Building the Cognitive Network: Pillars of AI-Native Wireless ecosystem

Supratik Bhattacharjee
Apple Inc.
San Diego, CA 92121
sbhattacharjee22@apple.com

Sharad Sambhwani
Apple Inc.
San Diego, CA 92121
ssambhwani@apple.com

Abstract

Classical wireless system design, which has underpinned generations up to 5G, depends on simplified or approximately linear mathematical models that are increasingly insufficient for the unprecedented complexity of future 6G and beyond networks. This paper proposes a paradigm shift to an AI-Native framework that makes the network inherently intelligent, rather than just managed by AI. This vision is built on six pillars, including reinforcement learning for self-optimization, predictive forecasting, a learned physical layer, generative models for digital twins, multi-modal data fusion, and hyper-local model management. We argue that this AI-Native approach represents not an incremental improvement, but a necessary evolution, transforming wireless systems into autonomous, adaptive ecosystems capable of meeting the demands of a hyper-connected future.

1 The Paradigm Shift: Beyond the Limits of Classical Design

For decades, wireless design has been a triumph of mathematical modeling. We build systems starting from well-understood models of the physical world, but suitably augment them with well thought out simplifications to make them mathematically tractable: the Gaussian channel, Rayleigh fading, and predictable interference patterns. These models enabled the creation of 2G through 5G, but they are reaching their limits. The future wireless environment—envisioned for 6G and beyond—is one of unprecedented complexity: trillions of connected devices, ultra-high frequencies (existing and new frequencies particularly around 7 GHz), new deployment scenarios with various architectural or topological options, and extreme demands for latency and reliability. Designing for this environment with static, human-derived models is like trying to navigate a bustling city with a hand-drawn map from a century ago. It is inefficient, brittle, and incapable of adapting to real-time conditions.

The vision is to transition from this model-based paradigm to an AI-Native one. [1], [2] The wireless network will cease to be a system that is merely *managed* by intelligence; it will become a system that is intelligent. It will learn, adapt, and optimize itself in real time, treating the radio environment not as a problem to be modeled, but as a sea of data to be understood and manipulated.

2 The Pillars of the AI-Native Wireless Vision

This transformation will be built on several key pillars, each leveraging a different facet of Machine Learning.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI and ML for Next-Generation Wireless Communications and Networking(AI4NextG).

2.1 Pillar 1: The Self-Optimizing Network: Real-Time Adaptation at Millisecond Scale

Today's networks are optimized offline for "average" or "worst-case" scenarios. The AI-native network will optimize itself continuously for the *current* scenario.

- **Core Concept:** The network operates as a *Reinforcement Learning (RL) agent*. Its "state" is the current channel conditions, user locations, and interference patterns. Its "actions" are resource allocation [3], beamforming angles, and power control. Its "reward" is maximizing throughput, minimizing latency, improving system efficiency and saving energy.
- **ML in Action:**
  - **Dynamic Spectrum Access:** An RL agent learns which frequency bands are free, combines information across multi-modality(RF,Camera,Lidar,radar) and allocates them to users in real-time, achieving spectral efficiency far beyond static allocation. [4]
  - **Intelligent Beamforming:** Instead of relying on imperfect channel state information (CSI), a Deep RL agent learns to form beams by directly observing the resulting data rates, finding optimal paths that classical models would miss, including reflections and diffractions.
  - **Interference Coordination:** In dense networks, multiple access points will learn cooperative strategies on their own, deciding which AP serves which user to nullify interference without a central controller dictating every move.
- **The Outcome:** A fluid, "liquid" network that molds itself around user demands and environmental dynamics, achieving unprecedented efficiency and reliability.

2.2 Pillar 2: The Predictive System: From Reactive to Proactive

Current networks react to events—a user moves, a channel fades, congestion occurs. The AI-native network will anticipate them.

- **Core Concept:** The system uses time-series forecasting models (like LSTMs and Transformers) to predict the future state of the environment both at the network and at the device level.
- **ML in Action:**
  - **Predictive Mobility Management:** By learning a user's movement patterns in a privacy-sensitive fashion the network and or device can pre-emptively hand over the connection to the next cell tower, ensuring zero interruption. [5]
  - **Proactive Resource Allocation:** The network predicts traffic hotspots (e.g., a stadium during a match) and allocates more resources to that area *before* congestion occurs.
  - **Intelligent Load Balancing:** Machine learning enables intelligent load balancing by training models to recognize and predict complex traffic patterns, allowing the network to proactively redistribute users from potential hotspots to underutilized cells. This preemptive action prevents congestion and maintains a high-quality user experience across the network.
  - **Channel Prediction:** ML models can predict how a channel will evolve over the next few milliseconds, allowing the transmitter and receiver to suitably adapt their processing for a future channel state, and thus dramatically improving spectral efficiency, conserve energy and increase reliability.
  - **Traffic Classification and Prediction:** ML Models can be used to *classify* different type of traffic at IP Level (Real-time Video, Streaming Playback, Gaming, Web Browsing) and *predict* the associated Burst arrival time, duration and volume. This information can then be used to adapt the end to end wireless protocol stack and thus delivering a better QOS to the user.[6]
- **The Outcome:** A system of network and devices that operates ahead of the curve, delivering a seamless, predictable, smooth and hyper-personalized user experience.

2.3 Pillar 3: The End-to-End Learned Physical Layer: Rethinking Communication from the Ground Up

The traditional physical layer (PHY) is a chain of distinct, separately optimized blocks: source coding, channel coding, modulation, etc. This is inherently suboptimal.

- **Core Concept:** We replace the entire transmitter/receiver chain with a two-sided deep neural network—an *autoencoder* The transmitter (encoder) learns the best way to represent information and map it to a waveform, while the receiver (decoder) learns the best way to interpret that waveform, all without human-specified rules. [7], [8]
- **ML in Action:**
  - **Joint Source-Channel Coding:** The autoencoder learns to create a "constellation" of signals that is optimized for both the specific data being sent (e.g., an image) and the specific channel it's passing through.
  - **Adaptive Modulation and Coding:** The system learns a communication scheme that is perfectly tailored to the live, non-linear, and "dirty" channel, outperforming standard schemes like QAM which assume a perfect, linear channel.
  - **Neural Receiver:** The receiver performs Joint channel estimation and demodulation and is inherently optimal as compared to individually optimized blocks.
  - **Non-Linear operating regimes:** Enable RF circuits to operate in non-linear regimes by taking advantage of ML based corrective schemes.
- **The Outcome:** A fundamental breakthrough in communication theory, where the system discovers novel communication strategies that are more robust and efficient than those conceived by humans.

2.4 Pillar 4: The Generative eco-system: AI as a Design Partner

The design and planning of wireless networks is a slow, expensive process involving complex simulations. Generative AI will change this.

- **Core Concept:** We use *Generative Adversarial Networks (GANs)* or *Diffusion Models* to create a "digital twin" of the entire wireless environment. This twin is not just a simulation; it is a generative model that can create infinite, realistic scenarios.
- **ML in Action:**
  - **Generative Channel Models:** Instead of relying on simplistic statistical models, a GAN can be trained on real-world channel measurements as well as real operation based experience (Loading effects based on time of the day, weather etc.) to generate new, ultra-realistic channel data for any environment, drastically improving the accuracy of simulations.[9], [10]
  - **Automated Network Planning:** An AI can generate and evaluate millions of potential network layouts (e.g., where to place cell towers or Reconfigurable Intelligent Surfaces) in a virtual environment to find the optimal design before a single piece of hardware is deployed.
  - **Cloud-based Design:** Digital twin records past and current parameter exchanges between devices and the network, analogous to a fingerprint, which can then be used as an additional input to make intelligent decisions.
  - **Generative Waveform and Protocol Design:** By leveraging their generative capabilities, LLMs can be used to invent and optimize novel, end-to-end communication schemes that transcend human-designed heuristics, enabling the creation of dynamic, hyper-efficient waveforms or semantic communication protocols tailored in real-time to specific channel conditions and application needs
- **The Outcome:** The wireless design cycle is accelerated from months to hours, leading to more innovative and cost-effective network deployments.

2.5 Pillar 5: Multi-Modality

Traditional wireless design has relied solely on RF inputs, often overlooking rich contextual information provided by sensors, user intent, past history, and more.

- **Core Concept:** Multi-modality in AI-based wireless design is the practice of creating models that can simultaneously process and learn from diverse data types, such as radio frequency (RF) signals, network traffic logs, user location, and device sensor data. By fusing these different inputs, the AI gains a more holistic and accurate understanding of the network environment, leading to more intelligent and robust decision-making. [11]
- **ML in Action:**
  - **Holistic Contextual Awareness:** By integrating diverse data streams—such as RF signals, network traffic patterns, and device sensor data—the AI gains a complete and context-rich understanding of the network environment. This comprehensive view allows the system to move beyond simple metrics and make decisions based on the "why" behind network events, not just the "what."
  - **Enhanced Robustness and Resilience:** Multi-modality makes the AI system more robust by reducing its dependence on any single data source. If one modality is noisy, corrupted, or unavailable (e.g., a weak GPS signal), the model can still make reliable inferences using the remaining data streams, ensuring consistent and stable network operation.
  - **Data Fusion and Cross-Modal Learning:** This is the core technical challenge, involving the fusion of heterogeneous data into a unified format that an AI model can process. Through cross-modal learning, the AI discovers complex relationships between different data types, enabling it to unlock novel insights and predictive capabilities that would be impossible with any single modality alone.
- **The Outcome:** This can go a long way in delivering new immersive user experience while utilizing sensing and context driven communications

2.6 Pillar 6: Hyper-local Life Cycle Management

Throughout its progression from 2G to 5G, the wireless ecosystem has adhered to a highly structured software update lifecycle. With the convergence of wireless and artificial intelligence, it is now critical to establish a parallel framework for AI model management. Such a framework is essential for providing robust monitoring, systematic updates, and necessary fallback procedures for deployed models.

- **Core Concept:** AI model lifecycle management addresses the fact that a model's performance can degrade after deployment as real-world data changes. It provides a structured framework for continuously monitoring, retraining, and redeploying models in a timely fashion and thus maintain their effectiveness and manage risk.[12]
- **ML in Action:**
  - **Model Monitoring** Model Monitoring is the continuous observation of a deployed AI model's performance and behavior in a live production environment. Its primary goal is to detect issues like data drift (when input data changes) or performance degradation, which indicate that the model is becoming less accurate and may need to be updated.
  - **Model Retraining** Model retraining involves updating an existing model by training it with new or more representative data. This action is triggered when monitoring detects a significant drop in performance, ensuring the model adapts to changes in the real world and regains its accuracy.
  - **Model Redeploying** Model Redeployment is the process of replacing the outdated model in the production environment with the newly retrained and validated version. This final step is carefully managed to ensure a seamless transition, allowing users to benefit from the improved performance of the updated model without service interruption.
- **The Outcome:** Efficient model lifecycle management in resource-constrained wireless systems is achieved by avoiding full, costly retraining. Instead, it utilizes techniques like parameter-efficient fine-tuning to generate only lightweight 'delta' updates. These minimal updates can then be cheaply deployed to keep on-device models current without significant computational overhead. A successful model lifecycle management framework transforms the wireless ecosystem into an agile platform for innovation. It allows operators

and manufacturers to rapidly and safely deploy, monitor, and enhance AI-driven features, accelerating the delivery of new, intelligent services to customers.

3 The Roadmap to Realization

This vision will not be realized overnight. It will unfold in phases:

1. **Phase 1 - ML Augmented (Present - 3 years):** ML is used as a powerful tool to solve specific, isolated problems within the existing framework (e.g., better channel estimation, anomaly detection).
2. **Phase 2 - ML Automated (3 - 7 years):** Key control loops within the network become fully automated by ML agents (e.g., self-optimizing network slices, autonomous beam management). Humans transition from being operators to becoming supervisors.
3. **Phase 3 - Fully Cognitive Network (7+ years):** The network operates as a holistic, cognitive system. End-to-end learning becomes viable, and the network can autonomously adapt to entirely new applications and environments. This is the true 6G/7G vision.

4 Conclusion

The integration of Machine Learning into wireless design is not an incremental improvement; it is a paradigm shift. It will transform wireless systems from static, human-engineered constructs into living, autonomous ecosystems. The future network will be *cognitive*: aware of its environment, predictive of the future, and capable of learning and evolving to meet demands we can't even yet imagine. This is the path to a truly intelligent and connected world.

References

[1] Zhijin Qin, Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang. Deep learning in physical layer communications. *IEEE Wireless Communications*, 26(2):93–99, 2019.

[2] Alessio Zappone, Marco Di Renzo, and Mérouane Debbah. Wireless networks design in the era of deep learning: Model-based, ai-based, or both? *IEEE Transactions on Communications*, 67(10):7331–7376, 2019.

[3] Le Liang, Hao Ye, Guanding Yu, and Geoffrey Ye Li. Deep-learning-based wireless resource allocation with application to vehicular networks. *Proceedings of the IEEE*, 108(2):341–356, 2019.

[4] Yue Xu, Jianyuan Yu, William C Headley, and R Michael Buehrer. Deep reinforcement learning for dynamic spectrum access in wireless networks. In *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, pages 207–212. IEEE, 2018.

[5] Pooyan Jahanmanesh, Amirfarhad Farhadi, and Azadeh Zamanifar. Mobility management with ai. *IEEE Access*, vol. 13, 2025.

[6] Jorge Gómez, Velssy Hernandez Riaño, and Gustavo Ramirez-Gonzalez. Traffic classification in ip networks through machine learning techniques in final systems. *IEEE Access*, vol. 11, 2023.

[7] Timothy O'shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

[8] José Miguel Mateos-Ramos, Jinxiang Song, Yibo Wu, Christian Häger, Musa Furkan Keskin, Vijaya Yajnanarayana, and Henk Wymeersch. End-to-end learning for integrated sensing and communication. In *ICC 2022-IEEE International Conference on Communications*, pages 1942–1947. IEEE, 2022.

[9] Xuanhao Luo, Zhizhen Li, Zhiyuan Peng, Mingzhe Chen, and Yuchen Liu. Denoising diffusion probabilistic model for radio map estimation in generative wireless networks. *IEEE Transactions on Cognitive Communications and Networking*, 2025.

[10] Xuanyu Liu, Shijian Gao, Boxun Liu, Xiang Cheng, and Liuqing Yang. Llm4wm: Adapting llm for wireless multi-tasking. *IEEE Transactions on Machine Learning in Communications and Networking*, 2025.

[11] Lu Cheng, Hongliang Zhang, Boya Di, Dusit Niyato, and Lingyang Song. Large language models empower multimodal integrated sensing and communication. *IEEE Communications Magazine*, 2025.

[12] Farhad Rezazadeh, Hatim Chergui, Luis Alonso, and Christos Verikoukis. Sliceops: Explainable mlops for streamlined automation-native 6g networks. *IEEE Wireless Communications*, 31(5):224–230, 2024.