International Journal of Engineering & Technology Sciences (ISSN: 2289-4152)

Volume 2025, 60 pages
Article ID: IJETS-2508132112961
HTPub, an imprint of Spot Publisher LLC

Review Article

A Comprehensive Review of Digital Signal Processing (DSP) Algorithms and Their Applications in Telecommunication and Wireless Communication Systems

Saman Sadeghi [a*]

Department of Electrical Engineering-Electronics, Mazandaran University of Science and Technology, Babol, Iran

Corresponding authors: Saman Sadeghi, Email: Samansdi30@gmail.com

Date Received: 01-11-2025, Date revised: 03-23-2025, Date of accepted: 05-30-2025;

Abstract

Digital Signal Processing (DSP) is critical in the development and optimization of modern telecommunication and wireless systems. The advent of complex standards like 5G and AI-native networks has increased the need to revisit classical and new DSP algorithms. This paper presents a systematic review of DSP techniques that include time-domain, frequency-domain, and advanced signal processing methods, as well as new AI-based approaches, such as Deep Reinforcement Learning (DRL), Neural Architecture Search (NAS), Graph Neural Networks (GNN), and Bayesian Optimization (BO). The survey explores basic DSP principles, historical developments, and key performance metrics like latency, throughput, and computational complexity. It compares classical and new algorithms, analysing their merits, limitations, and suitability for real-world applications, such as modulation schemes (QAM, OFDM), channel estimation, noise suppression, MIMO, and beamforming. Additionally, the paper addresses the integration with Software-Defined Radio (SDR), edge computing, and the Internet of Things (IoT), highlighting challenges related to real-time implementation and hardware acceleration using FPGA and ASIC platforms. Key research challenges and future directions are identified, such as the need for scalable, adaptive DSP in dynamic environments, power-efficient hardware implementation of AI models, and the growing importance of optimization in future wireless systems. This review is a comprehensive resource for researchers at the intersection of signal processing, wireless communication, and intelligent systems.

Keywords: Digital Signal Processing (DSP), Telecommunication Systems, Wireless Communication, 5G, Artificial Intelligence, Deep Reinforcement Learning (DRL), Neural Architecture Search (NAS), Graph Neural Networks (GNN), Bayesian Optimization (BO), Quadrature Amplitude Modulation (QAM), Orthogonal Frequency Division Multiplexing (OFDM), Channel Estimation, MIMO, Beamforming, Software-Defined Radio (SDR), Edge Computing, Internet of Things (IoT), FPGA, ASIC, Hardware Acceleration, Real-Time Systems, Adaptive DSP, Signal Processing Optimization

Introduction

Digital signal processing (DSP) serves as a foundational pillar in modern wireless and telecommunication systems, enabling essential operations such as filtering, spectral analysis, modulation, and channel equalization across high-speed communication networks. Decision-feedback equalization (DFE), for example, is widely implemented in systems such as WiFi, cable modems, and LTE to mitigate inter-symbol interference and dynamically adapt to fluctuating channel conditions (Farooq et al., 2025). Simultaneously, advanced DSP hardware architectures have emerged to meet the demands of 5G baseband processing and beyond, offering programmable multirate filtering capabilities and accommodating diverse modulation schemes at ultra-high sampling rates (Ghosh et al., 2023). Massive MIMO systems, integral to enhancing spectral efficiency, energy utilization, and network coverage, are central to current 5G and future 6G infrastructures (Kamil et al., 2025). Within these systems, DSP algorithms, particularly maximum likelihood detectors, are crucial for symbol detection in high-dimensional MIMO environments. Moreover, to preserve signal fidelity under noisy and dynamic conditions, adaptive signal-processing techniques are extensively employed. Adaptive equalizers, in particular, continuously adjust their parameters to counter multipath distortions, thereby enhancing communication reliability (Khan et al., 2023).

The evolving demands of modern communication systems have driven the advancement of sophisticated digital signal processing (DSP) techniques that go beyond traditional time-domain and frequency-domain filters. For instance, discrete wavelet transforms (DWT) have been integrated into multi-carrier modulation schemes such as OFDM to improve spectral containment and eliminate the need for a cyclic prefix, thereby enhancing spectral efficiency (Wils et al., 2023). Compressive sensing techniques exploit the inherent sparsity of signals in mmWave massive MIMO channels to substantially reduce the measurement burden during channel estimation (Oyerinde et al., 2024). In parallel, deep learning-based approaches have emerged, with recent "learnable DSP" frameworks treating the entire DSP chain as a unified neural network, enabling joint optimization of all processing blocks via backpropagation (Niu et al., 2024). Collectively, these advanced methodologies aim to improve estimation accuracy, accelerate convergence, and enhance robustness against noise and interference. A systematic review of such techniques helps practitioners understand their trade-offs, such as computational complexity, latency, throughput, and energy consumption, and make informed decisions when selecting algorithms that best meet specific performance targets (Pryamikov et al., 2024).

This review provides a comprehensive survey of digital signal processing (DSP) algorithms and their practical applications in telecommunication and wireless communication systems. We begin by categorizing the algorithms based on their operational domains, time-domain versus frequency-domain, as well as by methodological approaches such as linear, nonlinear, learning-based, sparse, and multirate techniques. The subsequent discussion addresses key application areas, highlighting multi-modulation transmitter architectures (Cartesian, polar, outphasing) that utilize reconfigurable DSP hardware with multistage, multirate filtering to satisfy 5G NR requirements (Ghosh et al., 2023). Additionally, channel estimation and equalization techniques in LTE and 5G networks are reviewed, with recent literature providing comprehensive taxonomies including pilot-assisted, blind, and decision-directed methods alongside emerging trends such as reconfigurable intelligent surfaces (RIS) and machine learning–assisted estimation (Tarafder et al., 2025). Noise suppression and distortion mitigation techniques, including advanced equalizers like decision-feedback equalizers (DFE) and maximum likelihood detection (MLD), are also examined due to their significance in high-order QAM modulation schemes (Farooq et al., 2025). Throughout the review, emphasis is placed on real-world hardware implementations using FPGA and ASIC DSP chains aligned with communication standards, considering the impact of hardware constraints such as area, power consumption, and platform flexibility on algorithm design (Ghosh et al., 2023; Tarafder et al., 2025). Finally, the review discusses current challenges, hardware limitations, and open research directions aimed at guiding future advances in AI-augmented DSP systems within telecommunication and wireless communication domains.

Recent studies demonstrate that the integration of deep learning with digital signal processing (DSP) has resulted in a remarkably effective, data-driven approach (Pryamikov et al., 2024). Emerging artificial intelligence techniques, including deep reinforcement learning (DRL) and graph neural networks (GNN), have been successfully applied in wireless DSP applications. For instance, a DRL-based joint beamforming strategy for reconfigurable intelligent surface (RIS)-aided massive multiple-input multiple-output (MIMO) systems has shown significant performance improvements over traditional methods (Zhou et al., 2024). Likewise, attention-based GNNs have been utilized to map channel information to beamforming vectors, achieving notable gains compared to benchmark algorithms in multi-antenna networks (Sun et al., 2025). Furthermore, adaptive DSP frameworks such as Bayesian hyperparameter tuning automatically optimize receiver parameters, resulting in substantial efficiency enhancements; one study reported up to a 48% reduction in complexity and approximately 1.21 dB improvement in Q-factor relative to manually tuned DSP systems (Niu et al., 2024). Collectively, these advancements demonstrate how DRL, neural architecture search (NAS)-inspired automated design, GNNs, and Bayesian optimization facilitate adaptive beamforming, interference cancellation, and self-organizing functionalities in contemporary wireless networks.

To meet the stringent real-time requirements of 5G and 6G networks, hardware acceleration is indispensable. FPGA and ASIC platforms can execute AI-enhanced DSP pipelines with ultra-low latency. For instance, an edge orchestration platform equipped with hardware acceleration has demonstrated reductions in AI inference times by 50–220%, alongside more than a 20-fold increase in throughput (Palomares et al., 2025). In a notable 6G case study, an FPGA implementation (Xilinx Virtex-7) of a reservoir-computing (Echo State Network, ESN) symbol detector achieved low bit-error rates and high throughput in massive MIMO scenarios, all while maintaining minimal hardware resource usage (Lin et al., 2024). Software-defined radios (SDRs) further facilitate agile DSP chains that can be dynamically reconfigured by AI algorithms. Shifting DSP processing to edge and IoT nodes offers the advantages of ultra-low latency and enhanced privacy; for example, a wide-area IoT system integrating low Earth orbit (LEO) satellites with edge computing employed a reinforcement learning-based offloading algorithm to minimize end-to-end delay and maximize throughput (Zhao et al., 2025). Concurrently, federated learning provides privacy-preserving distributed model training at the edge, safeguarding sensitive data while enabling network model learning (Zhan et al., 2025). Together, these converging trends, hardware-accelerated AI inference, SDR flexibility, edge computing, and federated learning, pave the way for an AI-native 6G paradigm where DSP algorithms and hardware are co-designed to deliver unprecedented performance, efficiency, and adaptability.

The accelerated evolution of wireless networks toward 6G introduces unprecedented demands for ultra-reliability, sub-millisecond latency, and intelligent adaptability, which require not only emerging technologies such as quantum links and optical fiber but also a deep integration of AI with digital signal processing (DSP) (Akbar et al., 2024; Shen et al., 2023). Recent studies highlight the need for real-time, low-error DSP systems capable of adapting to highly dynamic environments. FPGA and SoC platforms offer improved efficiency through parallel processing, while advanced denoising methods like GAN-CNN hybrids address complex propagation conditions in mmWave MIMO (Ghosh et al., 2023; She et al., 2021; Tarafder et al., 2025; Zhang et al., 2021). Simultaneously, advanced AI techniques such as Deep Reinforcement Learning (DRL), Neural Architecture Search (NAS), Graph Neural Networks (GNN), and Bayesian Optimization (BO) are being increasingly utilized for adaptive filtering, channel estimation, and intelligent feature extraction, marking a shift from traditional algorithmic design to intelligent, data-driven DSP. This survey aims to bridge classical and AI-enhanced DSP paradigms, underscore critical challenges, including energy efficiency, real-time adaptability, and model robustness, and outline a roadmap for next-generation communication systems. As illustrated in Figure 1, reconfigurable DSP processors are integrated between baseband signal generators and RF front-ends, enabling flexible and high-performance architectures in 5G and beyond communication systems (Ghosh et al., 2023).

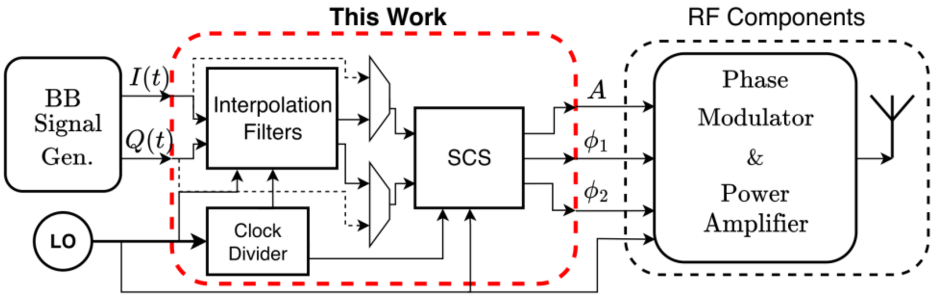

Figure 1. Block-level architecture of a reconfigurable DSP processor positioned between the baseband signal generator and the RF front-end in a 5G transmitter. (Ghosh et al, 2023)

Background and Fundamentals

Digital signals are produced by sampling analog inputs at a frequency at least twice the highest spectral component, as stipulated by the Nyquist–Shannon theorem (Oppenheim et al., 2009; Zeng et al., 2024), followed by quantization, which maps continuous amplitudes to discrete levels and inherently introduces quantization noise. Advanced quantization strategies, particularly in FFT-based architectures, can significantly enhance computational precision while minimizing hardware area and power consumption (Garrido et al., 2025). DSP filtering is predominantly realized through finite impulse response (FIR) filters, offering stability and linear phase, and infinite impulse response (IIR) filters, which achieve comparable selectivity with fewer coefficients but require careful pole placement to ensure stability. FIR convolution is often accelerated via the Fast Fourier Transform (FFT), reducing computational complexity from O(NM) to O((N+M) log (N+M)) for improved efficiency. Recent developments have embraced approximate computing, deliberately trading negligible accuracy loss for substantial reductions in energy, latency, and communication overhead, an approach particularly advantageous for IoT and edge DSP applications (Damsgaard et al., 2024). Techniques such as dynamic reconfiguration and low-precision arithmetic have enabled the development of ultra-low-power convolution engines, exemplified by the FPGA-based “DycSe” core employing a column-streaming architecture for resource-aware edge AI acceleration (Lin et al., 2023), while low-precision FFT cores and approximate convolution circuits have been tailored for wearable and audio devices to meet stringent power and area constraints without compromising functional requirements. (Garrido et al., 2025)

The evolution of digital signal processing (DSP) in communications began in the 1960s with the advent of the discrete Fourier transform (DFT) and its efficient implementation via the fast Fourier transform (FFT). By the 1970s and 1980s, the introduction of dedicated DSP chips equipped with multiply–accumulate units enabled on-chip filtering and modulation, making quadrature phase-shift keying (QPSK) and quadrature amplitude modulation (QAM) modems a commercial reality. Throughout the 1990s and 2000s, DSP architectures became central to digital cellular networks, supporting voice codecs, adaptive equalizers, and broadband technologies such as orthogonal frequency-division multiplexing (OFDM) for multicarrier modulation. By the 2010s, real-time DSP implementations facilitated advanced adaptive multiple-input multiple-output (MIMO) equalization and beamforming in 4G/LTE systems. In modern 5G deployments, massive MIMO and millimeter-wave beam steering demand high-performance hardware acceleration, with FPGA- and ASIC-based digital beamformers delivering sub-millisecond latency and extremely high throughput. FPGA architectures, in particular, achieve latency performance unmatched by CPUs and GPUs, while their reconfigurable fabric supports deep pipelining and parallelism essential for real-time DSP (Michon et al., 2025). These capabilities have propelled the development of software-defined radio (SDR) and adaptive radio platforms. For example, Komeylian et al. (2023) present an FPGA-based digital beamformer for a 10 GHz hybrid antenna array that integrates minimum variance distortionless response (MVDR) and linearly constrained minimum variance (LCMV) methods with a quadratic-surface support vector machine (QS-SVM) direction-of-arrival estimator to process nonlinearly separable data. The system effectively suppressed two interferers by placing deep nulls below −10 dB while preserving the desired signal, achieving millisecond-scale latency, over 90% efficiency, and nearly 100% throughput, as demonstrated in Figures 2–4 and Tables 1–2. Overall, each successive generation of wireless communication, from GSM to LTE to 5G, has relied heavily on DSP, with every new air interface, such as OFDM in 4G and massive MIMO in 5G, introducing increasingly complex DSP workloads that have been implemented in custom hardware.

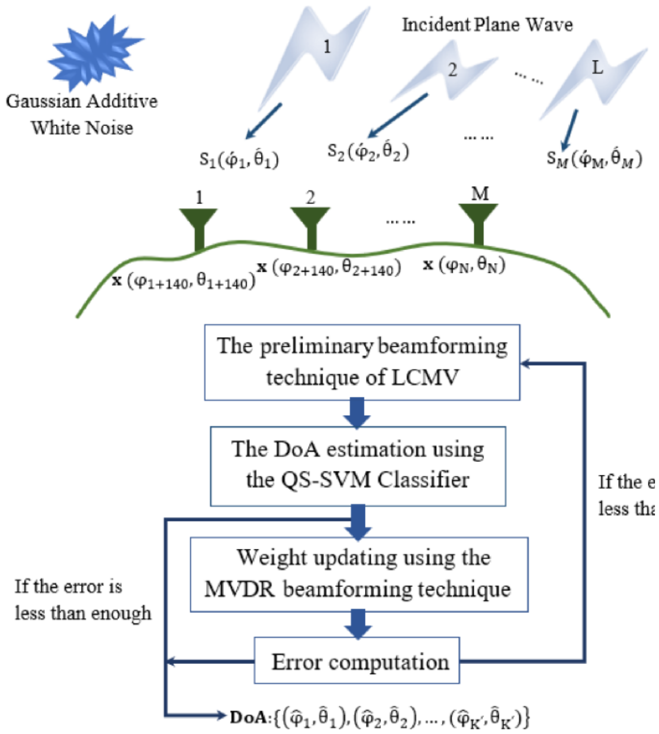

Figure 2. Procedures for the DoA estimation using the QS-SVM-based digital beamformer. (Komeylian et al, 2023)

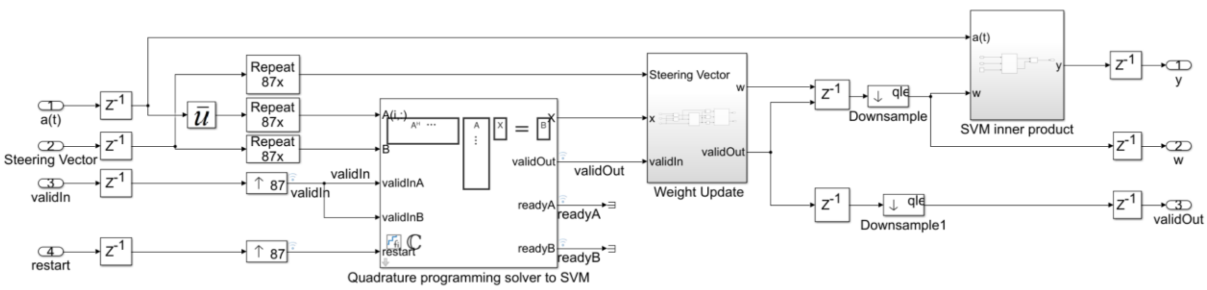

Figure 3. Hardware implementation of the proposed digital SVM-based beamformer on the FPGA platform. (Komeylian et al, 2023)

Table 1. The design parameters for the proposed hybrid antenna array. (Komeylian et al, 2023)

Parameters	Definition	Value
N_h	Number of elements of any circular loop	N_h=20
Q_h	Number of elements of any cylinder	Q_h=40
M_h	Total number of cylinders in the proposed array	M_h=3
P_h	Number of circular loops in the cylinder	P_h=2
d_v	Vertical spacing between two consecutive circular loops	d_v=0.5λ
d_r	Horizontal spacing between two consecutive circular loops	d_r=0.5λ
Φ, θ	Maximum scanning angles	ϕ=45°, θ=45°

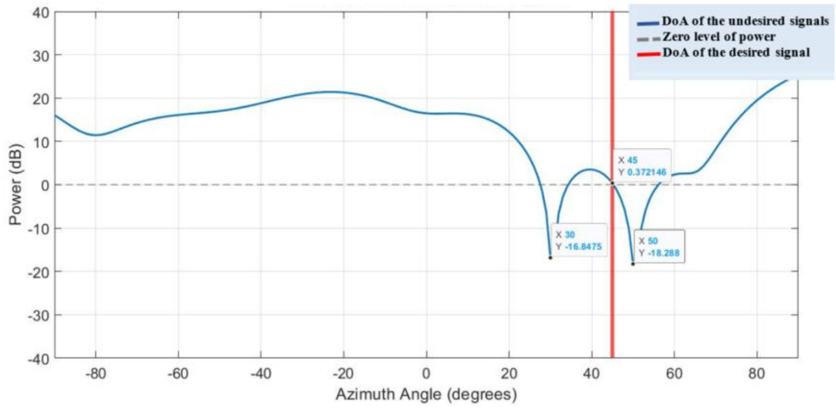

Figure 4. The spatial filtering performance of the QS-SVM-based digital beamformer on the FPGA board. (Komeylian et al, 2023)

Table 2. A comparison between the classification performance of the QS-SVM beamformer using the hybrid antenna array with bowtie elements. (Komeylian et al, 2023)

	Antenna array with bowtie elements	**Antenna array with dipole elements**
Performance efficiency of the proposed QS-SVM beamformer	**96%**	**75%**

Digital signal processing (DSP) algorithms are integral to virtually all modern communication systems, enabling robust and efficient transmission across diverse media. In receiver architectures, adaptive filters and equalizers continuously adjust to time-varying channel conditions, mitigating multipath distortion and suppressing co-channel interference. Modern blind adaptive equalizers, for instance, effectively counter inter-symbol interference (ISI) and significantly improve bit error rate (BER) performance; Silva et al. (2025) report that an adaptive equalizer can maintain BER at acceptable levels through continuous coefficient updates. Multiple-input multiple-output (MIMO) beamforming and spatial filtering are likewise implemented via DSP, often on FPGA or ASIC platforms, to dynamically steer beams or place nulls toward interference sources, with recent FPGA-based beamformers operating at gigahertz frequencies achieving millisecond-scale latencies and near-100% efficiency (Komeylian et al., 2023). Collectively, such capabilities enhance both link reliability and spectral efficiency, reducing BER while increasing throughput. Furthermore, DSP hardware underpins the execution of modern audio, video, and data services, including codecs, compression algorithms, noise suppression, and equalization in mobile handsets. Even high-speed optical communication systems rely extensively on DSP to compensate for channel impairments. Recent advances in learned DSP for fiber-optic links have demonstrated a 48% reduction in computational complexity without performance degradation, illustrating how algorithmic optimization in DSP directly translates into tangible system-level gains (Niu et al., 2024).

Critical performance metrics for digital signal processing (DSP) in communications include throughput, latency, and computational complexity. Throughput, measured in samples or bits processed per second, must align with the target data rates; for example, a beamforming or FFT accelerator must sustain multi-gigabit data streams. Latency, the time from input to output, often needs to be extremely low, typically in the microsecond-to-millisecond range, to enable real-time feedback and control in applications such as voice and video transmission or autonomous vehicle navigation. FPGA-based implementations have demonstrated performance improvements of up to an order of magnitude. For instance, Wang et al. (2025) introduced HP-FFT, a general FFT generator built using High-Level Synthesis that exploits hierarchical parallelism to enhance throughput and flexibility. The design supports multiple functionalities and customizable parallelism configurations, achieving performance that matches or surpasses state-of-the-art RTL and HLS FFT implementations while offering greater portability and resource efficiency (Wang et al., 2025; Provan et al., 2025). As shown in Figures 5, 6 and Table 3, these figures of merit guide DSP algorithm design, where real-time systems often balance precision and computational speed to meet throughput and latency requirements within the constraints of power-limited hardware.

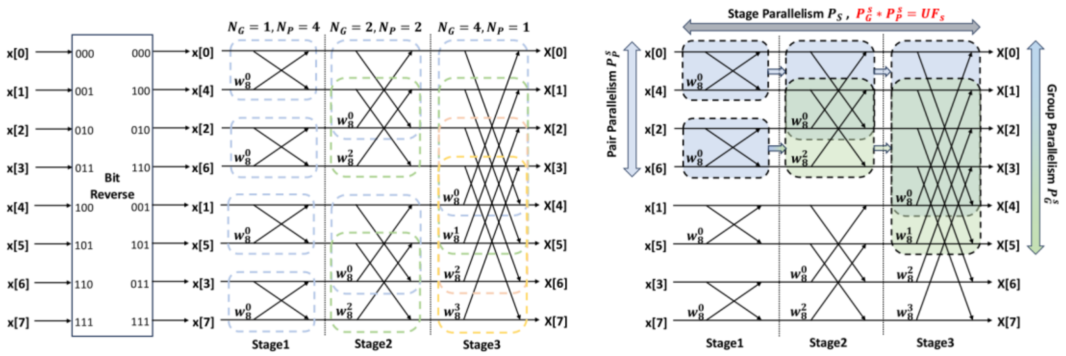

Figure 5. Data Flow Graph (left) and parallelism hierarchy (right) of 8-Point Radix-2 DITRN FFT. (Wang et al, 2025)

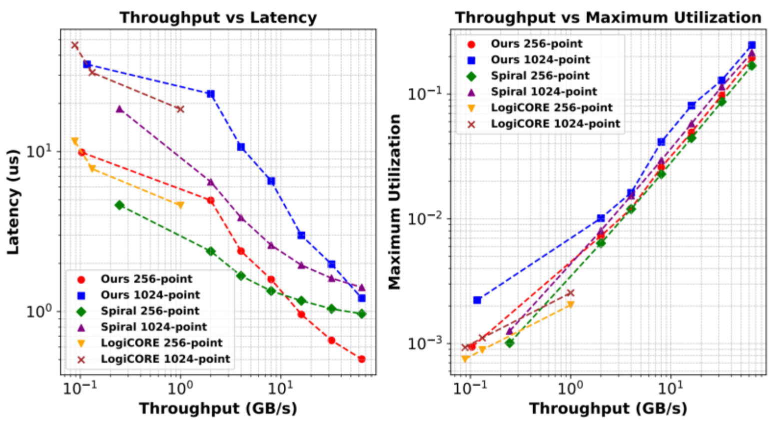

Figure 6. The design space of 256 and 1024-point Radix-2 DIT FFT architectures generated by different libraries/generators at 250 MHz. (Wang et al, 2025)

Table 3. Vitis/Vivado synthesis results at 250 MHz for 256-point Radix-2 DIT$_{RN}$ FFT in FP32, II and Latency are in cycle count. (Wang et al, 2025)

Design	II	Latency	BRAM	DSP	FF	LUT
Original C style	15603	15602	3.5	134	6401	11104
LogiCORE, no SP	1950	1950	3	6	3273	2995
LogiCORE, SP	256	1151	1	18	7335	6850
Xilinx SSR, SP	1791	2715	0	438	59553	38518
Spiral, no SP	1041	1157	5	4	3000	3000
Spiral, SP, UF=1	128	595	24	28	27983	21419
Spiral, SP, UF=2	64	418	42	48	52625	40227
Spiral, SP, UF=4	32	336	72	84	100173	76712
Spiral, SP, UF=8	16	292	120	156	197583	150414
Spiral, SP, UF=16	8	262	192	300	381743	291653
Spiral, SP, UF=32	4	242	288	588	743682	568370
HP-FFT, no SP	2085	2084	4.5	8	844	551
HP-FFT, SP, UF=1	128	1239	36	24	13421	18937
HP-FFT, SP, UF=2	64	597	48	44	27322	40551
HP-FFT, SP, UF=4	32	397	80	82	73916	86909
HP-FFT, SP, UF=8	16	240	128	158	119219	166047
HP-FFT, SP, UF=16	8	165	192	310	212171	328225
HP-FFT, SP, UF=32	4	126	256	618	401229	655458
SP represents **stage pipeline,** UF represents the constant **unrolling factor** in stages.						

FPGAs and ASICs are widely used to achieve aggressive DSP performance metrics (Sadeghi, 2024). FPGAs, in particular, provide deeply pipelined arithmetic units and extensive parallelism, enabling designers to customize data flows and precision, such as hybrid floating-point and fixed-point FFT cores, to optimize speed. Their reconfigurable fabric allows for maximum optimization through adaptable levels of parallelization and pipelining, resulting in custom DSP pipelines that reach peak performance (Wang et al., 2025). As a result, FPGA implementations consistently outperform CPUs and GPUs in DSP tasks, with experiments showing significantly lower latency for streaming signal processing (Michon et al., 2025). Advanced FPGA DSP cores now incorporate features like dynamic radix FFTs, mixed-precision datapaths, and run-time reconfiguration for on-the-fly optimization. MIMO equalizers and beamformers designed in VHDL and Verilog on Xilinx and Intel FPGAs demonstrate high energy efficiency per operation. High-end FPGA-based digital beamformers (e.g., Xilinx Versal or Intel Stratix families) can sustain multi-Gbps raw RF sample rates and achieve sub-millisecond end-to-end latency, leveraging high-speed serial I/O, streaming fixed-point datapaths, and extensive use of on-chip DSP slices and pipelining; for example, a single-FPGA implementation has achieved a sustained input rate of 4.83 Gsamples/s (Kou et al., 2023). These hardware-accelerated DSP solutions enable modern wireless standards like 5G and IoT by making real-time, high-throughput signal processing feasible on platforms constrained by power and area (Khan et al., 2025; Ney et al., 2024).

Moreover, emerging DSP classes such as sparse signal processing, wavelet-based transforms, and sub-band coding have attracted significant interest for resource-constrained wireless systems. In particular, adaptive sparse processing, which leverages inherent channel sparsity in RIS-assisted mmWave MIMO, enables accurate channel estimation from substantially fewer samples, offering a key advantage for bandwidth- and power-limited 5G and 6G systems (Shao et al., 2025; Shao & Lv, 2025). Sub-band and multirate filtering techniques, including polyphase filter banks, are extensively utilized in speech and audio coding as well as channelization for wideband modulation; polyphase implementations also provide computational and implementation benefits when applied to 2-D FIR filter-bank designs (Matei & Chiper, 2024; Xie et al., 2025). Wavelet transforms, with superior time–frequency localization compared to FFT, are particularly well suited for transient signal analysis and denoising, and have been explored as the foundation for alternative multi-carrier waveforms such as wavelet-OFDM and FBMC, which offer improved spectral containment and transient robustness relative to conventional OFDM (Gómez-Luna et al., 2024; Kiruluta et al., 2025). These advanced techniques necessitate revised performance metrics that consider not only complexity and latency but also convergence speed, sparsity exploitation, and learning adaptivity, especially when integrated with machine learning and deep learning models.

Classification of DSP Algorithms

Time-Domain Algorithms, Adaptive filters dynamically adjust their coefficients based on input signal statistics, making them indispensable in real-time applications such as channel equalization, noise cancellation, echo suppression, and biomedical signal enhancement. A widely used example is the Least Mean Squares (LMS) algorithm, which iteratively updates filter weights, utilizing the instantaneous error and input. LMS offers linear complexity O(N) and is valued for its simplicity (Khan et al., 2023), though it converges slowly when the input autocorrelation matrix is ill-conditioned. To address this limitation, the Normalized LMS (NLMS) normalizes the step size by the instantaneous input power, improving stability under nonstationary signal levels, while the Affine Projection Algorithm (APA) extends LMS by projecting the error onto a subspace formed by multiple recent input vectors, thereby accelerating convergence in colored-noise environments (Wan et al., 2024). For rapid adaptation in 5G MIMO channel tracking, the Recursive Least Squares (RLS) algorithm is preferred, as it minimizes a weighted least-squares cost with exponential memory and converges much faster than LMS (Al-Ibadi et al., 2021). Recent improvements enhance RLS tracking by adaptively adjusting the forgetting factor using low-complexity optimizers such as ADAM, and by integrating hybrid beamforming techniques to boost both convergence speed and accuracy in dynamic target scenarios (Zorkun et al., 2024). Hybrid approaches combining classic adaptive filters with machine learning or metaheuristics are also under active investigation; for example, particle swarm optimization (PSO) and reinforcement learning (PPO) have been used to adapt LMS and RLS parameters online. Since PSO and LMS/RLS share iterative parameter-adjustment mechanisms (Mohamed et al., 2023; Bereketoglu et al., 2025), PSO-enhanced filters can better track nonstationary environments. Practical hardware implementations focus on low-power designs, distributed arithmetic (DA) enables FIR filter realization without multipliers, and approximate computing techniques such as truncated or Booth-encoded DA units can significantly reduce area and latency. Notably, James et al. (2024) report that an FPGA-based adaptive filter implementation achieves over 90% logic reduction and approximately 90% speed-up by employing a radix-4 and radix-8 DA architecture with approximate arithmetic, demonstrating the feasibility of deploying high-performance adaptive filters in compact, real-time IoT and edge devices (James et al., 2024; Khan, 2024), as shown in Figures 7–9 and Table 4.

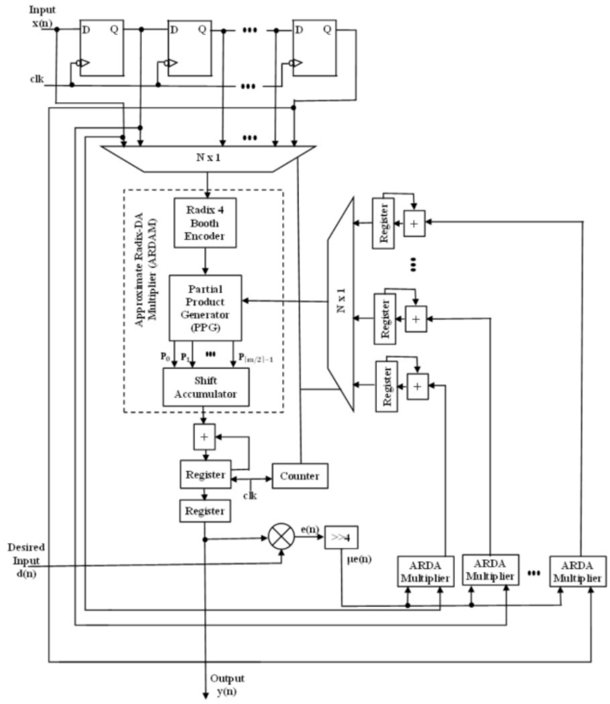

Figure 7. DAAFA with radix-4 architecture. (James et al, 2024)

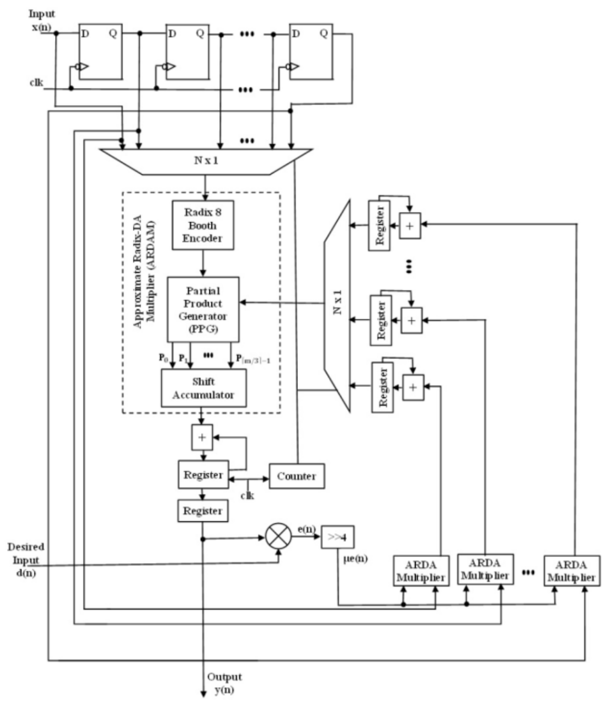

Figure 8. DAFFA with radix-8 architecture. (James et al, 2024)

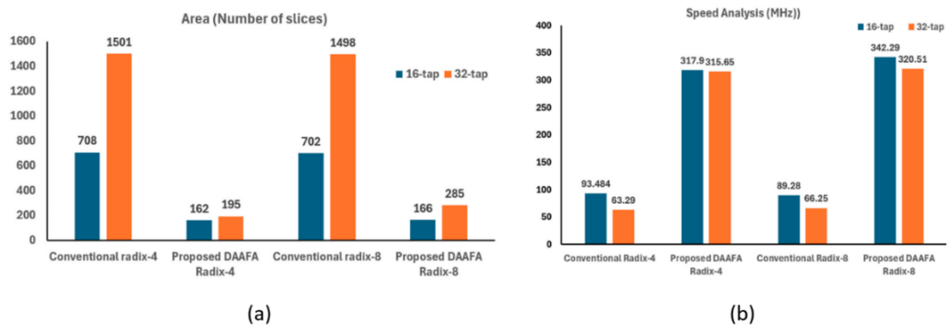

Figure 9. (a) Area (number of slices) analysis of proposed architecture with the reported works. (b) Speed comparison with the reported works. (James et al, 2024)

Table 4. Synthesis summary of the realized structure. (James et al, 2024)

Performance Measures	Existing Approximate DA-Radix-4		Proposed DAAFA-Radix4		Existing Approximate DA-Radix-8		Proposed DAAFA-Radix-8	
Device	Xilinx Virtex-5 device		Xilinx Virtex-5 device		Xilinx Virtex-5 device		Xilinx Virtex-5 device	
Filter length	16	32	16	32	16	32	16	32
Slices	708	1501	162	195	702	1498	166	285
MSP (ns)	10.697	15.8	3.146	3.168	11.20	15.1	2.92	3.12
MSF (MHz)	93.484	63.29	317.9	315.65	89.28	66.25	342.29	320.51

Echo Cancellation: Echo cancellation is crucial in contemporary audio communication for the elimination of undesired echoes created by acoustic feedback between microphones and speakers. These echoes severely reduce the perceived speech quality by making users hear delayed copies of their own voice. For instance, traditional AEC techniques utilize time-domain adaptive filters to model the echo path and recursively estimate the echo signal, which is then subtracted from the microphone input. More recently, deep learning-based AEC methods have been explored, enabling improved performance in complex acoustic environments such as surround sound systems. (Saremi et al., 2023; Li et al., 2023). This process retains the near-end speech and preserves full-duplex communication, as opposed to earlier mute-based echo suppression methods. Consequently, acoustic echo cancellation is widely considered a prominent challenge in speech enhancement for voice calls, teleconferencing, smart speakers, and other hands-free systems. Recent advances have incorporated metaheuristic optimization techniques, such as Grey Wolf Optimization (GWO) and Particle Swarm Optimization (PSO), to accelerate convergence and improve echo suppression performance in real-time applications. (Pichardo et al., 2024).

The conventional AEC architecture employs low-complexity adaptive filters (LMS and NLMS) that are driven from the far-end reference signal. Initial AEC implementations utilized the Least Mean Square (LMS) algorithm and its variations, and over time the normalized LMS (NLMS) and similar adaptive schemes became standard (Sameri et al., 2023; Kellermann, 2008). Adaptive filters iteratively update their coefficients by correlating the far-end (loudspeaker) signal with the microphone input to accurately estimate the acoustic echo path, enabling effective subtraction of the far-end echo component. Recent advances in deep learning techniques have further enhanced the modeling accuracy and echo cancellation performance (zhao et al., 2024). An essential adjunct in this loop is the double-talk detector (DTD), which halts adaptation whenever the near-end user speaks to prevent filter divergence. In practice, robust DTD and voice-activity detection (VAD) are employed to prevent misadjustment of the adaptive filter during near-end speech. Collectively, these mechanisms yield a stable AEC that suppresses far-end echo while preserving the desired near-end signal (Khanagha et al., 2024; Seidel et al., 2024; Zhao et al., 2025).

Linear adaptive filters perform adequately under simplified conditions but often face challenges when confronted with realistic distortions such as loudspeaker nonlinearity, room reverberation, and model mismatch. To address these limitations, recent acoustic echo cancellation (AEC) systems have augmented or replaced traditional filters with deep neural network–based approaches, including hybrid Neural-Kalman methods capable of learning nonlinear relationships. Deep architectures, particularly recurrent neural network (RNN) variants like LSTM and GRU, have been employed to capture long-term temporal dependencies in the echo path, thereby enhancing suppression of residual echoes as well as convergence speed and robustness (Seidel et al., 2025). Zhao et al., (2024) propose SDAEC, a signal-decoupling monaural AEC technique that estimates an energy-scaling factor from microphone and reference signals to rescale the reference prior to neural echo cancellation. This preprocessing step effectively mitigates energy mismatches under low signal-to-echo ratio conditions, simplifying the learning process for various neural network backbones, including CRN+LSTM and dual-path RNN, and improving suppression of residual echo and noise.

State-of-the-art echo cancellers utilize a multi-stage hybrid pipeline in which an initial linear AEC removes the dominant echo, followed by a neural post-filter, typically operating on time–frequency features, that further attenuates residual echo and noise. This two-stage approach achieves significant improvements in echo return loss enhancement (ERLE). For example, combining a linear filter with a deep convolutional-recurrent post-filter has been shown to increase ERLE beyond that of a baseline linear AEC alone (Shachar et al., 2023). In practical real-time devices such as smartphones and smart speakers, strict constraints apply: models must be extremely compact and low-latency, typically requiring less than 50 kB of memory and approximately 10 ms delay, to run efficiently on embedded DSPs. To meet these requirements, AEC solutions employ quantized, compact neural architectures and hardware-aware designs, enabling real-time echo suppression without audible delay. Overall, modern AEC systems in consumer products effectively combine classical adaptive filtering with highly optimized neural components to deliver robust echo suppression and high speech quality within practical resource constraints (Zhang et al., 2022; Cheng et al., 2022).

Frequency Domain Algorithms

Fast Fourier Transform (FFT): FFT has revolutionized digital signal processing (DSP) by reducing the computational complexity of the discrete Fourier transform (DFT) from $O(N^2)$ to $O(N \log N)$. For example, Vernon and Perera (2025) note that a direct computation of an *N-point* DFTrequires $O(N^2)$ operations, whereas FFT algorithms achieve $O(N \log N)$. They developed an FFT implementation for embedded systems using the Prime Factor Algorithm (PFA), demonstrating how FFT-based spectral analysis can be efficiently realized in real-time, resource-constrained environments. FFT's ability to convert signals between the time and frequency domains underpins modern communication standards such as LTE and 5G and Wi-Fi via orthogonal frequency-division multiplexing (OFDM). Khan et al. (2025) emphasize that although FFT dramatically reduces DFT complexity, its deployment in LTE and 5G systems still places substantial demands on hardware and processing resources to meet high-throughput requirements of up to 3 Gbps. OFDM, leveraging the FFT and IFFT pair, offers high spectral efficiency and resilience to multipath fading, performing robustly in frequency-selective fading channels (Wu et al., 2024). Beyond communications, FFT plays a pivotal role in radar, audio, and imaging. Vasilakis et al. (2025) present an efficient in-place radix-2 FFT method tailored for resource-constrained space and edge computing devices, achieving significant gains in memory efficiency, execution time, and spectral accuracy. In wireless systems, FFT and IFFT cores enable subcarrier modulation and channel equalization, while in broader DSP chains they support fast convolution, via overlap-add and overlap-save methods, and efficient filter bank implementations, all exploiting the inherently low computational complexity of the FFT.

To meet growing data demands, FFT hardware has evolved significantly by leveraging deep pipelining and parallelism in FPGA and ASIC designs. Wang et al. (2025) developed HP-FFT, a flexible High-Level Synthesis generator that exploits hierarchical parallelism within the FFT algorithm, achieving high throughput and resource efficiency. Their approach outperforms existing FFT implementations, enabling efficient, customizable hardware for modern applications. Similarly, high-end 5G baseband processors must handle multi-gigabit data rates with peak throughput ranging from 0.3 to 3 Gbps, and potentially up to 30 Gbps, where FFT throughput remains a primary bottleneck. To address this challenge, efficient FFT implementations are critical. Khan et al. (2025) propose a runtime reconfigurable butterfly structure capable of simultaneously performing Radix-8, Radix-4, Radix-3, and Radix-2 FFT computations, catering to the bandwidth requirements of current LTE, 5G, and beyond systems. This flexibility to adapt to different bandwidths is essential in 5G networks, and the configurable FFT engine optimizes resource utilization while delivering high throughput. The paper presents two FFT architectures based on the reconfigurable butterfly, offering designers the flexibility to select the most suitable approach for efficient, high-throughput solutions tailored to 5G and beyond. Additionally, Vasilakis et al. (2025) propose an in-place radix-2 FFT optimization that stores two FFT elements per memory address, enabling parallel data access. By optimizing floating point and block floating point configurations, they enhance FFT performance on resource-limited processors, improving signal-to-noise ratio and reducing memory usage by half. Tested on various edge and embedded devices, including FPGA and Raspberry Pi, their method significantly boosts execution speed for FFT sizes ranging from 1K to 16K points, as shown in Figures 10-12 and Tables 5, 6.

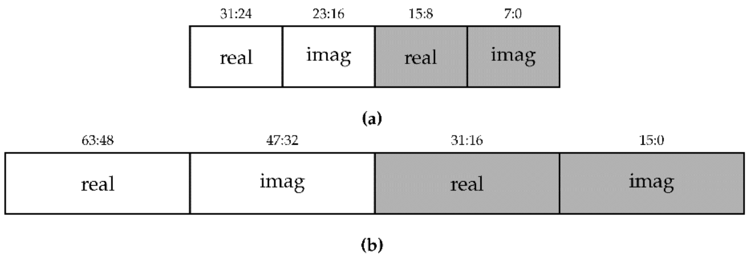

Figure 10. Organization of data in memory words of different lengths. (**a**) A 32-bit memory word, storing each part of the two data points with FP8 or BFP8 number representations. (**b**) A 64-bit memory word, storing each part of the two data points with FP16 or BFP16 number representations. (Vasilakis et al, 2025)

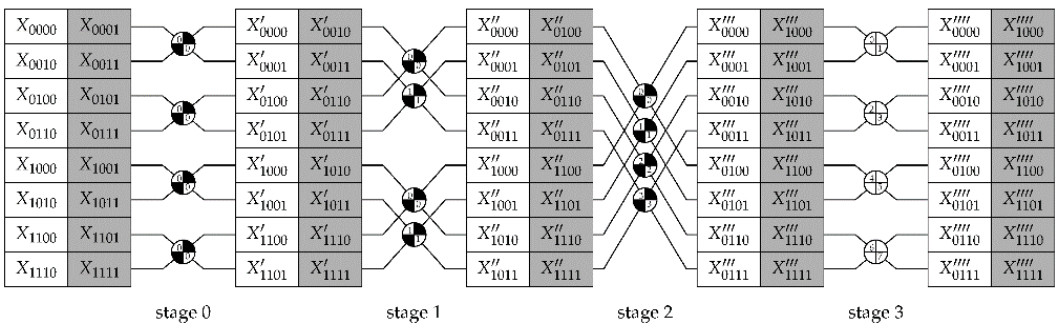

Figure 11. The Signal Flow Graph (SFG) of the In-Place Fast Fourier Transform (FFT), showing the indexes of the data points at each position in the memory and the permutation that follows each butterfly operation. (Vasilakis et al, 2025)

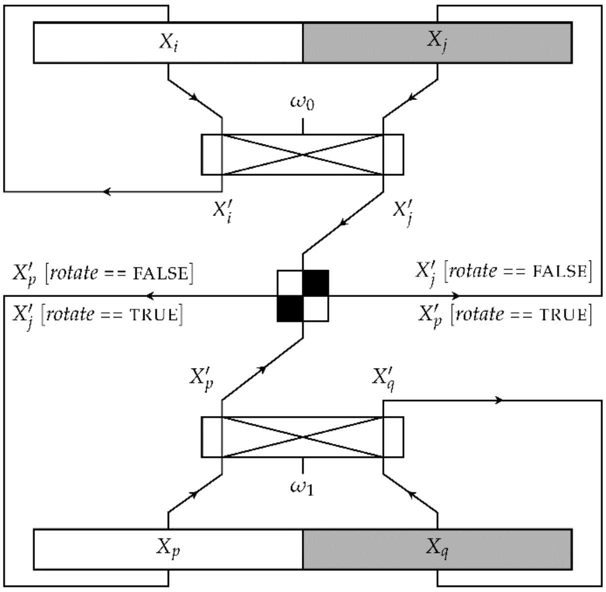

Figure 12. Depiction of MainOp. (Vasilakis et al, 2025)

Table 5. Signal-to-Noise Ratio (SNR), measured in dB for the different number representations which are compared to the double precision values of the numpy implementation. (Vasilakis et al, 2025)

FFT Points	AVR32, Myriad 2 (dB)		Zero 2W (dB)		FPGA (dB)	
	FP8	BFP8	FP16	BFP16	BFP8	BFP16
1024	20.94	25.56	63.59	73.67	25.56	73.68
2048	20.57	24.74	63.19	72.86	24.74	72.87
4096	20.23	23.86	62.82	71.98	23.87	72.00
8192	19.92	23.35	62.47	71.47	23.35	71.49
16,384	19.63	22.71	62.15	70.85	22.71	70.86

Table 6. Field Programmable Gate Array (FPGA) for 100 MHz clock. (Vasilakis et al, 2025)

FFT Points	Word Length (Bits)	LUT (%)	FF (%)	BRAM (%)	DSP (%)	Power (W)	Time (s)
1024	32	7.53	2.79	1.00	6.06	0.11	0.05
1024	64	10.44	3.70	2.00	6.06	0.12	0.05
2048	32	7.83	2.80	2.00	6.06	0.11	0.11
2048	64	12.66	3.70	4.00	6.06	0.13	0.11
4096	32	12.79	2.80	4.00	6.06	0.11	0.24
4096	64	15.24	3.71	8.00	6.06	0.13	0.24
8192	32	21.33	2.80	8.00	6.06	0.13	0.53
8192	64	28.10	3.72	15.00	6.06	0.14	0.53
16,384	32	41.73	2.81	16.00	6.06	0.14	1.14
16,384	64	41.83	3.74	32.00	6.06	0.16	1.14

Spectral Estimation: Spectral estimation aims to recover the frequency content of a signal from observed data. In this study, non-asymptotic analyses of classical methods, such as the Bartlett and Welch estimators, are presented for L-mixing time-series data with unknown means, providing performance guarantees within finite-sample regimes (Zheng & Lamperski, 2025). In other words, applications such as EEG analysis, radar return analysis, and radio channel analysis rely heavily on power spectral density (PSD) estimation to characterize signals and detect anomalies. This is exemplified by recent quantitative EEG (qEEG) studies on Alzheimer's disease, which employed the Welch method to analyze gamma-band power spectra (Simfukwe et al., 2025). Spectrum sensing is also essential in cognitive radio systems that monitor spectrum occupancy to exploit unused frequency bands efficiently. Furthermore, emerging wireless standards like 5G and 6G utilize real-time spectrum sensing for interference avoidance, with advanced techniques leveraging deep learning to enhance detection accuracy in 6G systems (Bandoudi & Boulouird, 2024; Muzaffar et al., 2024). Consequently, spectral estimation plays a pivotal role across diverse engineering fields, ranging from peak identification in biomedical spectra to dynamic spectrum access for locating vacant channels.

Spectral estimators fall into three main categories:

- **Nonparametric techniques:** These methods estimate the power spectral density (PSD) directly from observed data, encompassing approaches such as periodograms, Welch's method, and multitaper techniques. By employing windowing and averaging over segmented data, these techniques reduce variance, yielding estimates that are both computationally efficient and straightforward to implement. However, their spectral resolution deteriorates when applied to short-duration signals or scenarios with low signal-to-noise ratios (SNR). For instance, Welch's method reduces variance by averaging overlapping periodograms, which introduces a tradeoff with increased spectral leakage. Despite this bias-resolution compromise, nonparametric methods maintain considerable flexibility owing to their model-free nature, making them extensively utilized in practical signal processing applications (Zheng et al., 2025; Storica et al., 2005)
- **Parametric methods:** Parametric spectral estimation methods presuppose that the observed signal is generated by a linear time-invariant system, typically modeled as autoregressive (AR), autoregressive moving average (ARMA), or Prony's models, driven by white noise. These methods proceed by estimating the parameters of the assumed model, enabling higher spectral resolution with fewer samples compared to nonparametric techniques. Recent research underscores that parametric estimators effectively represent signals as outputs of linear systems excited by white noise, exhibiting superior performance particularly when low-order models sufficiently capture the signal dynamics (Mokrý & Rajmic, 2024; Zhang et al., 2025). For instance, Prony-based and related pole-fitting algorithms provide super-resolution capability for closely spaced spectral components and have been successfully adapted for contemporary radar signal processing and orthogonal time frequency space (OTFS) channel estimation (Jitsumatsu & Sun, 2025). Likewise, autoregressive spectral estimation, commonly implemented via the Levinson–Durbin recursion, can resolve spectral peaks that are otherwise smeared or indistinguishable in periodogram-based methods..
- **Subspace methods:** Algorithms such as MUSIC and ESPRIT take advantage of the eigenstructure of the signal autocorrelation. They are able to separate the signal subspace from noise and can resolve sinusoids or incident angles with super-resolution. Indeed, recent analyses show that classical subspace methods like MUSIC attain super-resolution for closely spaced line spectra, and that carefully designed algorithms can scale these

capabilities to large problems. By taking advantage of eigen-decomposition as seen in MUSIC and ESPRIT theory, these methods can resolve sources much closer than classical Fourier-based approaches. Subspace methods therefore play a significant role in array processing and MIMO radar, particularly in low-SNR and closely-spaced source scenarios (Fei & Zhang, 2025; Ding et al., 2024).

Recent advancements in deep learning have increasingly focused on training convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to map raw time-domain samples or precomputed power spectral densities (PSDs) directly to spectral features or occupancy decisions. Barker (2025) reviews the *DeepSense* family of approaches, emphasizing CNN architectures that process raw in-phase and quadrature (I/Q) waveforms for real-time detection of multiple occupied sub-bands. Studies also show that incorporating high-quality PSD representations, such as multitaper estimates, into convolutional architectures enhances sensitivity to spectral variations and robustness in low-SNR environments; for example, Zheng et al. (2025) propose a dual-representation CNN with multitaper PSD inputs, achieving substantial gains in detection probability and reductions in false alarms compared with I/Q-only or periodogram-based models. Broader developments include transformer-based and attention-driven frameworks, such as the Spectrum Transformer by Zhang et al. (2024), which improve wideband spectrum sensing through higher sample efficiency and resilience to nonstationarity. Overall, convolutional and recurrent architectures effectively learn complex spectral patterns and, in challenging low-SNR and multipath scenarios, can complement, or even surpass, traditional spectral estimators.

Hybrid approaches integrate model-based and data-driven estimation techniques to enhance robustness and interpretability in spectral analysis. For example, some methods utilize neural networks to estimate model parameters, such as autoregressive (AR) coefficients, which are subsequently incorporated into classical spectral estimators, thereby combining the theoretical rigor of parametric models with the adaptive learning capabilities of machine learning. Muzaffar and Sharqi (2024) and Dong et al. (2025) propose end-to-end neural identification frameworks for ARX models, wherein neural networks are trained to recover model terms and coefficients applicable for downstream spectral analysis, as illustrated in Figures 13-15 and Tables 7, 8. Additionally, recent studies demonstrate practical hybrid designs where deep neural networks (DNNs) preprocess signals or optimize hyperparameters and window functions through reinforcement learning prior to classical estimation methods like Levinson–Durbin recursion, effectively reducing bias and improving convergence under low-SNR conditions (Mehrabian et al., 2024). While still an emerging research area, early findings suggest that hybrid ML-classical frameworks can adaptively outperform purely classical techniques in both detection accuracy and convergence speed.

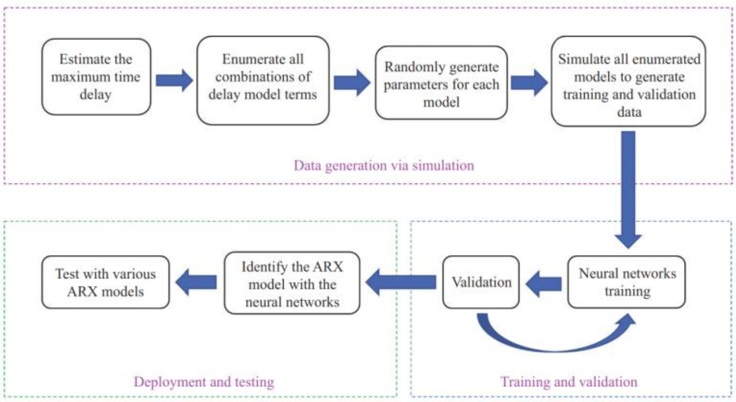

Figure 13. Flowchart of the end-to-end ARX model identification framework. (Dong et al, 2025)

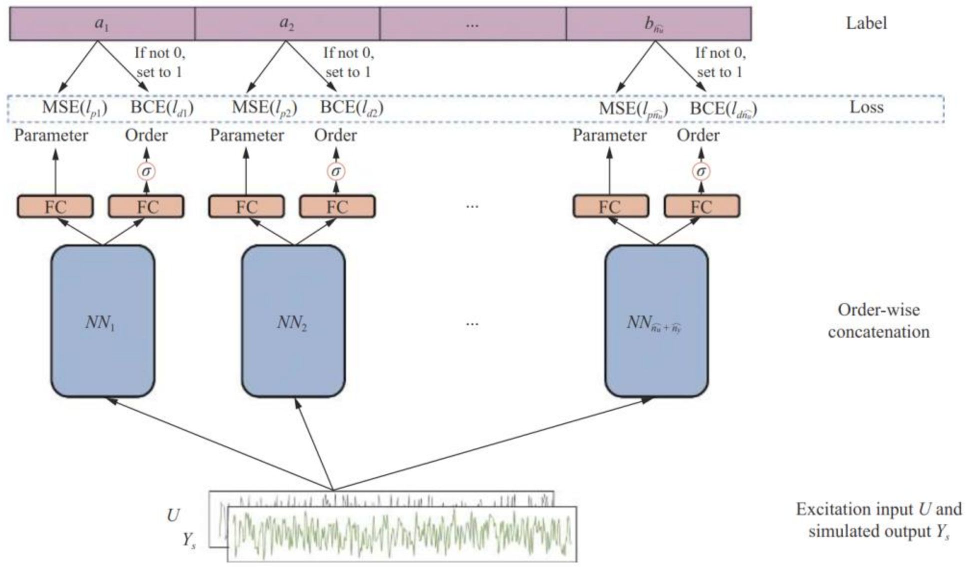

Figure 14. Training process of the order-wise neural networks. (Dong et al, 2025)

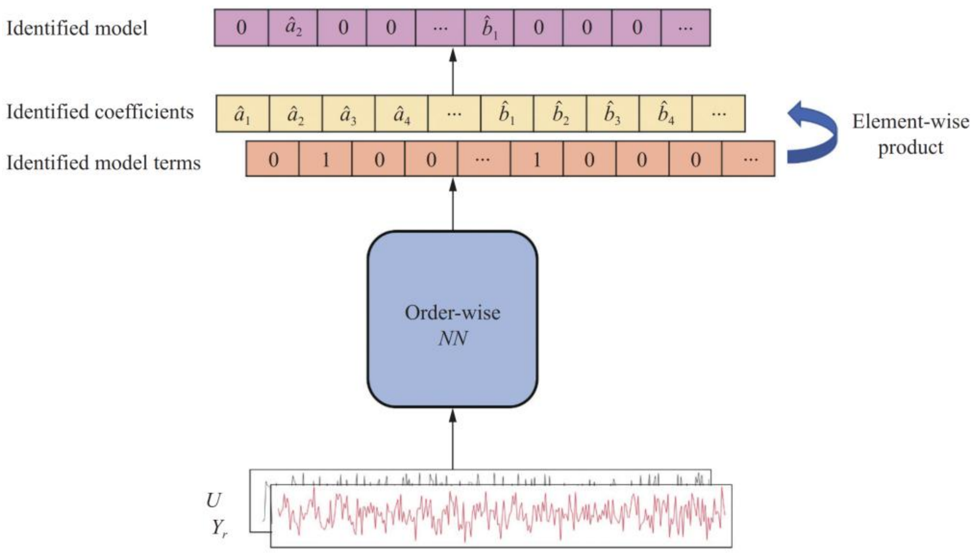

Figure 15. Deployment of the order-wise neural networks. (Dong et al, 2025)

Table 7. Performance of simple ARX model identification with a data length of 300 and various numbers of coefficient sets. (Dong et al., 2025)

Data length	Number of coefficient sets	Backbone type	Macro-averaged AUC	Accuracy	Coefficient estimation RMSE
		2-layer GRU	0.905 0	0.603 0	**0.064 3**
		Bi-GRU	0.921 0	0.691 0	0.069 6
300	50	2-layer LSTM	0.825 0	0.256 0	0.097 2
		Bi-LSTM	0.656 0	0.019 0	0.156 6
		TCN	**0.941 0**	**0.731 0**	0.102 5
		2-layer GRU	0.919 0	0.651 0	0.053 6
		Bi-GRU	0.948 0	**0.807 0**	**0.033 8**
300	100	2-layer LSTM	0.829 0	0.267 0	0.096 8
		Bi-LSTM	0.776 0	0.127 0	0.112 9
		TCN	**0.955 0**	0.802 0	0.095 7
		2-layer GRU	**0.989 0**	**0.956 0**	0.030 1
		Bi-GRU	0.988 0	0.950 0	0.031 1
300	300	2-layer LSTM	0.988 0	0.951 0	**0.030 0**
		Bi-LSTM	0.984 0	0.930 0	0.064 7
		TCN	0.978 0	0.903 0	0.099 5
		2-layer GRU	0.990 0	0.958 0	0.020 9
		Bi-GRU	0.991 0	0.963 0	0.025 9
300	500	2-layer LSTM	**0.992 0**	**0.965 0**	**0.021 4**
		Bi-LSTM	0.989 0	0.953 0	0.028 6
		TCN	0.981 0	0.919 0	0.088 3
		2-layer GRU	**0.991 0**	0.961 0	0.023 9
		Bi-GRU	**0.991 0**	**0.962 0**	0.025 9
300	1000	2-layer LSTM	**0.991 0**	0.961 0	**0.020 7**
		Bi-LSTM	0.989 0	0.954 0	0.027 7
		TCN	0.981 0	0.921 0	0.088 3

Table 8. Performance of simple ARX model identification with additive white noise. (Dong et al., 2025)

SNR	Model	Macro-averaged AUC	AUPR	Accuracy	Coefficients estimation RMSE
	2-layer GRU	0.947 0	0.972 0	0.802 0	0.060 8
	Bi-GRU	**0.949 0**	0.978 0	0.804 0	0.045 8
30 dB	**2-layer LSTM**	**0.949 0**	**0.981 0**	**0.807 0**	**0.041 2**
	Bi-LSTM	0.935 0	0.952 0	0.753 0	0.080 6
	TCN	0.927 0	0.931 0	0.719 0	0.091 7
20 dB	2-layer GRU	0.921 0	0.942 0	0.721 0	0.052 9
	Bi-GRU	**0.929 0**	0.951 0	0.741 0	0.052 9
	2-layer LSTM	0.925 0	**0.967 0**	**0.742 0**	**0.048 0**
	Bi-LSTM	0.915 0	0.933 0	0.703 0	0.053 9
	TCN	0.904 0	0.875 0	0.667 0	0.097 5
10 dB	**2-layer GRU**	**0.855**	**0.673**	**0.537**	**0.07**
	Bi-GRU	0.793	0.326	0.285	0.1044
	2-layer LSTM	0.841	0.537	0.476	0.08
	Bi-LSTM	0.811	0.281	0.222	0.0927
	TCN	0.843	0.528	0.473	0.1044

Advanced Signal Processing

Wavelet Transform: Wavelet transforms provide a time–frequency multiresolution analysis that is well suited for nonstationary signals (Silik et al., 2024). Unlike the Fourier transform, wavelets yield localized time–frequency representations, thus enabling efficient multi-level signal decompositions. For example, lifting-scheme implementations of the DWT allow for in-place filtering that is computationally efficient, incurring fewer operations and memory compared to conventional filter-bank implementations (Jiang et al., 2025). In practice, multi-scale wavelet decompositions separate coarse approximations from detailed high-frequency components, and they have proven instrumental in applications like signal compression and denoising. In particular, wavelet-based denoising methods often preserve transient features better than block-based transforms. This combination of DWT and NLM, enhanced by NOA, not only better preserves transient features but also outperforms traditional methods in reducing various types of noise (Chen et al., 2025).

In communications, wavelet-based multicarrier modulation (wavelet-OFDM or OWDM) exploits these characteristics for increased spectral efficiency and reduced PAPR. Since wavelet packet subcarriers can overlap in time–frequency, wavelet-OFDM systems tend to dispense with a cyclic prefix, improving bandwidth efficiency directly. Wavelet-OFDM schematics consistently exhibit narrower spectral sidelobes and lower PAPR in simulations, and a filtered wavelet-OFDM (F-OWDM) schematic achieves several dB lower PAPR than standard FFT-OFDM. These advantages enhance power-amplifier efficiency and lessen distortion for wireless transmitters, proving wavelet-based multicarrier designs can beat CP-OFDM in bandwidth utilization and linearity. This combination of wavelet-based multicarrier modulation and filtering not only improves bandwidth efficiency by eliminating the cyclic prefix but also achieves a significant reduction in PAPR, thereby enhancing the overall performance of communication systems (Almutairi & Krishna, 2022). Moreover, this wavelet-domain equalization approach not only mitigates Co-CFO effects but also enhances system performance by reducing computational complexity, aligning with the advancements discussed in recent literature (Ramadan, 2025).

Beyond communications, wavelet transforms are extensively employed in biomedical signal denoising. Hybrid methods that combine discrete wavelet transform (DWT) with advanced filtering techniques, such as Non-Local Means (NLM), have demonstrated notable noise reduction while preserving critical physiological features. For example, Zhu et al. (2025) introduced an enhanced DWT-based denoising approach incorporating adaptive thresholding and objective quality metrics, achieving approximately 1–5 dB improvements in signal-to-noise ratio (SNR) over conventional DWT methods under additive white Gaussian noise (AWGN) conditions, and achieving SNR enhancements by tens to hundreds of times in real electrocardiogram (ECG) noise scenarios. These findings were validated using fiber-optic micro-vibration sensor recordings. Collectively, such studies underscore the efficacy of wavelet-based denoisers in effectively eliminating broadband artifacts, including baseline wander, muscle artifacts, and electrode motion, while preserving the integrity of ECG and EEG waveforms.

Recent advances have adapted wavelet methods to more closely match data by combining them with machine-learning and optimization techniques. Hybrid DWT–NLM approaches have been developed for ECG denoising with optimizer-based threshold tuning (Chen et al., 2025). Enhanced DWT schemes with adaptive thresholding and objective quality metrics report improved AWGN performance approximatelly 1–5 dB SNR and orders-of-magnitude gains on real ECG noise, validated on fiber-optic micro-vibration recordings (Zhu et al., 2025). Related work integrates adaptive decompositions (e.g., VMD) with wavelet thresholding for machinery health monitoring (Jia et al., 2025), while wavelet pipelines coupled with learning models (PSO–SVM) and swarm and genetic optimizers demonstrate effective joint optimization of wavelet bases and denoising thresholds across diverse applications (Pal Shuvo et al., 2025; Xiong et al., 2025; Zhu et al., 2023).

Additionally, these studies indicate that tailoring wavelet filter shapes or thresholds via ML and optimization yields improved denoising and compression performance. At the same time, deep learning has been combined with wavelets via learned multiscale architectures; for instance, models such as learned ISTA or Wavelet Net effectively embed wavelet shrinkage in network layers. Such learned unfolding networks allow for fast, near-real-time sparse recovery by encapsulating sparsity priors from the wavelet domain in trainable network architectures. The combination of wavelet modeling with data-driven learning yields state-of-the-art denoisers and compressors, particularly on edge devices where the multiscale aspects of wavelets are especially well-suited to resource-limited settings. (Zheng et al., 2022; Frusque et al., 2024; Yu et al., 2024)

Sub-band Coding and Filter Banks: Sub-band coding splits a signal into several frequency bands using filter banks, allowing band wise processing and effective compression (Schuller, 2020; Kim & Skoglund, 2024; Ng et al., 2025). This forms the basis of almost all perceptual audio coders (e.g., MP3, AAC) and image coders (JPEG2000) (Herre et al., 2025). In audio, each sub-band can be quantized based on human auditory masking: perceptual models allocate bits preferentially to sensitive bands, allowing high compression without audible distortion. Modern coders therefore employ psychoacoustic bit-allocation to take advantage of the ear's frequency-dependent masking, greatly enhancing compression ratios under a quality constraint (Kim & Skoglund, 2024; Herre et al., 2025; Ng et al., 2025).

In wireless receivers (e.g., SDRs), polyphase channelizers realize M-channel filter banks by decomposing an M-tap prototype filter into M polyphase components and performing analysis and synthesis through a single M-point FFT. This architecture enables the efficient extraction of narrowband channels with minimal aliasing, low phase distortion, and reduced group delay, properties that are critical for latency-sensitive applications such as real-time audio transmission and VR streaming (AMD, 2024). For next-generation immersive systems, scalable sub-band audio coders employ time–frequency parameter control and multi-scale quantization to dynamically adapt both bitrate and quantization at the sub-band level. In particular, STFT-based codecs (STFTCodec) can adjust STFT frame, window, and hop parameters to balance bitrate against spectral resolution; multi-scale residual quantizers (SNAC) operate at varying temporal resolutions to achieve content-adaptive compression; and sparse-plus-residual expert vector quantization schemes (SwitchCodec) enhance embedding capacity and robustness under stringent bitrate constraints. Collectively, these techniques facilitate bandwidth-adaptive, perceptually weighted bit allocation, thereby preserving perceptual quality across fluctuating network conditions (Feng et al., 2025; Siuzdak et al., 2024; Wang et al., 2025).

Sub-band filter bank design prioritizes perfect or near-perfect reconstruction and the minimization of aliasing (Brislawn, 2024; Keerthana et al., 2024). Modern coders typically employ M-channel polyphase quadrature filter (PQF) banks and cosine-modulated filter banks, which provide tight spectral tiling with minimal distortion (Keerthana et al., 2024; Haider et al., 2025). Prototype filters are usually optimized, through windowing or iterative design methods, to maximize stop-band attenuation for a given short filter length (Keerthana et al., 2024). Such optimized designs ensure that the recombined sub-bands faithfully reconstruct the original signal (Brislawn, 2024). In practical sub-band coders, adaptive filtering or quantization techniques are also incorporated; for instance, audio codecs may switch to noise-fill or spectral band replication (SBR) in frequency-sparse frames. Ng et al. (2025) propose MUFFIN, a neural audio coding framework that uses perceptual bitrate allocation and advanced convolutional architectures to achieve high-quality, efficient audio compression with strong reconstruction performance (Ng et al., 2025; Choi et al., 2025), as illustrated in Figures 16, 17 and Tables

9-12. These adaptive approaches, combined with psychophysical bit allocation strategies, maintain high perceptual quality even under extreme compression scenarios.

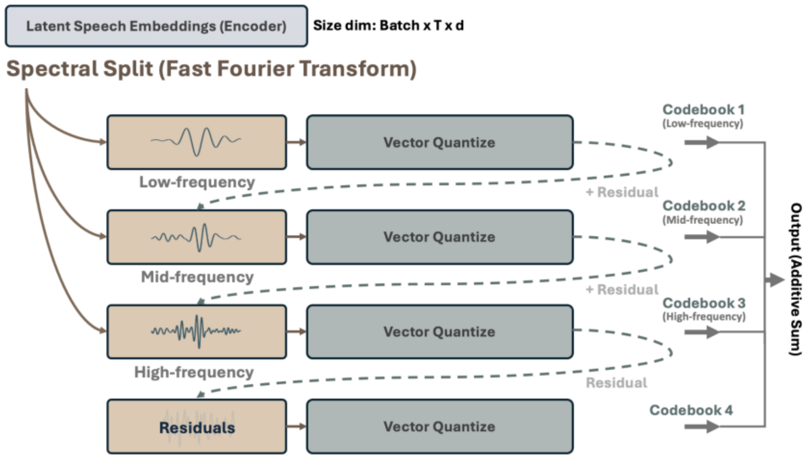

Figure 16. Illustration of the MBS-RVQ process. (Ng et al, 2025)

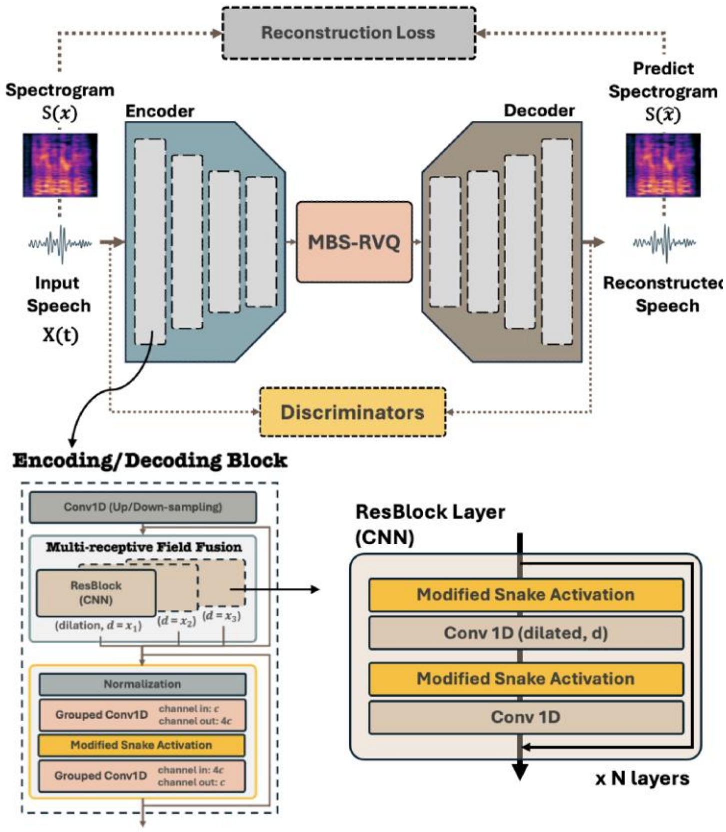

Figure 17. Architecture of MUFFIN incorporating a fully convolutional structure. (Ng et al, 2025)

Table 9. Test-Clean (LibriTTS). (Ng et al, 2025)

Model	STFT	MEL	PESQ	STOI	UTMOS	ViSQOL
GT	-	-	-	-	4.041	-
OPUS	5.728	2.796	1.132	0.715	1.264	2.878
Encodec	1.956	1.051	2.042	0.903	2.269	4.078
DAC	1.759	0.849	2.37	0.915	2.951	4.143
HiFi-Codec	1.618	0.765	2.712	0.943	3.831	4.41
Mimi ▲	2.488	1.706	1.715	0.62	2.966	3.791
MUFFIN	**1.555**	**0.692**	**2.996**	**0.954**	4.017	**4.516**
MUFFIN ▽	1.626	0.755	2.525	0.937	4.035	4.345
MUFFIN ▲	1.663	0.807	2.36	0.932	**4.074**	4.225

Table 10. Test-Other (Libri-TTS). (Ng et al, 2025)

Model	STFT	MEL	PESQ	STOI	UTMOS	ViSQOL
GT	-	-	-	-	3.453	-
OPUS	5.39	2.703	1.143	0.695	1.271	2.815
Encodec	1.998	1.119	1.96	0.888	2.026	4.017
DAC	1.813	0.913	2.22	0.897	2.497	4.053
HiFi-Codec	1.681	0.84	2.419	0.919	3.216	4.296
Mimi ▲	2.515	1.688	1.611	0.612	2.498	3.679
MUFFIN	**1.615**	**0.758**	**2.658**	**0.934**	3.444	**4.454**
MUFFIN ▽	1.681	0.817	2.232	0.914	3.516	4.276
MUFFIN ▲	1.725	0.875	2.086	0.904	**3.56**	4.129

Table 11. The table presents the WER of ASR performance on reconstructed speech from each NAC's codebook using the Whisperlarge V3 pre-trained model. (Ng et al, 2025)

	MUFFIN		**RVQ**		**Hifi-Codec**		**DAC**		**Encodec**	
GT	WER: 2.41; STOI: -									
	WER	*STOI*	*WER*	*STOI*	*WER*	*STOI*	*WER*	*STOI*	*WER*	*STOI*
All	**2.67**	0.94	2.72	0.93	3	0.919	3.53	0.901	3.15	0.9
Code 1	70.3	0.644	75.6	0.702	154	0.572	36.1	0.731	33.7	0.764
Code 2	114	0.379	141	0.426	139	0.454	132	0.148	159	0.199
Code 3	191	0.436	100	0.157	100	0.09	100	0.079	153	0.121
Code 4	107	0.082	101	0.086	112	0.129	100	0.049	147	0.094

Table 12. Zero-shot TTS results for VALL-E with three codecs. (Ng et al, 2025)

Systems	WER (%)	MOS	S-MOS	SECS
VALL-E w/ Encodec	21.05	3.91 ± 0.287	3.70 ± 0.368	0.5914
VALL-E w/ HiFi-Codec	32.35	4.00 ± 0.532	4.04 ± 0.333	0.5874
VALL-E w/ MUFFIN	**12.2**	**4.18 ± 0.278**	**4.19 ± 0.288**	**0.6099**

Sparse Signal Processing: Sparse signal processing capitalizes on the fact that many signals are inherently compressible in a suitable transform domain, such as wavelet, Fourier, or a learned dictionary. Compressive sensing (CS) theory demonstrates that a k-sparse signal of length N can often be reconstructed from substantially fewer than N samples by formulating and solving a sparse recovery problem. This reconstruction is typically performed via ℓ_1-norm minimization or through greedy pursuit algorithms, such as Orthogonal Matching Pursuit (OMP), to obtain the sparsest solution that remains consistent with the observed measurements. These methods offer provable recovery guarantees under mild assumptions regarding the sensing matrix. By exploiting signal sparsity, CS facilitates sub-Nyquist sampling through projections onto either random or structured measurement domains, followed by reconstruction using convex optimization or iterative greedy strategies (Jia et al., 2025; Zhang et al., 2025; Tian et al., 2025; Yi et al., 2025; Gao et al., 2025; Oyerin).

Compressive sensing (CS) has transformed a wide range of sensing applications by exploiting the inherent sparsity of signals and systems. In cognitive radio, for instance, the occupancy of a wide RF band is often sparse, only a small subset of frequencies is active, enabling rapid spectrum sensing from a limited set of radio samples (Tawfik et al., 2024). Subspace-based approaches have also been proposed to address the specific challenges of wideband MIMO cognitive radio networks (Morghare et al., 2024). Similarly, mmWave MIMO channels exhibit sparse angular supports, which facilitates the design of compressive pilot schemes that substantially reduce training overhead.

In radar and imaging, CS enables high-resolution reconstruction from significantly fewer measurements by leveraging the sparsity of the underlying scene, as demonstrated in compressive radar imaging and single-pixel camera systems. Recent advances in mmWave massive MIMO channel estimation formulate both uplink and downlink estimation as sparse recovery problems in the angular domain. Techniques such as simultaneous orthogonal matching pursuit (SOMP) and distributed compressive sensing (DCS) are employed to jointly recover angular-domain coefficients using far fewer pilots, thereby lowering pilot overhead while improving estimation accuracy (Oyerinde et al., 2024; Gao et al., 2024). These developments have yielded substantial performance gains, enabling millimeter-wave imaging systems to attain gigapixel-equivalent resolution with reduced sampling requirements and allowing cognitive sensors to detect spectral holes in real time.

The convergence of sparse modeling and machine learning has led to the emergence of learned sparse recovery (Deka et al., 2025). Deep-unfolding networks, such as ISTA-Net, map model-based iterative shrinkage algorithms (e.g., ISTA) into a fixed number of trainable phases, integrating the optimization framework with learnable nonlinearities and trainable step sizes (Zhang & Ghanem, 2018). These architectures provide provable optimization and generalization benefits in many scenarios and, in practice, can achieve near-optimal sparse reconstructions at substantially lower inference costs, enabling orders-of-magnitude speedups in real-time imaging and sensing applications on resource-constrained hardware (Shah et al., 2024; Machidon & Pejović, 2023).

The synergy between model structure and data-driven adaptation enhances robustness against measurement noise and model mismatch, offering greater parameter efficiency and improved resilience compared with fully black-box neural networks (Deka et al., 2025; Alhejaili et al., 2025). For example, both learned and unfolded estimators, as well as recent generative-prior methods, including diffusion-model priors, have been applied to recover sparse MIMO channels in real time and to reduce pilot overhead. Likewise, model-based end-to-end learning techniques have been developed to learn and mitigate hardware impairments in sensing and communication transceivers (Gao et al., 2024; Fesl et al., 2024; Zhou et al., 2024; Mateos-Ramos et al., 2023). Collectively, these hybrid sparse–deep approaches advance the state of the art in sensing and imaging, spanning wireless communications, medical imaging, and remote sensing, by enabling high-fidelity, low-latency recovery from limited measurements.

Emerging Algorithms

Deep Reinforcement Learning (DRL): Deep reinforcement learning integrates reinforcement-learning principles with deep neural networks, enabling agents to learn optimal policies by interacting with complex, time-varying environments (Saadi et al., 2025). In telecommunication-oriented digital signal processing (DSP), DRL has proven particularly effective for adaptive tasks such as dynamic resource allocation, power control, and interference management in non-stationary wireless channels, as highlighted in recent surveys and application studies on edge-resource scheduling and radio resource management (Ismail et al., 2025; Wang et al., 2024). Unlike many classical optimization approaches, DRL can adapt online to evolving network conditions without relying on explicit parametric channel models, making it well-suited for 5G and Open-RAN deployments and promising for 6G use cases, subject to practical considerations regarding sample efficiency and operational safety (Huang et al., 2024; Tran et al., 2024).

DRL-based adaptive DSP frameworks have been applied to optimize modulation and coding, dynamic spectrum access, and scheduling in self-organizing networks. For example, hierarchical and refined DRL agents have been developed for real-time resource scheduling and beamforming in multi-user massive MIMO and RIS-assisted systems, with strong simulation evidence demonstrating improvements in throughput and latency (Zargari et al., 2024; Wang et al., 2024). DRL also enables proactive interference mitigation and energy-efficient power control by leveraging long-term reward structures derived from network feedback, with several 2024 studies reporting improved energy–throughput trade-offs over classical heuristics in time-varying scenarios, including IWCL and accelerated DRL methods for power control (Tran et al., 2024). From an implementation perspective, recent work demonstrates edge and Open-RAN deployments using model compression and hardware-aware training workflows to meet strict latency and memory constraints. Combined with DRL-based automation, these strategies enhance self-organizing network (SON) capabilities, accelerating the practical integration of DRL into next-generation communication systems (Huang et al., 2024; Zhang et al., 2025).

Neural Architecture Search (NAS): Neural Architecture Search automates the design of neural network architectures by exploring large design spaces to discover models optimized for specific tasks and constraints (Wang & Zhu, 2024). In telecommunications and digital signal processing (DSP), NAS has been applied to develop efficient and compact models for physical-layer tasks such as channel estimation, CSI feedback, signal detection, and modulation recognition. Automated CSI-feedback architectures and channel-estimation–specific NAS frameworks have demonstrated superior reconstruction accuracy and reduced model complexity compared with manually designed baselines (Li et al., 2024; Shi & Huang, 2024). NAS techniques explicitly facilitate trade-offs among accuracy, latency, and computational complexity, a critical requirement for real-time and embedded DSP. Recent surveys and methodological studies emphasize multi-objective and predictor-based strategies that incorporate latency and hardware costs into the search objective (Wang & Zhu, 2024; Sinha et al., 2024; White et al., 2023).

Hardware-aware NAS approaches embed platform constraints, such as memory footprint, bit-width, inference latency, and bit operations (BOPs), directly into the optimization objective, employing latency and energy predictors or hardware-in-the-loop evaluations to ensure that the discovered architectures can be synthesized for FPGA, ASIC, or MCU targets (Ji et al., 2024; Sinha et al., 2024; Garavagno et al., 2024). In practice, NAS-derived networks have been used to optimize CNN-based modules for OFDM and MIMO channel estimation and CSI compression, achieving improved estimation accuracy under multipath fading while reducing inference time and parameter count (Shi & Huang, 2024; Li et al., 2024). NAS has also been employed to discover hybrid architectures, for example, convolutional–recurrent or CNN–attention combinations, that capture both temporal and spatial structures in massive MIMO and RIS-assisted scenarios, offering enhanced robustness to mobility and multipath dynamics compared with single-paradigm designs (Shi & Huang, 2024; Sinha et al., 2024). Finally, industrial- and edge-oriented NAS pipelines have shown that the search process can be executed with hardware constraints, even on-device or on IoT gateways, and co-designed with quantization and pruning to produce deployable models, an encouraging direction for automated physical-layer optimization and accelerating innovation cycles in wireless systems (Garavagno et al., 2024; Weitz et al., 2025; Fayyazi et al., 2025).

Graph Neural Networks (GNNs): Graph neural networks extend deep learning to graph-structured data by employing message-passing and aggregation mechanisms to learn relational representations among nodes and edges (Nguyen et al., 2024; Kose et al., 2024). This inductive bias, modeling devices, antennas, reflecting elements, and their interactions as graph nodes and edges, makes GNNs particularly well suited to wireless communication problems with inherent spatial and topological structure (Nguyen et al., 2024; Le et al., 2025). In MIMO and RIS-assisted systems, GNN architectures have been designed to jointly predict beamformers, infer CSI-related features, and allocate resources by capturing spatial dependencies among antennas, RIS elements, and users (Demirhan & Alkhateeb, 2024; Le et al., 2025). Several recent studies report that GNN-based beamforming and RIS phase-shift designs can match or exceed classical optimization baselines in sum-rate and spectral efficiency, while scaling more efficiently with network size due to permutation equivariance and localized message passing (Le et al., 2025; Zou et al., 2025).

Concretely, GNN models have been employed to jointly optimize base-station beamforming and RIS phase shifts, including multi-RIS and STAR-RIS configurations, demonstrating enhanced robustness to mobility and multipath dynamics while adhering to practical hardware constraints in simulation environments (Le et al., 2025; Lim & Vu, 2025). From an industry and edge-deployment perspective, recent surveys and system-level studies report active progress in developing computationally efficient and quantized GNNs, together with FPGA- and edge-accelerator–based toolchains that enable low-latency, on-device GNN inference for beamforming, user association, and interference management. As illustrated in Figures 18-20, these capabilities align closely with the ultra-reliable low-latency communication (URLLC) requirements and broader performance targets of 5G and 6G networks (Nguyen et al., 2024; Kose et al., 2024; Zou et al., 2025).

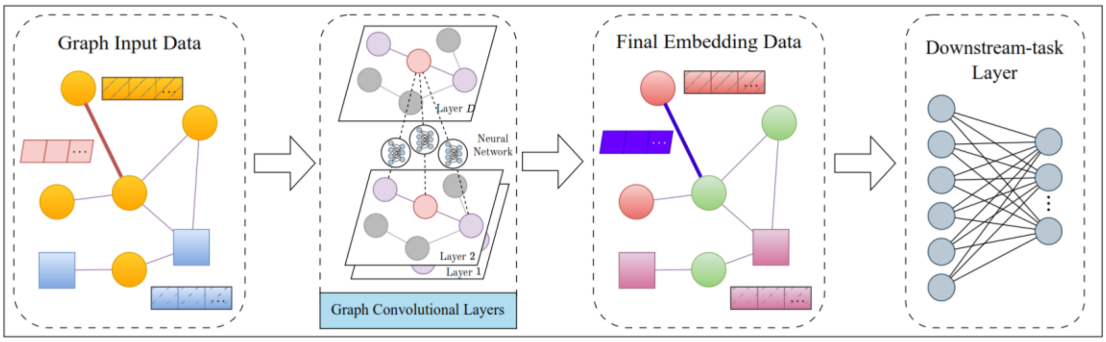

Figure 18. A general pipeline for solving wireless network problems by using GNNs. (Nguyen et al, 2024)

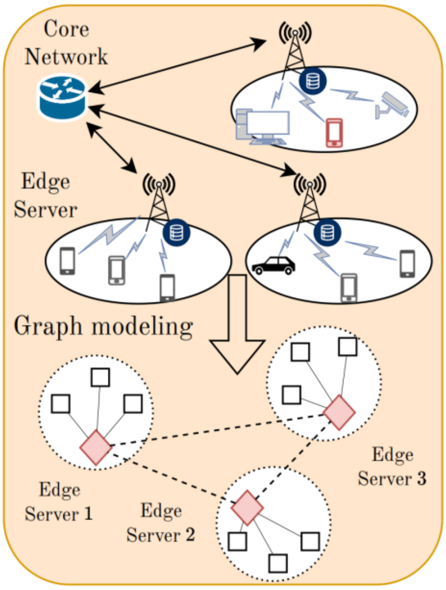

Figure 19. MEC-enable network. (Nguyen et al, 2024)

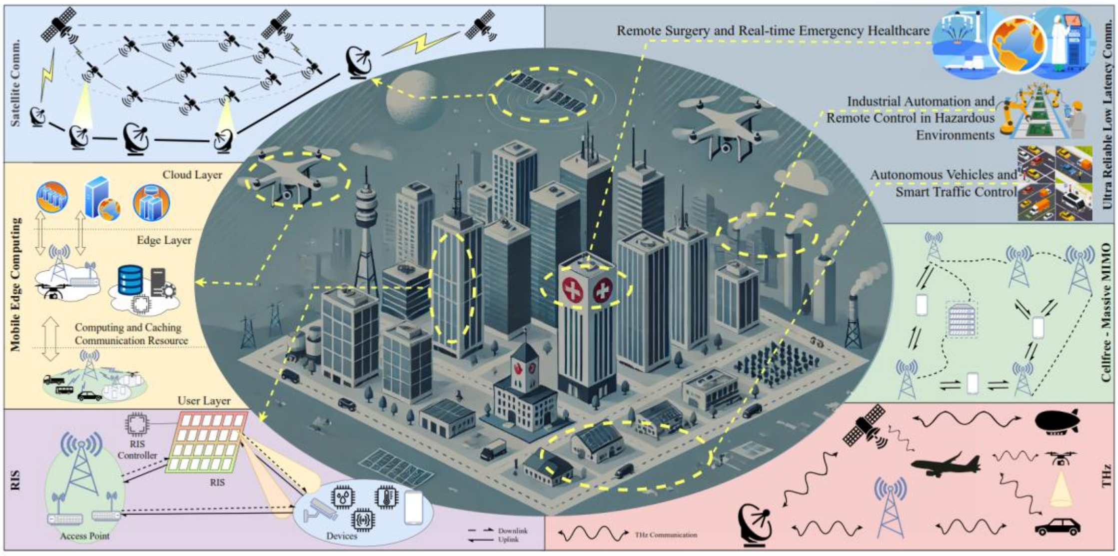

Figure 20. Next generation IoT networks. (Nguyen et al, 2024)

Bayesian optimization (BO): Bayesian optimization is a sample-efficient, probabilistic framework for the global optimization of expensive or noisy black-box functions. It constructs a surrogate model, most commonly a Gaussian process (GP), and employs an acquisition function to identify promising evaluation points, thereby minimizing the total number of costly experiments (González-Duque et al., 2024; Maus et al., 2025). In digital signal processing (DSP) for telecommunications, BO is extensively applied for hyperparameter tuning of complex adaptive algorithms, including filter coefficients, neural-network architectures, and waveform or configuration parameters, where conventional grid or random search methods are computationally prohibitive. Recent work has showcased BO workflows tailored for channel-simulation-based system configuration and for automated hyperparameter tuning of learning-based DSP modules (Sato & Suto, 2024; Onorato, 2024). By predicting performance metrics such as bit-error rate, spectral efficiency, or latency from candidate parameter settings, BO iteratively refines its surrogate models to converge on high-performance configurations with minimal evaluations. Advances in scalable GP approximations and pre-trained GP surrogates have further expanded BO's applicability to higher-dimensional DSP problems (González-Duque et al., 2024; Maus et al., 2025).

The efficiency and adaptability of BO make it well-suited for time-varying, near-real-time DSP tasks where rapid reconfiguration is essential, examples include pilot tuning, adaptive equalizer adjustments, and parameter selection based on costly channel simulations. Dynamic BO variants and simulation-driven frameworks have demonstrated substantial throughput and performance gains in wireless communication experiments (Bardou et al., 2024; Sato & Suto, 2024). Concrete telecommunications use cases include BO-assisted parameter searches for reconfigurable intelligent surfaces (RIS) and location-prediction pipelines, as well as BO-guided optimization of channel estimation strategies and configuration trade-offs. Experimental and simulation studies show that BO can significantly accelerate convergence and improve the robustness of equalizer and estimator parameter settings compared with traditional grid and random search approaches (Hu et al., 2024; Kato et al., 2025). These results underscore BO's capacity to navigate high-dimensional, non-convex design spaces in wireless systems while adapting to dynamic and uncertain channel conditions.

From an industrial and systems integration perspective, BO is increasingly embedded into broader automated machine learning (AutoML) pipelines, including BO combined with neural architecture search (NAS) and reinforcement learning (RL). In hardware-aware design loops, BO enables co-optimization of algorithmic hyperparameters and platform constraints such as power consumption, memory footprint, and latency. This integration is especially impactful in the design space exploration of FPGAs and ASICs, where latency and energy predictors, noise-aware objectives, and multi-objective BO techniques support balanced trade-offs between performance and resource efficiency. By unifying algorithm tuning with hardware-level optimization, BO has become a practical and powerful tool for deploying adaptive DSP solutions on embedded and edge devices (Yang et al., 2025; Padovano et al., 2024).

Comparative Analysis of Algorithms

Classical DSP algorithms: Classical DSP algorithms, including the least mean squares (LMS) family, the fast Fourier transform (FFT), Wiener and Kalman filters, and classical spectral estimators, remain foundational to signal processing because of their mathematical transparency, closed-form behavior, and provable convergence under well-specified assumptions. These methods typically offer compact computational footprints and predictable latency, which facilitates dependable real-time deployment on embedded and mission-critical platforms such as satellite and defense communications. Nevertheless, classical approaches often rely on restrictive assumptions (e.g., linearity, stationarity, Gaussian noise) and manual parameter tuning, which can limit their effectiveness in strongly non-stationary, nonlinear, or data-rich 5G and 6G and Internet-of-Things scenarios. Modern learning-based methods, exemplified by deep reinforcement learning (DRL), neural architecture search (NAS), graph neural networks (GNNs), and Bayesian optimization (BO), mitigate many of these limitations by providing adaptive behavior, automated model discovery, structured relational inference, and sample-efficient optimization. At the same time, data-driven approaches introduce new costs and constraints: substantial offline training and search overhead (notably for NAS and DRL), significant data or interaction requirements, and often reduced interpretability compared with classical models unless interpretable or model-aware designs are adopted. Deployment on resource-constrained edge platforms is feasible but typically requires complementary techniques such as quantization, pruning, hardware-aware neural-architecture design, or compression pipelines to meet latency and memory budgets. Consequently, classical DSP and modern ML-based approaches should be seen as complementary: retain classical methods where determinism, interpretability, and predictable resource use are paramount, and leverage learning-based architectures when managing complexity, nonlinearity, and highly dynamic environments at scale. (Deka et al., 2025; Tran et al., 2025; Fatima & Kondamuri, 2025)

Tables 13 and 14 provide a comprehensive comparison of traditional digital signal processing (DSP) algorithms and modern machine learning (ML)-driven approaches, including Deep Reinforcement Learning (DRL), Graph Neural Networks (GNN), Neural Architecture Search (NAS), and Bayesian Optimization (BO). Table 13 outlines key attributes such as adaptability, computational complexity, interpretability, scalability, robustness, and deployment readiness, highlighting the strengths and limitations of each technique. Table 14 complements this by mapping the features of classical and modern DSP methods to optimal application domains, such as OFDM modulation, channel estimation, beamforming, spectrum sensing, and resource allocation, offering a balanced view of their effectiveness in meeting the challenges of contemporary telecommunication networks.

Table 13. Comparative Scientific Analysis of Classical vs. Modern DSP Algorithms

Aspect	Classical DSP Algorithms	Modern ML-Based DSP Algorithms (DRL, GNN, NAS, BO)
Adaptivity	Limited (e.g., LMS, RLS); relies on static models or fixed adaptation	Highly adaptive to environment via learning (e.g., DRL policies, GNN-based updates)
Computational Complexity	Low complexity; real-time execution on DSP/FPGA platforms	High during training; optimized inference possible via NAS and model compression
Interpretability	Transparent; model parameters are physically and mathematically interpretable	Often opaque (black-box); interpretability is low, especially with deep learning models
Scalability	Limited in high-dimensional or complex topologies	Highly scalable using GNNs (spatial graphs), multi-agent DRL, and NAS-generated architectures
Robustness to Nonlinearity and Noise	Limited performance in nonlinear or low SNR conditions	High robustness; DRL and GNN learn from complex, noisy, and non-Gaussian environments
Automation and Design Flexibility	Requires manual tuning and expert knowledge; fixed structure	Fully automatic (NAS, BO); can evolve architectures and hyperparameters dynamically
Deployment Readiness	Highly optimized for embedded hardware (DSP cores, FPGAs)	Emerging; suitable with model quantization and pruning for edge and embedded deployment
Latency and Memory Footprint	Deterministic, very low latency, minimal memory usage	Acceptable with hardware-aware NAS; can achieve <10 ms latency with <100 kB memory on edge devices
Learning Capability	Static models; no environment learning	Dynamic learning from feedback, interactions, and data; adapts over time
Energy Efficiency	Highly efficient on fixed-point, low-power processors	Improving via quantization, pruning, and NAS-guided lightweight models
Best Use Cases	Fixed filtering, OFDM modulation and demodulation, classical channel estimation, echo cancellation	Beamforming, resource allocation, adaptive coding, RIS control, dynamic spectrum access

Table 14. Feature vs. Application Mapping in Wireless Communication Systems

Application Domain	Preferred Approach	Justification / Notes
OFDM Modulation/Demodulation	Classical (FFT/IFFT), NAS-enhanced CNNs	FFT is optimal for OFDM; NAS can improve neural OFDM decoder latency and accuracy for low-power systems
Channel Estimation	Classical (LS, MMSE), GNN, NAS	GNNs model spatial/user correlations; NAS selects optimal DNN-based estimators for dynamic channels
Beamforming (Massive MIMO)	GNN, DRL	GNNs model antenna-user topology; DRL learns dynamic beam patterns for user mobility
Spectrum Sensing	Sparse DSP, DRL, BO	DRL captures temporal occupancy patterns; BO optimizes sensing parameters under constraints
Adaptive Filtering	Classical (LMS/NLMS), BO-optimized DL filters	Classical filters are efficient; BO tunes DL filters for non-stationary, complex signals
Resource Allocation (5G/6G)	DRL	DRL learns optimal power and scheduling policies in dense, dynamic, multi-agent networks
RIS-Assisted Communication	GNN, BO	GNNs model RIS-device-channel relations; BO optimizes phase-shift configurations for reconfigurable surfaces
Error Correction Coding	Classical (LDPC, Turbo), NAS-designed DNN decoders	NAS enables design of ultra-light, low-latency neural decoders for real-time use
Denoising in Biomedical Signals	Wavelet + Classical Filtering, NAS + CNN	DWT is effective; NAS-optimized CNNs further improve real-time SNR without signal distortion
AI-Enhanced SDR Pipelines	DRL, BO, GNN	Enables reconfigurable SDR via AI policies, fast adaptation to spectral dynamics

Applications of DSP Algorithms

Modulation and Demodulation (e.g., QAM, OFDM): Classical digital signal-processing (DSP) primitives such as the fast Fourier transform (FFT) and matched filtering remain essential building blocks in modulation and demodulation chains, and are used directly in OFDM transceivers and matched-filter detectors. (Yin & He, 2024). These legacy algorithms deliver low computational overhead and implementable complexity, enabling efficient real-time (de)modulation and practical low-complexity receiver designs while retaining robustness across a wide range of channel conditions including noise, multipath fading and Doppler effects when combined with appropriate synchronization and framing strategies (Li et al., 2024). Because FFT and IFFT processing and matched-filter correlation directly implement frequency-domain demodulation and coherent detection, they remain indispensable primitives in contemporary communication systems and integrated sensing-and-communication (ISAC) designs.

Recent advances have established neural architecture search (NAS) as a practical framework for automatically designing neural networks tailored to communication tasks such as modulation recognition and demodulation (Salmani Pour Avval et al., 2025; Zhang et al., 2024; Karakoca et al., 2025). NAS pipelines enable the creation of adaptive, hardware-aware receiver architectures that address realistic channel impairments, such as multipath fading and low SNR, by identifying task-specific building blocks and quantization-aware configurations that satisfy strict real-time latency and memory constraints (Gao et al., 2024; Zhang et al., 2024). For handling transmitter nonlinearities (e.g., PA distortion), complementary RF-domain machine learning and compensation methods can also be incorporated into the NAS search objective (Chen et al., 2021; Xu et al., 2021).

Architectures discovered via progressive NAS (PNAS), often combined with quantization-aware search, pruning, or knowledge transfer, have been shown to enhance classification and demodulation robustness in challenging channel conditions. Dai and Liu (2025) reported that their PNAS-based automated modulation classification framework consistently outperformed classical feature-engineered DSP pipelines and manually designed neural networks, achieving higher accuracy and reduced complexity across diverse wireless scenarios. Furthermore, integrating classical signal-processing primitives (e.g., FFT, matched filtering, domain transforms) with NAS-discovered neural components, especially under hardware-aware NAS strategies, significantly improves the reliability and runtime efficiency of modern wireless receivers, enabling practical edge and embedded deployment.

Channel estimation: Channel estimation is a cornerstone of wireless communication systems, as accurate channel state information (CSI) underpins coherent detection, decoding, and key downstream operations such as demodulation, interference cancellation, and receiver combining (Li et al., 2025), as illustrated in Figures 21-24. Classical methods, most notably least squares (LS) and minimum mean square error (MMSE), remain standard due to their analytical tractability and well-understood bias–variance trade-offs under typical conditions (Arellano et al., 2024; Li et al., 2025). However, their performance degrades in highly dynamic environments (fast fading, high mobility) and in sparse, high-frequency channels like mmWave and THz, where limited scattering, rapid temporal variation, and excessive pilot overheads undermine their assumptions. In such scenarios, compressive sensing and sparsity-aware estimators are increasingly adopted to improve robustness and efficiency (Yao et al., 2024; Li et al., 2025).

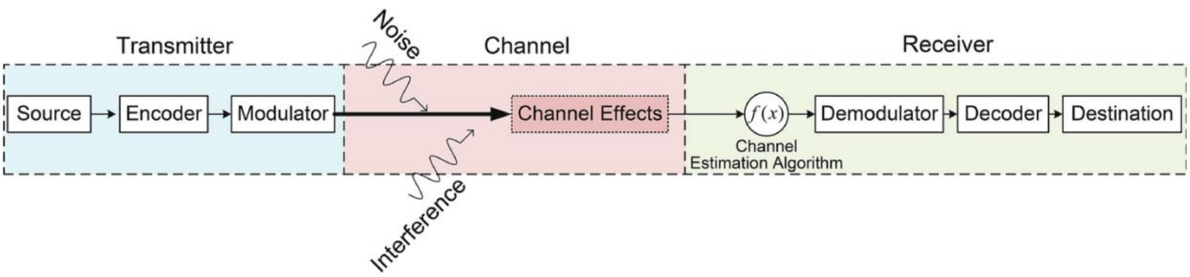

Figure 21. Wireless communication model. (Li et al, 2025)

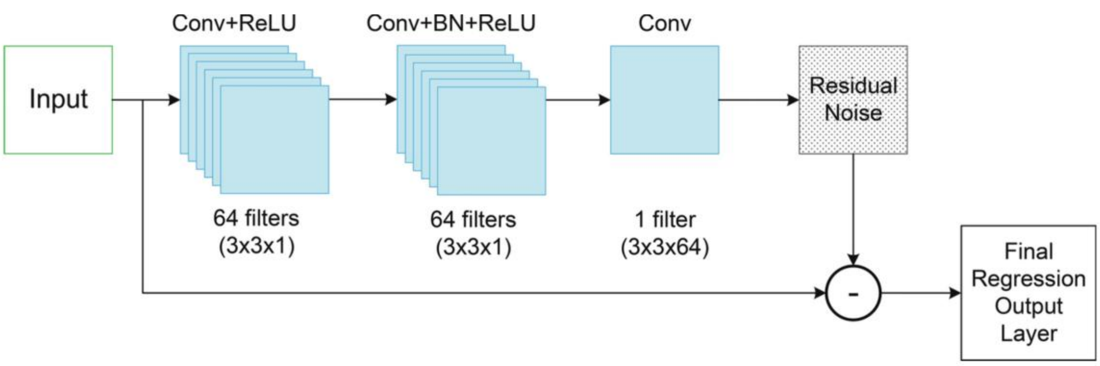

Figure 22. Typical CNN structure. (Li et al, 2025)

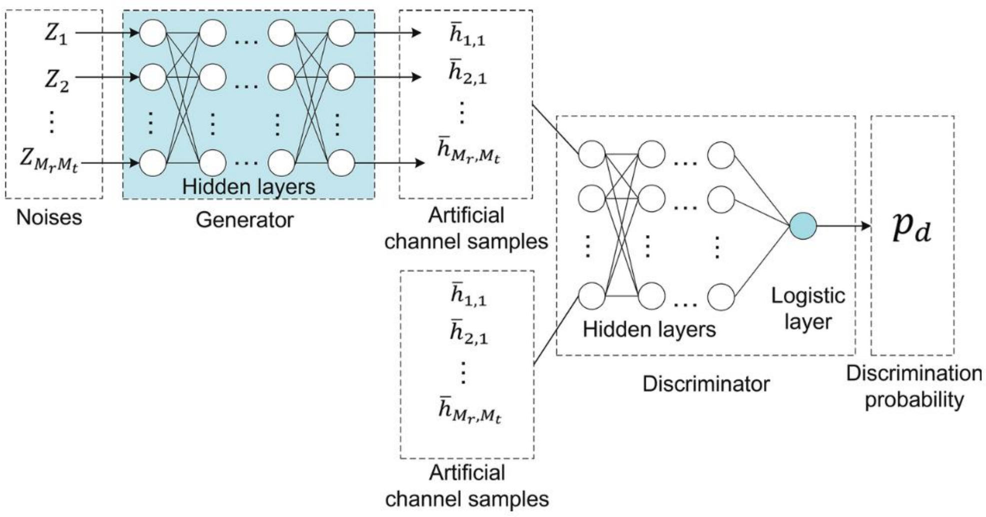

Figure 23. Typical GAN structure. (Li et al, 2025)

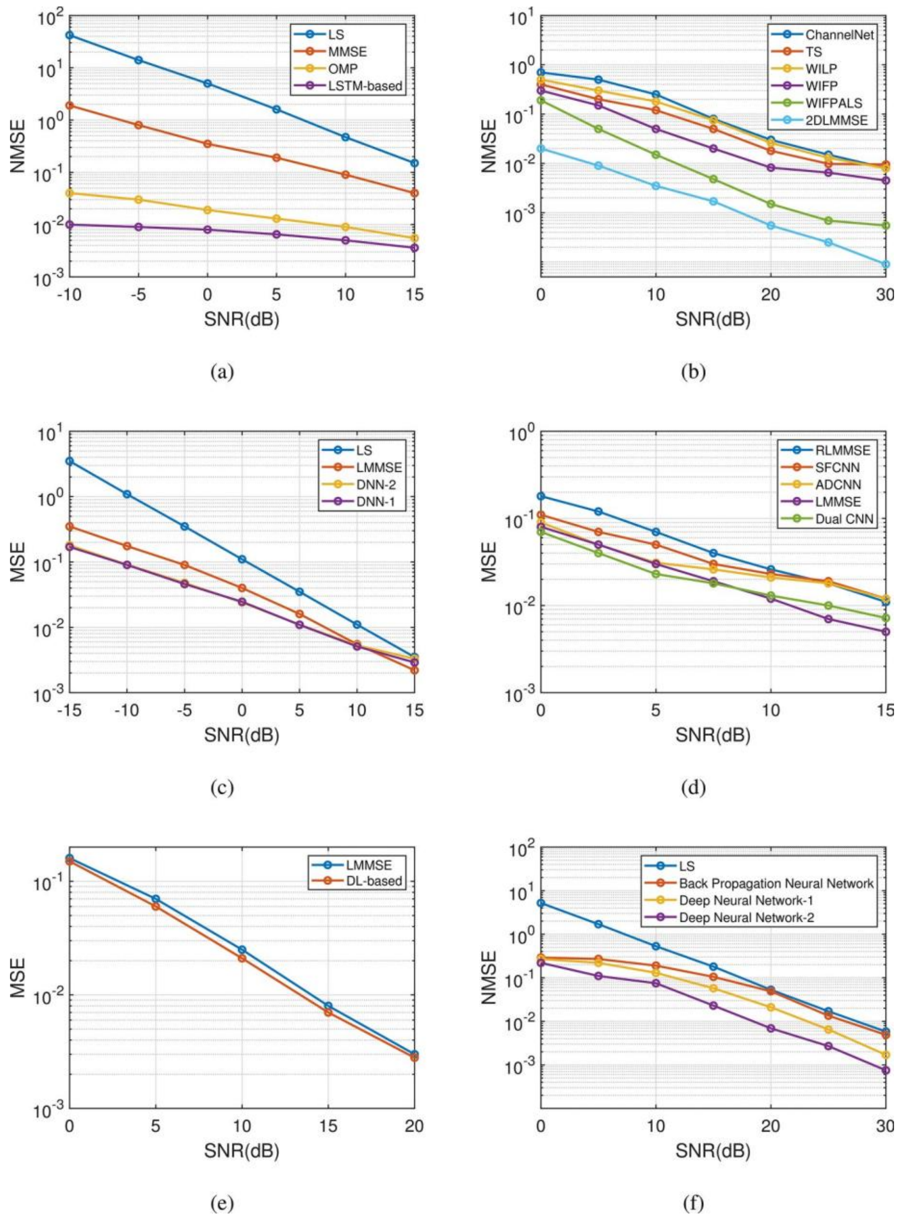

Figure 24. NMSE (MSE) performance of different channel estimation methods. (Li et al, 2025)

Advances in model-free and data-driven strategies have emerged as effective alternatives or complements to model-based estimation. Reinforcement learning (RL) and other sequential decision-making frameworks can optimize pilot allocation and beam selection without relying on explicit parametric channel models, thereby enabling real-time adaptation in nonstationary wireless environments (Li et al., 2025). Emerging diffusion-driven generative models combined with decision transformers (DTs) demonstrate the ability to map pilot observations to plausible channel realizations and adapt rapidly to changing channel distributions with minimal online samples, outperforming retraining-based approaches in adaptation speed (Zhang et al., 2024). In parallel, Neural Architecture Search (NAS) and hardware-aware NAS variants, such as quantization- and edge-aware NAS, provide systematic frameworks for designing compact, latency- and memory-efficient neural estimators suitable for real-time deployment on FPGAs, ASICs, and edge devices (Gao et al., 2024; Lu et al., 2024; Wang & Zhu, 2024). While most QA-NAS evidence originates from computer vision and general edge AI benchmarks, early RF-domain applications, including modulation classification and demodulation, report improved robustness and reduced complexity over manually designed networks and some classical detectors (Dai & Liu, 2025; Zhang et al., 2024). A promising near-term path is a hybrid approach: retain analytically tractable DSP primitives where effective (e.g., synchronization, coarse estimation) while embedding NAS-optimized, hardware-validated learned modules into components most sensitive to nonstationarity and sparsity. However, systematic RF-domain and hardware-in-the-loop validation remains essential, as many reported gains are still based on simulated or preprint-stage studies. (Li et al., 2025; Gao et al., 2024; Wang & Zhu, 2024)

Noise and Interference Reduction: Efficient noise suppression and interference mitigation are critical for achieving high-quality reception in dense, heterogeneous wireless networks, where interference and nonstationary noise directly constrain throughput, latency, and link reliability (Trabelsi et al., 2024). Classical techniques such as Wiener filtering and wavelet-based denoising provide well-established theoretical foundations and strong performance when noise statistics are stationary or approximately known (Benesty, 2024; Frusque et al., 2024). Nevertheless, their performance degrades in highly dynamic or spatially correlated interference environments, such as rapidly varying noise patterns across antenna arrays, highlighting the need for adaptive and learning-driven solutions. (Tulyakova & Trofymchuk, 2024; Trabelsi et al., 2024)

State-of-the-art advances in data-driven and automated-search methods have introduced adaptability and topology-aware modeling into noise-mitigation pipelines, delivering tangible improvements in both simulation and application-specific tests. Bayesian optimization (BO) has been employed to automate parameter tuning for denoisers and filters (e.g., wavelet thresholds, bilateral-filter parameters) and to accelerate beam and parameter selection in wireless systems, with variants tailored for dynamic objectives or beam tracking achieving notable throughput gains and reduced overhead in channel-simulation and 5G-compliant scenarios (Taassori, 2024; König et al., 2025; Maggi et al., 2023; Sato & Suto, 2024). Likewise, state-of-the-art graph neural networks (GNNs) leveraging spatial and topological relationships in MIMO, cell-free, RIS, and STAR-RIS systems enable distributed beamforming and interference suppression, often delivering scalable, near-optimal performance that surpasses conventional optimization approaches (Le et al., 2024; Salaün et al., 2024; Chan et al., 2024). In parallel, deep reinforcement learning (DRL) and multi-agent DRL frameworks have been applied to dynamic denoising, power control, and RIS configuration, supporting policies that adapt in real time to nonstationary interference and changing channel conditions (Pi et al., 2024; Nguyen, 2025). Together, BO, GNNs, and DRL form a complementary toolkit to classical filters, offering a hybrid path toward robust interference mitigation; however, many high-impact demonstrations remain simulation-based or adapted from other domains, making RF-domain, over-the-air, and hardware-in-the-loop validation essential for quantifying real-world gains and trade-offs (Taassori, 2024; König et al., 2024; Le et al., 2024; Maggi et al., 2023).

MIMO systems and beamforming: Massive MIMO and beamforming are cornerstone technologies for enhancing spectral efficiency and network capacity in 5G and beyond, especially in mmWave frequency bands where narrow, high-gain beams are essential to counteract severe path loss and enable spatial multiplexing (Shahjehan et al., 2024). Classical linear beamforming and precoding schemes such as zero-forcing (ZF) and minimum mean squared error (MMSE) serve as strong analytical baselines but encounter scalability, signaling-overhead, and robustness limitations in ultra-dense, highly dynamic deployments. These challenges become more severe under conditions such as rapid user mobility, pilot contamination, fronthaul constraints, and stringent latency requirements. (Ngo et al., 2024)

Graph neural networks (GNNs) provide a natural framework for representing the spatial and relational structure of antennas, users, and reflecting elements. By learning distributed message-passing rules, GNNs can jointly optimize active precoding and passive element control, such as RIS phase shifts, while effectively capturing complex channel correlations and interference patterns that are analytically intractable (Le et al., 2024). Recent model-based GNN beamforming studies report millisecond-level inference in simulations and introduce problem reformulations that reduce runtime complexity (Zhang et al., 2025). From a hardware perspective, FPGA-targeted GNN accelerators such as GraphAGILE and LL-GNN have demonstrated low-latency, resource-efficient inference on Xilinx Alveo platforms, achieving sub-millisecond to sub-microsecond latencies in domain-specific deployments (Zhang et al., 2023; Que et al., 2024). Surveys and accelerator studies further document compiler and hardware design patterns that consistently deliver millisecond-or-better end-to-end latency for GNN inference on edge platforms (Zhang & Xia, 2024; Procaccini et al., 2024). Collectively, simulation evidence and FPGA-based prototypes support the claim that GNN-based beamforming can outperform classical algorithms in many studied scenarios, though large-scale over-the-air (OTA) validation remains an essential step for practical deployment. (Shahjehan et al., 2024; Ngo et al., 2024; Le et al., 2024; Zhang et al., 2025; Zhang & Xia, 2024; Que et al., 2024)

AI-Powered DSP Applications

Deep Reinforcement Learning (DRL): Deep reinforcement learning equips model-free agents with the capability to learn adaptive resource-allocation and interference-management policies directly from online feedback, eliminating the dependence on explicit channel models and enabling robust operation in non-stationary cellular environments. Both single-agent and multi-agent DRL frameworks have demonstrated substantial empirical gains in power control, scheduling, and link adaptation across simulated and prototype wireless systems, frequently surpassing classical heuristic methods under dynamic traffic loads and high-mobility conditions (Sohaib et al., 2024; Malhotra et al., 2025; Li et al., 2024).

Neural Architecture Search (NAS): Neural Architecture Search has emerged as a key enabler for the automated design of compact, task-specific neural architectures in wireless communications, offering systematic mechanisms to balance accuracy, latency, and hardware costs for edge deployment scenarios (Gao et al., 2024; Lu et al., 2024). Recent domain-specific investigations, ranging from NAS frameworks tailored for modulation recognition under severe channel impairments to surveys on NAS combined with transfer learning for 6G applications, highlight the potential of hardware- and quantization-aware NAS variants to meet stringent power consumption and latency requirements of IoT and mobile receivers (Jiang et al., 2024; Orucu et al., 2024).

Graph Neural Networks (GNNs): Graph Neural Networks naturally model antennas, users, and reflecting elements as graph nodes, learning message-passing rules to jointly optimize active precoding and passive element control (e.g., RIS) and thus capture spatial correlations and interference structures that per-link methods often miss (Le et al., 2024; Gu et al., 2024). Model-aware and heterogeneous GNNs have demonstrated scalability and cross-user generalization in ISAC and ultra-dense MIMO scenarios, with model-based designs achieving millisecond-level inference in simulations (Zhang et al., 2024; Wang & Wong, 2024). For example, Zhang et al. (2024) introduced a model-based GNN combining MMSE and hybrid ZF–MRT priors to map CSI directly to beamforming vectors, using multi-head attention, residual connections, and a scheme-selection module trained in an unsupervised, multi-input regime. Results show enhanced energy efficiency, scalability, and low-latency performance with minimal accuracy loss, as illustrated in Figures 25-27 and Table 15. Coupled with FPGA and accelerator co-design, GNN inference can achieve low-latency, energy-efficient operation in prototypes, though large-scale over-the-air validation remains a key next step.

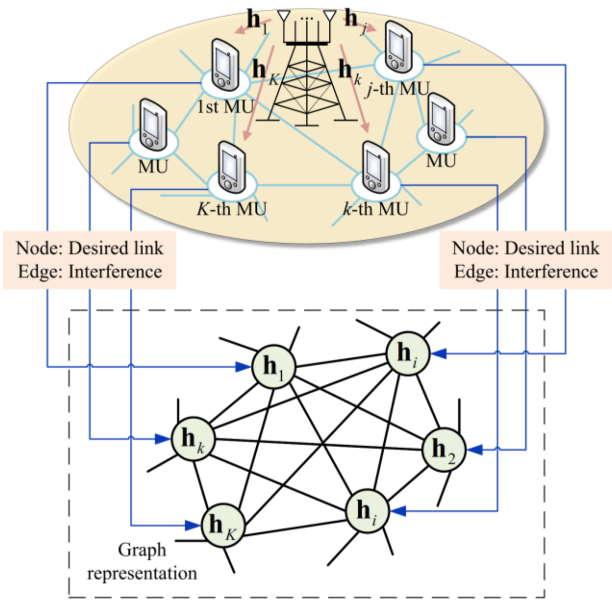

Figure 25. A graph representation of the MU-MISO system. (Zhang et al, 2024)

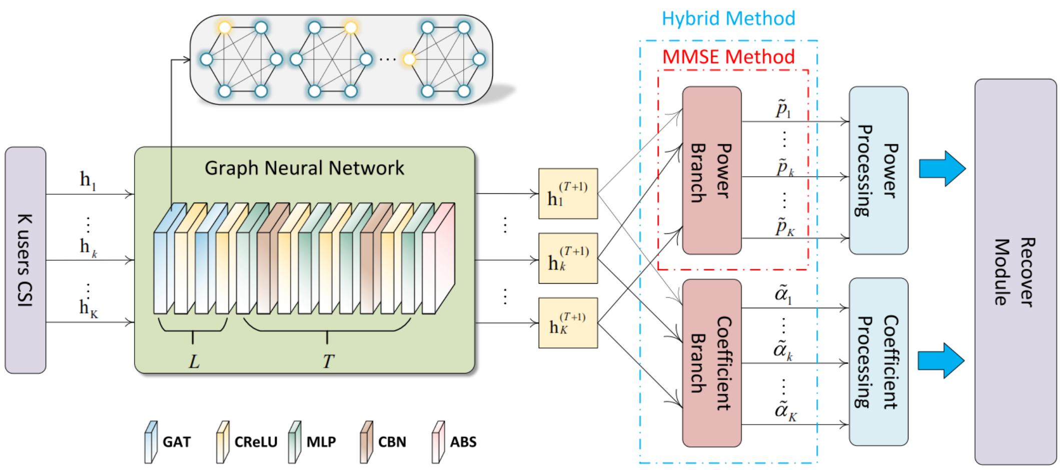

Figure 26. The illustration of the architecture of the model-based GNN. (Zhang et al, 2024)

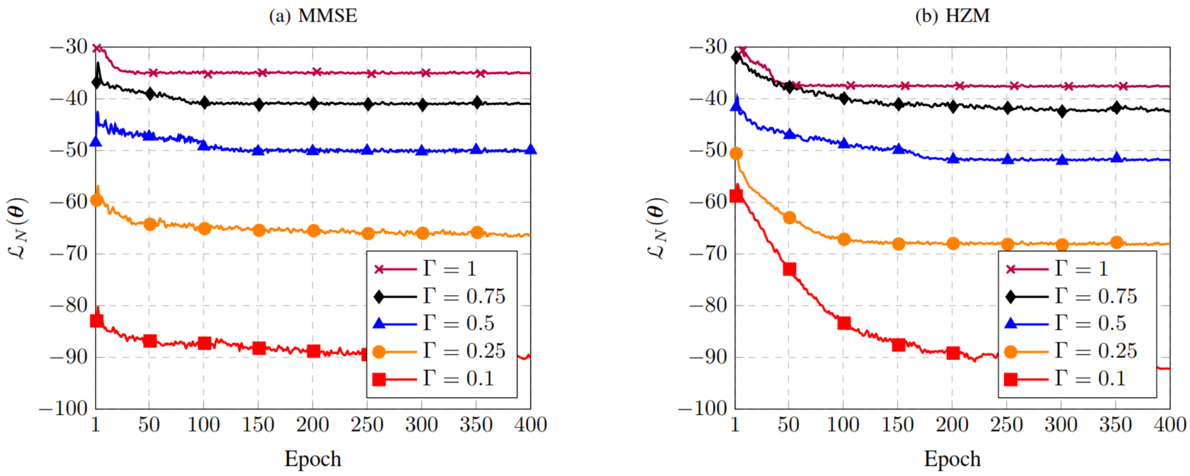

Figure 27. Convergence behavior of the model-based GNN with the MMSE and HZM scheme. (Zhang et al, 2024)

Table 15. Average inference time over the test set of dataset NO. 11. (Zhang et al, 2024)

Approach	CVX	MLP (MMSE)	MLP (HZM)	CNN (MMSE)	CNN (HZM)	GNN (MMSE)	GNN (HZM)
Inference time	11.4 s	1.60 ms	1.59 ms	25.4 ms	25.4 ms	4.94 ms	4.99 ms

Bayesian Optimization (BO): Bayesian Optimization has established itself as a sample-efficient framework for black-box optimization, enabling the tuning of hyperparameters, as well as filter and denoiser settings, under noisy or computationally costly evaluation regimes. By reducing the reliance on manual calibration, BO has proven valuable in adaptive filtering, classification, and related digital signal processing (DSP) pipelines (Taassori, 2024; Onorato, 2024). Recent methodological advances extend BO to scenarios with stringent resource constraints and dynamically evolving objectives, making it increasingly suitable for FPGA-based and edge-deployed wireless systems. For instance, resource-aware BO approaches have demonstrated hardware-constrained parameter tuning on soft processors and FPGA platforms (Guo et al., 2024), while adaptive experimental-design–driven BO frameworks have been applied to channel-simulation-assisted base-station placement and power allocation, addressing time-varying optimization challenges in practical wireless network design (Sato & Suto, 2024).

Table 16 compares major application domains, objectives, challenges, and industrial deployments, highlighting the integration of advanced AI-driven methods, DRL, NAS, GNNs, and BO, into next-generation systems to address the complexity, scalability, and adaptability requirements of 5G, 6G, and beyond.

Table 16. Key DSP Application Areas, Challenges, and AI-Enhanced Industrial Implementations in Wireless Communication

DSP Application Area	Key Objectives	Practical Challenges	Industrial Examples & Standards
Modulation and Demodulation (QAM, OFDM)	Improve spectral efficiency, reduce bit error rate (BER), ensure robust transmission	PAPR reduction, multipath fading, noise and synchronization issues	5G NR (Qualcomm Snapdragon), Wi-Fi 6/7, DVB-T2
Channel Estimation	Fast and accurate CSI acquisition for coherent decoding and MIMO optimization	Sparse mmWave channels, fast fading, noise sensitivity	5G Base Stations (Huawei, Ericsson), Open RAN frameworks
Noise and Interference Reduction	Suppress non-stationary noise and multi-user interference for clearer signal quality	Dynamic noise variation, algorithmic complexity, limited SNR	VoIP platforms (Zoom, Microsoft Teams), NOMA systems, industrial IoT
MIMO Systems and Beamforming	Maximize SNR, improve spatial reuse, enable focused energy transmission	Antenna coordination, channel tracking, hardware calibration	Massive MIMO in 5G, V2X, RIS-based beamforming, drone communications
Case Studies and Real-World Deployments	Evaluate DSP performance in practical and time-sensitive applications	Integration with real systems, scalability, low-latency constraints	URLLC, autonomous vehicles, AR/VR systems, satellite DSP units
AI-Driven DSP: Deep Reinforcement Learning (DRL)	Real-time adaptive decision-making for resource allocation and interference mitigation	Training cost, environment modeling, deployment on edge hardware	6G testbeds, O-RAN intelligent controllers, NVIDIA Aerial SDK
AI-Driven DSP: Neural Architecture Search (NAS)	Auto-discovery of task-optimized deep models for DSP on constrained hardware	Computational cost, architecture constraints, convergence speed	Edge-optimized AI chips, Google NAS Bench, Qualcomm AI Research
AI-Driven DSP: Graph Neural Networks (GNNs)	Exploit spatial structure for enhanced beamforming, channel estimation, interference suppression	Graph modeling overhead, dynamic topology adaptation	RIS-aided MIMO systems, vehicular mesh networks, federated MIMO
AI-Driven DSP: Bayesian Optimization (BO)	Efficient tuning of DSP model parameters under noisy or sparse feedback conditions	Sample efficiency, surrogate model quality, latency in online learning	Adaptive filtering in cognitive radios, low-power IoT signal classifiers

Emerging Trends and Integration with New Technologies

Artificial Intelligence and Machine Learning Combined with DSP: The integration of artificial intelligence (AI) and machine learning (ML) into traditional digital signal processing (DSP) pipelines is fundamentally reshaping the design and optimization of physical-layer and link-layer functions in modern wireless systems, enabling tighter co-design of algorithmic blocks and data-driven approaches (Cui et al., 2025; Wang et al., 2025). Unlike fixed analytical models, AI-empowered DSP systems learn statistical structures directly from measurements, leading to improved performance in practical tasks such as noise suppression (via learned denoisers), data-driven channel estimation, and automatic modulation and classification under realistic fading and interference conditions (Li et al., 2025; Wu et al., 2024; Zuo et al., 2024). Complementing these advances, Neshaastegaran and Jian (2025) demonstrated conditional and plug-in diffusion denoisers, along with general CoDiPhy-style diffusion frameworks, for physical-layer denoising and interference cancellation, further validating the potential of diffusion-based methods for robust wireless signal processing. In highly dynamic wireless environments characterized by rapid fading, mobility, or distributed and heterogeneous deployments, data-driven adaptivity is indispensable. Simulation and prototype studies reveal that pretrained models combined with online fine-tuning or few-shot adaptation significantly outperform static estimators in re-adaptation speed and robustness (Zhang et al., 2024; Wang et al., 2025). Survey research further documents on-device optimization and deployment techniques that enable practical implementation of these adaptive schemes (Wang et al., 2025).

Convolutional neural networks (CNNs), U-Net architectures, and sample-to-symbol neural equalizers exemplify these trends, having been successfully applied to channel equalization and signal reconstruction with demonstrated reductions in post-equalization error rates and improved bit error rates (BER) compared to classical linear equalizers in numerous experimental and simulation studies (Osadchuk et al., 2024; Xu, 2024). For example, Pasic et al. (2025) extend this evidence to millimeter-wave (mmWave) MIMO channel estimation, employing U-Net and CNN-based mmWave estimators that exploit sub-6 GHz side information, thereby highlighting their practical advantages in challenging propagation scenarios. From an industrial perspective, AI-augmented DSP functions are increasingly validated in operator testbeds, edge-AI stacks, and Open RAN frameworks (O-RAN nGRG), and are progressively adopted in edge and IoT deployments to enhance throughput and interference resilience. These developments rely on hardware-aware model design and rigorous testing pipelines to satisfy stringent real-time and reliability requirements (Alavirad et al., 2024). Industry initiatives and platforms such as the NVIDIA 6G Research Cloud provide critical infrastructure for large-scale testing and prototyping (Reuters, 2024). Complementing these efforts, Wang et al. (2025) offer comprehensive surveys on on-device AI, explicitly addressing hardware and latency trade-offs alongside best practices for effective deployment.

Deep Reinforcement Learning (DRL) as a Resource Management Technique in Self-Organizing Networks: Deep reinforcement learning (DRL) has quickly become a cornerstone data-driven approach for managing radio resources in self-organizing networks (SONs). Unlike traditional methods that rely on hand-crafted heuristics, DRL agents learn near-optimal control policies through direct interaction with the network environment. These techniques have been successfully demonstrated in key SON tasks such as dynamic spectrum access and channel selection (Safavinejad et al., 2024), latency-aware network slicing via xApps in O-RAN testbeds (Sever et al., 2025), and distributed transmit-power control across multi-cell configurations (Kim & So, 2025). Extensive emulation and cellular studies reveal that DRL consistently outperforms classical rule-based or static optimization approaches.

A significant strength of DRL in wireless systems lies in its ability to adapt policies online, accommodating user mobility, time-varying traffic, and partially observed channels without relying on explicit parametric channel models. This adaptability facilitates plug-and-play operation within near-real-time RAN Intelligent Controller (near-RT RIC) and O-RAN frameworks, which often incorporate digital twins for enhanced network management (Cui et al., 2024; Sever et al., 2025). However, despite these advantages, practical adoption faces challenges including poor sample efficiency, long training times, increased computational and signaling demands in large networks, and issues related to stability and generalization under out-of-distribution or adversarial network conditions (Eldeeb & Alves, 2025; Han et al., 2025; Safavinejad et al., 2024).

To overcome these limitations, current research emphasizes offline and distributional reinforcement learning to reduce risky online exploration (Eldeeb & Alves, 2025), federated and multi-agent training frameworks to lower communication overhead and enhance privacy (Wu & Fang, 2024), and model-based or simulator-assisted pretraining combined with risk-aware objectives to improve convergence and safety in deployment (Cui et al., 2024; Eldeeb & Alves, 2025; Sever et al., 2025). Industry and 6G research consortia are actively prototyping DRL-enabled SON components within O-RAN architectures to decrease operational expenditures (OPEX) and meet strict QoS requirements. Nonetheless, widespread deployment will require advances in explainability, standardization, and the development of compute-efficient, edge-capable learning pipelines (Ullah et al., 2025; Cui et al., 2024). With continued improvements in sample-efficient algorithms, distributional and offline training regimes, and hardware-aware federated learning, DRL is well-positioned to become a vital tool for autonomous 6G self-organizing networks, provided that remaining engineering and safety challenges are resolved (Eldeeb & Alves, 2025; Kim & So, 2025).

Neural Architecture Search (NAS) for AI-Native Physical Layer Optimization: Neural Architecture Search (NAS) automates the design of neural network topologies and hyperparameters, significantly reducing the need for manual trial-and-error and enabling the creation of reproducible, task-specific architectures without requiring extensive expert-driven engineering (Salmani Pour Avval et al., 2025). In physical-layer applications such as channel decoding and learned equalization, NAS has demonstrated the ability to discover compact and high-performing architectures that meet or surpass human-designed networks in communication-task accuracy, as shown in recent experimental studies (Zhu et al., 2025; Siddiqui et al., 2025).

Hardware-aware NAS (HW-NAS) methods extend conventional NAS approaches by explicitly incorporating constraints related to latency, memory, and energy consumption into the search objectives. This ensures that the resulting neural models are deployable on resource-constrained embedded targets, such as microcontrollers and edge system-on-chips (SoCs), which is particularly valuable for radio front-ends and edge digital signal processing accelerators (King et al., 2025; Aach et al., 2025). By leveraging differentiable and one-shot search paradigms, latency lookup tables, real-device latency measurements, and post-search techniques such as quantization and pruning, modern HW-NAS pipelines can produce models that achieve real-time inference while maintaining competitive accuracy and throughput for representative physical-layer tasks (King et al., 2025; Aach et al., 2025), as illustrated in Figures 28-31.

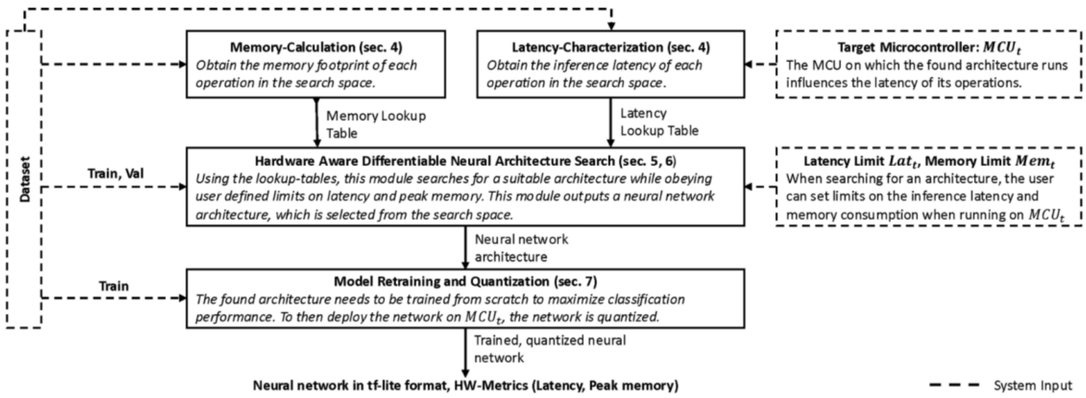

Figure 28. MicroNAS system overview. (King et al, 2025)

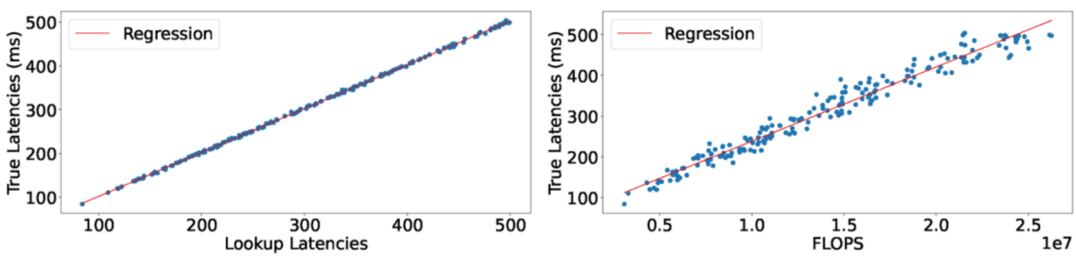

Figure 29. Execution latency of whole architectures from the search space. (King et al, 2025)

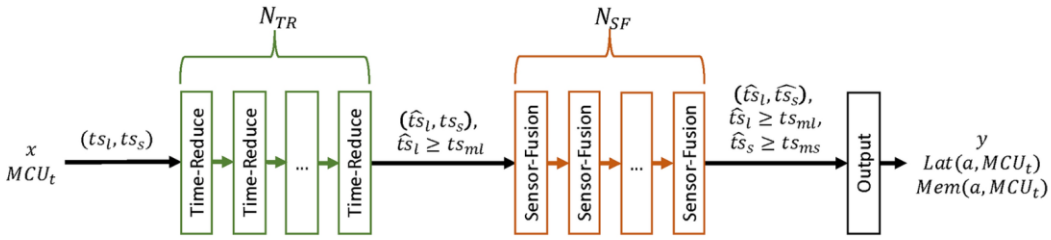

Figure 30. High-level overview over the search space. (King et al, 2025)

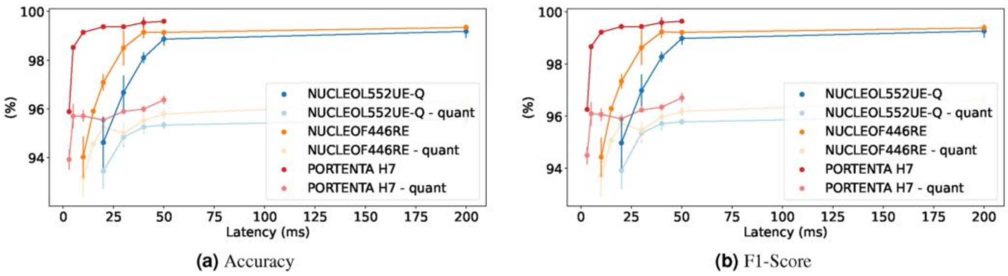

Figure 31. Latency trade-offs on the UCI-HAR dataset. (King et al, 2025)

The rise of AI-native physical-layer design, where entire signal-processing blocks are learned end-to-end, has been accelerated by NAS, enabling joint optimization of algorithmic performance and implementation cost. This paradigm shifts portions of traditional hand-crafted DSP pipelines toward jointly optimized learned modules, providing greater flexibility and efficiency (Zhu et al., 2025; Zhou et al., 2024). Nonetheless, challenges remain, including reducing the computational cost of the search process, improving architecture transferability across diverse channel conditions and hardware platforms, and co-optimizing NAS outputs with accelerator toolchains. Recent advances in efficient global search strategies and design-principle transfer directly address these issues, enhancing the practicality and adoption of NAS for physical-layer optimization (Siddiqui & Kyrkou, 2025; Zhou et al., 2024)

Graph Neural Networks (GNN) for Beamforming in Massive MIMO and RIS-Assisted Systems: Graph neural networks (GNNs) naturally align with wireless system modeling as they operate directly on graph-structured inputs, where nodes represent antennas, users, or reconfigurable intelligent surface (RIS) elements, and edges correspond to wireless channel links. This structure enables GNNs to leverage permutation equivariance and topology-aware message passing, which effectively reflect the underlying physical network relationships (He et al., 2025; Tam et al., 2024).

In massive MIMO scenarios, GNNs have demonstrated the ability to learn mappings from channel state information (CSI) to beamforming vectors that scale efficiently with array and user size. These models achieve enhanced spectral efficiency and improved interference suppression compared to conventional deep learning methods that do not incorporate network topology (Li et al., 2024; He et al., 2025; Li & Vu, 2025). For RIS-assisted systems, GNN-based architectures model users, RIS elements, and base stations as bipartite or heterogeneous graphs, enabling the joint optimization of transmit beamformers and RIS phase configurations. This approach efficiently captures multi-hop interactions, leading to substantial coverage and data rate improvements in multi-RIS deployments (Lim & Vu, 2025; Le et al., 2024; Nazar et al., 2025).

Recent advances focus on scalable GNN designs, including subgraph and message-passing strategies, attention mechanisms, and transfer learning or over-the-air (OTA) implementations. These innovations allow GNNs to generalize across varying topologies and produce near-real-time inference suitable for large antenna arrays and dense network deployments (He et al., 2025; Zhang et al., 2025; Sáez de Ocáriz Borde et al., 2024). Despite these promising developments, significant challenges remain in applying GNNs to wireless systems. Key issues include constructing robust graph representations from noisy or incomplete CSI and measurements, maintaining rapid adaptability to highly dynamic channels and mobility, and meeting stringent latency and computational constraints on edge and front-end hardware. Although recent lightweight and noise-aware GNN models have begun to address these challenges, comprehensive solutions are still under development (Norouzi et al., 2025; He et al., 2025; In et al., 2025).

Bayesian Optimization (BO) for Adaptive Signal Processing: Bayesian optimization (BO) is a sample-efficient, probabilistic framework for tuning expensive black-box functions and has been increasingly adopted for online configuration and management in wireless and edge signal-processing scenarios (Yan et al., 2024). BO constructs a probabilistic surrogate, often a Gaussian process or a Bayesian neural surrogate, of the performance landscape and uses an acquisition function to guide evaluations toward the most informative and promising configurations with minimal costly trials (Garrido-Merchán, 2025). Due to its sample efficiency, BO has become a leading method for automated hyperparameter tuning in modern neural networks and broader hyperparameter-optimization workflows within signal-processing pipelines (Franceschi et al., 2024). In practice, BO and its adaptive variants have been applied to tune controllers and low-level parameter sets analogous to adaptive filter coefficients, as well as to optimize hyperparameters and thresholds in data-driven spectrum-sensing models (König et al., 2024; Elmorsy et al., 2025). By automating parameter search, BO reduces manual trial-and-error, accelerates deployment, and enables continual re-tuning under changing operating conditions, benefits consistently reported across signal-processing and wireless-system studies (Suwandi, 2024; Yan et al., 2024).

However, standard BO techniques face challenges in high-dimensional parameter spaces, particularly when configurations involve mixed discrete and continuous variables, which can hinder scalability and sample efficiency (González-Duque et al., 2024; Garrido-Merchán, 2025). To address these issues, recent research advocates hybrid and hierarchical strategies that combine BO's global surrogate-based search with local optimizers, trust-region methods, or constrained solvers, thereby improving scalability and performance in complex, constrained optimization problems (Kyriakidis et al., 2025; Li et al., 2024). When integrated into adaptive signal-processing workflows, including mobile edge and 6G-era sensing systems, BO and its adaptive and hybrid variants measurably enhance algorithmic adaptability and real-world performance as operating conditions evolve (Yan et al., 2024; Cao et al., 2024).

Real-Time Implementation and Hardware Acceleration of AI-DSP Models on FPGA/ASIC: Efficient hardware-level implementations are critical to transitioning AI-enhanced digital signal processing (AI-DSP) algorithms from laboratory prototypes to real-world, latency-sensitive systems (Kang et al., 2024; Sadeghi, 2024). Field-programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs) have become common platforms for accelerating neural methods in DSP, including deep reinforcement learning (DRL) agents and graph neural networks (GNNs), delivering orders-of-magnitude improvements in throughput, latency, and energy efficiency compared to general-purpose processors in many embedded scenarios (Lin et al., 2024; Jha et al., 2025). To meet stringent power, area, and latency constraints, model compression techniques such as quantization, structured pruning, and knowledge distillation are routinely applied; when integrated with hardware-aware design, these approaches minimize resource consumption with negligible accuracy degradation (Wang et al., 2024; Zhang et al., 2025).

Hardware-aware model design, involving the co-design of network architectures, quantization formats, and mapping strategies with the target FPGA or ASIC fabric, streamlines deployment and enhances real-device performance beyond hardware-agnostic compression methods (Andronic & Constantinides, 2025; Zhang et al., 2025). Both industry and academia are increasingly prioritizing hardware–software co-design, reconfigurable FPGA-SoC runtimes, and, in some cases, domain-specific ASICs to address the stringent real-time throughput and deterministic performance demands of 5G, edge computing, and emerging 6G services. The ongoing trade-off between the flexibility of reconfigurable platforms, such as FPGAs, virtualized eFPGAs, and SoC fabrics, and the efficiency and lower per-unit cost of fixed ASIC accelerators continues to drive hybrid development flows, including FPGA prototyping followed by ASIC hardening or runtime-reconfigurable co-design, which are now common in production pipelines (Ramhorst et al., 2025; Boudjadar et al., 2025; Bartzoudis et al., 2024; Kang et al., 2024; Reuters, 2024), as illustrated in Figures 32-35 and Table 17.

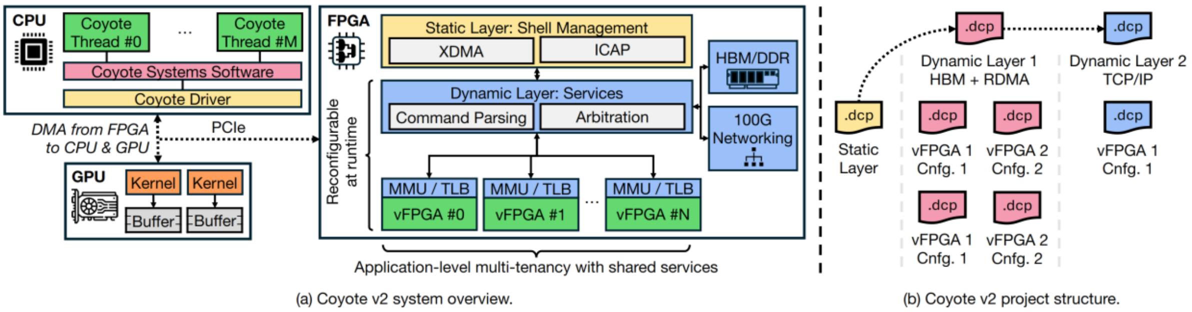

Figure 32. Coyote v2 system overview and project structure. (Ramhorst et al, 2025)

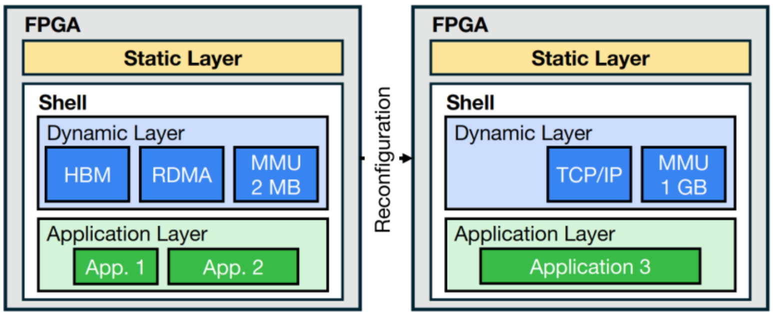

Figure 33. Shell reconfiguration example. (Ramhorst et al, 2025)

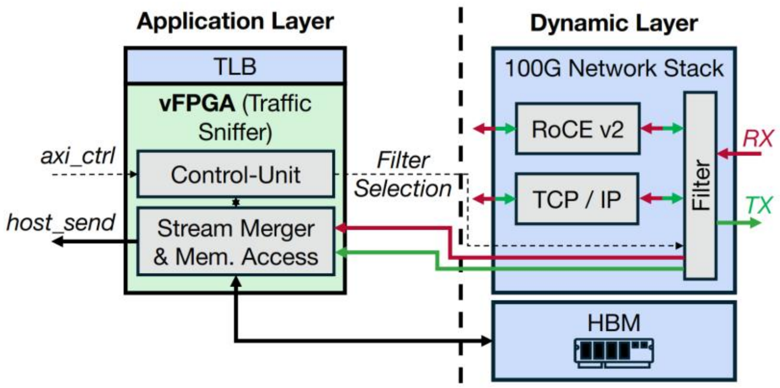

Figure 34. Schematic of the traffic sniffer design with the filter as service and vFPGA-backed application logic. (Ramhorst et al, 2025)

Table 17. Reconfiguration latency for various Coyote v2 shell configs. Average latency with STD reported from 5 trials. (Ramhorst et al, 2025)

Scenario	Coyote kernel latency [ms]	Coyote total latency [ms]	Vivado flow [ms]
#1	51.6 ± 0.0	536.2 ± 5.2	55922.5 ± 443.1
#2	72.3 ± 0.2	709.0 ± 9.1	63045.2 ± 78.0
#3	85.5 ± 0.7	929.1 ± 3.5	71417.9 ± 512.0

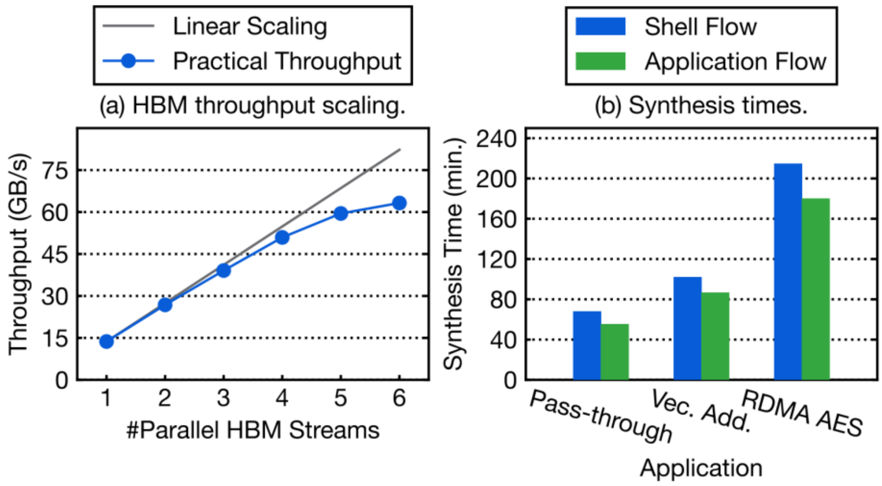

Figure 35. (a) Data transfer throughput scaling with number of HBM channels in one vFPGA. (b) Comparison of synthesis and implementation times with shell and app flow targetting an Alveo U250. (Ramhorst et al, 2025)

Edge Computing and Internet of Things (IoT) Applications: Edge computing brings AI and signal processing closer to data sources, enabling low-latency and privacy-sensitive processing essential for many IoT applications (Gill et al., 2025). Running AI models locally on devices or nearby edge nodes significantly reduces continuous cloud communication, conserving bandwidth and minimizing the risk of exposing raw data off-device (Heydari & Mahmoud, 2025). Federated learning and related privacy-preserving distributed training paradigms allow numerous heterogeneous edge devices to collaboratively improve a shared model while retaining local data on-node, addressing both privacy and communication costs in edge-IoT contexts (Li et al., 2024). Model classes such as deep reinforcement learning (DRL) and graph neural networks (GNNs) are increasingly deployed at the edge, often in hybrid or distributed forms, to dynamically optimize resource allocation, scheduling, and network control amid fluctuating device capabilities and network conditions (Ji et al., 2024; Gu et al., 2025).

Edge AI-enabled processing is critical in domains demanding strict latency, reliability, and privacy guarantees, including autonomous driving, smart-city sensing and control, and continuous health monitoring, where local inference and rapid closed-loop decisions are imperative (Trigka & Dritsas, 2025). Nevertheless, dominant practical constraints on edge deployment, limited compute, memory, and battery resources, drive active research into tiny-ML, model compression, efficient runtimes, and partitioned and distributed inference to render state-of-the-art models feasible on constrained hardware (Heydari & Mahmoud, 2025; Gill et al., 2025). Consequently, contemporary research and system designs emphasize lightweight model architectures, hardware- and topology-aware co-design, adaptive inference strategies such as early-exit and conditional computation, and privacy-preserving protocols to reconcile the conflicting demands of accuracy, latency, energy efficiency, and data privacy at the edge (Ficili et al., 2025; Li et al., 2024)

AI-Enabled Software-Defined Radios (SDRs): AI techniques have transformed software-defined radios (SDRs) into context-aware transceivers capable of real-time, data-driven adaptation in complex and congested spectrum environments, moving beyond fixed hardware architectures (Sabir et al., 2024). Modern AI-powered SDR stacks dynamically select modulation and coding schemes, adjust transmit power, and reconfigure waveform parameters on the fly to maximize link performance and coexistence with incumbent users (Ismail et al., 2025; Manikandan et al., 2025). These cognitive capabilities, driven by spectrum prediction, classification, and learned decision policies, enhance spectral efficiency and reduce harmful interference by enabling opportunistic and interference-aware resource utilization (Pan et al., 2024; Sabir et al., 2024). Prototype AI-SDR systems, along with Open RAN and private 5G testbeds, have demonstrated these benefits in practice, facilitating multi-standard interoperability and dynamic spectrum access in live or emulated 5G environments (Ferguson et al., 2025; Martian et al., 2025).

Despite these advances, key engineering challenges remain for deployed AI-SDRs, including meeting stringent real-time constraints to minimize processing and feedback delays and ensuring robust operation in rapidly time-varying channels and mobile scenarios (Lacava et al., 2025; Akbar et al., 2025). Equally critical is addressing security vulnerabilities: learning-based radio controllers are susceptible to adversarial manipulations, data poisoning, and jamming attacks. Protecting machine learning models and designing trustworthy, detection-capable systems thus remain active priorities in research and deployment efforts (Pelekis et al., 2025; Tan, 2024).

As wireless communication systems move toward 5G and 6G, merging Digital Signal Processing (DSP) with new technologies like Artificial Intelligence (AI), hardware acceleration, edge computing, and software-defined radio (SDR) has become the need of the hour. These mergers provide adaptive, real-time, and efficient processing for ever-evolving complex communication environments. Table 18 provides an overview of the functions, advantages, challenges, and most important applications of these technologies within the paradigm of contemporary DSP frameworks.

Table 18. Integration of DSP with Emerging AI and Hardware Technologies in Wireless Communication Systems

Technology Area	DSP Role and Integration	Key Benefits	Challenges	Examples / Use Cases
AI & ML (DRL, NAS, GNN, BO)	Advanced ML integrated with DSP for adaptive PHY, resource allocation, and model optimization	Adaptive, self-optimizing DSP; improved robustness	High training cost; interpretability; hardware deployment	AI beamforming, DRL spectrum management, NAS decoders, GNN RIS control
FPGA/ASIC Real-Time Processing	Hardware acceleration of DSP and AI models for low latency, high throughput	Ultra-low latency, energy efficiency	Power limits; design complexity	5G basebands, DRL/GNN accelerators, NAS hardware-aware models
Edge Computing & IoT	Local DSP with AI for low-latency, privacy-preserving inference and optimization	Reduced cloud dependency; low latency	Limited resources; edge-cloud partitioning	Smart sensors, UAV DRL optimization, federated learning
Software-Defined Radio (SDR)	Flexible DSP pipelines with AI for dynamic spectrum access and adaptation	High flexibility and real-time adaptation	Power efficiency; real-time constraints	DRL-based spectrum sensing, BO adaptive modulation, RIS cognitive radios

Challenges and Future Directions

Computational Complexity and Real-Time Constraints: Emerging DSP methodologies like Deep Reinforcement Learning (DRL), Neural Architecture Search (NAS), Graph Neural Networks (GNN), and Bayesian Optimization (BO) tend to have high computational expenses especially during training or search processes. This complexity is a show-stopper for real-time deployment in latency-constrained wireless communication systems. While such models provide high flexibility and learnability, their inference and retraining needs must be streamlined for dynamic low-power systems. Specifically, online training and adaptation in mobile setups continue to be energy-hungry and infeasible on most present platforms. Model minimization through pruning, quantization, and transfer learning is an essential avenue to alleviate these challenges.

Hardware Implementation Limitations: It is still difficult to implement sophisticated AI-based DSP algorithms on hardware platforms such as FPGAs and ASICs. DRL and GNN-based models tend to be demanding in terms of large memory footprints, intensive arithmetic operations, and strict latency requirements. Designing hardware-friendly architectures through model quantization, pruning, or hardware-aware NAS is instrumental in facilitating their deployment on mobile and embedded platforms. Further, power efficiency versus throughput tradeoff is nontrivial, particularly for handheld wireless devices. Emerging hardware-software co-design techniques are necessary to map learning-based DSP models to real-time accelerators.

Adaptivity in Time-Varying Wireless Environments: Wireless communication systems are dynamic in nature owing to user mobility, environmental changes, and interference. There is a growing demand for adaptive DSP algorithms that can adapt in real time to such changing environments. DRL, specifically, has promise in dynamic spectrum access and beamforming; however, robustness and convergence in non-stationary environments are still challenges. Future systems need to integrate model-based priors with learning-based policies to provide stability and learning rate boosts. Real-time feedback loops and online fine-tuning methods are also essential for sustenance of performance in unstable environments.

Research Gaps and Open Problems: Despite the growing convergence of artificial intelligence (AI) with digital signal processing (DSP), several key research gaps remain. One of the main challenges is the limited application of Bayesian Optimization (BO) for adaptive signal processing tasks and real-time hyperparameter tuning in non-stationary environments. Similarly, the existing Neural Architecture Search (NAS) techniques are often computationally demanding, making them impractical for real-time DSP system design. There is a pressing need for lightweight, domain-tailored NAS solutions tailored to communication tasks. In addition, the lack of standardized benchmarking procedures hinders objective comparison and evaluation between conventional DSP algorithms and modern AI-based approaches under the same testing conditions. Another poorly explored area is cross-layer optimization, where AI models jointly tackle the interdependence between the physical (PHY), MAC, and application layers for end-to-end system optimizations. Moreover, the robustness of AI models under low-data and noisy scenarios continues to be a significant concern, limiting the generalizability of learned models across heterogeneous wireless environments.

To counter the new challenges in digital signal processing (DSP) for contemporary wireless communication systems, we need to identify the main challenges and look for potential future research opportunities. Table 19 recapitulates the main challenges from computational requirements and hardware limitations to the necessity of adaptive algorithms and robust evaluation frameworks and proposes solutions and active research areas, particularly in the direction of incorporating innovative AI methods like DRL, NAS, GNN, and Bayesian Optimization.

Table 19. Summary of Key Challenges and Future Directions in DSP for Communication Systems

Challenge	Description	Impact Area	Future Directions / Solutions
Computational Complexity	High processing load from AI-based DSP in real-time scenarios	Baseband DSP, massive MIMO	Model pruning, quantization, efficient NAS, low-complexity inference
Hardware Constraints	Difficulty deploying DRL, GNN, NAS on FPGAs/ASICs under power/area limits	SDR, embedded/edge platforms	HLS, fixed-point design, hardware-aware NAS, custom accelerators
Adaptivity in Dynamic Environments	Static algorithms fail in fast-changing channels (e.g., UAVs, mobile IoT)	Beamforming, channel estimation	DRL, GNN for dynamic graphs, online learning, adaptive filtering
Real-Time Processing Limits	Latency issues for AI inference in DSP pipelines	Low-latency systems, edge AI	Pipelined FPGA/ASIC design, lightweight DRL, co-optimization
Lack of Datasets and Benchmarks	Scarcity of open, realistic data for training/testing AI-DSP models	Evaluation, reproducibility	Open datasets, simulators, federated learning, synthetic data tools
Interpretability and Robustness	Black-box nature of AI models limits trust and reliability	PHY security, mission-critical DSP	Explainable AI, uncertainty modeling, interpretable GNNs, BO-guided tuning
Cross-Layer Optimization	Missing frameworks for AI-driven joint PHY-MAC optimization	6G, SONs, full-stack DSP	Multi-agent DRL, GNNs across layers, BO for multi-objective trade-offs

Conclusion

Digital Signal Processing (DSP) remains a foundational element of modern telecommunication and wireless systems, bridging classical signal-processing theory with emerging AI-driven methodologies. Traditional techniques, such as FFT, matched filtering, and adaptive filtering, continue to provide reliable, low-latency solutions for modulation, channel estimation, and noise suppression, while newer approaches including Deep Reinforcement Learning (DRL), Neural Architecture Search (NAS), Graph Neural Networks (GNN), and Bayesian Optimization (BO) enable adaptive, data-driven optimization in complex, dynamic wireless environments.

Despite these advances, practical deployment faces significant challenges. Computational complexity, hardware constraints, and limited adaptivity in time-varying channels pose obstacles for real-time implementation. Many AI-based DSP methods require substantial memory and computational resources, making them difficult to deploy on resource-constrained platforms. Efficient mapping onto FPGA and ASIC accelerators demands hardware-aware model design, aggressive compression (e.g., pruning and quantization), and co-design strategies balancing throughput, latency, and energy efficiency. Furthermore, ensuring robust convergence of learning algorithms in non-stationary channels remains an open problem, limiting the potential for online adaptation and continual learning.

Looking ahead, future research in DSP should prioritize developing lightweight, hardware-friendly AI architectures and hardware-aware NAS procedures tailored to communication tasks; establishing standardized benchmarks and open datasets that evaluate latency, power, throughput, and robustness under realistic mobile and edge conditions; and designing efficient online adaptation and federated learning methods capable of operating under constraints such as limited data, energy budgets, and adversarial or noisy environments. Integrating explainability, security, and privacy considerations into algorithmic and system design will be essential to ensure trustworthiness and scalability.

By uniting principled signal-processing models with resource-aware learning strategies and standardized testing frameworks, the research community can deliver scalable, energy-efficient, and resilient DSP solutions that meet the stringent latency, power, and reliability requirements of next-generation systems, including 6G. Sustained collaboration among academia, industry, and hardware vendors will be key to translating these research directions into practical, deployable solutions that enable intelligent, ubiquitous connectivity.

References

Farooq, U., & Miliou, A. (2025). Advanced DFE, MLD, and RDE equalization techniques for enhanced 5G mm-Wave A-RoF performance at 60 GHz. Photonics, 12, Article 496.

Ghosh, A., Spelman, A., Cheung, T. H., et al. (2023). Reconfigurable signal processing and DSP hardware generator for 5G and beyond transmitters. IEEE Transactions on Very Large-Scale Integration (VLSI) Systems, PP, 112.

Kamil, A. S., Misran, N., & Oleiwi, T. A. (2025). Demystifying 5G: The strategic evolution to massive MIMO for enhanced network capacity and economic gains. Power Technology, 49(2), May.

Khan, A., Shafi, I., Khawaja, S. G., de la Torre Díez, I., Flores, M. A. L., Galván, J. C., & Ashraf, I. (2023). Adaptive filtering: Issues, challenges, and best-fit solutions using particle swarm optimization variants. Sensors (Basel), 23.

Wils, M., Sharma, M., & Moonen, M. (2023). Per-wavelet equalization for discrete wavelet transform based multi-carrier modulation systems. IEEE Open Journal of Signal Processing, 4, 52–60.

Oyerinde, O. O., Flizikowski, A., Marciniak, T., Zelenchuk, D., & Ngatched, T. M. N. (2024). Compressive sensing-based channel estimation for uplink and downlink reconfigurable intelligent surface-aided millimeter wave massive MIMO systems. Electronics, 13.

Niu, Z., Yang, H., Li, L., et al. (2024). Learnable digital signal processing: A new benchmark of linearity compensation for optical fiber communications. Light: Science & Applications, 13.

Pryamikov, A. (2024). Deep learning as a highly efficient tool for digital signal processing design. Light: Science & Applications, 13.

Khan, A., Shafi, I., Khawaja, S. G., de la Torre Díez, I., Flores, M. A. L., Galván, J. C., & Ashraf, I. (2025). Channel estimation in 5G-and-beyond wireless communication: A comprehensive survey. Electronics, 14.

Zhou, X., Chen, X., Tong, L., et al. (2024). Attention-deep reinforcement learning jointly beamforming based on tensor decomposition for RIS-assisted V2X mmWave massive MIMO system. Complex & Intelligent Systems, 10, 145–160.

Sun, Q., Fang, Z., Li, Y., et al. (2025). Jumping knowledge graph attention network for resource allocation in wireless cellular system. Scientific Reports, 15.

Palomares, J., Coronado, E., Tzenetopoulos, A., et al. (2025). Hardware-accelerated edge AI orchestration on the multi-tier edge-to-cloud continuum. Journal of Network and Systems Management, 33.

Lin, C., Azmine, M. F., Liang, Y., & Yi, Y. (2024). Leveraging neuro-inspired AI accelerator for high-speed computing in 6G networks. Frontiers in Computational Neuroscience, 18.

Zhao, D., Ding, R., & Song, B. (2025). Satellite-assisted 6G wide-area edge intelligence: Dynamics-aware task offloading and resource allocation for remote IoT services. Science China Information Sciences, 68.

Zhan, S., Huang, L., Luo, G., Zheng, S., Gao, Z., & Chao, H.-C. (2025). A review on federated learning architectures for privacy-preserving AI: Lightweight and secure cloud–edge–end collaboration. Electronics, 14.

Akbar, M. S., Hussain, Z., Ikram, M., Sheng, Q. Z., & Mukhopadhyay, S. C. (2025). On challenges of sixth-generation (6G) wireless networks: A comprehensive survey of requirements, applications, and security issues. Journal of Network and Computer Applications, 233.

Shen, L.-H., Feng, K.-T., & Hanzo, L. (2023). Five facets of 6G: Research challenges and opportunities. ACM Computing Surveys, 55.

She, C., et al. (2021). A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning. Proceedings of the IEEE, 109, 204–246.

Tarafder, P., Chun, C., Ullah, A., Kim, Y., & Choi, W. (2025). Channel estimation in 5G-and-beyond wireless communication: A comprehensive survey. Electronics, 14.

Zhang, Y., Zhang, Z., Zhang, Y., & Zhang, T. (2021). Blind denoiser-based beam space channel estimation with GAN in millimeter-wave systems. In Proceedings of IEEE Vehicular Technology Conference (VTC-Fall) (pp. 1–5). Norman, OK, USA.

Oppenheim, A. V., & Schafer, R. W. (2009). Discrete-time signal processing (3rd ed.). Upper Saddle River, NJ: Pearson.

Zeng, Z., Liu, J., & Yuan, Y. (2024). A generalized Nyquist-Shannon sampling theorem using the Koopman operator. IEEE Transactions on Signal Processing, 72, 3595–3610.

Garrido, M., Bautista, V. M., Portas, A., & Hormigo, J. (2025). Advanced quantization schemes to increase accuracy, reduce area, and lower power consumption in FFT architectures. IEEE Transactions on Circuits and Systems I: Regular Papers, 72, 203–213.

Damsgaard, H. J., Grenier, A., Katare, D., Taufique, Z., Shakibhamedan, S., Troccoli, T., Chatzitsompanis, G., Kanduri, A., Ometov, A., Ding, A. Y., Taherinejad, N., Karakonstantis, G., Woods, R., & Nurmi, J. (2024). Adaptive approximate computing in Edge AI and IoT applications: A review. Journal of Systems Architecture, 150.

Lin, W., Zhu, Y., & Arslan, T. (2023). DycSe: A low-power, dynamic reconfiguration column streaming-based convolution engine for resource-aware edge AI accelerators. Journal of Low Power Electronics and Applications, 13.

Michon, R., Ducceschi, M., Cochard, P., Skare, T., Webb, C. J., & Russo, R. (2025). Evaluating CPU, GPU, and FPGA performance in the context of modal reverberation: A comparative analysis. Frontiers in Signal Processing, 5.

Komeylian, S., & Paolini, C. (2023). Implementation of the digital QS-SVM-based beamformer on an FPGA platform. Sensors, 23.

Silva, C. A. D., & Fernandes, M. A. C. (2025). New approach of blind adaptive equalizer based on genetic algorithms. Telecom, 6.

Sadeghi, S. (2024). Classifying FPGA technology in digital signal processing: A review. International Journal of Engineering Technology Science, 10.

Khan, D., Jan, L., & Zafar, M. H. (2025). Optimized FFT designs for high-performance LTE and 5G networks. Arabian Journal for Science and Engineering.

Ney, J., & Wehn, N. (2024). Achieving high throughput with a trainable neural-network-based equalizer for communications on FPGA. arXiv preprint arXiv:2407.02967.

Shao, S., Lv, T., Du, S., Lin, Z., et al. (2025). Adaptive sparse channel estimation for RIS-assisted mmWave massive MIMO systems. IEEE Transactions on Vehicular Technology, PP, 1–13.

Shao, S., Lv, T., & Huang, P. (2025). Adaptive filtering for channel estimation in RIS-assisted mmWave systems. *Sensors, 25*(2), 297.

Matei, R., & Chiper, D. F. (2024). Design and polyphase implementation of rotationally invariant 2D FIR filter banks based on maximally flat prototype. Electronics, 13.

Gómez-Luna, E., Cuadros-Orta, D. E., Candelo-Becerra, J. E., & Vasquez, J. C. (2024). The development of a novel transient signal analysis: A wavelet transform approach. Computation, 12.

Kiruluta, A., Burity, P., & Williams, S. (2025). Learnable multi-scale wavelet transformer: A novel alternative to self-attention. *arXiv*.

Xie, R. (M.), Sinclair, A. K., Burgoyne, J., Chapman, S., & Huber, A. (2024). CCAT: Multi-rate DSP for sub-mm astronomy: Polyphase synthesis filter bank on FPGA for enhanced MKID readout. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy XII*, Proceedings of SPIE (Vol. 13102, p. 1310213).

Mohamed, M. A. A., Yu, T. F., & Grandjean, T. (2023). PSO-tuned variable forgetting factor recursive least square estimation of 2RC equivalent circuit model parameters for lithium-ion batteries. In *2023 IEEE Vehicle Power and Propulsion Conference (VPPC)* (pp. 1–6). Milan, Italy.

Bereketoglu, A. B. (2025). Composite reward design in PPO-driven adaptive filtering. *arXiv*.

Wan, Y., & Wang, W. (2024). Combined-step-size affine projection Andrew's sine estimate for robust adaptive filtering. Information, 15.

Al-Ibadi, M., & Mahmood, F. E. (2021). Beam and channel tracking for 5G communication systems using adaptive filtering techniques: A comparison study. Journal of Communications Software and Systems, 17(4), 360–368.

Zhao, H., & Li, C. (2025). Transient error analysis of the LMS and RLS algorithm for graph signal estimation. arXiv:2506.00403.

James, B. P., Leung, M.-F., Vaithiyanathan, D., & Mariammal, K. (2024). Optimal realization of distributed arithmetic-based MAC adaptive FIR filter architecture incorporating radix-4 and radix-8 computation. Electronics, 13.

Khan, M. T. (2024). Application of distributed arithmetic to adaptive filtering algorithms: Trends, challenges and future. *arXiv*.

Saremi, A., Ramkumar, B., Ghaffari, G., et al. (2023). An acoustic echo canceller optimized for hands-free speech telecommunication in large vehicle cabins. Journal of Audio Speech Music Processing, 39.

Li, G., Zheng, C., Ke, Y., & Li, X. (2023). Deep learning-based acoustic echo cancellation for surround sound systems. Applied Sciences, 13.

Pichardo, E., Avalos, J. G., Sánchez, G., Vazquez, E., & Toscano, L. K. (2024). A high-speed acoustic echo canceller based on Grey Wolf Optimization and Particle Swarm Optimization algorithms. Biomimetics, 9.

Kellermann, W. (2008). Echo cancellation. In D. Havelock, S. Kuwano, & M. Vorländer (Eds.), Handbook of Signal Processing in Acoustics (Vol. 1, pp. 883–895). New York: Springer.

Zhao, F., Zhang, C., He, S., & Liu, J. (2024). Deep echo path modeling for acoustic echo cancellation. In Proceedings of Interspeech 2024.

Seidel, E., Enzner, G., Mowlaee, P., & Fingscheidt, T. (2025). Neural Kalman filters for acoustic echo cancellation: Comparison of deep neural network-based extensions. IEEE Signal Processing Magazine, 42, 104–117.

Zhang, H., & Wang, D. (2018). Deep learning for acoustic echo cancellation in noisy and double-talk scenarios. In Proceedings of the 19th Annual Conference of the International Speech Communication Association (Interspeech) (pp. 3239–3243). Hyderabad, India.

Zhao, F., Liu, J., & Zhang, X. (2024). SDAEC: Signal decoupling for advancing acoustic echo cancellation. In Proceedings of Interspeech 2024.

Khanagha, V., Koutsaidis, D., Kalgaonkar, K., & Srinivasan, S. (2024). Interference aware training target for DNN based joint acoustic echo cancellation and noise suppression. In *Interspeech 2024*.

Seidel, E., Mowlaee, P., & Fingscheidt, T. (2024). Efficient high-performance Bark-scale neural network for residual echo and noise suppression. *arXiv*.

Zhao, F., He, S., & Zhang, X. (2025). Room impulse response as a prompt for acoustic echo cancellation. *arXiv*.

Shachar, E., Cohen, I., & Berdugo, B. (2023). Acoustic echo cancellation with the normalized sign-error least mean squares algorithm and deep residual echo suppression. Algorithms, 16.

Zhang, H., & Wang, D. (2022). Neural cascade architecture for multi-channel acoustic echo suppression. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 2326–2336.

Cheng, L., Zheng, C., Li, A., Peng, R., & Li, X. (2022). A deep complex network with multi-frame filtering for stereophonic acoustic echo cancellation. In Proceedings of the 23rd Interspeech Conference (pp. 2508–2512). Songdo Convention, Incheon, Korea.

Vernon, J., & Perera, D. G. (2025). Spectrum analysis with the prime factor algorithm on embedded systems. arXiv preprint arXiv:2501.10864.

Wu, M., Yu, Y., Wan, F., et al. (2024). Decoding of polar-coded OFDM systems over multipath fading channels with interleaving and channel state information. Scientific Reports, 14.

Vasilakis, C., Tsagkaropoulos, A., Koutoulas, I., & Reisis, D. (2025). Improving the fast Fourier transform for space and edge computing applications with an efficient in-place method. Software, 4.

Wang, C., Zhang, J., Wu, Y., & Cong, J. (2025). HP-FFT: A general high-performance FFT generator using high-level synthesis. In Proceedings of the 2025 IEEE 33rd Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM) (pp. 19–23). Fayetteville, AR, USA.

Zheng, Y., Lamperski, A. (2025). Non-asymptotic analysis of classical spectrum estimators for L-mixing time-series data with unknown means. arXiv preprint arXiv:2504.00217.

Muzaffar, M. U., Sharqi, R. (2024). A review of spectrum sensing in modern cognitive radio networks. Telecommun. Syst., 85, 347–363.

Simfukwe, C., An, S. S. A., & Youn, Y. C. (2025). Investigating gamma frequency band PSD in Alzheimer's disease using qEEG from eyes-open and eyes-closed resting states. *Journal of Clinical Medicine, 14*(12), 4256.

El Bandoudi, A., & Boulouird, M. (2024). Intelligent spectrum sensing for 6G wireless communication systems. In *International Workshop on Information Technology and Communication (IWITIC)*, Marrakesh, Morocco.

Mokrý, O., & Rajmic, P. (2025). Tweaking autoregressive methods for inpainting of gaps in audio signals. *EUSIPCO 2025* (accepted). [Preprint]. arXiv.

Jitsumatsu, Y., & Sun, L. (2025). Two-stage Prony-based estimation of fractional delay and Doppler shifts in OTFS modulation. *arXiv* preprint arXiv:2506.17599.

Stoica, P., Moses, R. (2005). Spectral Analysis of Signals. Upper Saddle River, NJ, USA: Pearson/Prentice Hall.

Fei, Z., Zhang, H. (2025). SCAN-MUSIC: An efficient super-resolution algorithm for large-scale single snapshot line spectral estimation. Numer. Algor.

Ding, Z., Epperly, E. N., Lin, L., & Zhang, R. (2024). The ESPRIT algorithm under high noise: Optimal error scaling and noisy super-resolution. In *FOCS 2024* (preprint). arXiv:2404.03885.

Yang, Z., Mo, Y.-L., Tang, G., & Xu, Z. (2022). Separation-free spectral super-resolution via convex optimization. *arXiv* preprint arXiv:2211.15361.

Barker, R. (2025). From DeepSense to Open RAN: AI/ML advancements in dynamic spectrum sensing and their applications. *arXiv* preprint arXiv:2502.02889.

Zhang, W., Wang, Y., Chen, X., Cai, Z., & Tian, Z. (2024). Spectrum transformer: An attention-based wideband spectrum detector. *IEEE Transactions on Wireless Communications, 23*(9), 12343–12353.

Dong, A., Starr, A., & Zhao, Y. (2025). End-to-end identification of autoregressive with exogenous input (ARX) models using neural networks. *Machine Intelligence Research, 22*(4), 117–130.

Mehrabian, A., Sabbaghian, M., & Yanikomeroglu, H. (2024). RL-based hyperparameter selection for spectrum sensing with CNNs. *IEEE Transactions on Communications*, January 2024.

Zheng, S., Ye, Z., Zhang, L., Yue, K., Zhao, Z. (2025). Deep learning-based wideband spectrum sensing with dual-representation inputs and subband shuffling augmentation. arXiv preprint arXiv:2504.07427.

Silik, A., Noori, M., Wu, Z., et al. (2024). Wavelet-based vibration denoising for structural health monitoring. Urban Lifeline, 2, 15.

Jiang, C., Yang, W., Wang, Y., et al. (2025). GNSS signal processing based on improved lifting wavelet transform with prior constraint. Scientific Reports, 15, 20019.

Chen, W., Zhang, Y., Yu, K., et al. (2025). Hybrid DWT NLM method with NOA optimization for ECG signal denoising. Scientific Reports, 15, 24258.

Almutairi, A. F., Krishna, A. (2022). Filtered-orthogonal wavelet division multiplexing (F-OWDM) technique for 5G and beyond communication systems. Scientific Reports, 12, 4607.

Ramadan, K. (2025). Improved DWT-OFDM system performance with wavelet domain equalization for Co-CFO mitigation. *Telecommunication Systems, 88*(3), 1–28.

Zhu, P., Feng, L., Yu, K., et al. (2025). Enhanced DWT for denoising heartbeat signal in non-invasive detection. Sensors, 25, 1743.

Jia, G., Yang, J., Liang, H. (2025). A combined denoising method of adaptive VMD and wavelet threshold for gear health monitoring. Structural Durability & Health Monitoring, 19, 1057–1072.

Pal Shuvo, S., Pal Shibazee, S., Paul, G., et al. (2025). A novel integration of Hodrick–Prescott filter (Hp-filter) and wavelet transform (WT) with optimize support vector machine (PSO-SVM) in predicting solar radiation. Scientific Reports, 15, 14410.

Xiong, Z., Huang, S., Ren, S., Lin, Y., Li, Z., Li, D., Deng, F. (2025). Particle swarm optimization support vector machine-based grounding fault detection method in distribution network. Algorithms, 18, 259.

Zhu, J., Fu, Z., Li, K., Su, A. (2023). Chromatography denoising with improved wavelet thresholding based on modified genetic particle swarm optimization. Electronics, 12, 4249.

Zheng, Z., Dai, W., Xue, D., Li, C., Zou, J., Xiong, H. (2022). Hybrid ISTA: Unfolding ISTA with convergence guarantees using free-form deep neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Frusque, G., Fink, O. (2024). Robust time series denoising with learnable wavelet packet transform. Advanced Engineering Informatics, 62, 102669.

Yu, H., Guo, P., Sano, A. (2024). AdaWaveNet: Adaptive wavelet network for time series analysis. arXiv preprint arXiv:2405.11124.

Nason, G. P., Wei, J. L. (2024). Leveraging non-decimated wavelet packet features and transformer models for time series forecasting. arXiv preprint arXiv:2403.08630.

Necciari, T., Holighaus, N., Balazs, P., Průša, Z., Majdak, P., Derrien, O. (2018). Audlet filter banks: A versatile analysis/synthesis framework using auditory frequency scales. Applied Sciences, 8, 96.

Sivapatham, S., Kar, A., Græsbøll Christensen, M. (2022). Gammatone filter bank-deep neural network-based monaural speech enhancement for unseen conditions. Applied Acoustics, 194, 108784.

Hao, X., Wen, S., Su, X., Liu, Y., Gao, G., Li, X. (2020). Sub-band knowledge distillation framework for speech enhancement. arXiv preprint arXiv:2005.14435.

Lv, S., Hu, Y., Zhang, S., Xie, L. (2021). DCCRN+: Channel-wise subband DCCRN with SNR estimation for speech enhancement. arXiv preprint arXiv:2106.08672.

Schuller, G. (2020). Filter banks and audio coding. Springer.

Kim, M., & Skoglund, J. (2024). Neural speech and audio coding: Modern AI technology meets traditional codecs. *IEEE Signal Processing Magazine, 41*(6), 85–93.

Ng, D., Zhou, K., Chao, Y.-W., Xiong, Z., Ma, B., & Chng, E. S. (2025). Multi-band frequency reconstruction for neural psychoacoustic coding. *arXiv preprint arXiv:2505.07235*.

Herre, J., Quackenbush, S., Kim, M., & Skoglund, J. (2025). Perceptual audio coding: A 40-year historical perspective. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2025*. IEEE.

Brislawn, C. M. (2024). Factoring perfect reconstruction filter banks into causal lifting matrices: A Diophantine approach. *Journal of Computational Algebra, 12*, 100024.

Keerthana, B., et al. (2024). Designing optimal prototype filters for maximally decimated cosine modulated filter banks with rapid convergence. *Heliyon, 10*(11), e31912.

Haider, D., Perfler, F., Balazs, P., Hollomey, C., & Holighaus, N. (2025). ISAC: An invertible and stable auditory filter bank with customizable kernels for ML integration. In *Proceedings of the IEEE International Conference on Sampling Theory and Applications (SampTA) 2025*. IEEE.

Choi, W., Kim, B. H., Lim, H., Jang, I., & Kang, H.-G. (2025). Neural spectral band generation for audio coding. In *Proceedings of Interspeech 2025*.

AMD (2024). Polyphase Channelizer Tutorial. Vitis AI Engine Development.

Feng, T., Zhao, Z., Xie, Y., Ye, Y., Luo, X., Guan, X., Li, Y. (2025). STFTCodec: High-fidelity audio compression through time-frequency domain representation. arXiv preprint arXiv:2503.16989.

Siuzdak, H., Grötschla, F., Lanzendörfer, L. A. (2024). SNAC: Multi-scale neural audio codec. arXiv preprint arXiv:2410.14411.

Wang, J., Jiang, W., Wang, X. (2025). SwitchCodec: A high-fidelity neural audio codec with sparse quantization. arXiv preprint arXiv:2505.24437.

Provan, R., Crockett, L., Stewart, R., & Weiss, S. (2025, July 2–5). *Farrow-based true time-delay FPGA wideband digital beamforming architecture*. In *Proceedings of the 30th IEEE Symposium on Computers and Communications (ISCC)* (pp. 1–6). IEEE. Bologna, Italy.

Kou, Z., You, Q., Kim, J., Dong, Z., Lowerison, M. R., Vaithiyalingam Chandra Sekaran, N., Llano, D. A., Song, P., & Oelze, M. L. (2023). High-level synthesis design of scalable ultrafast ultrasound beamformer with single FPGA. *IEEE Transactions on Biomedical Circuits and Systems, 17*(3), 446–457.

Jia, J., Prater-Bennette, A., Shen, L., et al. (2025). Sparse recovery: The square of ℓ_1/ℓ_2 norms. J. Sci. Comput., 102, 24.

Zhang, X., Xiong, H. (2025). Knockoff-guided compressive sensing: A statistical machine learning framework for support-assured signal recovery. arXiv preprint arXiv:2505.24727.

Tian, Z., Hu, T., Niu, C., Wu, D., Wang, S. (2025). Sampling innovation-based adaptive compressive sensing. arXiv preprint arXiv:2503.13241.

Yi, R., Han, S., Liu, P., Zhang, B., Liu, H. (2025). Methods of sparse measurement matrix optimization for compressed sensing. IET Signal Processing.

Gao, Y., Liu, J., Peng, H., Chen, S. (2025). Compressive sensing networks based on attention mechanism reconfiguration. Digital Signal Processing, 166, 105413.

Oyerinde, O. O., Flizikowski, A., Marciniak, T. (2024). Iterative hybrid compressive sensing-based channel estimation method for intelligent reflecting surface-supported millimeter wave systems. AEU - Int. J. Electron. Commun., 184, 155415.

Tawfik, A. A., Abuelenin, S. M., Abdelkader, M. F. (2024). An adaptive approach for integrating primary user behavior in compressive spectrum sensing. J. Wireless Commun. Netw., 2024, 74.

Morghare, G., Kushwah, A. S., Bhadauria, S. S. (2024). Wideband spectrum sensing technique based on subspace compression for MIMO cognitive radio networks. Sādhanā, 49, 293.

Gao, Z., Mei, Y., Qiao, L. (2024). Compressive sensing sparse channel estimation in broadband millimeter-wave massive MIMO systems. In Sparse Signal Processing for Massive MIMO Communications, Springer, Singapore, ch. 5.

Deka, S., Deka, K., Nguyen, N. T., Sharma, S., Bhatia, V., & Rajatheva, N. (2025). *Comprehensive review of deep unfolding techniques for next-generation wireless communication systems* (Version 3). arXiv:2502.05952 [eess.SP].

Zhang, J., & Ghanem, B. (2018). *ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing* (Version 2). arXiv:1706.07929 [cs.CV].

Shah, S. B., Pradhan, P., Pu, W., Randhi, R., Rodrigues, M. R. D., & Eldar, Y. C. (2023). *Optimization guarantees of unfolded ISTA and ADMM networks with smooth soft-thresholding* (Version 1). arXiv:2309.06195 [cs.LG].

Machidon, A. L., & Pejović, V. (2023). Deep learning for compressive sensing: A ubiquitous systems perspective. *Artificial Intelligence Review, 56*, 3619–3658.

Alhejaili, R., Alfarraj, M., & Al-Shaikhi, A. (2025). Listen to your gradients: Integrating gradients into deep unfolding networks. *Pattern Recognition Letters, 192*, 99–105.

Gao, J., Chen, X., & Li, G. Y. (2024). Deep unfolding based channel estimation for wideband terahertz near-field massive MIMO systems. *Frontiers of Information Technology & Electronic Engineering, 25*, 1162–1172.

Fesl, B., Baur, M., Strasser, F., Joham, M., & Utschick, W. (2024). *Diffusion-based generative prior for low-complexity MIMO channel estimation* (Version 1). arXiv:2403.03545 [eess.SP].

Zhou, X., Liang, L., Zhang, J., Jiang, P., Li, Y., & Jin, S. (2025). *Generative diffusion models for high dimensional channel estimation* (Version 2). arXiv:2408.10501 [cs.IT].

Mateos-Ramos, J. M., Häger, C., Keskin, M. F., Le Magoarou, L., & Wymeersch, H. (2024). *Model-based end-to-end learning for multi-target integrated sensing and communication under hardware impairments* (Version 2). arXiv:2307.04111 [eess.SP].

Li, H., Xu, J., Song, H., Wang, Y. (2024). Deep unfolding for sparse aperture ISAR imaging via compressed sensing: CIST+. IET Conf. Proc., 47.

Li, Z., Wu, K., Guo, Y., Zhang, C. (2024). Learned ISTA with error-based thresholding for adaptive sparse coding. In Proc. 2024 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP).

Saadi, A., Abghour, N., Chiba, Z., & others. (2025). A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control. *Journal of Big Data, 12*, 84.

Huang, C., Chen, X., Chen, G., Xiao, P., Li, G. Y., & Huang, W. (2025). *Deep reinforcement learning-based resource allocation for hybrid bit and generative semantic communications in space-air-ground integrated networks* (Version 3). arXiv:2412.05647 [cs.IT].

Zargari, S., Galappaththige, D., & Tellambura, C. (2024). *Refined-deep reinforcement learning for MIMO bistatic backscatter resource allocation* (Version 2). arXiv:2405.14046 [cs.IT].

Wang, M., Liu, X., Wang, F., & others. (2024). Spectrum-efficient user grouping and resource allocation based on deep reinforcement learning for mmWave massive MIMO-NOMA systems. *Scientific Reports, 14*, 8884.

Wang, Y., Li, X., Yi, X., Matthaiou, M., & Jin, S. (2025). Joint AP scheduling and precoding in RIS-aided distributed MIMO networks: A hierarchical DRL framework. *IEEE Transactions on Vehicular Technology, 74*(3), 5264–5269.

Tran, D.-H., Huynh, N. V., Kaada, S., Chatzinotas, S., et al. (2025). Network energy saving for 6G and beyond: A deep reinforcement learning approach. In 2025 IEEE Wireless Communications and Networking Conference (WCNC).

Zhang, K., Li, G., Lu, N., Yang, P., & Tang, K. (2025). *Hardware-aware DNN compression for homogeneous edge devices* (Version 1). arXiv:2501.15240 [cs.LG].

Wang, X., & Zhu, W. (2024). Advances in neural architecture search. *National Science Review, 11*, nwae282.

Shi, H., Huang, Y., Jin, S., & Wang, Z. (2024). Automatic high-performance neural network construction for channel estimation in IRS-aided communications. *IEEE Transactions on Wireless Communications, PP*(99), 1–1.

Ji, M., Chang, Y., Zhang, B., & Al-Ars, Z. (2024). *NASH: Neural architecture search for hardware-optimized machine learning models* (Version 2). arXiv:2403.01845 [cs.LG].

Sinha, N., Rostami, P., Shabayek, A. E. R., Kacem, A., & Aouada, D. (2024). *Multi-objective hardware aware neural architecture search using hardware cost diversity* (Version 1). arXiv:2404.12403 [cs.LG].

Garavagno, A. M., Ragusa, E., Frisoli, A., & Gastaldo, P. (2024). An affordable hardware-aware neural architecture search for deploying convolutional neural networks on ultra-low-power computing platforms. *IEEE Sensors Letters, 8*(5), 1–4.

Weitz, J., Demler, D., McDermott, L., Tran, N., & Duarte, J. (2025). *Neural architecture codesign for fast physics applications* (Version 1). arXiv:2501.05515 [cs.LG].

Fayyazi, A., Kamal, M., & Pedram, M. (2025). *MARCO: Hardware-aware neural architecture search for edge devices with multi-agent reinforcement learning and conformal prediction filtering* (Version 1). arXiv:2506.13755 [cs.LG].

Nguyen, X. T., Le, T. G., Bui, D. S., Jeong, S. G., Chien, T. V., Hwang, W. J., & Hanzo, L. (2024). *Graph neural networks for next-generation IoT: Recent advances and open challenges* (Version 2). arXiv preprint arXiv:2412.20634.

Kose, H. T., Nunez-Yanez, J., Piechocki, R., & Pope, J. (2024). A survey of computationally efficient graph neural networks for reconfigurable systems. *Information, 15*(7), 377.

Le, H. A., Chien, T. V., & Choi, W. (2025). Graph neural network based active and passive beamforming for distributed STAR-RIS-assisted multi-user MISO systems. *IEEE Transactions on Communications*.

Demirhan, U., & Alkhateeb, A. (2024). *Learning beamforming in cell-free massive MIMO ISAC systems*. arXiv:2409.18237 [cs.IT].

Zou, L., Pan, Z., & El-Hajjar, M. (2025). *Graph neural network aided beamforming for holographic millimeter wave MIMO systems*. *IEEE Transactions on Vehicular Technology, PP* (99), 1–14.

Lim, B., & Vu, M. (2025). *Graph neural network-based beamforming and RIS reflection design in a multi-RIS assisted wireless network*. *arXiv preprint* arXiv:2501.14987.

González-Duque, M., Michael, R., Bartels, S., Zainchkovskyy, Y., Hauberg, S., & Boomsma, W. (2024). *A survey and benchmark of high-dimensional Bayesian optimization of discrete sequences*. *arXiv preprint* arXiv:2406.04739.

Maus, N., Kim, K., Pleiss, G., Eriksson, D., Cunningham, J. P., & Gardner, J. R. (2025). *Approximation-aware Bayesian optimization*. *arXiv preprint* arXiv:2406.04308.

Onorato, G. (2024). *Bayesian optimization for hyperparameters tuning in neural networks* [Bachelor's thesis, Optimization for Machine Learning]. arXiv.

Hu, X., Tian, Y., Kho, Y. H., Xiao, B., Li, Q., Yang, Z., Li, Z., & Li, W. (2024). Location prediction using Bayesian optimization LSTM for RIS-assisted wireless communications. *IEEE Transactions on Vehicular Technology, 73*(10), 15156–15171.

Kato, T., Iimori, H., Pradhan, C., Malomsoky, S., & Ishikawa, N. (2025). Maximizing spectrum efficiency of data-carrying reference signals via Bayesian optimization. *arXiv preprint* arXiv:2502.14453.

Yang, C.-Y., Zhou, M., Ponzina, F., et al. (2025). Multi-objective software-hardware co-optimization for HD-PIM via noise-aware Bayesian optimization. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design (ICCAD '24)*. Association for Computing Machinery.

Padovano, D., Carpegna, A., Savino, A., & Di Carlo, S. (2024). SpikeExplorer: Hardware-oriented design space exploration for spiking neural networks on FPGA. *arXiv preprint* arXiv:2404.03714.

Deka, S., Deka, K., Nguyen, N. T., Sharma, S., Bhatia, V., & Rajatheva, N. (2025). *Comprehensive review of deep unfolding techniques for next-generation wireless communication systems*. arXiv:2502.05952 [eess.SP].

Tran, H. Q., Pham, V.-T., & Ngo, S. (2025). A survey on the applications of machine learning, deep learning, and reinforcement learning in wireless communications. *Communications Technology and Engineering*, 3(1), Article 3170.

Fatima, S., & Kondamuri, S. (2025). Machine learning in future wireless communications: Innovations, applications, and open challenges. *Wireless Personal Communications, 140*(2), 1269–1313.

Yin, L., & He, J. (2024). Modulated-symbol domain matched filtering scheme for photonic-assisted integrated sensing and communication system based on a single OFDM waveform. *Optics Letters, 49*(8), 2153–2156.

Li, L., Han, Y., Li, Z., Li, H., Lv, J., & Li, Y. (2024). Robust symbol and frequency synchronization method for burst OFDM systems in UAV communication. *Drones, 8*(9), 425.

Salmani Pour Avval, S., Eskue, N. D., Groves, R. M., & others. (2025). Systematic review on neural architecture search. *Artificial Intelligence Review, 58*, 73.

Karakoca, E., Çevik, H., Hökelek, İ., & Görçin, A. (2025). Novel deep neural OFDM receiver architectures for LLR estimation. *arXiv:2503.20500 [eess.SP]*.

Zhang, X., Lu, G., Wang, J., Gui, G., *et al.* (2024). Efficient modulation recognition with minimal samples leveraging architecture search and knowledge transfer in combined radar-communication environments. *In 2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*.

Chen, P.-Y., Chen, H., Tsai, Y.-M., Cheng, C.-M., *et al.* (2021, June). Learning to compensate: A deep neural network framework for 5G power amplifier compensation. *In ICC 2021 – IEEE International Conference on Communications.*

Xu, L., Gao, F., Zhang, W., & Ma, S. (2021, May). Model-aided deep learning-based MIMO OFDM receiver with nonlinear power amplifiers. *In 2021 IEEE Wireless Communications and Networking Conference (WCNC)*.

Gao, T., Guo, L., Zhao, S., Xu, P., Yang, Y., Liu, X., Wang, S., Zhu, S., & Zhou, D. (2024). QuantNAS: Quantization-aware neural architecture search for efficient deployment on mobile device. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Wang, B., Yang, H., & Fang, T. (2024). Modulation recognition of underwater acoustic communication signals based on deep learning. *EURASIP Journal on Advances in Signal Processing, 2024*, 103.

Dai, S., & Liu, Y. (2025). An automated modulation classification approach utilizing progressive neural architecture search. *In 2025 IEEE 7th International Conference on Communications, Information System and Computer Engineering (CISCE)*.

Li, B., Zheng, Q., Tian, X., et al. (2025). A survey of artificial intelligence enabled channel estimation methods: Recent advance, performance, and outlook. *Artificial Intelligence Review, 58*, 187.

Arellano, J. F., Altamirano, C. D., Carvajal Mora, H. R., Orozco Garzón, N. V., & Almeida García, F. D. (2024). On the performance of MMSE channel estimation in massive MIMO systems over spatially correlated Rician fading channels. *Wireless Communications and Mobile Computing*.

Yao, X., An, J., Huang, G., Liu, H., Gan, L., & Yuen, C. (2024). Sparse channel estimation for stacked intelligent metasurface-assisted mmWave communications. *2024 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS)*, 1–5. IEEE.

Zhang, J., Li, J., Wang, Z., Han, Y., Shi, L., & Cao, B. (2024). Decision transformer for IRS-assisted systems with diffusion-driven generative channels. *arXiv preprint* arXiv:2406.19769.

Lu, Y., Rayo Torres Rodriguez, H., Vogel, S., van de Waterlaat, N., & Jancura, P. (2024). Scaling up quantization-aware neural architecture search for efficient deep learning on the edge. *arXiv preprint* arXiv:2401.12350.

Trabelsi, N., Chaari Fourati, L., & Chen, C. S. (2024). Interference management in 5G and beyond networks. *arXiv preprint* arXiv:2401.01608.

Benesty, J., Paleologu, C., Stanciu, C.-L., Costea, R.-L., Dogariu, L.-M., & Ciochină, S. (2024). Wiener filter using the conjugate gradient method and a third-order tensor decomposition. *Applied Sciences, 14*(6), 2430.

Tulyakova, N., & Trofymchuk, O. (2024). Adaptive nonlinear filtering algorithms for removal of non-stationary noise in electronystagmographic signals. *Computers in Biology and Medicine, 183*, 109229.

Taassori, M. (2024). Enhanced Wavelet-Based Medical Image Denoising with Bayesian-Optimized Bilateral Filtering. Sensors, 24(21), 6849. https://doi.org/10.3390/s24216849

König, C., Krishnadas, R., Balta, E. C., & Rupenyan, A. (2025). Adaptive Bayesian optimization for high-precision motion systems. *IEEE Transactions on Automation Science and Engineering, 22*, 15627–15637.

Maggi, L., Koblitz, R., Zhu, Q., & Andrews, M. (2023). Tracking the best beam for a mobile user via Bayesian optimization. *arXiv*.

Sato, K., & Suto, K. (2024). Bayesian optimization framework for channel simulation-based base station placement and transmission power design. *IEEE Networking Letters*, 6(10), 640–644.

Salaun, L., Yang, H., Mishra, S., & Chen, C. S. (2022). A GNN approach for cell-free massive MIMO. *2022 IEEE Global Communications Conference (GLOBECOM)*, 3053–3058. IEEE.

Chan, K. L., Chien, F. T., & Chang, R. Y. (2024). Beamforming and load-balanced user association in RIS-aided systems via graph neural networks. *ICC 2024 – IEEE International Conference on Communications*, 1–6. IEEE.

Pi, Y., Zhang, W., Zhang, Y., Huang, H., Rao, B., Ding, Y., & Yang, S. (2024). Applications of multi-agent deep reinforcement learning communication in network management: A survey. *arXiv preprint* arXiv:2407.17030.

Nguyen, D. N. T. (2025). Deep reinforcement algorithms in RIS-empowered wireless communication systems. In *Reconfigurable intelligent surfaces for 6G and beyond wireless networks* (Chap. 14).

Shahjehan, W., Rathore, R. S., Shah, S. W., Aljaidi, M., Sadiq, A. S., & Kaiwartya, O. (2024). A review on millimeter-wave hybrid beamforming for wireless intelligent transport systems. *Future Internet, 16*(9), 337.

Ngo, H. Q., Interdonato, G., Larsson, E. G., Caire, G., & Andrews, J. G. (2024). Ultra-dense cell-free massive MIMO for 6G: Technical overview and open questions. *Proceedings of the IEEE*. (In press).

Zhang, B., Zeng, H., & Prasanna, V. (2023). GraphAGILE: An FPGA-based overlay accelerator for low-latency GNN inference. *arXiv preprint* arXiv:2302.01769.

Que, Z., Fan, H., Loo, M., Li, H., Blott, M., Pierini, M., Tapper, A., & Luk, W. (2024). LL-GNN: Low latency graph neural networks on FPGAs for high energy physics. *ACM Transactions on Embedded Computing Systems, 23*(2), 17.

Zhang, X., Xia, Y., Kang, Z., Du, X., & Gao, J. (2024). A review of FPGA-based graph neural network accelerator architectures. *In Proceedings of the 2024 7th International Conference on Signal Processing and Machine Learning (SPML '24)* (pp. 335–343).

Procaccini, M., Sahebi, A., & Giorgi, R. (2024). A survey of graph convolutional networks (GCNs) in FPGA-based accelerators. *Journal of Big Data, 11*, 163.

Zhang, R., Lu, Y., Chen, W., Ai, B., & Ding, Z. (2025). Model-based GNN enabled energy-efficient beamforming for ultra-dense wireless networks. *IEEE Transactions on Wireless Communications, 24*(4), 3333–3345.

Ismail A. A., Khalifa N. E., El-Khoribi R. A. (2025). A survey on resource scheduling approaches in multi-access edge computing environment: a deep reinforcement learning study. Cluster Comput., 28, 184.

Zhao J., Liu C., Liao J., Wang D. (2024). Deep learning in wireless communications for physical layer. Physical Communication, 67, 102503.

Sohaib, R. M., Shah, S. T., Onireti, O., Sambo, Y., Abbasi, Q. H., & Imran, M. A. (2024). DRL-based joint resource scheduling of eMBB and URLLC in O-RAN. *arXiv preprint* arXiv:2407.11558.

Malhotra, S., Yashu, F., Saqib, M., Mehta, D., Jangid, J., & Dixit, S. (2025). Deep reinforcement learning for dynamic resource allocation in wireless networks. *arXiv preprint* arXiv:2502.01129.

Li, X., Zhang, X., Li, J., et al. (2024). Blocklength allocation and power control in UAV-assisted URLLC system via multi-agent deep reinforcement learning. *International Journal of Computational Intelligence Systems*, 17, 138.

Jiang, Z., Zhang, J., Wang, T., & Wang, H. (2024). Modulation recognition of underwater acoustic communication signals based on neural architecture search. *Applied Acoustics, 225,* 110155.

Orucu, A., Moradi, F., Ebrahimi, M., & Johnsson, A. (2024). Towards neural architecture search for transfer learning in 6G networks. *arXiv preprint* arXiv:2406.02333.

Li X., Guo J., Wen C.-K., Jin S. (2024). Auto-CsiNet: Scenario-customized automatic neural network architecture generation for massive MIMO CSI feedback. IEEE Trans. Wireless Commun., 23(10), 14759–14775.

White C., Safari M., Sukthanker R., Ru B., Elsken T., Zela A., Dey D., Hutter F. (2023). Neural architecture search: insights from 1000 papers. arXiv preprint arXiv:2301.08727.

Wu M.-T., Tsai C.-W. (2024). Training-free neural architecture search: a review. ICT Express, 10(1), 213–231.

Gu, Y., She, C., Bi, S., Quan, Z., & Vucetic, B. (2024). Graph neural network for distributed beamforming and power control in massive URLLC networks. *IEEE Transactions on Wireless Communications, 23*(8), 9099–9112.

Wang, Z., & Wong, V. W. S. (2024). Heterogeneous graph neural network for cooperative ISAC beamforming in cell-free MIMO systems. In *Proceedings of the 3rd ACM MobiCom Workshop on Integrated Sensing and Communications Systems (ISACom)*, Washington, DC, USA.

Guo, C., Wu, H., & Luk, W. (2024). Resource-constraint Bayesian optimization for soft processors on FPGAs. *In Proceedings of the 14th International Symposium on Highly Efficient Accelerators and Reconfigurable Technologies (HEART 2024)*. Association for Computing Machinery.

Wang, X., Tang, Z., Guo, J., Meng, T., Wang, C., Wang, T., & Jia, W. (2025). Empowering edge intelligence: A comprehensive survey on on-device AI models. *ACM Computing Surveys, 57*(9), Article 228, 1–39.

Li, B., Zheng, Q., Tian, X., et al. (2025). A survey of artificial intelligence enabled channel estimation methods: Recent advance, performance, and outlook. *Artificial Intelligence Review, 58*, 187.

Wu, T., Chen, Z., He, D., Qian, L., Xu, Y., Tao, M., & Zhang, W. (2024). CDDM: Channel denoising diffusion models for wireless semantic communications. *IEEE Transactions on Wireless Communications, 23*(9_Part_1), 11168–11183.

Zuo, X., Yang, Y., Yao, R., Fan, Y., & Li, L. (2024). An automatic modulation recognition algorithm based on time–frequency features and deep learning with fading channels. *Remote Sensing, 16*(23), 4550.

Neshaastegaran, P., & Jian, M. (2025). CoDiPhy: A general framework for applying denoising diffusion models to the physical layer of wireless communication systems. *arXiv preprint* arXiv:2503.10297.

Ai, X., Deng, A., Li, Z., & Wu, S. (2025). Channel estimation for high-speed mobile MIMO-OFDM systems using deep learning. In *Proceedings of the 2025 3rd International Conference on Communication Networks and Machine Learning (CNML '25)* (pp. 291–297). ACM.

Osadchuk, Y., Jovanovic, O., Ranzini, S. M., Dischler, R., Aref, V., Zibar, D., & Da Ros, F. (2024). Low-complexity samples versus symbols-based neural network receiver for channel equalization. *Journal of Lightwave Technology, 42*(15), 5167–5174.

Xu, Z., Ji, T., Wu, Q., Lu, W., Ji, H., Yang, Y., Qiao, G., Tang, J., Cheng, C., Liu, L., Wang, S., Liang, J., Sun, Z., Fan, L., Wei, J., & Shieh, W. (2024). Advanced neural network-based equalization in intensity-modulated direct-detection optical systems: Current status and future trends. *Photonics, 11*(8), 702.

Faruk Pasic, L., Eller, S., Schwarz, M., Rupp, M., & Mecklenbräuker, C. F. (2025). Deep learning-based mmWave MIMO channel estimation using sub-6 GHz channel information: CNN and UNet approaches. *2025 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*.

Alavirad, M., Raghothaman, B., Garcia-Ulloa, D., Beattie, J. G. Jr., & Harris, P. (2024). Principles and methodologies for AI/ML testing in next generation networks (O-RAN next Generation Research Group contributed research report, RR-2024-04). VIAVI Solutions, Keysight Technologies, Dell Technologies.

Reuters. (2024). Nvidia launches cloud-based research platform for 6G research and testing.

Safavinejad, R., Chang, H.-H., & Liu, L. (2024). Deep reinforcement learning for dynamic spectrum access: Convergence analysis and system design. IEEE Transactions on Wireless Communications, 23(12, Part 2), 18888–18902.

Sever, O., Salan, O., Hokelek, I., & Gorcin, A. (2025). A practical demonstration of DRL-based dynamic resource allocation xApp using OpenAirInterface. arXiv preprint arXiv:2501.05879.

Kim, H., & So, J. (2025). Distributed multi-agent deep reinforcement learning-based transmit power control in cellular networks. *Sensors, 25*(13), 4017.

Cui Q., You X., Wei N., et al. (2025). Overview of AI and communication for 6G network: fundamentals, challenges, and future research opportunities. Sci. China Inf. Sci., 68, 171301.

Eldeeb, E., & Alves, H. (2025). Offline and distributional reinforcement learning for wireless communications. *arXiv preprint* arXiv:2504.03804.

Han, L., Zeng, T., Zhao, J., Bao, X., Liu, G., & Liu, Y. (2025). Queue stability-constrained deep reinforcement learning algorithms for adaptive transmission control in multi-access edge computing systems. *Algorithms, 18*(8), 498.

Wu, J., & Fang, X. (2024). Collaborative optimization of wireless communication and computing resource allocation based on multi-agent federated weighting deep reinforcement learning. *arXiv preprint* arXiv:2404.01638.

Ullah, I., Arishi, A., Singh, S. K., Alharbi, F., Ibrahim, A. H., Islam, M., Daradkeh, Y. I., & Choi, C. (2025). Autonomous network management for 6G communication: A comprehensive survey. *Digital Communications and Networks*. Advance online publication.

Zhu, F., Wang, X., Zhu, C., Gong, T., Yang, Z., Huang, C., Chen, X., Zhang, Z., & Debbah, M. (2025). Robust deep learning-based physical layer communications: Strategies and approaches. *IEEE Network*.

Siddiqui, S., Kyrkou, C., & Theocharides, T. (2025). Efficient global neural architecture search. *arXiv preprint* arXiv:2502.03553.

King, T., Zhou, Y., Röddiger, T., et al. (2025). MicroNAS for memory and latency constrained hardware aware neural architecture search in time series classification on microcontrollers. *Scientific Reports, 15*, 7575.

Aach, M., Blanc, C., Lintermann, A., & De Grave, K. (2025). Optimizing edge AI models on HPC systems with the edge in the loop. *arXiv preprint* arXiv:2505.19995.

Zhou, X., Wu, X., Feng, L., Lu, Z., & Tan, K. C. (2024). Design principle transfer in neural architecture search via large language models. *arXiv preprint* arXiv:2408.11330.

He, C., Lu, Y., Ai, B., Dobre, O. A., Ding, Z., & Niyato, D. (2025). ICGNN: Graph neural network enabled scalable beamforming for MISO interference channels. *IEEE Transactions on Mobile Computing*.

Tam, P., Ros, S., Song, I., Kang, S., & Kim, S. (2024). A survey of intelligent end-to-end networking solutions: Integrating graph neural networks and deep reinforcement learning approaches. *Electronics, 13*(5), 994.

Li, Y., Lu, Y., Ai, B., Dobre, O. A., Ding, Z., & Niyato, D. (2024). GNN-based beamforming for sum-rate maximization in MU-MISO networks. *IEEE Transactions on Wireless Communications, 23*(8, Part 1), 9251–9264.

B. Li and M. Vu, "Graph Neural Network Based Hybrid Beamforming Design in Wideband Terahertz MIMO-OFDM Systems," *2024 IEEE International Symposium on Phased Array Systems and Technology (ARRAY)*, Boston, MA, USA, 2024, pp. 1-6, doi: 10.1109/ARRAY58370.2024.10880428.

Lim, B., & Vu, M. (2023). Graph neural network-based beamforming and RIS reflection design in a multi-RIS assisted wireless network. In *2023 IEEE Statistical Signal Processing Workshop (SSP)* (pp. 120–124). IEEE.

Nazar, A. M., Selim, M. Y., & Qiao, D. (2025). Revolutionizing RIS networks: LiDAR-based data-driven approach to enhance RIS beamforming. *Sensors, 25*(1), 75.

Sáez de Ocáriz Borde, H., Lukoianov, A., Kratsios, A., Bronstein, M. M., & Dong, X. (2024). Scalable message passing neural networks: No need for attention in large graph representation learning. *ICLR submission*.

In, Y., Yoon, K., Yun, S., Kim, K., Kim, S., & Park, C. (2025). Training robust graph neural networks by modeling noise dependencies. *arXiv preprint* arXiv:2502.19670.

Yan J., Lu Q., Giannakis G. B. (2024). Bayesian optimization for online management in dynamic mobile edge computing. IEEE Trans. Wireless Commun., 23(4), 3425–3436.

Garrido-Merchán E. C. (2025). Information-theoretic Bayesian optimization: Survey and tutorial. arXiv preprint arXiv:2502.06789.

Franceschi, L., Donini, M., Perrone, V., Klein, A., Archambeau, C., Seeger, M., Pontil, M., & Frasconi, P. (2025). Hyperparameter optimization in machine learning. *arXiv*.

Elmorsy, S. M. A., Osman, S. M., & Gamel, S. A. (2025). Enhanced spectrum sensing for 5G and LTE signals using advanced deep learning models and hyperparameter tuning. *Scientific Reports, 15*, 24825.

Suwandi, R. C. (2024, September 9). Optimize your signal processing with Bayesian optimization. *IEEE Signal Processing Society* [Blog].

Kyriakidis, L., Bähr, M., & Mendez, M. A. (2025). Enhanced hybrid algorithm based on Bayesian optimization and Interior Point OPTimizer for constrained optimization. *Optimization and Engineering*.

Li, W., Wang, T., & Yoneki, E. (2024). HiBO: Hierarchical Bayesian optimization via adaptive search space partitioning. *ICLR 2025 submission*. OpenReview.

Cao, X., et al. (2024). AI-empowered multiple access for 6G: A survey of spectrum sensing, protocol designs, and optimizations. *Proceedings of the IEEE, 112*(9), 1264–1302.

Kang, B. J., Lee, H. I., Yoon, S. K., Kim, Y. C., Jeong, S. B., O, S. J., & Kim, H. (2024). A survey of FPGA and ASIC designs for transformer inference acceleration and optimization. *Systems Architecture, 155*(C), Article 103247.

Lin, Y.-C., Xu, Z., & Prasanna, V. (2025). xBS-GNN: Accelerating billion-scale GNN training on FPGA. In *Proceedings of the SC '24 Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis* (pp. 659–666).

Jha, A., Dewangan, T., Lokhande, M., & Vishvakarma, S. K. (2025). QForce-RL: Quantized FPGA-optimized reinforcement learning compute engine. *arXiv*.

Wang, W., Chen, W., Luo, Y., Long, Y., Lin, Z., Zhang, L., Lin, B., Cai, D., & He, X. (2024). Model compression and efficient inference for large language models: A survey. *arXiv*.

Andronic, M., & Constantinides, G. A. (2025). NeuraLUT-Assemble: Hardware-aware assembling of sub-neural networks for efficient LUT inference. In *2025 IEEE 33rd Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)* (pp. 208–216). Fayetteville, AR, USA.

Bartzoudis, N., Rubio Fernández, J., López-Bueno, D., Román Villarroel, A., & Antonopoulos, A. (2024). Agile FPGA computing at the 5G edge: Joint management of accelerated and software functions for open radio access technologies. *Electronics, 13*(4), 701.

Boudjadar, J., Ul Islam, S., & Buyya, R. (2025). Dynamic FPGA reconfiguration for scalable embedded artificial intelligence (AI): A co-design methodology for convolutional neural networks (CNN) acceleration. *Future Generation Computer Systems, 169*, Article 107777.

Ramhorst, B., Korolija, D., Heer, M. J., Dann, J., Liu, L., & Alonso, G. (2025). Coyote v2: Raising the level of abstraction for data center FPGAs. *arXiv*.

Gill, S. S., Golec, M., Hu, J., et al. (2025). Edge AI: A taxonomy, systematic review and future directions. *Cluster Computing, 28*, 18.

Heydari, S., & Mahmoud, Q. H. (2025). Tiny machine learning and on-device inference: A survey of applications, challenges, and future directions. *Sensors, 25*(10), 3191.

Li, H., Ge, L., & Tian, L. (2024). Survey: Federated learning data security and privacy-preserving in edge-Internet of Things. *Artificial Intelligence Review, 57*, 130.

Ji, M., et al. (2025). Graph neural networks and deep reinforcement learning-based resource allocation for V2X communications. *IEEE Internet of Things Journal, 12*(4), 3613–3628.

Gu, Y., Liu, Z., Dai, S., Liu, C., Wang, Y., Wang, S., Theodoropoulos, G., & Cheng, L. (2025). Deep reinforcement learning for job scheduling and resource management in cloud computing: An algorithm-level review. *arXiv*.

Trigka, M., & Dritsas, E. (2025). Edge and cloud computing in smart cities. *Future Internet, 17*(3), 118.

Ficili, I., Giacobbe, M., Tricomi, G., & Puliafito, A. (2025). From sensors to data intelligence: Leveraging IoT, cloud, and edge computing with AI. *Sensors, 25*(6), 1763.

Sabir, B., Yang, S., Nguyen, D., Wu, N., Abuadbba, A., Suzuki, H., Lai, S., Ni, W., Ming, D., & Nepal, S. (2024). Systematic literature review of AI-enabled spectrum management in 6G and future networks. *arXiv*.

Ismail, F., Khawaja, S. G., Khan, A. M., et al. (2025). Cognitive link adaptation via modulation scheme classification in narrowband networks under AWGN and SUI channel conditions. Scientific Reports, 15, 27604.

Manikandan, A., Babu, R. T. S., Ganesh, S. J., et al. (2025). Generative adversarial networks based adaptive modulation and coding for next-generation 5G communication systems. Discover Applied Sciences, 7, 91.

Pan, G., Yau, D. K. Y., Zhou, B., & Wu, Q. (2025). Deep learning for spectrum prediction in cognitive radio networks: State-of-the-art, new opportunities, and challenges. *IEEE Network*.

Ferguson, A. E., Pawar, U., Wang, T., & Marina, M. K. (2025). Campus5G: A campus scale private 5G Open RAN testbed. *arXiv*.

Martian, A., Trifan, R.-F., Stoian, T.-C., Vochin, M.-C., & Li, F. Y. (2025). Towards Open RAN in beyond 5G networks: Evolution, architectures, deployments, spectrum, prototypes, and performance assessment. Computer Networks, 259, Article 111087.

Lacava, A., Bonati, L., Mohamadi, N., Gangula, R., Kaltenberger, F., Johari, P., D'Oro, S., Cuomo, F., Polese, M., & Melodia, T. (2025). dApps: Enabling real-time AI-based Open RAN control. *Computer Networks, 269*, Article 111342.

Pelekis, S., Koutroubas, T., Blika, A., et al. (2025). Adversarial machine learning: A review of methods, tools, and critical industry sectors. *Artificial Intelligence Review, 58*, 226.

Tan, W. L. (2024). Adversarial machine learning attacks in wireless networks (Technical report). Griffith University.