HELENA: High-Efficiency Learning-based channel Estimation using dual Neural Attention

Miguel Camelo Botero, Esra Aycan Beyazıt, Nina Slamnik-Kriještorac, Johann M. Marquez-Barja
University of Antwerp - imec, IDLab, Antwerp, Belgium

***Abstract*—Accurate channel estimation is critical for high-performance Orthogonal Frequency-Division Multiplexing (OFDM) systems, particularly at low signal-to-noise ratios and under stringent latency constraints. This paper presents High-Efficiency Learning-based channel Estimation using dual Neural Attention (HELENA), a compact deep learning model that combines a lightweight convolutional backbone with two efficient attention mechanisms: patch-wise multi-head self-attention for capturing global dependencies and a squeeze-and-excitation block for local feature refinement. Compared to CEViT, a state-of-the-art vision transformer-based estimator, HELENA reduces inference time by 45.0% (0.175ms vs. 0.318ms), achieves comparable accuracy (-16.78dB vs. -17.30dB), and requires $8\times$ fewer parameters (0.11M vs.0.88M), demonstrating its suitability for low-latency, real-time OFDM-based wireless systems.**

***Index Terms*—Channel Estimation, 5G-NR, OFDM, Deep Learning, Neural Attention, Neural Network Acceleration.**

I. Introduction

Accurate estimation of Channel State Information (CSI) is crucial for the effectiveness of Orthogonal Frequency-Division Multiplexing (OFDM)-based wireless communication systems, such as 5G New Radio (5G-NR), as it enables optimal resource allocation, beamforming, and adaptive modulation, key factors that directly influence system capacity and reliability. Channel Estimation (CE), which refers to estimating CSI from received and reference signals (e.g., pilots), faces practical limitations: Least Squares (LS) performs poorly in noisy or high-mobility scenarios, while Minimum Mean Square Error (MMSE) depends on unavailable channel statistics and incurs high computational complexity, restricting its deployment in real-time wireless systems.

To address these challenges, recent research has investigated both model-based and data-driven approaches for improving CE under realistic 5G and beyond conditions. Among model-based techniques, several recent methods have demonstrated competitive performance by leveraging signal structure and auxiliary information, such as pilot-free estimation for high-Doppler Orthogonal Time Frequency Space (OTFS) scenarios [1] and sensing-assisted denoising of LS estimates [2]. In parallel, Deep Learning (DL)-based methods have emerged as powerful tools for enhancing CE by learning complex propagation patterns directly from data. These approaches lay the foundation for Artificial Intelligence (AI)-native physical layer design [3] and the development of the next generation of intelligent radios [4], which is the focus of this paper.

Architectures such as ChannelNet [5], Enhanced Deep Super-Resolution (EDSR) [6], Attention mechanism and Residual Network (AttRNet) [7], Progressive Estimation Network (ProEsNet) [8], and the more recent Transformer-based Efficient Parallel Transformer (EPformer) [9] and Channel Estimator Vision Transformer (CEViT) [10] achieve superior estimation accuracy over conventional methods. However, their high computational complexity limits their deployment in real-time 5G systems, which are constrained by strict latency, power, and hardware budgets [11]. This has led to the development of lightweight architectures such as LS-augmented interpolated Deep Neural Network (LSiDNN) [12], which aim to balance accuracy with computational efficiency. However, these models tend to achieve lower complexity at the expense of estimation accuracy or real-time performance.

To address these limitations, we introduce **High-Efficiency Learning-based channel Estimation using dual Neural Attention (HELENA)**, a lightweight hybrid DL architecture for pilot-based CE in OFDM systems without interpolation. It combines convolutional feature extraction with dual attention: 2D Convolutional (Conv2D) layers for local features, patch-wise Multi-Head Self-Attention (MHSA) for global context, and a post-attention Squeeze-and-Excitation (SE) block for channel refinement. A linear reconstruction head and residual connection preserve structural information and support convergence, especially under high-Signal-to-Noise Ratio (SNR) conditions. In this paper, we show that HELENA (i) achieves better accuracy–efficiency trade-offs than several state-of-the-art DL models, (ii) reveals that computational cost does not always correlate with inference time, and (iii) is fully reproducible via open-source code and data[1].

II. System Model and Problem Statement

We consider a downlink Single Input Single Output (SISO) OFDM system compliant with 5G-NR. The main notation used throughout the paper is summarized in Table I for clarity.

The received signal at subcarrier i and OFDM symbol k is given by

$$Y_{i,k} = H_{i,k}X_{i,k} + Z_{i,k}, \tag{1}$$

where $X_{i,k}$ is the transmitted symbol, $H_{i,k}$ the channel coefficient, and $Z_{i,k} \sim \mathcal{CN}(0, \sigma^2)$ represents additive white Gaussian noise. The full channel matrix $\mathbf{H} \in \mathbb{C}^{N_S \times N_D}$ over subcarriers and OFDM symbols must be estimated from a

[1]https://github.com/miguelhdo/HELENA_Channel_Estimation

TABLE I
SUMMARY OF NOTATION

Symbol	Description
$\mathbf{H}$	True channel matrix over subcarriers and OFDM symbols
$\hat{\mathbf{H}}$	Estimated (reconstructed) channel matrix
$\mathbf{H}_{\text{LR}}$	Low-resolution channel input (pilot-based or interpolated)
$\mathbf{H}_p$	Pilot-based channel observations
$\mathcal{P}$	Set of pilot positions in the time-frequency grid
f_Θ	Deep learning-based channel estimation model
Θ	Learnable model parameters
$\mathbf{X}$	Transmitted OFDM signal matrix
$\mathbf{Y}$	Received OFDM signal matrix
$\mathbf{Z}$	Additive white Gaussian noise matrix
$X_{i,k}$	Transmitted symbol at subcarrier i and OFDM symbol k
$Y_{i,k}$	Received symbol at subcarrier i and OFDM symbol k
$H_{i,k}$	Channel coefficient at subcarrier i and OFDM symbol k
N_S	Number of subcarriers
N_D	Number of OFDM symbols
p	Patch height along the frequency dimension
N	Number of patches (tokens), $N = N_S/p$
σ^2	Noise variance
LS	Least Squares channel estimation
MMSE	Minimum Mean-Square Error estimation
LI	Linear interpolation
SR	Super-resolution-based channel reconstruction
$\|\cdot\|_F$	Frobenius norm
$T_{\text{inf}}(f_\Theta)$	Inference time of model f_Θ
$T_{\max}$	Maximum allowable inference latency
MSE	Mean Squared Error
NMSE	Normalized Mean Squared Error
$\mathbb{E}[\cdot]$	Expectation operator

sparse set of pilot observations $\mathcal{P}$. Classical methods such as LS and MMSE [13] rely on pilot-based interpolation [14] to reconstruct the full channel, but each has drawbacks. While LS is computationally efficient, it is sensitive to noise. MMSE provides higher accuracy by using prior channel statistics, but is computationally intensive and difficult to implement in dynamic environments. To address this, recent approaches reformulate CE as a Super-resolution (SR) problem, aiming to recover the full-resolution channel $\hat{\mathbf{H}}$ from sparse pilot observations $\mathbf{H}_p$, modeled as:

$$\hat{\mathbf{H}} = f_\Theta(\mathbf{H}_{LR}), \tag{2}$$

where f_Θ denotes a DL model parameterized by Θ that learns a mapping from a low-resolution input $\mathbf{H}_{LR}$ to the full channel estimate $\hat{\mathbf{H}}$. The input $\mathbf{H}_{LR}$ can be obtained either directly from raw pilot observations ($(\mathbf{H}_{LR} = \mathbf{H}_p)$) or from LS-based estimates combined with Linear Interpolation (LI), providing a partial yet denoised representation of the true channel $\mathbf{H}$. This formulation is analogous to SR in image processing, where high-resolution details are reconstructed from degraded or sparse inputs. The objective is to design $f\Theta$ such that it accurately reconstructs $\hat{\mathbf{H}}$ while meeting the low-latency requirements of real-time 5G systems.

The problem can be formalized as the following constrained optimization:

$$\min_{\Theta} \ \mathbb{E}\left[\|f_\Theta(\mathbf{H}_{LR}) - \mathbf{H}\|_F^2\right] \quad \text{s.t.} \quad T_{\text{inf}}(f_\Theta) \le T_{\max}, \tag{3}$$

where $\|\cdot\|_F$ is the Frobenius norm, and the objective corresponds to the Mean Squared Error (MSE) between the predicted and true channel matrices. $T_{\text{inf}}(f_\Theta)$ denotes the model's inference time, and $T_{\max}$ is the allowable latency budget. This formulation captures the need to jointly optimize estimation accuracy and computational efficiency for real-time 5G systems. While MSE serves as the training objective, we report performance using the Normalized Mean Squared Error (NMSE), defined as

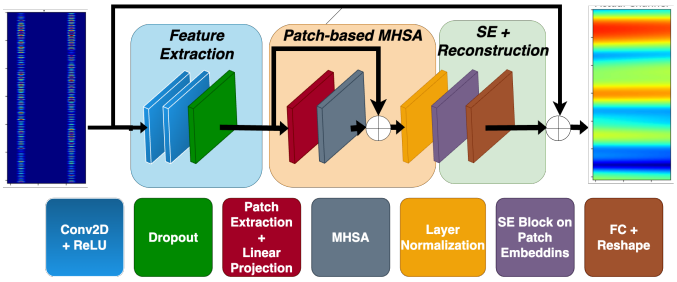

Fig. 1. The proposed HELENA architecture.

$$\text{NMSE} = \frac{\mathbb{E}\left[\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2\right]}{\mathbb{E}\left[\|\mathbf{H}\|_F^2\right]}, \tag{4}$$

which offers a scale-invariant measure of estimation quality across varying signal and channel conditions. All results are reported in Decibels (dB).

Prior work such as ChannelNet [5], EDSR [6], and more recently ProEsNet [8], apply deep Convolutional Neural Networks (CNNs) inspired by image SR to enhance CE accuracy, albeit with increased latency. Super Resolution Convolutional Neural Network (SRCNN), uses as first stage of ChannelNet, reduces complexity but suffers in performance. AttRNet [7] improves efficiency using SE blocks, and LSiDNN [12] explicitly targets low complexity with dense layers. EPformer [9] and CEViT [10] leverages Transformer-based architecture for top accuracy, but incurs high computational cost. These limitations motivate HELENA, which combines shallow convolution, dual attention, and compact design to jointly optimize accuracy, complexity, and latency.

III. HELENA ARCHITECTURE

HELENA is a lightweight DL model for pilot-based CE in 5G-NR OFDM systems, designed to balance accuracy and inference time. It combines convolutional feature extraction with dual attention: patch-wise MHSA for long-range context and a low-cost SE block for channel-wise recalibration. The input is a sparse LS-based estimate $\mathbf{H}_{\text{LR}} \in \mathbb{R}^{N_S \times N_D \times 2}$, where values are present only at pilot positions and set to zero elsewhere, while the output is a full-resolution estimate $\hat{\mathbf{H}} \in \mathbb{R}^{N_S \times N_D \times 2}$ covering the entire time-frequency grid. The following subsection outlines the main architectural components and their design rationale.

A. Feature Extraction via Shallow CNN

In order to extract local spatial features from $\mathbf{H}_{\text{LR}}$, the model uses two 2D convolutional layers with ReLU activations. The first convolutional layer applies a kernel of size

$f_1 \times t_1$ and uses C_1 filters, producing an intermediate feature map $\mathbf{F}^{(1)}$. The second layer uses a kernel of size $f_2 \times t_2$ with C filters, resulting in $\mathbf{F}^{(2)}$. These layers have associated learnable weights $\mathbf{W}_1, \mathbf{W}_2$ and biases $\mathbf{b}_1, \mathbf{b}_2$, respectively. A dropout layer is applied to improve generalization:

$$\mathbf{F}^{(1)} = \text{ReLU}\left(\text{Conv2D}_1(\mathbf{H}_{\text{LR}}; \mathbf{W}_1, \mathbf{b}_1)\right) \tag{5}$$

$$\mathbf{F}^{(2)} = \text{ReLU}\left(\text{Conv2D}_2(\mathbf{F}^{(1)}; \mathbf{W}_2, \mathbf{b}_2)\right) \tag{6}$$

$$\mathbf{F}_{\text{drop}} = \text{Dropout}(\mathbf{F}^{(2)}) \tag{7}$$

The output tensor $\mathbf{F}_{\text{drop}} \in \mathbb{R}^{N_S \times N_D \times C}$ is then partitioned into non-overlapping patches of height p, giving $N = N_S/p$ patches. This patching strategy enables localized spatial aggregation while reducing the sequence length, making the subsequent attention computation more efficient. Each patch is then flattened into a vector:

$$\mathbf{F}^{(i)}_{\text{patch}} \in \mathbb{R}^{p \times N_D \times C}, \quad \mathbf{P}_i = \text{Flatten}(\mathbf{F}^{(i)}_{\text{patch}}) \tag{8}$$

B. Patch Embedding and Multi-Head Self-Attention

Each patch vector $\mathbf{P}_i$ is projected into a shared d-dimensional embedding space using a learnable weight matrix $\mathbf{W}_e \in \mathbb{R}^{(pN_DC) \times d}$ and bias $\mathbf{b}_e \in \mathbb{R}^d$:

$$\mathbf{Z}_i = \mathbf{P}_i \mathbf{W}_e + \mathbf{b}_e, \quad \mathbf{Z} \in \mathbb{R}^{N \times d} \tag{9}$$

In HELENA, each flattened and embedded patch vector $\mathbf{Z}_i$, also referred to as a *token*, following transformer terminology, represents a localized region of the time-frequency channel grid, corresponding to a group of neighboring subcarriers and OFDM symbols. These patches are designed to capture local variations in the wireless channel, such as those introduced by multipath delay and Doppler shift. The resulting N tokens are stacked into a matrix $\mathbf{Z} \in \mathbb{R}^{N \times d}$, where each row is a token embedding of dimension d. This sequence is fed into the MHSA block, which models global context and long-range dependencies in the time–frequency domain. Each patch embedding $Z \in \mathbb{R}^{N \times d}$ is linearly projected into query, key, and value vectors, enabling the computation of attention weights that capture contextual relevance between patches. These weights guide how information from other patches is aggregated to refine each token representation. For each head j, Z is projected to queries, keys, and values as:

$$Q_j = ZW_j^Q, \quad K_j = ZW_j^K, \quad V_j = ZW_j^V \tag{10}$$

$$\text{head}_j = \text{Softmax}\left(\frac{Q_j K_j^\top}{\sqrt{d_k}}\right) V_j \tag{11}$$

Outputs from all heads are concatenated and projected using $W^O \in \mathbb{R}^{hd_k \times d}$, followed by a residual connection and layer normalization to produce the final MHSA output:

$$\text{MHSA}(Z) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O \tag{12}$$

$$Z_{\text{att}} = \text{LayerNorm}(Z + \text{MHSA}(Z)) \tag{13}$$

C. Channel-wise Attention and Reconstruction

While MHSA captures long-range dependencies and aggregates contextual information across the full time-frequency grid, it treats all embedding channels equally during its output projection. However, not all channels (i.e., feature dimensions) contribute equally to the task of CE, e.g., some may carry more informative patterns depending on the propagation environment and pilot configuration. To recalibrate channel importance, HELENA applies a lightweight SE block [15] across the token embeddings. First, a global descriptor $\mathbf{s} \in \mathbb{R}^d$ is computed via average pooling:

$$\mathbf{s} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_{\text{att},i} \tag{14}$$

Then, two fully connected layers with weights $\mathbf{W}_{se1} \in \mathbb{R}^{d \times d/r}, \mathbf{W}_{se2} \in \mathbb{R}^{d/r \times d}$ and nonlinearities (ReLU and sigmoid) produce a scaled excitation vector $\mathbf{e} \in \mathbb{R}^d$:

$$\mathbf{e} = \sigma\left(\mathbf{W}_{\text{se2}} \cdot \text{ReLU}(\mathbf{W}_{\text{se1}} \cdot \mathbf{s} + \mathbf{b}_{\text{se1}}) + \mathbf{b}_{\text{se2}}\right) \tag{15}$$

$$\mathbf{Z}_{\text{scaled},i} = \mathbf{Z}_{\text{att},i} \odot \mathbf{e} \tag{16}$$

This post-attention recalibration step enables the network to focus on semantically meaningful representations by learning a global importance weighting over the embedding dimensions and selectively amplifying the most relevant features, all with minimal computational overhead. Then, each scaled token is projected back to the original patch space using reconstruction weights $\mathbf{W}_r \in \mathbb{R}^{d \times (pN_D \cdot 2)}$ and bias $\mathbf{b}_r \in \mathbb{R}^{pN_D \cdot 2}$, and the resulting set of patch vectors $\{\mathbf{P}'_i\}_{i=1}^N$ is reshaped to form the final output estimate $\hat{\mathbf{H}} \in \mathbb{R}^{N_S \times N_D \times 2}$:

$$\mathbf{P}'_i = \mathbf{Z}_{\text{scaled},i} \mathbf{W}_r + \mathbf{b}_r \tag{17}$$

$$\hat{\mathbf{H}} = \text{Reshape}(\{\mathbf{P}'_i\}_{i=1}^N) \tag{18}$$

Finally, a global residual connection [16] is included to improve convergence and ensure preservation of coarse structural information from the initial LS estimates [7]:

$$\hat{\mathbf{H}} = \mathbf{H}_{\text{LR}} + \text{HELENA}(\mathbf{H}_{\text{LR}}) \tag{19}$$

The resulting DL architecture, HELENA (equivalent to f_Θ in Eq. 2), combines local and global feature learning with low complexity. Its dual attention mechanism, MHSA and SE, enables accurate reconstruction without interpolation or side information, achieving high accuracy at low computational and memory cost as shown in Section IV.

IV. Evaluation Results

A. Dataset Description

The dataset was synthetically generated using MATLAB's 5G Deep Learning Data Synthesis framework[2], with minor modifications, and emulates pilot-based CE in realistic OFDM systems. Key parameters, including Physical Downlink Shared Channel (PDSCH) configuration, carrier frequency, Subcarrier Spacing (SCS), Cyclic Prefix (CP) type, number of RBs,

[2]https://nl.mathworks.com/help/5g/ug/deep-learning-data-synthesis-for-5g-channel-estimation.html

TABLE II
CHANNEL AND 5G-NR PARAMETERS FOR DATASET GENERATION

Parameter	Value/Description
Channel Profiles	TDL-A to TDL-E
Fading Distribution	Rayleigh
Antennas (Tx, Rx)	1, 1
Carrier Configuration	51 RB, 30 kHz SCS, Normal CP
Sub-carriers per Resource Block (RB)	12
Symbols per Slot	14
Slots per Subframe	2
Slots per Frame	20
Frame Duration	10 ms
NFFT	1024
Transmission Direction	Downlink
PDSCH Configuration	PRB: 0–50, All symbols, Type A, 1 layer
Modulation	16QAM
DM-RS Configuration	Ports: 0, Type A Position: 2 Length: 1, Config: 2
Delay Spread	1–300 ns
Doppler Shift	5–400 Hz
SNR	0–20 dB (2 dB steps)
Sample Rate	30.72 MHz
Noise	AWGN
Interpolation	Linear
Dataset Shape	[11264, 612, 14, 2] = [samples, SCS×RB, symbols, real & imaginary components]

code rate, and modulation, are summarized in Table II. The SNR range was extended to 0–20,dB (11 values, 1024 samples each), yielding 11,264 samples split into 70% training, 15% validation, and 15% testing. Ground truth labels are derived from full CSI, with models' input varying by method: HELENA, LSiDNN, and ProEsNet use LS estimates at pilot positions, while ChannelNet, EDSR, AttRNet, and CEViT use LI-interpolated LS. The dataset includes 3GPP TDL-A–E profiles and a range of Doppler shifts and SNR levels, enabling robust generalization across diverse propagation conditions.

B. Baseline Methods, Models and Experimental Setup

We compare HELENA against seven DL-based baselines, SRCNN, ChannelNet [5], EDSR [6], ProEsNet [8], AttRNet [7], LSiDNN [12], and CEViT [10], as well as two classical estimators: the LS method with LI, and a practical 5G-NR variant with denoising and averaging[3][4].

All models were reimplemented with adaptations for deployment and fair comparison. ChannelNet uses 32 Denoising Convolutional Neural Network (DnCNN) filters (vs. 64). EDSR includes 32 filters and 16 residual blocks, restructured for joint I/Q input using 2D-CNNs. ProEsNet omits upsampling blocks from the original design, as it operates directly on the full resolution resource grid without requiring interpolation from sparse pilot positions. AttRNet uses the AttResNet-Conv variant (32 filters). LSiDNN is evaluated with 48 and 1024 neurons to reflect input size. CEViT uses `Conv2DTranspose` in place of inverse patch embedding for TensorRT compatibility. All models use `padding='same'` and omit upsampling. We also include HELENA-MHSA (without SE) to isolate the contribution of dual attention.

[3]https://www.mathworks.com/help/5g/ref/nrchannelestimate.html

[4]https://github.com/srsran/srsRAN_Project/blob/main/lib/phy/upper/signal_processors/port_channel_estimator_average_impl.cpp

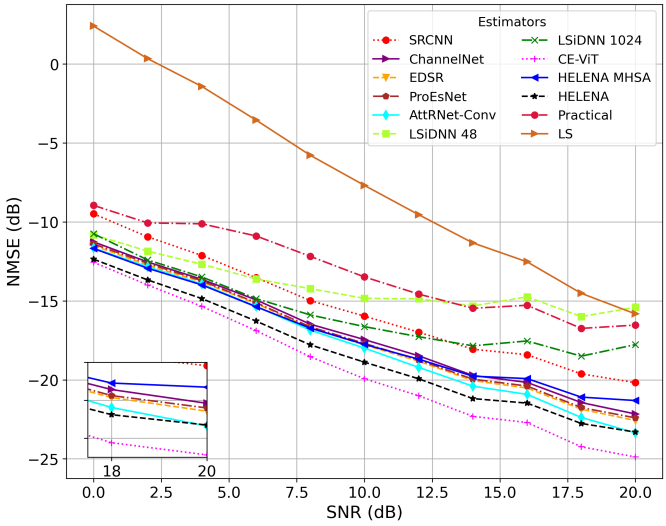

Fig. 2. NMSE vs. SNR for various CE methods.

TABLE III
COMPARISON OF NMSE (DB), FLOPS, AND INFERENCE TIME USING HELENA AS BASELINE. LOWER VALUES ARE BETTER.

Model	Params ($\times 10^6$)	FLOPS ($\times 10^9$)	NMSE (dB)		T_{inf} (ms)	
			Value	Δ	Value	Δ
SRCNN	0.014	0.241	-13.828	17.60%	0.120	-31.43%
ChannelNet	0.184	3.108	-15.507	7.59%	0.293	67.43%
EDSR	0.306	5.245	-15.773	6.03%	0.388	121.71%
ProEsNet	0.170	2.915	-15.684	6.54%	0.466	166.29%
AttRNet-Conv	0.075	1.288	-15.993	4.70%	0.293	67.43%
LSiDNN 48	1.662	0.003	-13.546	19.28%	0.0738	-57.83%
LSiDNN 1024	35.11	0.070	-14.834	11.61%	0.158	-9.71%
CE-ViT	0.880	0.053	-17.303	-3.11%	0.318	81.71%
HELENA MHSA	0.114	0.072	-15.839	5.62%	0.172	-1.71%
HELENA	0.116	0.077	-16.782	—	0.175	—

Experiments were run on the GPULab testbed[5] using 4 vCPUs, 64 GB RAM, and an NVIDIA Tesla V100-SXM3 (32 GB). The software stack included CUDA 12.2, TensorFlow 2.15, and TensorRT 8.6.1. Training used Adam optimizer and MSE loss (batch size 64), with early stopping (patience 50) and learning rate 0.01, reduced by 0.8 every 40 epochs without improvement (min 1×10^{-5}). Checkpoints were selected based on validation loss. Inference time was averaged over 100 single-sample runs. Results refer to TensorRT-optimized models, as non-optimized versions showed similar accuracy but are unsuitable for deployment [17].

C. Design Decisions and Parameter Selection in HELENA

The parameters of HELENA were chosen based on the structural characteristics of the input data and validated empirically to optimize the trade-offs between estimation accuracy, inference time, and model complexity. The convolutional layers use 12×2 and 6×7 kernels to exploit the

[5]https://doc.ilabt.imec.be/ilabt/gpulab/

time-frequency layout of the OFDM grid while remaining lightweight. A patch size of 12 along the frequency axis yields 51 tokens, matching the 51 RBs in the dataset. Each token is embedded into a 64-dimensional space, which offers sufficient capacity for expressive representation while avoiding overfitting or excessive compute. The MHSA block uses 4 heads, allowing parallel modeling of diverse attention patterns without excessive overhead while maintaining high accuracy. A reduction ratio of 4 in the SE block avoids aggressive bottlenecking, enabling effective channel-wise recalibration at low cost. Dropout (0.1) and residual connections further enhance generalization and convergence.

D. Model Accuracy

Fig. 2 shows the variation of NMSE, as defined in Eq. 4, across different SNR levels. CEViT achieves the best NMSE (-17.303 dB), followed closely by HELENA (-16.782 dB), which requires no explicit SNR, Doppler, or latency inputs and avoids interpolation. Compared to LS (-3.56 dB) and the practical estimator (-12.25 dB), HELENA improves NMSE by 95.3% and 65.5%, respectively. Relative to other DL baselines, HELENA improves over ProEsNet by 6.54%, EDSR by 6.03%, AttRNet by 4.70%, and ChannelNet by 7.59%. Compared to the lighter models, the gain is even higher: 17.60% over SRCNN, 19.28% over LSiDNN-48, and 11.61% over LSiDNN-1024, highlighting the trade-off between model complexity and diminishing performance returns. Finally, removing the SE block degrades accuracy by 5.62%, confirming the benefit of dual attention, yet the model still outperforms all lighter baselines and the most accurate SR-based models (ProEsNet, EDSR, ChannelNet), in average across all SNRs.

E. Computational and Memory Cost vs. Inference Time

To meet the stringent latency requirements of 5G-NR, CE, equalization, and decoding must complete within the Hybrid Automatic Repeat Request (HARQ) deadline, which allows up to three Transmission Time Intervals (TTIs) (0.5 ms per TTI at 30 kHz SCS) [11]. This implies a tight budget $T_{\max}$ of roughly 0.5 ms for CE. As shown in Table III, all models satisfy the latency constraint $T_{\inf}(f_\Theta) \leq T_{\max}$ from Eq. 3, albeit with varying margins.

HELENA achieves a strong trade-off: it delivers the second-best accuracy, just 3.1% lower than CEViT, and has the lowest inference time among the top-performing models (excluding HELENA MHSA). Specifically, CEViT and AttRNet are 81.71% and 67.43% slower, respectively (0.318 ms and 0.293 ms vs. 0.175 ms). Compared to the lighter models, HELENA is 31.43% slower than SRCNN (0.120ms) and LSiDNN-48 (0.0738ms, 57.83% fastest) but they achieve it at the cost of higher prediction error. HELENA MHSA, without the SE block, performs similarly at 0.172ms. ProEsNet, while achieving comparable accuracy to ChannelNet and EDSR, exhibits the highest latency among top models (0.466ms), consuming over 90% of the latency budget. In general, while all models satisfy Eq. 3, HELENA uses only 35%, highlighting its suitability for real-time deployment.

We can see that runtime does not scale linearly with computational complexity (in Floating-point operations per second (FLOPS)[6]). For example, HELENA uses 66.7% fewer FLOPS than SRCNN (0.08 vs. 0.24 G), yet runs 38.5% slower. AttRNet and EDSR require 16× and 65× more operations than HELENA, but inference is only 65.6% and 96.5% slower, respectively. ChannelNet also runs 37.9% slower, despite requiring 39× more FLOPS. In contrast, ProEsNet performs 2.9 GFLOPS per inference, over 36× more than HELENA, yet is only 2.7× slower, reinforcing the decoupling between FLOPS and latency due to implementation and hardware bottlenecks.

The number of parameters, which relates to the memory required at inference time, shows limited correlation with runtime or estimation accuracy. HELENA reaches -16.87 dB NMSE with only 0.11M parameters, outperforming ChannelNet (0.18M) and LSiDNN-1024 (35.11M) while being faster. CEViT achieves the best NMSE with 0.88M parameters, 8× more than HELENA, but is slower. ProEsNet has a moderate footprint of 0.17M parameters. These results emphasize the need to jointly consider FLOPS, parameter count, and inference time when optimizing models for real-time CE, as actual latency is shaped by hardware–software co-design aspects such as memory access, operator fusion, and backend scheduling, which extend beyond theoretical complexity.

V. Conclusion and Future Works

This paper presented HELENA, a DL-based model for efficient and accurate pilot-based CE in OFDM systems. HELENA integrates shallow convolutional layers with dual attention—patch-wise MHSA and channel-wise SE—to balance estimation accuracy, computational complexity, and inference latency. Experimental results demonstrate that HELENA achieves near state-of-the-art performance with low computational and memory requirements while maintaining fast inference, making it well suited for real-time, latency-sensitive wireless systems. The observed weak correlation between FLOPS, parameter count, and actual inference time further highlights the importance of holistic evaluation criteria when designing practical DL-based CE models. Although HELENA builds upon established architectural components, this work shows that a careful and lightweight combination of attention mechanisms can yield deployment-ready performance under strict latency constraints.

Future research will explore a broader architectural design space through systematic ablation studies, including the impact of patch size, embedding dimension, and the number of attention heads, in order to better characterize accuracy-complexity trade-offs. In addition, although the current evaluation relies on synthetically generated datasets, validating HELENA using real wireless measurements and hardware testbeds remains an important next step to assess robustness under practical RF impairments. The framework is also expected to be extended to Multiple-Input Multiple-Output (MIMO), massive MIMO,

[6]Measured using TensorFlow's profiling tool: https://www.tensorflow.org/api_docs/python/tf/compat/v1/profiler/ProfileOptionBuilder

and beamformed systems, which will require adaptations to efficiently exploit spatial correlations across antennas. Further directions include the visualization and analysis of attention maps to improve model interpretability, as well as the incorporation of statistical performance indicators such as variance and confidence intervals alongside average NMSE. Finally, deployment on hardware-constrained platforms, including FPGA and edge AI accelerators, will be investigated to further evaluate real-time feasibility.

ACKNOWLEDGMENTS

This research was supported by the 6G-TWIN project under the SNS JU Horizon Europe program with Grant Agreement No. 101136314. This research is partially funded by the imec.icon project Rapidness, which is co-financed by imec and Flanders Innovation and Entrepreneurship under project nr HBC.2024.0772. The 6G-TWIN support relates to energy-efficient data-driven functional models for Network Digital Twins, while RAPIDNESS supported the development of a hardware-efficient, low-latency learning-based channel estimation model.

REFERENCES

[1] C. Qing, Z. Liu, G. Ling *et al.*, "Channel estimation in otfs systems by leveraging differential modulation," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 5, pp. 6907–6918, 2025.

[2] C. Qing, W. Hu, Z. Liu *et al.*, "Sensing-aided channel estimation in ofdm systems by leveraging communication echoes," *IEEE Internet of Things Journal*, vol. 11, no. 23, pp. 38 023–38 039, 2024.

[3] J. Hoydis, F. A. Aoudia, A. Valcarce *et al.*, "Toward a 6g ai-native air interface," *IEEE Communications Magazine*, vol. 59, no. 5, pp. 76–81, 2021.

[4] M. Camelo, R. Mennes, A. Shahid *et al.*, "An ai-based incumbent protection system for collaborative intelligent radio networks," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 16–23, 2020.

[5] M. Soltani, V. Pourahmadi, A. Mirzaei *et al.*, "Deep learning-based channel estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, 2019.

[6] D. Maruyama, K. Kanai, and J. Katto, "Performance evaluations of channel estimation using deep-learning based super-resolution," in *2021 IEEE 18th Annual Consumer Communications and Networking Conference (CCNC)*, 2021, pp. 1–6.

[7] W. Gao, W. Zhang, L. Liu *et al.*, "Deep residual learning with attention mechanism for ofdm channel estimation," *IEEE Wireless Communications Letters*, vol. 14, no. 2, pp. 250–254, 2025.

[8] Y. Zhang, J. Hou, and H. Liu, "Deep learning based fully progressive image super-resolution scheme for channel estimation in ofdm systems," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 6, pp. 9021–9025, 2024.

[9] J. Guo, G. Liu, Q. Wu *et al.*, "Parallel attention-based transformer for channel estimation in ris-aided 6g wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 11, pp. 15 927–15 940, 2024.

[10] F. Liu, P. Jiang, J. Zhang *et al.*, "Pd-cevit: A novel pilot pattern design and channel estimation network for ofdm systems," *IEEE Transactions on Communications*, pp. 1–1, 2024.

[11] S. A. Damjancevic, E. Matus, D. Utyansky *et al.*, "Channel estimation for advanced 5g/6g use cases on a vector digital signal processor," *IEEE Open Journal of Circuits and Systems*, vol. 2, pp. 265–277, 2021.

[12] A. Sharma, S. A. U. Haq, and S. J. Darak, "Low complexity deep learning augmented wireless channel estimation for pilot-based ofdm on zynq system on chip," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 71, no. 5, pp. 2334–2347, 2024.

[13] A. B. Singh and V. K. Gupta, "Performance evaluation of mmse and ls channel estimation in ofdm system," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 15, no. 1, pp. 39–43, 2014.

[14] J. Zhang, K. Qiu, Y. Li *et al.*, "Channel estimation based on linear interpolation algorithm in ddo-ofdm system," in *Asia Communications and Photonics Conference and Exhibition*, 2010, pp. 605–606.

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[16] K. He, X. Zhang, S. Ren *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778. [Online]. Available: https://arxiv.org/abs/1512.03385

[17] D. Góez, E. A. Beyazıt, N. Slamnik-Krješ̆torac *et al.*, "Computational efficiency of deep learning-based super resolution methods for 5g-nr channel estimation," in *2024 IEEE Latin-American Conference on Communications (LATINCOM)*, 2024, pp. 1–7.