day conditioning label
$\mathbf{x}_\tau \in \mathbb{R}^{N_{\mathrm{AP}} \times F}$	AP-level network-state snapshot
$G_\theta,\ p_\theta(\cdot \mid h)$	Generator and conditional distribution
p_{data}	Real-data distribution
$\hat{\mathbf{x}}$	Generated network-state snapshot
Diffusion process	
$t \in \{1, \ldots, T_{\mathrm{diff}}\}$	Diffusion timestep (T_{diff} total)
$\mathbf{x}_t \in \mathbb{R}^{N_{\mathrm{AP}} \times F}$	Noisy snapshot at diffusion step t
$\bar{\alpha}_t$	Cumulative noise-schedule coefficient
$\epsilon_\theta(\mathbf{x}_t, t, \cdot)$	Noise-prediction network
$\hat{\mathbf{x}}_0$	Tweedie estimate of clean snapshot
Conditioning and sampling	
$\mathbf{z}^{(\ell)},\ \mathbf{z}^{(L)}$	Hidden state at layer ℓ (final L)
w	Classifier-free guidance scale
Training losses	
$\mathcal{L}_{\mathrm{total}}$	Composite training objective
$\mathcal{L}_{\mathrm{diff}}$	Diffusion denoising loss
$\mathcal{L}_{\mathrm{physics}}$	SNR–density domain-prior loss
$\mathcal{L}_{\mathrm{cls}}$	Auxiliary hour-classification loss
$\mathcal{L}_{\mathrm{var}}$	Variance-matching loss
$\mathcal{L}_{\mathrm{KL}}$	KL regulariser
$\mathcal{L}_{\mathrm{contrast}}$	Contrastive loss
$\lambda_{\mathrm{phys}}, \ldots, \lambda_{\mathrm{con}}$	Weights of auxiliary losses
$\rho_0 = -0.3$	SNR–density feasibility threshold
τ_c	Contrastive temperature
Evaluation	
ρ_{SD}	SNR–density correlation
ρ_{FC}	Feature (cross-feature) correlation

$$\rho(\mathrm{UE}_{\mathrm{density}}, \mathrm{SNR}) \leq \rho_0, \qquad \rho_0 < 0, \tag{3}$$

where $\rho(\cdot, \cdot)$ denotes the correlation coefficient and ρ_0 defines a feasibility threshold. Importantly, this relationship is introduced as a modelling prior for interference-limited high-load operation rather than as a universal propagation law or a correlation estimated from the Blackpool measurements. Its purpose is to prevent the generator from simply extrapolating deployment-specific observational correlations into high-load what-if scenarios for which the available measurement record provides limited evidence. The specific implementation of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

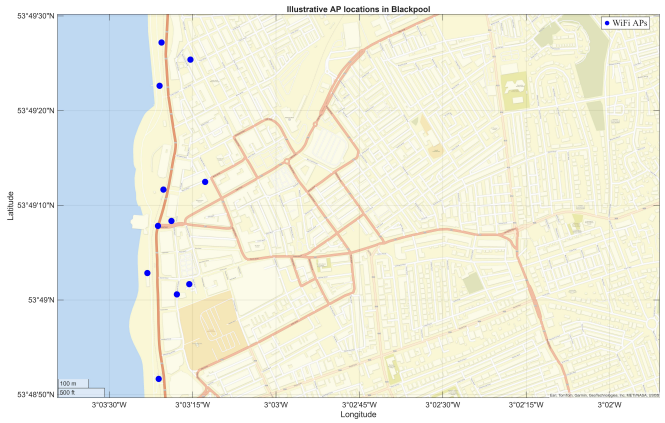


FIGURE 1. Illustrative spatial distribution of the 11 WiFi access points in the Blackpool deployment area. For privacy and security reasons, slight perturbations have been applied to the displayed coordinates while preserving the relative spatial structure and inter-AP proximity.

this prior as a differentiable physics-guided training loss is presented in Section B.

The resulting generation problem therefore seeks a balance between (i) fidelity to the observed network statistics, (ii) controllability with respect to the desired temporal condition, and (iii) compliance with an explicit wireless-domain prior when generating beyond the well-sampled operating regime. The remainder of this section describes the observational data and establishes the empirical characteristics that motivate these requirements.

B. Blackpool WiFi Telemetry Dataset

The case study is based on a municipal WiFi deployment in Blackpool, UK, comprising 11 geographically distributed APs, as shown in Fig. 1. The dataset spans approximately five weeks of continuous operation and is sampled at a temporal resolution of 10 seconds, yielding several million client-level telemetry records. Each record corresponds to a single client device connected to AP a and is represented as $\mathbf{r}^{(c,a)} \in \mathbb{R}^{50}$, where the 50 raw features include identifiers, control-plane states, traffic counters, retransmission statistics, radio measurements, and administrative attributes. This client-centric representation captures cross-layer network behaviour over extended operational periods. It must be noted that the dataset contains no private information of any users or their data, and only contains network measurements.

While the original schema contains 50 columns, many variables are either low-variance identifiers, sparse administrative flags, or highly collinear counters that contribute little to modelling network dynamics. To construct a tractable and operationally meaningful generative space, we retain a subset of 14 representative features that preserve the essential cross-layer characteristics of the WiFi system. These 14 columns were selected from the original telemetry schema to span four essential cross-layer categories: identifiers and connection state (timestamp, AP name, client MAC, connection state, spatial stream count), traffic counters (bytes and packets received/transmitted, data retries), radio measurements (RSSI, SNR), and idle-state indicators (idle timeout, idle timestamp). This selection preserves the cross-layer representativeness necessary for modelling network dynamics while excluding low-variance administrative flags, redundant identifiers, and highly collinear counters that contribute little to downstream optimisation tasks.

C. Feature Engineering and Network-State Representation

The retained client-level telemetry is transformed into the AP-level network-state representation of Eq. (1) through a multi-stage feature-engineering pipeline. Client records are first aggregated into fixed-duration 10-second bins and grouped by AP to obtain load, traffic, radio-quality, and mobility statistics. These quantities are subsequently augmented with cyclical temporal encodings, first-order temporal differences, and spatial-context descriptors.

The resulting representation contains $F = 29$ engineered features for each AP, organised into seven categories:

- *Load indicators* (4): user equipment count, active user count, idle user ratio, and mobility user count;
- *Traffic statistics* (4): total bytes, total packets, average packet size, and retransmission rate;
- *Signal quality metrics* (4): mean RSSI, mean SNR, RSSI standard deviation, and poor signal ratio;
- *Mobility indicators* (4): handover rate, arrival rate, departure rate, and mean session duration;
- *Cyclical temporal encodings* (6): sine and cosine encodings of hour-of-day and day-of-week, weekend indicator, and peak-hour indicator;
- *First-order temporal differences* (4): changes in user count, total bytes, mean RSSI, and retransmission rate relative to the preceding window;
- *Spatial context features* (3): mean load and mean RSSI of neighbouring APs, and spatial load gradient.

These engineered feature categories map directly onto the network behaviours relevant to subsequent optimisation and what-if analysis. Load and traffic features characterise demand intensity; signal-quality metrics describe radio conditions; mobility indicators represent user dynamics; temporal encodings capture diurnal and weekly context; first-order differences describe short term changes; and spatial features encode relationships between neighbouring APs.

The resulting AP-level feature vector is $\mathbf{x}_\tau^{(a)} \in \mathbb{R}^{29}$, and the corresponding network state is therefore $\mathbf{x}_\tau \in \mathbb{R}^{11\times 29}$. The complete record is partitioned chronologically using one-hour sliding windows containing $T = 360$ observations at 10-second resolution, with a stride of 60 samples, yielding $N = 5{,}017$ windows. These windows are used to construct the chronological train/validation/test partitions and to associate observations with their temporal context.

Importantly, the generative problem defined in Section A concerns AP-level network-state snapshots rather than explicit generation of a 360-step sequence. Accordingly, each

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

modelled state has dimension $\mathbb{R}^{N_{\mathrm{AP}} \times 29}$. Temporal information is carried by the hour-of-day condition, cyclical temporal features, and first-order differences. Explicit sequence generation across multiple successive time steps is outside the scope of the present work and is identified as a future extension in Section VII.

D. Statistical Characterisation and Generative Requirements of the Dataset

Having defined the conditional generation problem and the AP-level network-state representation, we next examine the observational data to determine the statistical structures that a high-fidelity synthetic generator should preserve. The characterisation is organised into three complementary levels: marginal distribution analysis, cross-feature dependency analysis, and spatio-temporal AP-level analysis. Table 2 summarises the main observations and links each empirical property to the corresponding generative requirement introduced above.

For marginal behaviour, traffic-related variables were examined using histograms, kernel density estimation, skewness, and excess kurtosis, which quantify asymmetry and tail concentration in the empirical distributions. The traffic counters exhibit long right tails and high excess kurtosis, indicating bursty and heavy-tailed behaviour caused by heterogeneous user demand and intermittent activity patterns in the municipal WiFi network. Radio-related variables, including RSSI and SNR, were analysed using Gaussian Mixture Models (GMMs), with the number of components selected using the Bayesian Information Criterion (BIC). The BIC-selected multi-component models indicate that radio measurements are better described by multiple propagation regimes than by a single unimodal distribution. A valid generator should therefore reproduce heavy-tailed traffic marginals and multimodal radio characteristics rather than relying on simple Gaussian or unimodal approximations [42].

Beyond marginal distributions, cross-feature dependencies were analysed using Pearson correlation, Spearman correlation, and mutual information (MI). Pearson correlation captures linear dependence, Spearman correlation captures monotonic dependence, and MI captures more general nonlinear dependence. MI was estimated using a k-nearest-neighbour estimator and validated through permutation testing. Several traffic, retransmission, and signal-quality feature pairs show statistically significant dependence ($p < 0.01$), reflecting known interactions between physical- and MAC-layer behaviour in wireless networks [43]. Some feature pairs exhibit high MI while having only moderate or weak Pearson correlation, indicating nonlinear coupling that cannot be fully captured by second-order statistics alone [44], [45]. This motivates a generator capable of preserving joint feature structure, not only individual feature distributions.

To assess spatial and temporal structure, AP-level behavioural fingerprints were constructed for segmented time-of-day windows. Each fingerprint aggregates load, traffic, retransmission, and radio-quality statistics for an AP within a given temporal segment. These fingerprints were compared using similarity matrices, clustering analysis, and principal component projection. The resulting structure indicates that APs do not behave as independent and identically distributed sources: their statistical profiles vary across deployment locations and across morning, peak, and late-night periods. This spatio-temporal heterogeneity motivates AP-aware and hour-conditioned generation.

Fig. 2 provides representative visual examples of these properties. Fig. 2(a) illustrates the multimodal SNR distribution; Fig. 2(b) shows the cross-layer correlation structure among retained features; and Figs. 2(c)–(d) illustrate nonlinear joint relationships between uplink and downlink traffic and between transmitted traffic and retransmissions. These plots serve as visual summaries of the statistical analyses, while Table 2 provides the corresponding evidence-to-requirement mapping.

Overall, the empirical analysis identifies the key generative requirements: preserving heavy-tailed traffic, multimodal radio behaviour, nonlinear cross-feature dependencies, and AP-specific temporal variation. Together with temporal controllability and the high-load domain prior, these properties guide the design and evaluation of PhysDiff-Net.

IV. Proposed Framework

PhysDiff-Net builds on the problem formulation and feature representation of Section III, adding three components to a conditional denoising diffusion backbone: variational hour conditioning, physics-guided training losses, and VGDS. Fig. 3 shows the overall architecture.

The framework is organised around four functional components. The Transformer-based diffusion backbone models dependencies among AP-level cross-layer measurements; variational hour conditioning provides explicit control over time-dependent network states; physics-guided regularisation incorporates the high-load wireless-domain prior defined in Section III; and VGDS addresses the variance distortion introduced by strong conditional guidance at inference time. The following subsections describe the design rationale and formulation of each component.

A. Conditional Diffusion Backbone

The diffusion backbone is designed to learn the joint distribution of AP-level network states while retaining dependencies across APs and cross-layer features. We adopt a Transformer-based denoiser because self-attention allows each AP representation to interact directly with all other AP representations, providing a natural mechanism for capturing the cross-AP dependencies identified in Section III. The Transformer is embedded within the DDPM framework because diffusion models provide stable likelihood-based training and strong coverage of complex continuous distributions.

PhysDiff-Net extends the Denoising Diffusion Probabilistic Model (DDPM) framework [46], incorporating a

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

TABLE 2. Summary of statistical characteristics used to motivate the generative requirements.

Property	What it captures	Statistical evidence (How Measured)	Key result (Blackpool WiFi)	Generative requirement
Heavy-tailed traffic	Bursty and heterogeneous user demand.	Histograms, KDE, skewness, and excess kurtosis of traffic counters.	Traffic counters show long right tails and high kurtosis.	**Preserve bursty traffic marginals.**
Multimodal radio states	Multiple propagation regimes and mobility conditions.	GMM model selection using BIC for RSSI and SNR.	BIC favours multi-component models for radio features.	**Capture multimodal radio distributions.**
Cross-layer dependence	Coupling between traffic intensity, retransmissions, and signal quality.	Pearson/Spearman correlation and kNN mutual information with permutation testing.	Several traffic–radio–retransmission pairs show significant dependence ($p < 0.01$).	**Preserve joint cross-layer structure.**
Nonlinear coupling	Dependencies not fully captured by linear correlation.	Comparison of mutual information with Pearson correlation for selected feature pairs.	Some feature pairs show high MI but weak linear correlation.	**Model nonlinear dependencies beyond second-order statistics.**
Spatio-temporal heterogeneity	AP behaviour varies across locations and time periods.	AP-level fingerprints, similarity matrices, clustering, PCA, and time-of-day segmentation.	APs exhibit distinct profiles and time-dependent variance changes.	**Support AP-aware and hour-conditioned generation.**

Transformer encoder backbone to capture complex spatio-temporal dependencies across access points. The forward diffusion process gradually adds Gaussian noise to data over T_{diff} steps according to $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$, where $\{\beta_t\}_{t=1}^{T_{\text{diff}}}$ is a variance schedule. Using the notation $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, we can sample $\mathbf{x}_t$ directly from $\mathbf{x}_0$ via $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$. The reverse process learns to denoise by predicting the noise $\boldsymbol{\epsilon}$ added at each step:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, h) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t, h), \sigma_t^2\mathbf{I}) \tag{4}$$

where the mean is parameterized as:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t, h) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, h)\right) \tag{5}$$

The noise prediction network $\boldsymbol{\epsilon}_\theta$ employs a 6-layer Transformer encoder architecture. Each layer consists of multi-head self-attention (8 heads, 256-dimensional embeddings) and feed-forward networks (1024 hidden units), with residual connections and layer normalization, yielding approximately 5.6 million trainable parameters.[2] Each AP's feature vector, concatenated with the hour embedding, is projected to a 256-dimensional embedding space and combined with learnable position embeddings to distinguish between access points. We define the forward process over $T_{\text{diff}} = 1000$ noise schedule steps with a cosine schedule [47], which provides more uniform noise levels across timesteps compared to the linear schedule of [46], and apply dropout with rate 0.1.

[2]Parameter count: 6 layers $\times$ ($4d^2$ attention + $2d \cdot d_{ff}$ FFN) + embeddings $\approx$ 5.6M, where $d = 256$ and $d_{ff} = 1024$.

B. Variational Hour Conditioning

To enable temporal controllability, we condition the diffusion model on hour-of-day labels. Although the underlying data is aggregated at 10-second resolution, we adopt hour-of-day as the conditioning granularity because it aligns with the dominant diurnal periodicity in WiFi traffic, demand patterns (e.g., daytime visitor peaks, evening leisure activity, overnight lulls) are primarily governed by human activity cycles that repeat on an hourly basis. Finer conditioning (e.g., per-minute) would fragment the training data into sparse bins, while coarser conditioning (e.g., morning/afternoon/night) would obscure operationally relevant intra-day variation. The hour label $h \in \{0, \ldots, 23\}$ is embedded via a variational formulation. A learnable encoder maps each hour to the parameters of a Gaussian distribution, $\boldsymbol{\mu}_h = \mathbf{W}_\mu \cdot \text{OneHot}(h) + \mathbf{b}_\mu$ and $\log \boldsymbol{\sigma}_h^2 = \mathbf{W}_\sigma \cdot \text{OneHot}(h) + \mathbf{b}_\sigma$, from which the hour embedding is sampled using the reparameterisation trick [48]:

$$\mathbf{h}_{\text{emb}} = \boldsymbol{\mu}_h + \boldsymbol{\sigma}_h \odot \boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{6}$$

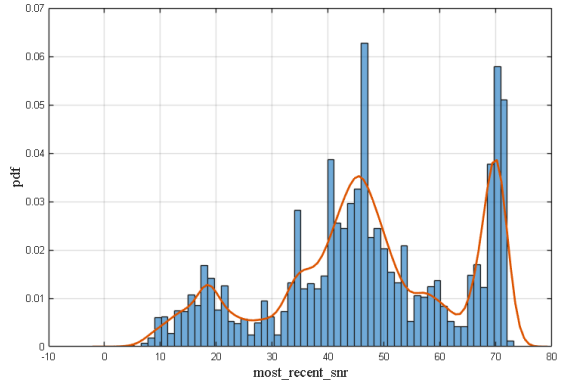
This stochastic embedding is concatenated with the input feature vectors prior to the Transformer encoder. The diffusion timestep t is encoded via sinusoidal positional embeddings followed by a multi-layer perceptron, producing a timestep representation $\mathbf{t}_{\text{emb}}$ that is added to the hidden states through residual connections at each Transformer layer:

$$\mathbf{z}^{(\ell)} = \text{TransformerBlock}^{(\ell)}([\mathbf{x}_t \,\|\, \mathbf{h}_{\text{emb}}]) + \mathbf{t}_{\text{emb}} \tag{7}$$

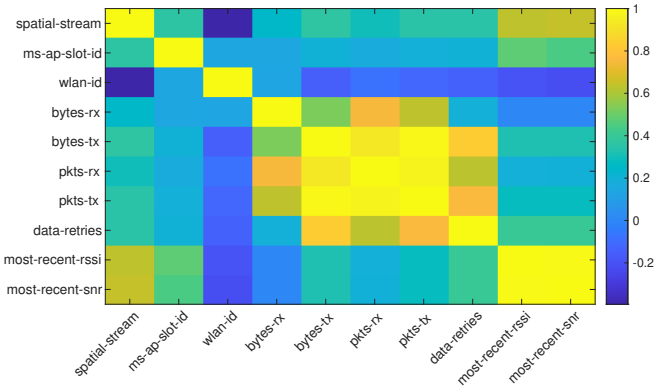
where $[\cdot\|\cdot]$ denotes concatenation along the feature dimension. This design decouples the two conditioning signals: the hour embedding steers the input distribution towards time-of-day-specific traffic patterns, while the timestep embedding

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

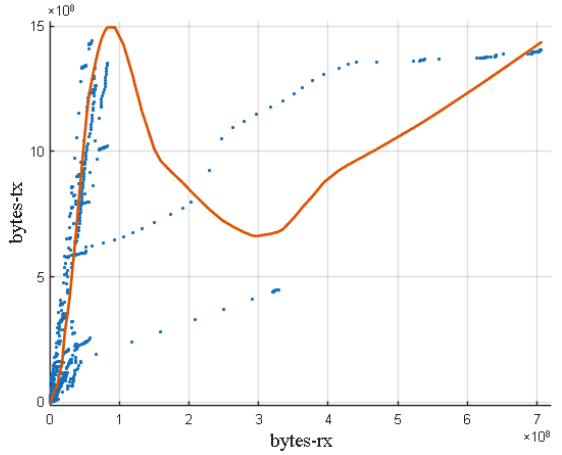
This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046



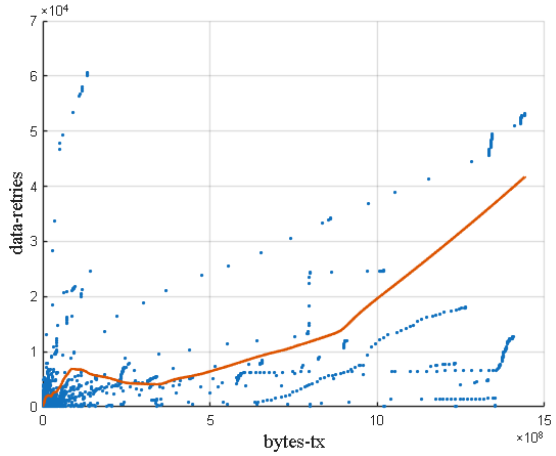
(a) Distribution of most_recent_snr



(b) Pearson correlation heatmap



(c) Joint: bytes_rx vs bytes_tx



(d) Joint: bytes_tx vs data_retries

FIGURE 2. Empirical statistical characteristics of the Blackpool WiFi telemetry dataset. (a) Multimodal SNR distribution capturing diverse radio propagation conditions. (b) Cross-layer correlation structure among retained features. (c) Joint distribution of uplink and downlink traffic volumes, indicating heterogeneous traffic regimes. (d) Nonlinear relationship between transmitted traffic and retransmissions, reflecting coupled load–reliability dynamics.

informs the denoising network of the current noise level. To enforce strong temporal pattern encoding, we introduce an auxiliary classifier C_ϕ that predicts the hour label from intermediate representations, yielding the classification loss $\mathcal{L}_{\text{cls}} = -\mathbb{E}_{(\mathbf{x},h)}[\log C_\phi(h|\mathbf{z}^{(L)})]$. During inference, we employ classifier-free guidance [49] with the guided noise prediction:

$$\tilde{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t, t, h) = (1+w)\cdot\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, h) - w\cdot\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \varnothing) \quad (8)$$

where $w \geq 0$ is the guidance scale controlling the strength of conditional generation, and $\varnothing$ denotes the null embedding. When $w = 0$, the model performs standard conditional sampling, while larger values of w progressively strengthen the alignment with the specified hour's traffic patterns by increasing the difference between conditional and unconditional predictions. To enable this guidance mechanism, the model must learn both conditional and unconditional noise prediction during training. This is achieved by randomly replacing the hour embedding $\mathbf{h}_{\text{emb}}$ with a learnable null embedding $\mathbf{h}_\varnothing \in \mathbb{R}^{d_h}$, initialised to zeros, with probability $p_{\text{uncond}} = 0.3$ at each training step. The null embedding is treated as an additional trainable parameter that the model optimises jointly with all other weights, allowing it to learn a meaningful unconditional representation rather than relying on a fixed placeholder. The dropout probability $p_{\text{uncond}} = 0.3$ governs the trade-off between conditional and unconditional prediction quality: lower values bias the model towards stronger conditional generation but yield a weaker unconditional estimator, reducing the effectiveness of CFG at inference time, while higher values improve the unconditional baseline at the expense of conditional fidelity. We selected $p_{\text{uncond}} = 0.3$ following established practice in the diffusion literature [49], which identifies the range $[0.1, 0.5]$ as generally effective, and verified through preliminary experiments that this value yielded stable guidance behaviour across scales $w \in [1, 10]$.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

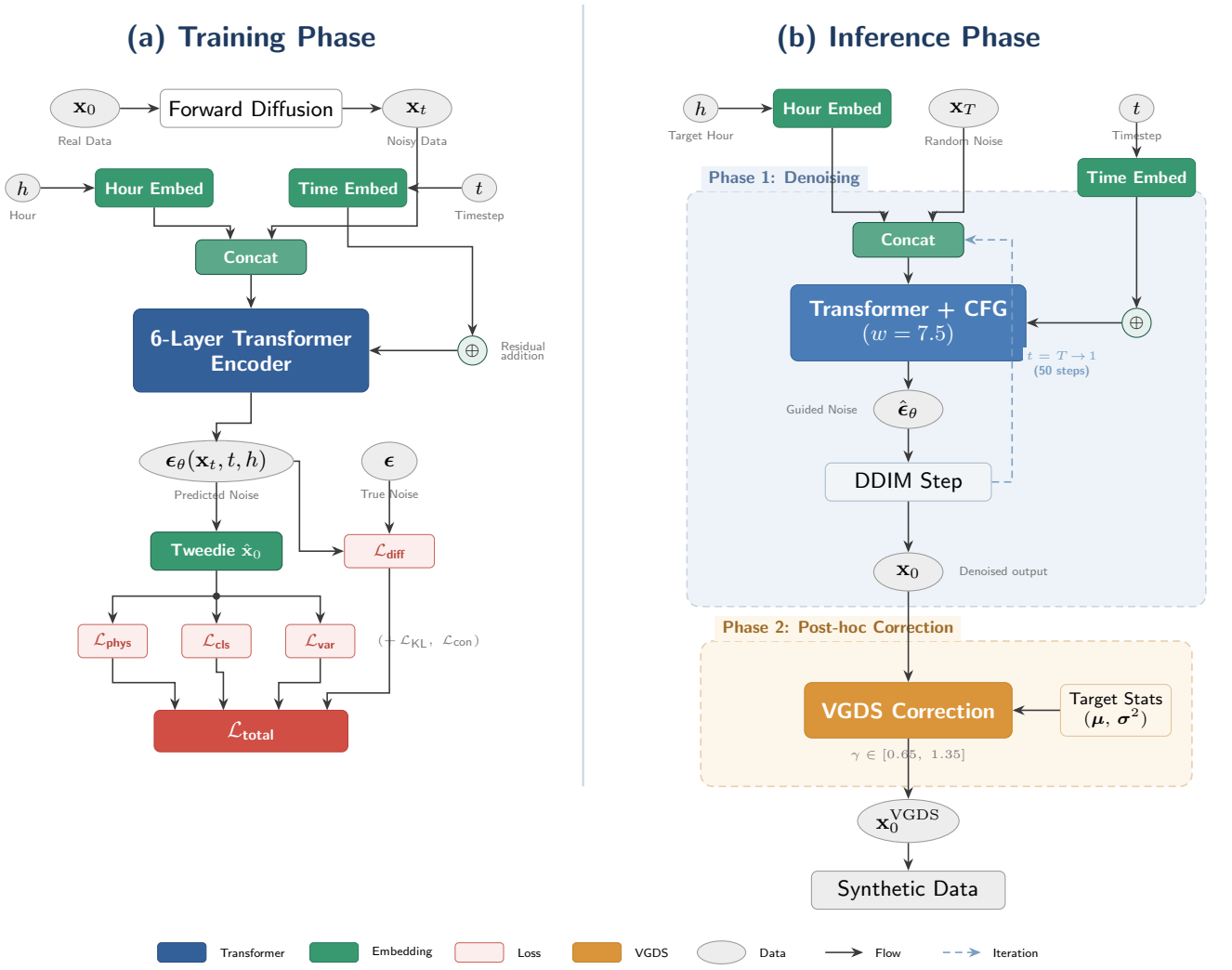


FIGURE 3. **PhysDiff-Net architecture overview. (a) Training Phase: Real data $\mathbf{x}_0$ undergoes forward diffusion to produce noisy samples $\mathbf{x}_t$. A 6-layer Transformer encoder with variational hour conditioning and residual timestep embedding predicts the noise conditioned on hour embedding $\mathbf{h}_{\text{emb}}$ and timestep embedding $\mathbf{t}_{\text{emb}}$. The total loss comprises diffusion loss $\mathcal{L}_{\text{diff}}$, physics-guided loss $\mathcal{L}_{\text{physics}}$ computed via Tweedie's formula on the predicted $\hat{\mathbf{x}}_0$, hour classification loss $\mathcal{L}_{\text{cls}}$, and variance regularization $\mathcal{L}_{\text{var}}$. (b) Inference Phase: Starting from Gaussian noise $\mathbf{x}_{T_{\text{diff}}}$, samples are iteratively denoised using classifier-free guidance (CFG) with scale $w = 7.5$. The VGDS module applies adaptive correction using target statistics $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ from real data, enabling temporal controllability while preserving realistic variance.**

C. Physics-guided mechanism and domain prior.

The physics-guided component is designed to inject wireless-domain knowledge into the learned reverse diffusion process. In the present implementation, the domain prior represents an interference-limited high-load operating regime: as the number of simultaneously active users increases, greater contention and co-channel interference can degrade radio-link quality. We capture this behaviour through an SNR–user-density anti-correlation constraint. Importantly, this relationship is introduced as a modelling prior for an operating regime that is under-represented in the available measurements; it is not claimed to be a universal propagation law inferred from the Blackpool dataset.

The mechanism by which this prior influences generation is as follows. At each training iteration, the diffusion model receives the noisy state $\mathbf{x}_t$ and predicts the added noise $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, h)$. Because the physical relationship cannot be meaningfully evaluated directly on the noisy state $\mathbf{x}_t$, the predicted noise is first mapped to an estimate of the underlying clean network state $\hat{\mathbf{x}}_0$ using Tweedie's reconstruction. The user-density and SNR features are then extracted from $\hat{\mathbf{x}}_0$, and violations of the prescribed domain relationship contribute to the differentiable physics-guided loss. Since $\hat{\mathbf{x}}_0$ depends on $\boldsymbol{\epsilon}_\theta$, gradients from this loss propagate through the reconstruction into the denoising network. The constraint therefore acts as a domain-informed regulariser on the learned reverse diffusion process rather than as a post-processing rule applied after generation.

Generative models trained only on observational network data naturally reproduce the statistical relationships present

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

in that record, including relationships shaped by deployment configuration and operating conditions. In the Blackpool dataset, SNR and user density are weakly positively correlated: $\rho \approx +0.17$ when pooled across APs and $+0.148$ after removing per-AP means, with a 95% cluster-bootstrap confidence interval of $[+0.107, +0.246]$; 10 of the 11 APs exhibit a positive within-AP association. This measured relationship is consistent with the specific deployment and operating regime captured during the five-week observation period, for example, busier APs may also correspond to better-provisioned radio locations. The measurement record, however, contains relatively few interference-limited high-load conditions. For synthetic stress-testing and capacity-planning scenarios that extend into this regime, simply extrapolating the observed positive association may therefore be undesirable. We consequently introduce the SNR–density anti-correlation as an explicit high-load domain prior.

The corresponding physics-guided loss is

$$\mathcal{L}_{\text{physics}} = \max\left(0, \rho(\hat{\mathbf{u}}, \hat{\mathbf{s}}) - \rho_0\right), \tag{9}$$

where $\hat{\mathbf{u}}$ and $\hat{\mathbf{s}}$ denote the predicted user-density and SNR features, respectively, and $\rho_0 < 0$ defines the desired feasibility region for the high-load domain prior.

A key challenge is that the denoising network predicts noise rather than clean network states. We therefore evaluate the domain constraint on the Tweedie estimate of the clean sample rather than on the noisy intermediate state. Specifically,

$$\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, h)}{\sqrt{\bar{\alpha}_t}}, \tag{10}$$

which provides a differentiable estimate of the clean network state at any diffusion timestep. This step creates the explicit link between the wireless-domain constraint and the diffusion dynamics:

$$\mathbf{x}_t \to \boldsymbol{\epsilon}_\theta \to \hat{\mathbf{x}}_0 \to \rho(\hat{\mathbf{u}}, \hat{\mathbf{s}}) \to \mathcal{L}_{\text{physics}} \to \nabla_\theta. \tag{11}$$

Thus, when the reconstructed state violates the prescribed high-load relationship, $\mathcal{L}_{\text{physics}}$ produces a gradient that updates the parameters of the noise-prediction network. Through repeated training updates, this biases the learned reverse diffusion trajectory towards network states satisfying the domain prior while the standard diffusion objective continues to preserve the statistical structure of the observational data.

At each training step, we extract the user density and SNR features from $\hat{\mathbf{x}}_0$ and compute their batch-level Pearson correlation using the differentiable formulation

$$\rho(\hat{\mathbf{u}}, \hat{\mathbf{s}}) = \frac{\sum_i (\hat{u}_i - \bar{\hat{u}})(\hat{s}_i - \bar{\hat{s}})}{\sqrt{\sum_i (\hat{u}_i - \bar{\hat{u}})^2 \sum_i (\hat{s}_i - \bar{\hat{s}})^2}}, \tag{12}$$

where $\hat{u}_i$ and $\hat{s}_i$ denote the predicted user-density and SNR values for the i-th sample in the batch. Since all operations in Eqs. (10)–(12) are differentiable, gradients propagate from $\mathcal{L}_{\text{physics}}$ through the correlation computation and Tweedie reconstruction into the parameters of $\boldsymbol{\epsilon}_\theta$.

The hinge formulation in Eq. (9) is deliberately used instead of forcing the generated data to match a fixed correlation coefficient. It defines a feasible region $\rho(\hat{\mathbf{u}}, \hat{\mathbf{s}}) \le \rho_0$; once this condition is satisfied, the physics-guided penalty becomes zero. Consequently, the prior specifies the direction and minimum strength of the desired high-load relationship but leaves the exact correlation to be determined jointly with the diffusion objective. In this work, we use $\rho_0 = -0.3$. The magnitude $|\rho_0| = 0.3$ corresponds to the conventional boundary of a moderate correlation [50]; hence it is used as a design threshold rather than as a physical constant or a quantity estimated from the Blackpool observations. The negative direction reflects the intended interference-limited high-load regime, in which increasing active-user density can increase contention and co-channel interference and consequently reduce link quality [43].

We additionally introduce a variance regularization term $\mathcal{L}_{\text{var}} = |\text{Var}_f(\hat{\mathbf{x}}_0)/\text{Var}_f(\mathbf{x}_{\text{data}}) - 1|$, computed per feature and averaged, where $\hat{\mathbf{x}}_0$ is the Tweedie estimate from Eq. 10, to prevent mode collapse. In practice, the optimised variance ratio remains within $[0.8, 1.2]$.

The complete training objective combines these components with additional regularization terms:

$$\begin{aligned}\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{diff}} + \lambda_{\text{phys}}\mathcal{L}_{\text{physics}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} \\ + \lambda_{\text{var}}\mathcal{L}_{\text{var}} + \lambda_{\text{KL}}\mathcal{L}_{\text{KL}} + \lambda_{\text{con}}\mathcal{L}_{\text{contrast}}\end{aligned} \tag{13}$$

where $\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, h)\|^2]$ is the standard diffusion loss, $\mathcal{L}_{\text{KL}}$ provides latent regularization, and $\mathcal{L}_{\text{contrast}}$ encourages temporal discrimination between different hours. The weighting coefficients $\lambda_{\text{phys}} = 0.1$, $\lambda_{\text{cls}} = 0.5$, $\lambda_{\text{var}} = 0.1$, $\lambda_{\text{KL}} = 0.01$, and $\lambda_{\text{con}} = 0.1$ were determined through ablation studies.

The KL divergence term $\mathcal{L}_{\text{KL}} = D_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}_h, \text{diag}(\boldsymbol{\sigma}_h^2)) \,\|\, \mathcal{N}(\mathbf{0}, \mathbf{I}))$ regularises the variational hour embedding by penalising deviation of the learned posterior from the standard Gaussian prior, preventing the representations from collapsing to a narrow region or diverging unboundedly. This regularisation ensures that the conditioning space remains well-structured and that interpolation between hour conditions produces meaningful intermediate traffic patterns. The contrastive loss $\mathcal{L}_{\text{contrast}}$ encourages temporal discrimination by pulling representations of samples from the same hour closer together while pushing apart those from different hours. Specifically, for a batch of intermediate representations $\{\mathbf{z}_i^{(L)}\}$ with corresponding hour labels $\{h_i\}$, we adopt a supervised contrastive formulation [51]:

$$\mathcal{L}_{\text{contrast}} = -\sum_i \frac{1}{|P(i)|} \sum_{j \in P(i)} \log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau_c)}{\sum_{k \neq i} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau_c)} \tag{14}$$

where $P(i) = \{j : h_j = h_i, j \neq i\}$ is the set of positive pairs that share the same hour label, $\text{sim}(\cdot,\cdot)$ denotes cosine similarity and $\tau_c = 0.1$ is a temperature parameter. This loss complements $\mathcal{L}_{\text{cls}}$ by operating directly on the representation geometry rather than through a classification

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

head, encouraging the model to learn hour-specific feature clusters that improve the quality of conditional generation.

D. Variance-Guided Diffusion Sampling

Conditional diffusion models exhibit a fundamental trade-off: strong classifier-free guidance improves temporal controllability but leads to unrealistic output variance. This effect can be understood analytically from the CFG formulation (Eq. 8). The guided noise prediction $\tilde{\boldsymbol{\epsilon}}_\theta = (1+w)\boldsymbol{\epsilon}_\theta^{(c)} - w\boldsymbol{\epsilon}_\theta^{(u)}$ is a linear extrapolation beyond the conditional prediction, and its variance expands as:

$$\mathrm{Var}(\tilde{\boldsymbol{\epsilon}}_\theta) = (1+w)^2\sigma_c^2 + w^2\sigma_u^2 - 2w(1+w)\rho_{cu}\sigma_c\sigma_u \quad (15)$$

where σ_c^2 and σ_u^2 denote the variances of the conditional and unconditional predictions, respectively, and ρ_{cu} is their correlation coefficient. Under the reasonable assumption that $\sigma_c \approx \sigma_u \approx \sigma_\epsilon$, this simplifies to $\mathrm{Var}(\tilde{\boldsymbol{\epsilon}}_\theta) \approx [1 + 2(1 - \rho_{cu})(w + w^2)]\sigma_\epsilon^2$. For moderate to large guidance scales ($w \geq 1$), the quadratic term dominates. Since the DDPM reverse step mean is a linear function of $\tilde{\boldsymbol{\epsilon}}_\theta$ (Eq. 5), the variance inflation in noise predictions propagates proportionally to the generated samples, giving $\sigma^2_{\text{gen}}(w)/\sigma^2_{\text{real}} \approx 1 + \eta w^2$ where $\eta = 2(1 - \rho_{cu})$ is model-dependent. We validate this relationship in Section C, where the measured variance ratios across guidance scales $w \in \{1, 3, 5, 7.5, 10\}$ closely follow the predicted quadratic trend (Fig. 7). This variance inflation degrades sample quality and violates statistical fidelity. Previous approaches address this by reducing the guidance scale, but this sacrifices temporal controllability. We propose VGDS to resolve this trade-off without requiring model retraining.

VGDS applies an adaptive linear transform as a post-hoc correction after the full denoising process is completed. Given a batch of generated samples $\mathbf{x}_0$, we compute batch-level statistics and apply a per-feature correction:

$$\mathbf{x}_0^{\text{VGDS}} = \boldsymbol{\mu}_{\text{target}} + \boldsymbol{\gamma} \odot (\mathbf{x}_0 - \boldsymbol{\mu}_{\text{batch}}) \quad (16)$$

where $\boldsymbol{\mu}_{\text{target}}$ and $\boldsymbol{\mu}_{\text{batch}}$ are the target and batch means respectively, and $\boldsymbol{\gamma} \in [0.65, 1.35]$ is an adaptive correction factor computed per feature (Eq. 17). The clipping bounds allow both contraction ($\gamma < 1$) when generated variance is too large and mild expansion ($\gamma > 1$) when variance is suppressed, providing bidirectional correction. Since the variance of the corrected samples scales as γ^2, the effective variance adjustment ranges from $0.42\times$ to $1.82\times$ the batch variance, preventing overcorrection while preserving temporal patterns and physics correlations.

The complete VGDS procedure is presented in Algorithm 1. The algorithm operates in two phases: first, standard classifier-free guided denoising generates an initial batch of samples; second, a post-hoc batch-level correction adjusts the variance and mean of the generated samples to match target statistics. This two-phase process produces samples that jointly satisfy temporal conditioning and statistical constraints.

Algorithm 1 Variance-Guided Diffusion Sampling (VGDS)

Require: Target statistics $(\boldsymbol{\mu}_{\text{target}}, \boldsymbol{\sigma}^2_{\text{target}})$, guidance scale w, hour h

Ensure: Generated samples $\mathbf{x}_0$

// Phase 1: Standard CFG-guided denoising

$\mathbf{x}_{T_{\text{diff}}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

for t in DDIM subsequence of $S = 50$ uniformly spaced steps **do**

  $\tilde{\boldsymbol{\epsilon}} \leftarrow (1 + w) \cdot \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, h) - w \cdot \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \varnothing)$

  $\mathbf{x}_{t-1} \leftarrow \text{DDIMStep}(\mathbf{x}_t, \tilde{\boldsymbol{\epsilon}}, t)$

end for

// Phase 2: Post-hoc variance correction

$\boldsymbol{\mu}_{\text{batch}} \leftarrow \text{Mean}(\mathbf{x}_0)$

$\boldsymbol{\sigma}^2_{\text{batch}} \leftarrow \text{Var}(\mathbf{x}_0)$

$\boldsymbol{\gamma} \leftarrow \text{clip}\Big(\sqrt{\boldsymbol{\sigma}^2_{\text{target}} / (\boldsymbol{\sigma}^2_{\text{batch}} + \epsilon)},\ 0.65,\ 1.35\Big)$

$\mathbf{x}_0 \leftarrow \boldsymbol{\mu}_{\text{target}} + \boldsymbol{\gamma} \odot (\mathbf{x}_0 - \boldsymbol{\mu}_{\text{batch}})$

return $\mathbf{x}_0$

The adaptive correction factor is computed per feature as:

$$\gamma^{(f)} = \text{clip}\left(\sqrt{\frac{{\sigma^2_{\text{target}}}^{(f)}}{{\sigma^2_{\text{batch}}}^{(f)} + \epsilon}},\ 0.65,\ 1.35\right) \quad (17)$$

PhysDiff-Net is trained end-to-end using AdamW with an initial learning rate of 10^{-4} and cosine learning rate decay with a maximum training budget of 600 epochs and batch size 64. We apply gradient clipping with maximum norm 1.0. During inference, we use 50 denoising steps uniformly subsampled across the T_{diff}-step noise schedule following denoising diffusion implicit models (DDIM) [52], with classifier-free guidance at scale $w = 7.5$. Target statistics are computed from training data for the specified hour. VGDS is then applied to calibrate the generated feature variance under classifier-free guidance. PhysDiff-Net integrates three main components: variational hour conditioning provides temporal controllability, physics-guided domain regularisation encourages consistency with the prescribed high-load wireless-domain prior, and VGDS mitigates the output-variance distortion introduced by strong classifier-free guidance. Together, these components provide a controllable framework for generating synthetic cross-layer wireless telemetry for network digital-twin and O-RAN-oriented what-if studies.

V. Implementation

This section describes the dataset, model architecture, training procedure, baseline implementations, and computational environment used in our experiments. The source code and dateset for the implementation used in our experiments is available at [1].

A. Model Architecture

PhysDiff-Net employs a Transformer-based denoising network with variational hour conditioning and residual timestep embedding. Table 3 details the architectural hyper-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

TABLE 3. PhysDiff-Net architecture hyperparameters.

Component	Configuration
Transformer Encoder	
Number of layers	6
Hidden dimension	256
Attention heads	8
Feed-forward dimension	1024
Dropout rate	0.1
Normalization	LayerNorm
Conditioning	
Hour embedding dimension	64
AP embedding dimension	32
Timestep embedding	$256 \to 1024 \to 256$
Diffusion Process	
Noise schedule	Cosine
Diffusion steps T_{diff}	1000
β clipping range	$[1 \times 10^{-4},\ 0.02]$

TABLE 4. PhysDiff-Net parameter breakdown by component.

Component	Parameters	Share
Transformer Blocks (6 layers)	4.74M	84.4%
Time Embedding	0.53M	9.4%
Input Projection	0.10M	1.7%
Positional Encoding	0.09M	1.6%
Hour Classifier	0.08M	1.4%
Output Projection	0.07M	1.3%
Hour Conditioning	0.01M	0.2%
Total	**5.62M**	**100%**

parameters, and Table 4 shows the parameter breakdown by component.

The model receives a noisy input $\mathbf{x}_t \in \mathbb{R}^{11\times 29}$, concatenated with the hour embedding $\mathbf{h}_{\text{emb}}$ along the feature dimension. The timestep embedding $\mathbf{t}_{\text{emb}}$ is added to the hidden representations via residual connections at each Transformer layer, enabling the network to adapt its denoising behaviour for different noise levels while the hour conditioning steers generation towards time-specific traffic patterns.

B. Training Procedure

We train PhysDiff-Net using the composite loss function (Eq. 13), comprising diffusion, physics, classification, variance, KL, and contrastive terms. The physics loss components are computed on Tweedie estimates $\hat{\mathbf{x}}_0$ obtained from the predicted noise at each training step.

Table 5 summarizes the training hyperparameters. We use the AdamW optimizer with cosine annealing learning rate schedule. Training employs Projecting Conflicting Gradients (PCGrad) to handle multi-task optimization, and classifier-free guidance (CFG) dropout with probability 0.3 during training.

TABLE 5. Training hyperparameters.

Hyperparameter	Value
Optimizer	AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning rate	$1 \times 10^{-4} \to 1 \times 10^{-6}$
Weight decay	1×10^{-2}
Batch size	64
Training epochs	600 max
LR schedule	Cosine annealing
Physics warmup	100 epochs
CFG dropout probability	0.3
Loss weights	
λ_{phys} (physics)	0.1
λ_{cls} (hour classification)	0.5
λ_{var} (variance)	0.1
λ_{KL} (KL)	0.01
λ_{con} (contrastive)	0.1

TABLE 6. Inference configuration.

Parameter	Value
Sampling steps	50 (DDIM)
CFG guidance scale w	7.5
VGDS correction	Enabled
VGDS strength	0.35

During inference, we employ classifier-free guidance (CFG) at scale $w = 7.5$ to enhance hour-conditional generation. After the denoising process completes, the VGDS module applies a post-hoc batch-level correction using per-feature adaptive scaling factors $\boldsymbol{\gamma} \in [0.65, 1.35]$ (Eq. 17) to match target statistics $(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ for the specified hour, improving temporal steerability without additional training. The output is a batch of synthetic network state snapshots $\hat{\mathbf{x}} \in \mathbb{R}^{N_{\text{AP}}\times F}$ that preserve the marginal distributions, temporal patterns, and physics-compliant correlations of the target hour's real traffic data.

C. Baseline Implementations

We compare against three representative baselines spanning different generative paradigms:

VAE: A variational autoencoder with MLP encoder-decoder architecture operating on flattened inputs ($360 \times 11 \times 29 = 114{,}840$ dimensions). The model uses a 64-dimensional latent space with 3-layer encoder and decoder networks (hidden dimension 256). Although the latent representation is compact, the flattened high-dimensional input results in 59.24M trainable parameters. The VAE is therefore retained as a representative conventional generative baseline rather than as an upper bound on the performance achievable by a task-specific VAE architecture.

TimeGAN: A GAN-based approach specifically designed for time-series generation [53]. We use a compact archi-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

TABLE 7. Baseline model comparison.

Method	Architecture	Parameters
PhysDiff-Net	Transformer + VGDS	5.62M
Vanilla Diffusion	Transformer (unconstrained)	5.62M
VAE	3-layer MLP	59.24M
TimeGAN	2-layer GRU	1.13M

tecture with 2-layer GRU networks (hidden dimension 128, latent dimension 32), resulting in 1.13M parameters.

Vanilla Diffusion: An unconstrained conditional-diffusion baseline using the identical Transformer backbone as PhysDiff-Net (5.62M parameters). It is trained using only the standard diffusion loss $\mathcal{L}_{\text{diff}}$, without the auxiliary objectives $\mathcal{L}_{\text{physics}}$, $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{\text{var}}$, $\mathcal{L}_{\text{KL}}$, and $\mathcal{L}_{\text{contrast}}$, and without VGDS at inference. A reduced guidance scale of $w = 3.0$ is used, compared with $w = 7.5$ for PhysDiff-Net, because stronger guidance without VGDS produced unstable output variance. The baseline also employs a linear β schedule, whereas PhysDiff-Net uses the cosine schedule described in Section IV. Consequently, Vanilla Diffusion should be interpreted as an architecture-matched unconstrained reference rather than as a single-term ablation of $\mathcal{L}_{\text{physics}}$. Differences between the two configurations therefore reflect the aggregate effect of the proposed auxiliary objectives and sampling strategy and are not attributed to the physics-guided loss alone.

Table 7 compares the parameter counts across all methods. The parameter counts are reported descriptively to indicate model size; they are not used as a measure of computational efficiency across architectures with different input representations and modelling assumptions.

NetDiff and PacketDiff generate flow- and packet-level traces, respectively, whereas MCDiff synthesises grid-level mobile-traffic maps conditioned on additional urban context. None directly generates an AP-level cross-layer network state matching the Blackpool telemetry representation considered here. A direct application of these methods would therefore require substantial modification of their input/output representations and conditioning mechanisms, making the resulting comparison dependent on adaptation choices rather than providing a like-for-like evaluation. We consequently discuss NetDiff, PacketDiff, and MCDiff in Section II as closely related diffusion-based approaches, while retaining Vanilla Diffusion as the principal conditional-diffusion reference on the Blackpool dataset.

All baselines use the same data preprocessing, train/validation/test splits, and evaluation protocol as PhysDiff-Net.

D. Evaluation Protocol

We generate 500 synthetic samples per method for primary evaluation. For the feature correlation metric, we use $N_s = 5{,}000$ samples to ensure statistical stability ($N_s \gg F$). Each sample is generated conditioned on a randomly selected hour $h \in \{0, 1, \ldots, 23\}$. Evaluation metrics are computed on the test set:

- **Pattern Score**: Percentage of the three evaluation criteria (variance ratio $\in [0.8, 1.2]$, hourly UE range > 1.5, and SNR–density $\rho < -0.3$) satisfied by the generated data at the dataset level; attainable values are therefore 0, 33.3, 66.7, and 100%.
- **Variance Ratio**: Ratio of generated to real data variance, aggregated across features.
- **Hourly UE Range**: Difference between maximum and minimum mean UE count across 24 hours.
- **SNR–Density Correlation**: Pearson correlation between SNR and user density, used as a diagnostic of whether the imposed high-load domain prior is present in the generated samples.
- **MMD**: Maximum Mean Discrepancy with an RBF kernel, used as a sample-based measure of discrepancy between the real and synthetic multivariate distributions.
- **Feature Correlation**: Pearson correlation between upper triangular elements of real and generated 29×29 feature correlation matrices, where each matrix captures pairwise correlations among the 29 engineered features across all access points, using $N_s = 5{,}000$ samples for statistical stability ($N_s \gg F$).

The Pattern Score is used as a compact summary of whether the three prescribed behavioural criteria are satisfied; it is not interpreted as an independent measure of overall generative fidelity. All criteria are computed from statistics aggregated over the complete generated dataset across all APs rather than from individual samples.

MMD is selected as the primary sample-based distributional discrepancy metric because it directly compares high-dimensional continuous real and synthetic samples without requiring explicit probability-density estimation. In contrast, applying KL divergence in this setting would require estimating the underlying multivariate densities, making the result dependent on the density-estimation procedure and potentially sensitive to regions with limited distributional overlap. MMD is complemented by the feature-correlation metric, which evaluates preservation of the broader cross-feature dependency structure. Importantly, neither MMD nor the feature-correlation metric appears in the training objective, and these two quantities therefore provide the principal independent evaluation of generated-data fidelity.

Each Pattern-Score threshold has an empirical or conventional basis. The hourly-UE-range threshold of 1.5 is a conservative floor: across the real training and validation windows, the difference between the maximum and minimum mean per-AP user-equipment count over the 24 hours is 9.84, so the threshold demands roughly 15% of the observed diurnal swing and rejects only generators that produce near-uniform temporal profiles. The variance-ratio band $[0.8, 1.2]$ is calibrated against natural sampling variability: over 300

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

TABLE 8. Computational characteristics of the evaluated methods.

Method	Parameters (M)	Training (hours)	Memory (GB)
PhysDiff-Net	5.62	11.6	15
VAE	59.24	2.1	8
TimeGAN	1.13	4.5	6
Vanilla Diffusion	5.62	10.8	14

random partitions of the real windows into disjoint halves, the variance ratio between halves ranges from 0.909 to 1.130 (5th–95th percentile 0.929–1.065), so the band contains the full real-vs-real spread with margin while still excluding the deviations observed in the baselines (0.537 for VAE, 2.751 for Vanilla Diffusion).

The SNR–density threshold differs from the first two criteria because it is an explicit domain-prior design target rather than a data-derived bound. Its negative direction represents the intended interference-limited high-load regime, in which increasing active-user density can increase contention and co-channel interference and thereby degrade radio quality [43]. Its magnitude, $|\rho_0| = 0.3$, corresponds to the conventional boundary of a moderate correlation [50], providing a feasibility threshold without prescribing a specific target correlation. The three Pattern-Score quantities should also be distinguished from the independent fidelity metrics. SNR–density correlation appears explicitly in $\mathcal{L}_{\text{physics}}$, variance ratio is directly encouraged through $\mathcal{L}_{\text{var}}$, and temporal differentiation is encouraged through the hour-conditioning objectives $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{contrast}}$. Consequently, these quantities are interpreted primarily as diagnostics of whether the intended statistical, temporal, and domain-guided behaviours have been realised. Independent evidence of real–synthetic distributional quality is instead provided by MMD and feature-correlation preservation, neither of which is directly included in the training objective.

E. Computational Environment

All experiments are conducted on a single NVIDIA A40 GPU (48GB). Table 8 reports the computational characteristics of each method. These quantities are provided descriptively and are not used to claim computational superiority across models with different architectures and input representations.

PhysDiff-Net requires 11.6 hours of training compared with 10.8 hours for Vanilla Diffusion under the reported experimental configuration, reflecting the additional auxiliary objectives and PCGrad optimisation. Because the evaluated methods differ in architecture and representation, parameter count, training time, and memory usage are interpreted only as implementation characteristics rather than as normalized measures of computational efficiency. PhysDiff-Net inference uses 50 DDIM sampling steps followed by the VGDS correction described in Section IV.

TABLE 9. Main comparison results for synthetic WiFi data generation. PhysDiff-Net is the only method to achieve 100% Pattern Score. ✓ indicates target met.

Method	Pattern	Var.	UE	SNR-ρ	MMD	FC-ρ
PhysDiff-Net	100%	1.122 ✓	2.87 ✓	−0.835 ✓	0.210	0.856 ✓
VAE	0%	0.537	1.34	+0.113	0.045	0.820
TimeGAN	0%	0.747	0.01	+0.331	0.346	0.078
Vanilla Diff.	0%	2.751	0.52	+0.012	0.069	−0.007
Target	100% (3/3)	0.8–1.2	>1.5	<−0.3	→0	>0.5

F. Reproducibility

We fix random seeds across all experiments for reproducibility. Our implementation uses PyTorch 2.0 with the following key dependencies: `torch`, `einops`, `scipy`, and `scikit-learn`. Code and pretrained models will be released upon publication.

VI. Results

A fundamental challenge in synthetic network data generation is balancing statistical consistency with physics compliance. We evaluate this trade-off by comparing PhysDiff-Net against VAE, TimeGAN, and Vanilla Diffusion across six metrics: the composite Pattern Score, its three physics-compliance constraints (SNR–density ρ_{SD}, variance ratio, hourly UE range), distributional proximity (MMD), and cross-feature coherence (feature correlation ρ_{FC}). Table 9 shows that PhysDiff-Net is the only method to satisfy all three physics constraints (100% Pattern Score) while preserving cross-feature correlation structure ($\rho_{\text{FC}} = 0.856$) across the 29 engineered features. VAE attains the lowest MMD (0.045) but violates the SNR–density constraint ($\rho_{\text{SD}} = +0.113$ against the $\rho_{\text{SD}} < -0.3$ target), a divergence that reflects the difference between distributional proximity and physics compliance.

A. Physics-Distribution Trade-off

We first analyze the trade-off between physics compliance and distribution matching. Fig. 4(a) shows the SNR-Density correlation (ρ) for each method. PhysDiff-Net achieves $\rho = -0.835$, comfortably satisfying the target ($\rho < -0.3$), which indicates a strong negative correlation between signal strength and user density — a well-established physical relationship in wireless networks. In contrast, VAE ($\rho = +0.113$), TimeGAN ($\rho = +0.331$), and Vanilla Diffusion ($\rho = +0.012$) all exhibit positive or near-zero correlations, violating the expected physics.

Fig. 4(b) reports the MMD scores, computed on the held-out test set: VAE attains the lowest value (0.045) and PhysDiff-Net a higher value (0.210). MMD measures the discrepancy between marginal feature distributions and is insensitive to the cross-feature dependence and physical relationships that the SNR–density and feature-correlation metrics capture. A model can match the marginals closely while still violating the SNR–density relationship, which is

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

TABLE 10. Ablation study results. Removing the physics-guided loss causes complete failure across all metrics.

Configuration	SNR-ρ	Variance	Hourly UE	Pattern
Full Model (Ours)	-0.835	1.122	2.87	100%
w/o Physics Loss	$+0.012$	2.751	0.52	0%
Δ (Degradation)	$+0.847$	$+1.629$	-2.35	-100%

what VAE does ($\rho = +0.113$). Low MMD and physics compliance are therefore distinct criteria: a low MMD alone does not establish that a model has recovered the physical structure of the data.

Fig. 4(c) plots the two axes jointly, with the physics-compliant region ($\rho < -0.3$) shaded. PhysDiff-Net is the only method that falls in this region, at a moderate MMD (0.210) rather than the lowest: VAE and Vanilla Diffusion reach lower MMD (0.045 and 0.069) but positive SNR–density correlations ($+0.113$ and $+0.012$), and TimeGAN fails on both axes. PhysDiff-Net trades higher marginal discrepancy for physics compliance and does not dominate VAE on MMD.

B. Ablation Study: Necessity of Physics-Guided Loss

To quantify the contribution of our physics-guided loss, we compare PhysDiff-Net against Vanilla Diffusion, which employs the same backbone architecture but excludes the physics constraints. Fig. 5, Fig. 6, and Table 10 present the results.

Removing the physics loss results in substantial degradation across all metrics. The SNR-Density correlation shifts from $\rho = -0.835$ to $\rho = +0.012$ (Fig. 5(a)), a change of $+0.847$ that reverses the expected physical relationship. Variance Ratio increases from 1.122 to 2.751 (Fig. 5(b)), substantially exceeding the acceptable range of 0.8–1.2. Temporal controllability, measured by Hourly UE Range, decreases from 2.87 to 0.52 (Fig. 6(a)), falling below the utility threshold of 1.5. Consequently, the overall Pattern Score decreases from 100% to 0% (Fig. 6(b)).

These results demonstrate that the diffusion backbone alone is insufficient for generating physically valid network data. The physics-guided loss is essential for enabling PhysDiff-Net to capture the underlying physical relationships rather than reproducing statistical patterns from the data.

C. Variance-Law Validation and VGDS Ablation

Section D derives the relation $\sigma^2_{\text{gen}}(w)/\sigma^2_{\text{real}} = 1 + \eta w^2$ for guided sampling without correction. Fig. 7 measures it. With VGDS disabled and 500 snapshots drawn at each of $w \in \{1, 3, 5, 7.5, 10\}$, the variance ratio rises from 1.010 at $w = 1$ to 1.996 at $w = 10$. A least-squares fit through the unit intercept returns $\eta = 0.00991$ and $R^2 = 0.9972$, averaged over three seeds with per-scale standard deviation below 0.02. The ratio crosses the upper edge of the $[0.8, 1.2]$ band near $w \approx 4.5$ and reaches 1.515 at the deployed scale, so the inflation that VGDS removes is present at the operating point and not only under aggressive guidance.

TABLE 11. Ablation of VGDS at the deployed guidance scale $w = 7.5$. Each entry is the mean over three seeds (500 generated snapshots per seed); the value after $\pm$ is the standard deviation across seeds.

Configuration	Variance ratio	ρ_{SD}	Hourly UE range
VGDS disabled	1.521 ± 0.014[a]	-0.833 ± 0.004	2.887 ± 0.014
VGDS enabled	1.122 ± 0.004	-0.834 ± 0.002	2.871 ± 0.008

[a] The sweep in Fig. 7 and this ablation draw samples independently; the uncorrected $w = 7.5$ ratio therefore differs marginally between them (1.515 vs. 1.521, within one standard deviation).

The physics-loss ablation in Section B isolates the training objective; the same logic applies to VGDS, which acts only at sampling time. Table 11 isolates it at $w = 7.5$ by toggling the post-hoc correction with the checkpoint, guidance scale, and sampling schedule held fixed. Without the correction the variance ratio is 1.521, outside the target band; with it the ratio is 1.122, matching the full-model entry in Table 9. The SNR–density correlation changes by 0.001 across the two settings, inside the three-seed spread, and the hourly UE range is unchanged within the same margin. The correction moves generated variance toward the real reference and leaves both the physics relationship and the temporal profile intact, as the derivation in Section D predicts.

D. Cross-Feature Coherence: Feature Correlation Preservation

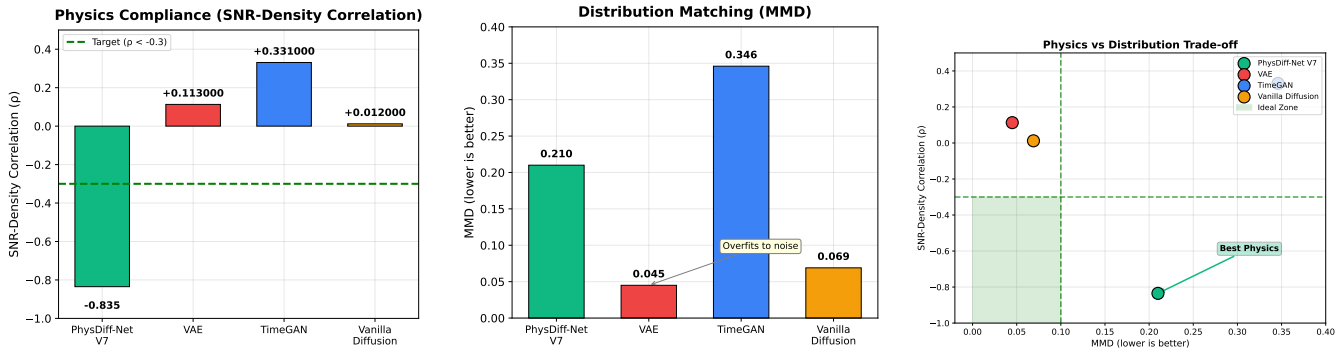
A valid network digital twin must preserve not only individual AP statistics but also the cross-feature correlations that characterise network behaviour. We evaluate this by computing the correlation-of-correlations metric: the Pearson correlation between the upper triangular elements of the real and generated 29×29 feature correlation matrices, using $N_s = 5{,}000$ samples for statistical stability ($N_s \gg F$). This metric captures whether the model preserves the complex interdependencies among load, traffic, signal quality, and spatial features that define the operational state of the network.

Fig. 8(a) presents the feature correlation preservation results. PhysDiff-Net achieves $\rho = 0.856$, the highest among all methods and well above the target threshold of 0.5. VAE follows closely with $\rho = 0.820$, indicating that purely statistical approaches can capture cross-feature structure when physics is not required. TimeGAN ($\rho = 0.078$) and Vanilla Diffusion ($\rho = -0.007$) produce near-zero feature correlation, losing the inter-feature dependencies that characterise realistic network states.

The feature-correlation–physics trade-off is shown in Fig. 8(b). VAE preserves feature correlation but fails physics compliance, while TimeGAN and Vanilla Diffusion fail both. PhysDiff-Net is the only method in the high-feature-

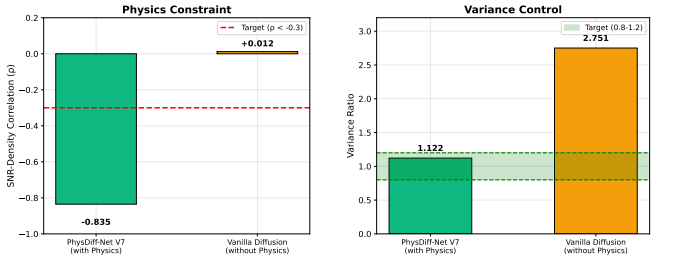
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046



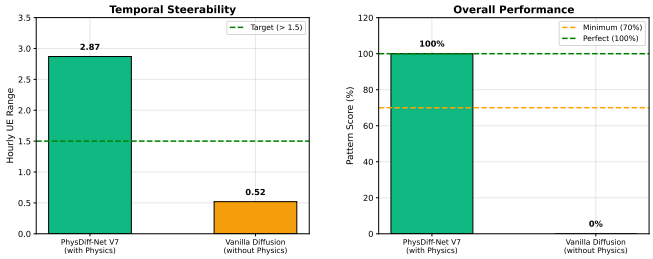
(a) Physics Compliance (b) Distribution Matching (c) Physics vs Distribution Trade-off

FIGURE 4. Trade-off analysis between physics compliance and distribution matching across generative models. (a) SNR-density correlation coefficients, where PhysDiff-Net is the only method satisfying the physics constraint ($\rho < -0.3$), reflecting the expected inverse relationship between signal quality and user density. (b) Maximum Mean Discrepancy (MMD), a measure of marginal-distribution proximity; VAE attains the lowest MMD and PhysDiff-Net a higher one. (c) Joint view of MMD and SNR–density correlation: PhysDiff-Net is the only method in the physics-compliant region ($\rho < -0.3$), at a moderate rather than minimal MMD.



(a) Physics Constraint (b) Variance Control

FIGURE 5. Ablation study on physics compliance metrics illustrating the necessity of physics-guided losses. (a) SNR-density correlation, which should be negative to reflect physical interference effects, inverts from the expected -0.835 to an implausible $+0.012$ when physics losses are removed. (b) Variance ratio between synthetic and real data deviates significantly from the target range of $[0.8, 1.2]$, indicating loss of distributional fidelity without physics guidance.



(a) Temporal Steerability (b) Overall Performance

FIGURE 6. Ablation study on utility metrics demonstrating the critical role of physics-guided losses. (a) Temporal steerability, measured as the ratio of inter-hour to intra-hour variance, falls below the acceptability threshold of 1.5 when physics losses are removed. (b) Pattern Score, the percentage of the three physics-aware constraints satisfied, falls from 100% (all three) to 0% (none) when the physics losses are removed.

correlation, physics-compliant quadrant ($\rho > 0.5$ and SNR $\rho < -0.3$).

Fig. 8(c) presents the multi-metric comparison as a radar chart over five axes (feature correlation, physics compliance, variance control, Pattern Score, temporal controllability). PhysDiff-Net leads on all five, while VAE matches it only on feature correlation. MMD, the one axis on which PhysDiff-Net does not lead, is reported in Table 9 and discussed in Section A.

E. Downstream Utility: What-If Network Simulations

The primary purpose of synthetic network data is to enable what-if simulations for network planning and optimization. We evaluate this downstream utility through three measures: temporal controllability, overall quality score, and scenario coverage.

Fig. 9(a) presents the Hourly UE Range metric, which quantifies a model's capability to generate diverse temporal patterns. PhysDiff-Net achieves a range of 2.87, substantially above the utility threshold of 1.5. This result indicates that PhysDiff-Net can generate synthetic data representing distinct periods of the day, from low-traffic late-night periods to high-demand peak hours. VAE (1.34) falls marginally below the threshold, while TimeGAN (0.01) and Vanilla Diffusion (0.52) produce temporally uniform outputs that are unsuitable for time-varying simulations.

Fig. 9(b) reports the Pattern Score, the percentage of the three physics-aware constraints (variance ratio, hourly UE range, SNR–density) satisfied at the dataset level. PhysDiff-Net is the only method to reach 100%, satisfying all three; the baselines satisfy none and score 0%.

Fig. 9(c) presents scenario coverage across four representative network conditions: Normal Day, Concert (high-density event), Late Night (minimal traffic), and Morning Rush (commuter peak). Coverage score is the fraction of scenario-conditioned samples whose aggregated statistics satisfy all three physics-aware constraints (variance ratio, hourly UE range, and SNR–density), evaluated separately for

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

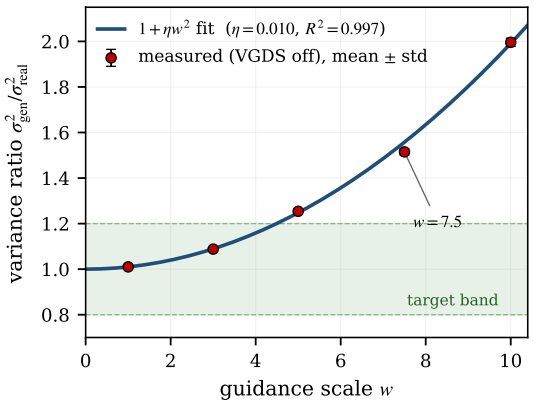
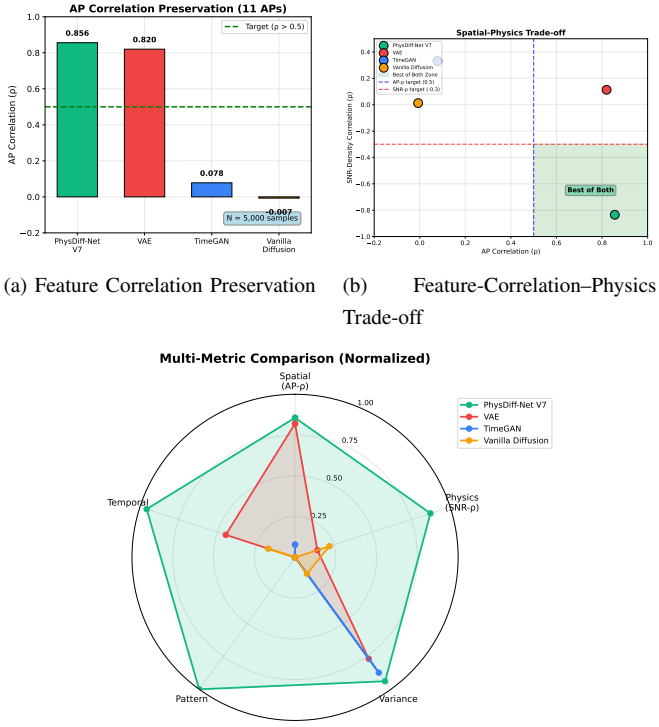


FIGURE 7. Generated-to-real variance ratio $\sigma^2_{\text{gen}}/\sigma^2_{\text{real}}$ against classifier-free guidance scale w with VGDS disabled. Markers are measured ratios (500 snapshots per scale, averaged over three seeds); the curve is the least-squares fit $1+\eta w^2$ with $\eta = 0.00991$ and $R^2 = 0.9972$. The shaded region is the $[0.8, 1.2]$ target band. The ratio leaves the band above $w \approx 4.5$ and reaches 1.515 at the deployed scale $w = 7.5$.



(a) Feature Correlation Preservation (b) Feature-Correlation–Physics Trade-off

(c) Multi-Metric Radar

FIGURE 8. Feature correlation evaluation ($N_s = 5{,}000$ samples). (a) Feature correlation preservation showing PhysDiff-Net achieves the highest value ($\rho = 0.856$). (b) Feature-correlation–physics trade-off, where PhysDiff-Net is the only method in the high-correlation, physics-compliant quadrant. (c) Multi-metric radar over the five comparison axes; PhysDiff-Net leads on each, with MMD (its one weaker metric) reported separately in Table 9.

each scenario. PhysDiff-Net consistently achieves coverage scores above 0.9 across all scenarios, demonstrating its capability to generate valid synthetic data for diverse operational conditions. VAE shows moderate performance on Normal Day (0.7) but degrades substantially for extreme scenarios. TimeGAN and Vanilla Diffusion fail to generate satisfactory samples for any scenario.

These findings demonstrate that PhysDiff-Net is the only evaluated method suitable for supporting reliable what-if simulations in network digital twins, allowing operators to investigate hypothetical scenarios, including concert events, infrastructure failures, and capacity planning, without deploying physical resources.

F. Summary of Comparative Results

Across the six metrics in Table 9, PhysDiff-Net is the only method that satisfies all three physics constraints (100% Pattern Score), and the only one that pairs a compliant SNR–density correlation ($\rho = -0.835$) with strong cross-feature preservation ($\rho = 0.856$). VAE matches PhysDiff-Net on feature correlation ($\rho = 0.820$) and reaches lower MMD, but violates the SNR–density constraint ($\rho = +0.113$). TimeGAN and Vanilla Diffusion fail the physics, variance, and feature-correlation criteria together. Vanilla Diffusion shares PhysDiff-Net's backbone and differs only in the physics-guided loss, so its failure isolates that loss as the component responsible for physics compliance, consistent with the ablation in Section B.

G. Nearest-Neighbour Distribution Consistency Analysis

To check that the generated samples reproduce the local structure of the real data without memorising training instances, we compare two distance distributions: distances between real samples, and distances from each synthetic sample to its nearest real sample. Nearest-neighbour distances probe the fine-scale geometry of the data rather than its marginals, and a synthetic-to-real distribution that collapsed toward zero would indicate memorisation [54].

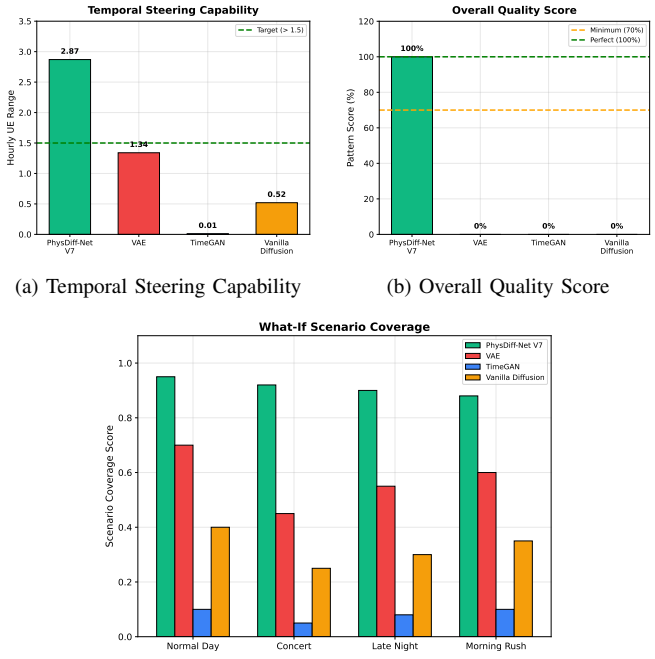
Fig. 10 shows the empirical CDFs of both distances. The real–real (blue) and synthetic–real (orange) curves track each other across the full range, with no spike in the synthetic–real curve near zero distance, so the generated samples are not disproportionately close to individual real samples and the model does not reproduce training points. The synthetic–real curve sits slightly above the real–real curve over most of the range: synthetic samples are on average marginally farther from real samples than real samples are from one another, the behaviour expected from a generator that captures the distribution while retaining diversity.

VII. Conclusion and Future Work

This paper presented PhysDiff-Net, a physics-guided conditional diffusion framework for generating synthetic WiFi network traffic data. The proposed approach addresses three fundamental requirements for network digital twins: statistical consistency, temporal controllability, and physics compliance. PhysDiff-Net integrates three key innovations: (i) variational hour conditioning with an auxiliary classification

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046



(a) Temporal Steering Capability (b) Overall Quality Score

(c) What-If Scenario Coverage

FIGURE 9. **Downstream utility evaluation for network what-if simulations. (a) Temporal steering capability showing PhysDiff-Net's ability to generate diverse hourly patterns (threshold: 1.5). (b) Overall quality score where only PhysDiff-Net achieves 100% Pattern Score. (c) Scenario coverage across four representative network conditions demonstrating PhysDiff-Net's versatility for operational planning.**

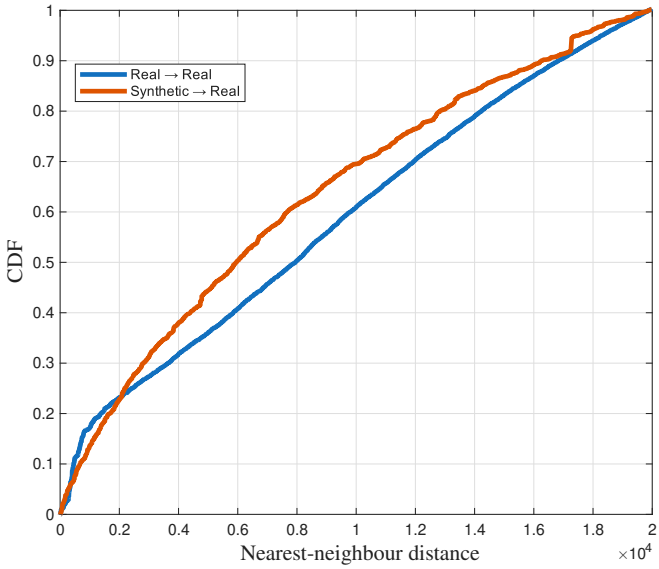


FIGURE 10. **Empirical CDF of nearest-neighbour distances for real–real and synthetic–real sample pairs. The close alignment of the curves indicates that synthetic samples follow the same local geometric structure as real data, while the slightly larger synthetic–real distances indicate distribution learning without memorisation.**

loss that enables the generation of distinct traffic profiles corresponding to different times of day; (ii) physics-guided training losses that impose the SNR–density anti-correlation of interference physics as a domain prior, so that generated data conforms to propagation principles in the high-load regimes that the observational record under-represents; and (iii) VGDS, an inference-time mechanism that resolves the trade-off between strong classifier-free guidance and appropriate output variance.

Experimental evaluation on real-world WiFi measurements from the Blackpool deployment showed that PhysDiff-Net provides the strongest overall balance across the evaluated statistical, temporal, and domain-prior criterias. PhysDiff-Net achieved a 100% Pattern Score, the only method to satisfy all physics constraints, while maintaining strong feature correlation preservation ($\rho_{\mathrm{FC}} = 0.856$) across the 29 engineered features characterising the 11-AP network. The physics-guided loss enforced the expected SNR-density anti-correlation ($\rho_{\mathrm{SD}} = -0.835$), well past the $\rho_{\mathrm{SD}} < -0.3$ target. The ablation study confirmed that the physics-guided loss is the component responsible for physics compliance: without it, the same backbone reproduces the near-zero correlation of an unconstrained fit rather than the imposed interference relationship.

The physics-guided loss here targets the SNR–density relationship as one instance of embedding wireless-domain knowledge into the generative process, but the formulation is modular. It can incorporate further wireless-domain relationships such as path-loss consistency across access points, interference coupling between neighbouring cells, SINR–throughput relations derived from capacity bounds, or load–latency trade-offs, which would extend the framework to richer domain knowledge for network digital twins and data-driven optimisation.

Several directions remain for future investigation. First, extending PhysDiff-Net to multi-site generation would support training across heterogeneous deployments and transfer between geographical locations. Second, generating temporally consistent sequences rather than independent snapshots would broaden the simulation scenarios the model can support. Third, integration with O-RAN intelligent controllers would enable online data augmentation within reinforcement-learning-based network optimisation. Finally, distilling the sampler to a few steps, for example with consistency models or DPM-Solver, could reduce inference time below the current 50-step DDIM cost for latency-sensitive deployment.

The proposed framework provides a foundation for trustworthy synthetic data generation in network digital twins. By bridging data-driven generative modeling with domain-specific physical constraints, PhysDiff-Net demonstrates that physics-aware deep learning can produce both statistically valid and physically plausible network data, supporting the development and evaluation of intelligent algorithms for beyond-5G and 6G wireless systems.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

ACKNOWLEDGMENT

This work was supported in part by the UK Engineering and Physical Sciences Research Council (EPSRC), Hub for All Spectrum Communications, O-RAN intelligent adaptive load balancing and efficiency in highly dense deployments (Orlando) project, under Grant No. EP/X040569/1.

REFERENCES

[1] "PhysDiff-Net: Physics-guided diffusion for synthetic wireless traffic generation," GitHub repository, 2026. [Online]. Available: https://github.com/kfatehi66/PhysDiff-Net

[2] Y. Shi, L. Lian, Y. Shi, Z. Wang, Y. Zhou, L. Fu, L. Bai, J. Zhang, and W. Zhang, "Machine learning for large-scale optimization in 6G wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2088–2132, 2023.

[3] L. Liu, X. Chen, and W. Zhang, "OpenPathNet: An open-source rf multipath data generator for AI-driven wireless systems," *arXiv preprint arXiv:2512.17286*, 2025.

[4] J. L. Guerra, C. Catania, and E. Veas, "Datasets are not enough: Challenges in labeling network traffic," *Computers & Security*, vol. 120, p. 102810, 2022.

[5] W. O'Brien, A. Dooley, M. Penica, S. McGrath, and E. O'Connell, "Data-driven, real-time diagnostics of 5G and Wi-Fi networks using mobile robotics," *Journal of Sensor and Actuator Networks*, vol. 14, no. 6, p. 110, 2025.

[6] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems," *Ieee Access*, vol. 8, pp. 165 130–165 150, 2020.

[7] Z. Lin, A. Jain, C. Wang, G. Fanti, and V. Sekar, "Using gans for sharing networked time series data: Challenges, initial promise, and open questions," in *Proceedings of the ACM internet measurement conference*, 2020, pp. 464–483.

[8] Y. Huangfu, J. Wang, S. Dai, R. Li, J. Wang, C. Huang, and Z. Zhang, "Wair-d: Wireless AI research dataset," *arXiv preprint arXiv:2212.02159*, 2022.

[9] K.-C. Wang, R. Hutchins, and A. Nikolich, "Data for AI in network systems workshop report," 2023.

[10] M. A. Mohsin, J. Ahmad, M. H. Nawaz, and M. A. Jamshed, "Towards 6G intelligence: The role of generative AI in future wireless networks," *arXiv preprint arXiv:2508.19495*, 2025.

[11] G. Aceto, F. Giampaolo, C. Guida, S. Izzo, A. Pescapè, F. Piccialli, and E. Prezioso, "Synthetic and privacy-preserving traffic trace generation using generative AI models for training network intrusion detection systems," *Journal of Network and Computer Applications*, vol. 229, p. 103926, 2024.

[12] K. Muhammad, T. David, G. Nassisid, and T. Farus, "Integrating generative AI with network digital twins for enhanced network operations," *arXiv preprint arXiv:2406.17112*, 2024.

[13] E. Mikołajewska, D. Mikołajewski, T. Mikołajczyk, and T. Paczkowski, "Generative AI in AI-based digital twins for fault diagnosis for predictive maintenance in industry 4.0/5.0," *Applied Sciences*, vol. 15, no. 6, p. 3166, 2025.

[14] R. Liang, B. Yang, Z. Yu, B. Guo, X. Cao, M. Debbah, H. V. Poor, and C. Yuen, "DiffSG: A generative solver for network optimization with diffusion model," *IEEE Communications Magazine*, vol. 63, no. 6, pp. 16–24, 2025.

[15] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[16] N. Sivaroopan, D. Bandara, C. Madarasingha, G. Jourjon, A. P. Jayasumana, and K. Thilakarathna, "Netdiffus: Network traffic generation by diffusion models through time-series imaging," *Computer Networks*, vol. 251, p. 110616, 2024.

[17] C. Zhang, Y. Zhang, Q. Shao, B. Li, Y. Lv, X. Piao, and B. Yin, "Chattraffic: Text-to-traffic generation via diffusion model," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[18] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687*, 2022.

[19] A. Celik and A. M. Eltawil, "At the dawn of generative AI era: A tutorial-cum-survey on new frontiers in 6G wireless intelligence," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2433–2489, 2024.

[20] T.-H. Vu, S. K. Jagatheesaperumal, M.-D. Nguyen, N. Van Huynh, S. Kim, and Q.-V. Pham, "Applications of generative AI (GAI) for mobile and wireless networking: A survey," *IEEE Internet of Things Journal*, 2024.

[21] T. J. Anande and M. S. Leeson, "Generative adversarial networks (GANS): a survey of network traffic generation," *International Journal of Machine Learning and Computing*, vol. 12, no. 6, pp. 333–343, 2022.

[22] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, and C.-X. Wang, "Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 22–27, 2019.

[23] T. J. Anande, S. Al-Saadi, and M. S. Leeson, "Generative adversarial networks for network traffic feature generation," *International Journal of Computers and Applications*, vol. 45, no. 4, pp. 297–305, 2023.

[24] G. Kakkavas, M. Kalntis, V. Karyotis, and S. Papavassiliou, "Future network traffic matrix synthesis and estimation based on deep generative models," in *2021 international conference on computer communications and networks (ICCCN)*. IEEE, 2021, pp. 1–8.

[25] K. C. van der Deijl, "Synthetic 5G traffic generation: A machine learning approach," Ph.D. dissertation, Delft University of Technology, 2025.

[26] D. A. Ammara, J. Ding, and K. Tutschku, "Synthetic network traffic data generation: A comparative study," *arXiv preprint arXiv:2410.16326*, 2024.

[27] H. Chai, T. Jiang, and L. Yu, "Diffusion model-based mobile traffic generation with open data for network planning and optimization," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 4828–4838.

[28] P. Yan, H. Zeng, and Y. T. Hou, "xDiff: Online diffusion model for collaborative inter-cell interference management in 5G O-RAN," *IEEE Transactions on Networking*, 2025.

[29] S. Zhang, T. Li, D. Jin, and Y. Li, "Netdiff: a service-guided hierarchical diffusion model for network flow trace generation," *Proceedings of the ACM on Networking*, vol. 2, no. CoNEXT3, pp. 1–21, 2024.

[30] S. Zhang, H. Chai, Y. Li, B. Qiu, L. Yue, and R. Pan, "Packetdiff: A flow guided diffusion model for network packet trace generation," *IEEE Internet of Things Journal*, 2025.

[31] H. Chai, T. Li, F. Jiang, S. Zhang, and Y. Li, "Knowledge guided conditional diffusion model for controllable mobile traffic generation," in *Companion Proceedings of the ACM Web Conference 2024*, 2024, pp. 851–854.

[32] H. Chai, X. Qi, and Y. Li, "Spatio-temporal knowledge driven diffusion model for mobile traffic generation," *IEEE Transactions on Mobile Computing*, 2025.

[33] H. Chai, B. Qiu, X. Mo, R. Pan, Z. Wang, and Y. Li, "MCDiff: Mobile traffic and user generation with multimodal context-aware diffusion model," *IEEE Transactions on Networking*, 2026.

[34] Q. Yu, C. Yan, X. Liu, and H. Wang, "FlowDiffGAN: High-fidelity network flow generation via diffusion-guided adversarial training," in *Proceedings of the 2025 2nd International Conference on Image Processing, Intelligent Control and Computer Engineering*, 2025, pp. 383–388.

[35] C.-L. Wu, Y.-Y. Chen, P.-Y. Chou, and C.-Y. Wang, "Synthetic packet traffic generative adversarial networks in multi agents with peer-to-peer and global priority queue generation," *IEEE Transactions on Network Science and Engineering*, 2026.

[36] B. Böck, A. Oeldemann, T. Mayer, F. Rossetto, and W. Utschick, "Physics-informed generative modeling of wireless channels," *arXiv preprint arXiv:2502.10137*, 2025.

[37] K. Qiu, S. Bakirtzis, I. Wassell, H. Song, K. Lin, and J. Zhang, "IRDM: A generative diffusion model for indoor radio map interpolation," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 01–06.

[38] X. Gong, X. Liu, A.-A. Lu, X. Gao, X.-G. Xia, C.-X. Wang, and X. You, "Digital twin of channel: Diffusion model for sensing-assisted statistical channel state information generation," *IEEE Transactions on Wireless Communications*, 2025.

[39] X. Luo, L. Zhizhen, Z. Peng, X. Dongkuan, and Y. Liu, "Rm-gen: Conditional diffusion model-based radio map generation for wireless

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046

networks," in *2024 IFIP Networking Conference (IFIP Networking)*. IEEE, 2024, pp. 543–548.
[40] H. Jia, W. Chen, Z. Huang, H. Xiao, N. Jia, K. Wu, S. Lai, and Y. Yue, "RMDM: Radio map diffusion model with physics informed," *arXiv e-prints*, pp. arXiv–2501, 2025.
[41] Z. Liu, S. Zhang, Q. Liu, H. Zhang, and L. Song, "Wifi-diffusion: Achieving fine-grained wifi radio map estimation with ultra-low sampling rate by diffusion models," *IEEE Journal on Selected Areas in Communications*, 2025.
[42] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless lan," in *Proceedings of the 2002 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, 2002, pp. 195–205.
[43] B. Fu, Y. Xiao, H. Deng, and H. Zeng, "A survey of cross-layer designs in wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 110–126, 2013.
[44] R. F. Nalewajski, "Elements of information theory," in *Perspectives in Electronic Structure Theory*. Springer, 2011, pp. 371–395.
[45] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 6, p. 066138, 2004.
[46] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.
[47] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning (ICML)*. PMLR, 2021, pp. 8162–8171.
[48] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations (ICLR)*, 2014.
[49] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.
[50] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
[51] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 18 661–18 673.
[52] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations (ICLR)*, 2021.
[53] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.
[54] M. Niedoba, D. Green, S. Naderiparizi, V. Lioutas, J. W. Lavington, X. Liang, Y. Liu, K. Zhang, S. Dabiri, A. Ścibior *et al.*, "Nearest neighbour score estimators for diffusion generative models," *arXiv preprint arXiv:2402.08018*, 2024.



Kavan Fatehi received the M.Sc. degree from Yazd University, Yazd, Iran, in 2014, and the Ph.D. degree in computer science from the University of Nottingham, Nottingham, U.K., in 2024, with a focus on self-supervised learning for low-resource automatic speech recognition. He was a Postdoctoral Research Associate at the Institute for Safe Autonomy, University of York, York, U.K. He is now with the Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, U.K. His research interests include explainable reinforcement learning, multi-agent systems, self-supervised representation learning, machine learning for wireless networks, and multimodal speech analysis for mental health.



Mostafa Rahmani Ghourtani (Member, IEEE) received the B.Sc. degree in electrical engineering from Shiraz University, Shiraz, Iran, in 2009; the M.Sc. degree in communication systems engineering from Tarbiat Modares University, Tehran, Iran, in 2012; and the Ph.D. degree in communication systems engineering from Shiraz University of Technology, Shiraz, Iran, in 2023. He is currently a Research Fellow at the University of Surrey, Guildford, U.K. His research focuses on the design, analysis, and performance evaluation of advanced wireless communication systems for 5G and beyond, including 6G networks. His interests include MIMO and near-field MIMO, cell-free massive MIMO, Open RAN (O-RAN), and non-terrestrial networks. He also investigates machine-learning and deep-reinforcement-learning methods for radio resource allocation and network optimisation, as well as physical-layer techniques for high-mobility settings, including railway communications and train-to-infrastructure backhaul. His work spans analytical modelling, algorithm development, and link-level and system-level simulation. He is particularly interested in connecting theoretical methods with practical wireless system design and evaluation.



David Grace (SM, IEEE) received the Ph.D. degree from the University of York, in 1999. His Ph.D. dissertation was titled "Distributed Dynamic Channel Assignment for the Wireless Environment." Since 1994, he has been a member with the Department of Electronic Engineering, University of York, where he is currently a Professor (Research), the Head of the Communication Technologies Research Group, and the Director of the Centre for High Altitude Platform Applications. In 2000, he jointly founded SkyLARC Technologies Ltd. and was one of its directors. From 2014 to 2018, he was the Non-Executive Director of Stratospheric Platforms Ltd. He is a Lead Investigator on H2020 MCSA SPOTLIGHT, U.K., Government-funded MANY, dealing with 5G trials in rural areas, and HiQ investigating quantum key distribution from high-altitude platforms. He was the Technical Lead on the 14-partner FP6 CAPANINA Project that dealt with broadband communications from high-altitude platforms. He is the author of more than 280 articles and the author/editor of two books. His current research interests include aerial platform-based communications, the application of artificial intelligence to wireless communications, 5G system architectures, dynamic spectrum access, and interference management. He is a Founding Member of the IEEE Technical Committee on Green Communications and Computing. He was the Former Chair of the IEEE Technical Committee on Cognitive Networks, from 2013 to 2014.



Mahnaz Arvaneh is a Professor of Intelligent Human–Machine Interfaces at the University of Sheffield, U.K., and Director of the Physiological Signals and Systems Laboratory. Her research interests include brain–computer interfaces, EEG signal processing, machine learning, physiological computing, human–machine interaction, and AI-enabled neurotechnologies for healthcare and rehabilitation. She has authored more than 90 peer-reviewed publications and has led multiple interdisciplinary research projects in neurotechnology, artificial intelligence, and rehabilitation engineering. She serves as Associate Editor for the IEEE Transactions on Neural Systems and Rehabilitation Engineering, IEEE Transactions on Biomedical Engineering, and Scientific Reports. She was a contributor to the Royal Society report on Neural Interface Technologies and is a co-editor of the IET book Machine Learning and Signal Processing in Brain–Machine Interfaces.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Machine Learning in Communications and Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMLCN.2026.3741046



Kin K Leung (Fellow, IEEE) received his M.S. and Ph.D. degrees from University of California, Los Angeles. He worked AT&T Bell Labs and its successor companies in New Jersey for 19 years. Since 2004, he has been the Tanaka Chair Professor in the Electrical and Electronic Engineering (EEE), and Computing Departments at Imperial College. He is the Head of Communications and Signal Processing Group in the EEE Department at Imperial. His research focuses on networking, protocols, optimization and modeling of wireless broadband, sensor and ad-hoc networks. He also works on multi-antenna systems and cross-layer optimization of these networks. He received the Distinguished Member of Technical Staff Award from AT&T Bell Labs in 1994, and was a co-recipient of the 1997 Lanchester Prize Honorable Mention Award. He was elected as an IEEE Fellow in 2001. He received the Royal Society Wolfson Research Merits Award from 2004 to 2009 and became a member of Academia Europaea in 2012. Along with his co-authors, he also received best paper awards at major conferences, including the IEEE ICC 2019, ICDCS 2013 and PIMRC 2012. He was a member (2009-11) and the chairman (2012-15) of the IEEE Fellow Evaluation Committee for Communications Society. He has served on the editorial boards for 10 leading technical journals.



Hamed Ahmadi (SM'15, IEEE) s a Professor in Digital Engineering, at the School of Physics, Engineering and Technology, University of York, UK. He received his Ph.D. from National University of Singapore in 2012 where he was an SINGA PhD scholar at Institute for Infocomm Research, A-STAR. Since then he worked at different academic and industrial positions in the Republic of Ireland and UK. Prof. Ahmadi has published a book and more than 100 peer reviewed book chapters, journal and conference papers. He was the Associate Editor in Chief of IEEE Communication Standards magazine (2022-2025), and Associate Editor in IEEE System, IEEE Access, and Wireless Networks. He is now serving as an Associate Editor at IEEE Transactions on Green Communications and Networking, and IEEE Networking Letters. He is a senior member of IEEE, and Fellow of UK Higher Education Academy. He has been the Networks working group chair of COST Actions CA15104 (IRACON) and CA20120 (INTERACT). He had chairing roles in organising and technical programme committee of several IEEE flagship conferences including IEEE ICC 2024, EUCNC 2025 and 2019, and PIMRC 2019, 2024, and 2026. He and his team were the recipients of IEEE PIMRC 2025, and WINCOM2024 best paper awards. His current research interests include the application of machine learning in wireless networks, Open Radio Access and Networking, green and sustainable networks, airborne networks, Digital twins of networks, and Internet-of-Things.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/