APPENDIX FOR "EFFICIENT TIME SERIES FORECASTING VIA HYPER-COMPLEX MODELS AND FREQUENCY AGGREGATION"

A NOTATIONS & SYMBOLS

A.1 NOTATION

We provide a detailed table of the involved notation in this paper:

Symbol	Description
B	Batch size.
L	Lookback window size.
D	Number of features for each time step.
T	Length of the prediction horizon.
E	Embedding size.
M	Number of frequencies to select from all the frequencies using the top M magnitudes.
X	Multivariate time series with a lookback window of size L at timestamps t, where $X \in \mathbb{R}^{L\times D}$.
X_t	Multivariate values of D distinct series at timestamp t, where $X_t \in \mathbb{R}^D$.
$X_{t,i}$	The value of the i-th feature of the distinct series at timestamp t, where $X_{t,i} \in \mathbb{R}$.
$\hat{X}$	Ground truth target values, where $\hat{X} \in \mathbb{R}^{T\times D}$
σ	activation function
P	Number of windows in the STFT.
N_{FFT}	Number of frequency bins in each window of the STFT.
ω	Window function for the STFT.
X_E	X after traversing through the embedding layer. $X_E \in \mathbb{R}^{L\times D\times E}$
X_{Rec}	The reconstructed X after the frequency alteration. $X_{Rec} \in \mathbb{R}^{L\times D\times E}$
c_i^t	The i-th window of the input in the time domain $c_i^t \in \mathbb{C}^{D\times 2N_{FFT}\times E}$.
C_i	The i-th window of the STFT containing N_{FFT} frequency bins $C_i \in \mathbb{C}^{D\times N_{FFT}\times E}$.
C_i^{in}	The i-th window of the STFT, retaining the top M frequency components based on magnitude $C_i^{'} \in \mathbb{C}^{D\times M\times E}$.
C_i^{out}	The i-th window of the STFT after the WM-MLP/WHC has been applied $C_i^M \in C^{D\times M\times E}$.
$W_{i\to j}$	The weights that capture the frequency energy shift between window i and j, defined as $W_{i\to j} = W_{i\to j}^{\text{Real}} + jW_{i\to j}^{\text{Img}}$, where $W_{i\to j} \in \mathbb{C}^{E\times E}$.
$B_{i\to j}$	The bais that capture the frequency energy shift between windows i and j, defined as $B_{i\to j} = B_{i\to j}^{\text{Real}} + jB_{i\to j}^{\text{Img}}$, where $B_{i\to j} \in \mathbb{C}^E$.

Table 4: Table of Symbols and Descriptions

A.2 DIMENSIONS

The following table summarizes the dimensions of the data tensor in every step of the FIA-Net.

B ADDITIONAL EXPERIMENTAL DETAILS

B.1 DATASET DESCRIPTIONS

In our experiments, we utilized thirteen real-world datasets to assess the effectiveness of models for long-term TSF. Below, we provide the details of these datasets, categorized by their forecasting horizon.

- **Exchange**: This dataset includes daily exchange rates for eight countries (Australia, Britain, Canada, Switzerland, China, Japan, New Zealand, and Singapore) from 1990 to 2016.

Symbol	Dimension
X	$\mathbb{R}^{B\times L\times D}$
X_E	$\mathbb{R}^{B\times L\times D\times E}$
C_i	$\mathbb{C}^{B\times N_{FFT}\times D\times E}$
C_i^M	$\mathbb{C}^{B\times M\times D\times E}$
$C_i^{\text{in/out}}$	$\mathbb{C}^{B\times M\times D\times E}$
X_{Rec}	$\mathbb{R}^{B\times L\times D\times E}$
$\hat{X}$	$\mathbb{R}^{B\times T\times D}$

Table 5: Table of Symbols and Dimension

- **Weather**: This dataset gathers 21 meteorological indicators, including humidity and air temperature, from the Weather Station of the Max Planck Biogeochemistry Institute in Germany in 2020. The data is collected every 10 minutes.
- **Traffic**: For long-term forecasting, this dataset includes hourly traffic data from 862 freeway lanes in San Francisco, with data collected since January 1, 2015.
- **Electricity**: For long-term forecasting, this dataset covers electricity consumption data from 321 clients, with records starting from January 1, 2011, and a sampling interval of 15 minutes.
- **ETT**: This dataset is sourced from two electric transformers, labeled ETTh1 and ETTm1, with two different resolutions: 15 minutes and 1 hour. These are used as benchmarks for long-term forecasting.

Datasets	Weather	Traffic	Electricity	ETTh1	ETTm1	Exchange Rates
Features	21	862	321	7	7	8
Timesteps	52696	17544	26304	17420	69680	7588
Frequency	10m	1h	1h	1h	15m	1d
Lookback Window	96	48	96	96	96	96
Prediction Length	96, 192, 336, 720	96, 192, 336, 720	96, 192, 336, 720	96, 192, 336, 720	96, 192, 336, 720	96, 192, 336, 720

Table 6: Long Term Datasets Parameters

B.2 BASELINES

We employ a selection of SoTA representative models for our comparative analysis, focusing on Transformer-based architectures and other popular models. The models included are as follows:

- **Informer**: Informer enhances the efficiency of self-attention mechanisms to effectively capture dependencies across variables. The source code was obtained from GitHub, and we utilized the default configuration with a dropout rate of 0.05, two encoder layers, one decoder layer, a learning rate of 0.0001, and the Adam optimizer.
- **Reformer**: Reformer combines the power of Transformers with efficient memory and computation management, especially for long sequences. The source code was sourced from GitHub, and we employed the recommended configuration for our experiments.
- **Autoformer**: Autoformer introduces a decomposition block embedded within the model to progressively aggregate long-term trends from intermediate predictions. The source code was accessed from GitHub, and we followed the recommended settings for all experiments.
- **FEDformer**: FEDformer introduces an attention mechanism based on low-rank approximation in the frequency domain combined with a mixture of expert decomposition to handle distribution shifts. The source code was retrieved from GitHub. We utilized the Frequency Enhanced Block (FEB-f) and selected the random mode with 64 as the experimental configuration.

- **LTSF-Linear**: LTSF-Linear is a minimalist model employing simple one-layer linear models to learn temporal relationships in time series data. We used it as our baseline for long-term forecasting, downloading the source code from GitHub, and adhered to the default experimental settings.
- **PatchTST**: PatchTST is a Transformer-based model designed for TSF, introducing patching and a channel-independent structure to enhance model performance. The source code was obtained from GitHub, and we used the recommended settings for all experiments.
- **FreTS**: FRETS is a sophisticated model tailored for efficient TSF by exploiting a frequency domain approach. The implementation is available on GitHub, and we utilized the default configuration as recommended by the authors. In our work, FRETS serves as the foundational model. We address its limitations, particularly its handling of non-stationary data, while adapting its strengths, such as its complex frequency learner. To fully grasp the contributions of this paper, we recommend reviewing FRETS in detail first.

B.3 Implementation Details

Table 7 lists the hyperparameter values used in the FIA-Net implementation. Both WM-MLP and HC-MLP backbones are implemented with the same hyperparameter values, except for p, the number of STFT windows.

DataSets	Weather	Traffic	Electricity	ETTh1	ETTm1	Exchange rate
Batch Size	16	4	4	8	8	8
Embed Size	128	32	64	128	128	128
Hidden Size	256	256	256	256	256	256
NFF	16	32	32	6	48	32
STFT Windows	7	13	13	33	4	13
S-M	10	M_{max}	4	4	4	M_{max}
Epoch	10	10	10	10	10	10

Table 7: Hyperparameter Settings for Long-Term Datasets for the WM-MLP and HC-MLP

B.4 Evaluation Metrics

In this study, we use the Mean Squared Error (MSE) as the loss function during training. However, for evaluation, we report both the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE).

which are defined as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2, \quad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}, \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

Where:

- Y_i represents the true target values,
- $\hat{Y}_i$ represents the predicted values,
- n is the total number of samples.

B.5 Normalization Methods

In this study, similar to the FRETS model Yi et al. (2023b), we apply min-max normalization to standardize the input data to the range between 0 and 1. This method helps in ensuring that all features contribute equally to the model and prevents any specific feature from dominating due to differences in scale. The formula for min-max normalization is given by:

$$X_{Norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

By normalizing the data, we ensure that all input features are within the same range, which can improve model convergence and performance.

C ADDITIONAL INFORMATION ON HC NUMBERS AND MODELS

In this section we extend the discussion on HC numbers, considering additional values of p beyond $p = 4$. We couple the presentation with the construction of the corresponding HC-MLP in the considered base. Recall that the base of a HC number, i.e., the number of its components is given by $b = 2p$. While hyper-complex number can be defined for any value of b, most research has been performed on b that is given by a power of 2, as the resulting structure of the (algebraic) field. The addition of two HC numbers is simply given by the component-wise summation. In what follows, we focus on HC multiplication and additional properties. For more information on the HC number system,, we refer the reader to Kantor & Solodovnikov (1989).

C.1 BASE 2 - COMPLEX NUMBERS

When $b = 2$, the resulting field is the complex plane $\mathbb{C}$. We describe $\mathbb{C}$ for completeness of presentation. Given two complex numbers $C_1 = \alpha_1 + j\alpha_2$ and $C_2 = \beta_1 + j\beta_2$, where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are real numbers, their complex multiplication is defined as:

$$C_1 \cdot C_2 = (\alpha_1\beta_1 - \alpha_2\beta_2) + j(\alpha_1\beta_2 + \alpha_2\beta_1)$$

The norm of a complex number is given by:

$$|C_1|_{\mathbb{C}} = \sqrt{\alpha_1^2 + \alpha_2^2},$$

which is preserved under multiplication, i.e.,

$$|C_1 \cdot C_2|_{\mathbb{C}} = |C_1|_{\mathbb{C}} \cdot |C_2|_{\mathbb{C}}.$$

Since the STFT with a single window ($p = 1$) is equivalent to the standard FFT, applying our method for hyper-complex number MLP results in the following equation:

$$C_{\text{in}} = \text{FFT}(X)$$

$$C^{\text{out}} = \sigma(C^{\text{in}}_{\text{Real}} \cdot W_{1,\text{Real}} - C^{\text{in}}_{\text{Imag}} \cdot W_{1,\text{Imag}} + B_{1,\text{Real}}) + \sigma(j(C^{\text{in}}_{\text{Real}} \cdot W_{1,\text{Imag}} + C^{\text{in}}_{\text{Imag}} \cdot W_{1,\text{Real}} + B_{1,\text{Imag}})$$

Here, $W_i \in \mathbb{C}^{E\times E}$ denotes the layer weights, $B \in \mathbb{C}^E$ represents the bias term, and the multiplication occurs across the embedding dimension. Note that for $b = 2$ the HV formulation boils down to the one from Yi et al. (2023b). Thus, the HC-MLP can be considered as an HC generalization of the FD-MLP. which allows for efficient window aggregation.

C.2 BASE 4 - QUATERNIONS

Denote the field of Quaternions with $\tilde{\mathbb{Q}}$. We represent Quatenions with a couple of Complex number, i.e., for $H_1, H_2 \in \tilde{\mathbb{Q}}$, $H_1 = (\alpha_1, \alpha_2)$ and $H_2 = (\beta_1, \beta_2)$, their multiplication is defined as

$$H_1 \cdot H_2 = (\alpha_1\beta_1 - \overline{\alpha_2}\beta_2, \quad \alpha_2\overline{\beta_1} + \alpha_1\beta_2)$$

The norm of a quaternion is given by:

$$|q|_{\tilde{\mathbb{Q}}} = \sqrt{|\alpha_1|^2_{\mathbb{C}} + |\alpha_2|^2_{\mathbb{C}}}$$

The norm is preserved under multiplication, meaning:

$$|q_1 \cdot q_2|_{\tilde{\mathbb{Q}}} = |q_1|_{\tilde{\mathbb{Q}}} \cdot |q_2|_{\tilde{\mathbb{Q}}}$$

For our model, the corresponding HC-MLP (which we denote QuatMLP) operating on $C^{\text{in}} = (C_1^{\text{in}}, C_2^{\text{in}}) \in \tilde{\mathbb{Q}}$, is given by,

$$C^{\text{out}} = \text{QuatMLP}(C^{\text{in}}) = \sigma(C^{\text{in}} \cdot W + B)$$

where:

$$C_1^{\text{out}} = \sigma(C_1 \cdot W_1 - \overline{C_2} \cdot W_2 + B_1), \quad C_2^{\text{out}} = \sigma(C_2 \cdot \overline{W_1} + C_1 \cdot W_2 + B_2).$$

Here, $W_i \in \mathbb{C}^{E\times E}$, $i = 1, 2$ denote the layer weights, $B \in \mathbb{C}^E$ represents the bias term, and the multiplication involves complex MLP operations across the embedding dimension.

C.3 BASE 16 - SEDENIONS

Elements on the Sedenions field, denoted SS, are denoted with 8-tuples of complex numbers. Given two sedenions represented by complex numbers $S_1, S_2 \in SS$, $S_1 = (\alpha_1, \alpha_2, \ldots, \alpha_8)$ and $S_2 = (\beta_1, \beta_2, \ldots, \beta_8)$, their multiplication is given by

$$S_1 \cdot S_2 = \begin{pmatrix} \alpha_1\beta_1 - \alpha_2\overline{\beta_2} - \alpha_3\overline{\beta_3} - \alpha_4\overline{\beta_4} - \alpha_5\overline{\beta_5} - \alpha_6\overline{\beta_6} - \alpha_7\overline{\beta_7} - \alpha_8\overline{\beta_8} \\ \alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\overline{\beta_4} - \alpha_4\overline{\beta_3} + \alpha_5\overline{\beta_6} - \alpha_6\overline{\beta_5} + \alpha_7\overline{\beta_8} - \alpha_8\overline{\beta_7} \\ \alpha_1\beta_3 - \alpha_2\overline{\beta_4} + \alpha_3\beta_1 + \alpha_4\beta_2 + \alpha_5\overline{\beta_7} - \alpha_6\overline{\beta_8} - \alpha_7\overline{\beta_5} + \alpha_8\overline{\beta_6} \\ \alpha_1\beta_4 + \alpha_2\beta_3 - \alpha_3\beta_2 + \alpha_4\beta_1 + \alpha_5\overline{\beta_8} + \alpha_6\overline{\beta_7} - \alpha_7\overline{\beta_6} - \alpha_8\overline{\beta_5} \\ \alpha_1\beta_5 - \alpha_2\overline{\beta_6} - \alpha_3\overline{\beta_7} - \alpha_4\overline{\beta_8} + \alpha_5\beta_1 + \alpha_6\beta_2 + \alpha_7\beta_3 + \alpha_8\beta_4 \\ \alpha_1\beta_6 + \alpha_2\beta_5 - \alpha_3\overline{\beta_8} + \alpha_4\overline{\beta_7} - \alpha_5\beta_2 + \alpha_6\beta_1 - \alpha_7\beta_4 + \alpha_8\beta_3 \\ \alpha_1\beta_7 + \alpha_2\beta_8 + \alpha_3\beta_5 - \alpha_4\beta_6 - \alpha_5\beta_3 + \alpha_6\beta_4 + \alpha_7\beta_1 - \alpha_8\beta_2 \\ \alpha_1\beta_8 - \alpha_2\beta_7 + \alpha_3\beta_6 + \alpha_4\beta_5 - \alpha_5\beta_4 - \alpha_6\beta_3 + \alpha_7\beta_2 + \alpha_8\beta_1 \end{pmatrix}$$

where each component follows the rules of Complex multiplication. The norm of a sedenion is given by:

$$|S|_{SS} = \sqrt{\sum_{j=1}^{8} |\alpha_j|_{\mathbb{C}}^2}$$

Unlike nase 2, 4 and 8, Sedenions *do not* preserve the norm under addition and multiplication.

The base-16 HC-MLP, denoted SedMLP, operating on an input C^{in} from the STFT with multiple windows $C^{\mathsf{in}} = (C_j^{\mathsf{in}})_{j=1}^{8}$, is given by

$$C^{\mathsf{out}} = \text{SedMLP}(C^{\mathsf{in}}) = \sigma(C^{\mathsf{in}} \cdot W + B)$$

where:

$$\begin{aligned}
C_1^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_1 - C_2^{\mathsf{in}}\overline{W_2} - C_3^{\mathsf{in}}\overline{W_3} - C_4^{\mathsf{in}}\overline{W_4} - C_5^{\mathsf{in}}\overline{W_5} - C_6^{\mathsf{in}}\overline{W_6} - C_7^{\mathsf{in}}\overline{W_7} - C_8^{\mathsf{in}}\overline{W_8} + B_1\right) \\
C_2^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_2 + C_2^{\mathsf{in}}W_1 + C_3^{\mathsf{in}}\overline{W_4} - C_4^{\mathsf{in}}\overline{W_3} + C_5^{\mathsf{in}}\overline{W_6} - C_6^{\mathsf{in}}\overline{W_5} + C_7^{\mathsf{in}}\overline{W_8} - C_8^{\mathsf{in}}\overline{W_7} + B_2\right) \\
C_3^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_3 - C_2^{\mathsf{in}}\overline{W_4} + C_3^{\mathsf{in}}W_1 + C_4^{\mathsf{in}}W_2 + C_5^{\mathsf{in}}\overline{W_7} - C_6^{\mathsf{in}}\overline{W_8} - C_7^{\mathsf{in}}\overline{W_5} + C_8^{\mathsf{in}}\overline{W_6} + B_3\right) \\
C_4^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_4 + C_2^{\mathsf{in}}W_3 - C_3^{\mathsf{in}}W_2 + C_4^{\mathsf{in}}W_1 + C_5^{\mathsf{in}}\overline{W_8} + C_6^{\mathsf{in}}\overline{W_7} - C_7^{\mathsf{in}}\overline{W_6} - C_8^{\mathsf{in}}\overline{W_5} + B_4\right) \\
C_5^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_5 - C_2^{\mathsf{in}}\overline{W_6} - C_3^{\mathsf{in}}\overline{W_7} - C_4^{\mathsf{in}}\overline{W_8} + C_5^{\mathsf{in}}W_1 + C_6^{\mathsf{in}}W_2 + C_7^{\mathsf{in}}W_3 + C_8^{\mathsf{in}}W_4 + B_5\right) \\
C_6^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_6 + C_2^{\mathsf{in}}W_5 - C_3^{\mathsf{in}}\overline{W_8} + C_4^{\mathsf{in}}\overline{W_7} - C_5^{\mathsf{in}}W_2 + C_6^{\mathsf{in}}W_1 - C_7^{\mathsf{in}}W_4 + C_8^{\mathsf{in}}W_3 + B_6\right) \\
C_7^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_7 + C_2^{\mathsf{in}}W_8 + C_3^{\mathsf{in}}W_5 - C_4^{\mathsf{in}}W_6 - C_5^{\mathsf{in}}W_3 + C_6^{\mathsf{in}}W_4 + C_7^{\mathsf{in}}W_1 - C_8^{\mathsf{in}}W_2 + B_7\right) \\
C_8^{\mathsf{out}} &= \sigma\left(C_1^{\mathsf{in}}W_8 - C_2^{\mathsf{in}}W_7 + C_3^{\mathsf{in}}W_6 + C_4^{\mathsf{in}}W_5 - C_5^{\mathsf{in}}W_4 - C_6^{\mathsf{in}}W_3 + C_7^{\mathsf{in}}W_2 + C_8^{\mathsf{in}}W_1 + B_8\right)
\end{aligned}$$

Here, $W_i \in \mathbb{C}^{E \times E}$, $i = 1, \ldots, 8$ denotes the layer weights, $B \in \mathbb{C}^E$ represents the bias term, and the multiplication involves complex MLP operations across the embedding dimension.

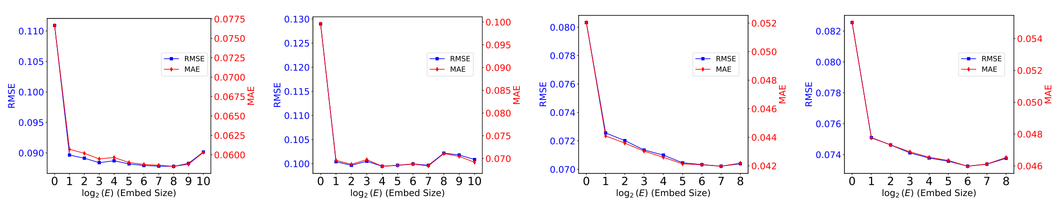
D Additional Ablation Studies

This section presents additional ablation studies, expanding on the findings reported in Section 5.3. We analyze the impact of FFT resolution, embedding size, and the number of STFT windows on WM-MLP performance. Additionally, we include further results for the frequency compression, sequence length, and Real vs. Imaginary component discussions. Furthermore, we provide a comparative analysis of various hyper-complex fields (Octonions, Quaternions, and Sedenions) for the HC-MLP and report the corresponding results.

D.1 Parameter Sensitivity

In this section, we conduct a parameter sweep to examine the effects of different hyperparameters on model performance. To accomplish this, we utilize two datasets: the ETTh1 dataset and the electricity dataset. Each section presents four graphs illustrating the results on the two datasets for a configuration of $I/O = 96 \times 96, 336$. Except for the specific experiment sweep, the embedding size is set to 128 for the ETTh1 dataset and 64 for the electricity dataset, with M set to 0 for all datasets.

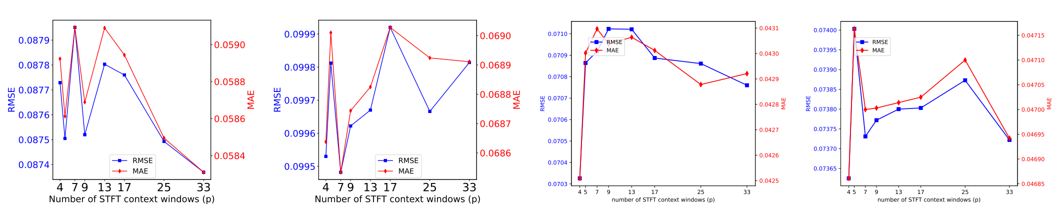
Embed Size In this section, we evaluate the influence of embedding size on the model's performance. We conducted experiments with embedding dimensions $E \in \{1, 2, 4, 8, 16, 32, 64, 128, 256, 512\}$, while keeping the following parameters fixed: $N_{\text{FFT}} = 16$, $B = 8$, $p = 13$, and $M = M_{\text{max}}$.We can observe that as we increase the embedding size, the loss decreases until we reach a certain point (which is dependent on the dataset). This is likely because a larger embedding size enables the model to capture more features; however, an excessively high embedding size may lead to overfitting.



(a) T=96 on ETTh1 (b) T=336 on ETTh (c) T=96 on electricity (d) T=336 on electricity

Figure 7: Comparison of MSE and MAE across different values of E for varying Ton the ETTh1 and Electricity datasets.

Amount of Windows (High Dim) In this section, we evaluate the influence of the number of windows (p) on the model's performance. We conducted experiments with different window counts $p \in \{3, 6, 14, 17, 25, 33\}$, while keeping the following parameters fixed: $B = 8$, $M = M_{\text{max}}$, and the overlap between windows is 50%.

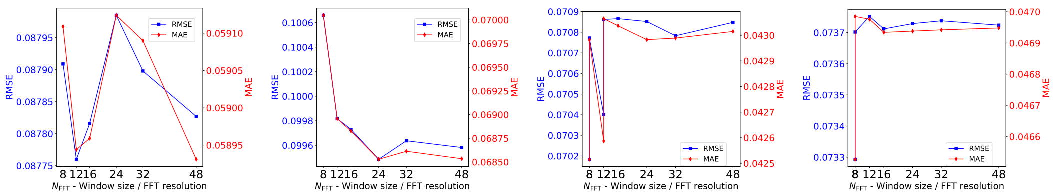


(a) T=96 on ETTh1 (b) T=336 on ETTh1 (c) T=96 on electricity (d) T=336 on electricity

Figure 8: Comparison of MSE and MAE across different values of p for varying T on the ETTh1 and Electricity datasets.

FFT Resolution (NFFT) In this section, we evaluate the influence of the FFT resolution (N_{FFT}) on the model's performance. We conducted experiments with different $N_{\text{FFT}} \in$

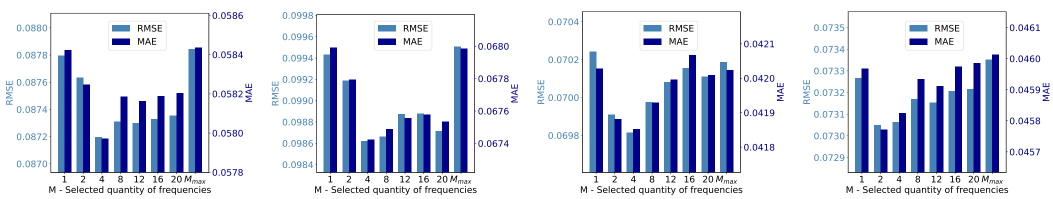
$\{6, 8, 12, 16, 24, 32, 48\}$, while keeping the following parameters fixed: $p = 25$, $B = 8$, $M = M_{\text{max}}$.



(a) T=96 on ETTh1 (b) T=336 on ETTh1 (c) T=96 on electricity (d) T=336 on electricity

Figure 9: Comparison of MSE and MAE across different values of N_{FFT} for varying T on the ETTh1 and Electricity datasets.

Frequency Choose Max (M) In this section, we provide additional results for various datasets and prediction lengths T regarding the discussion on frequency compression 5.

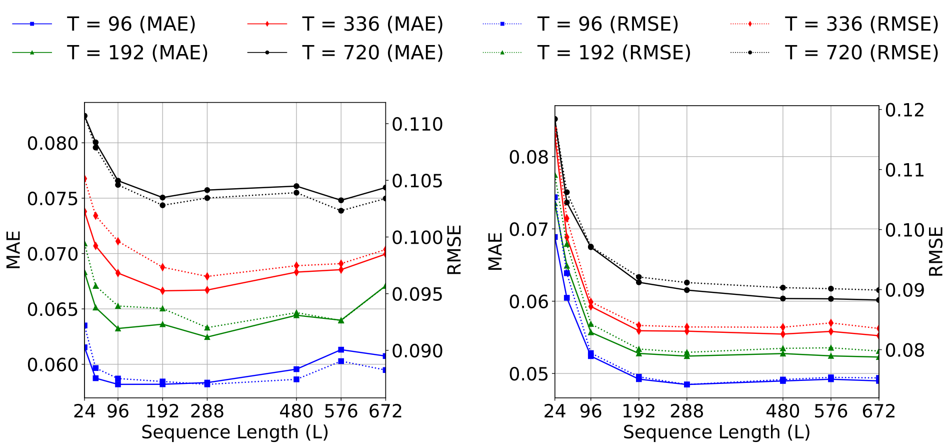


(a) T=96 on ETTh1 (b) T=336 on ETTh1 (c) T=96 on electricity (d) T=336 on electricity

Figure 10: Comparison of MSE and MAE across different values of M for various T on the ETTh1 and Electricity datasets.

D.2 Different LookBack Window

In this section, we present additional results for various lookback windows on the ETTh1 and ETTm1 datasets.



(a) ETTh1 Dataset (b) ETTm1 Dataset

Figure 11: MAE and RMSE in relation to the Lookback Window L for varying prediction lengths $T \in \{96, 192, 336, 720\}$ for the ETTh1 and ETTm1 datasets.

D.3 Real Vs imaginary Components

This section provides additional information regarding the real versus imaginary experiment discussed in Section 5.3.3.

Dataset	I/O	96/96		96/192		96/336		96/720	
	Metric	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
ETTm1	X^{Real}	0.0522	0.0797	0.0560	0.0850	0.0597	0.0888	0.0658	0.0958
	X^{Imag}	0.0521	0.0792	0.0562	0.0844	0.0592	0.0879	0.0684	0.0976
	W^{Real}	0.0522	0.0791	0.0557	0.0843	0.0588	0.0875	0.0669	0.0964
	W^{Imag}	0.0526	0.0801	0.0560	0.0849	0.0596	0.0888	0.0651	0.0953
	$W^{\mathsf{Imag}}, X^{\mathsf{Imag}}$	0.0523	0.0798	0.0560	0.0849	0.0592	0.0884	0.0644	0.0947
	$W^{\mathsf{Real}}, X^{\mathsf{Real}}$	0.0522	0.0791	0.0557	0.0843	0.0588	0.0887	0.0669	0.0930
	Normal	0.0522	0.0791	0.0565	0.0848	0.0592	0.0878	0.0685	0.0975
ETTh1	X^{Real}	0.0584	0.0877	0.0638	0.0944	0.0684	0.0997	0.0767	0.1047
	X^{Imag}	0.0582	0.0879	0.0634	0.0943	0.0679	0.0997	0.0756	0.1041
	W^{Real}	0.0586	0.0880	0.0644	0.0948	0.0685	0.0998	0.0759	0.1039
	W^{Imag}	0.0584	0.0880	0.0646	0.0951	0.0694	0.1008	0.0781	0.1065
	$W^{\mathsf{Imag}}, X^{\mathsf{Imag}}$	0.0586	0.0880	0.0644	0.0947	0.0685	0.0998	0.0759	0.1040
	$W^{\mathsf{Real}}, X^{\mathsf{Real}}$	0.0587	0.0882	0.0642	0.0948	0.0690	0.1005	0.0765	0.1050
	Normal	0.0586	0.0878	0.0639	0.0945	0.0684	0.0998	0.0765	0.1043

Table 8: Performance comparison on the ETTm1, ETTh1, and Electricity datasets for $I/O = 96 \times \{96, 192, 336, 720\}$ with different modes. X^{Real} and X^{Imag} refer to hiding the real and imaginary parts of the input, respectively. W^{Real} and W^{Imag} denote zeroing the real and imaginary weights, respectively. The cases where both the real and imaginary components are completely ignored (i.e., both weights and inputs are zeroed) are represented by $W^{\mathsf{Imag}}, X^{\mathsf{Imag}}$ and $W^{\mathsf{Real}}, X^{\mathsf{Real}}$. MAE and RMSE are reported, where lower values indicate better performance.

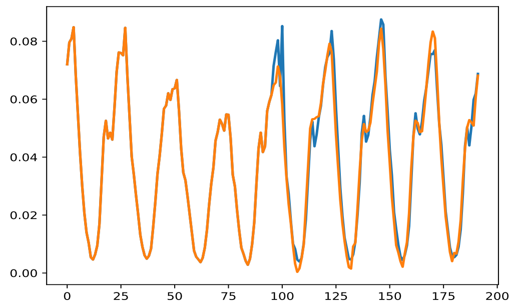
E HC-MLP Experimental Results With For Various Values of p

In this section, we present additional results on the HC-MLP for various bases. We provide results for the Complex base (p=1 - FreTS), Quaternion base (p=2 - QuatMLP), Octonion base (p=4 - OctMLP), and Sedin base (p=8 - SedMLP).

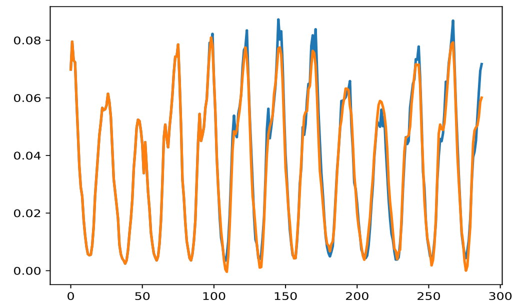
		Traffic				**ETTh1**				**ETTm1**			
	Metric	96	192	336	720	96	192	336	720	96	192	336	720
SedenionMLP ($p = 8$)	RMSE	0.0340	0.0346	0.0351	0.0363	0.0896	0.0948	0.0999	0.1047	0.0814	0.0857	0.0894	0.0977
	MAE	0.0168	0.0169	0.0173	0.0186	0.0598	0.0640	0.0685	0.0767	0.0542	0.0573	0.0609	0.0682
OctontionMLP ($p = 4$)	RMSE	0.0335	0.0343	0.0349	0.0361	0.0834	0.0874	0.0941	0.1017	0.0739	0.0831	0.0888	0.0967
	MAE	0.0166	0.0167	0.0172	0.0185	0.0579	0.0635	0.0676	0.0759	0.0496	0.0556	0.0603	0.0673
QuaternionMLP ($p = 2$)	RMSE	0.0335	0.0343	0.0350	0.0362	0.0874	0.0938	0.0997	0.1059	0.0796	0.0847	0.0887	0.0974
	MAE	0.0165	0.0167	0.0172	0.0184	0.0580	0.0633	0.0687	0.0783	0.0526	0.0564	0.0603	0.0678

Table 9: Comparison between the WM-MLP and the HC-MLP of RMSE and MAE on ETTh1, Traffic, and ETTm1 datasets

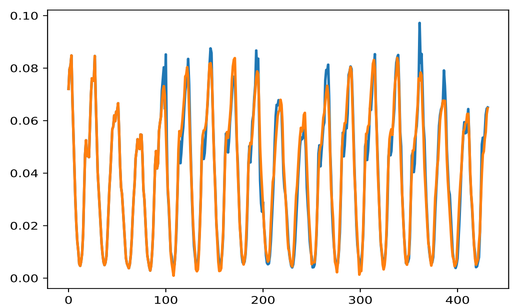
E.1 VISUALIZATIONS



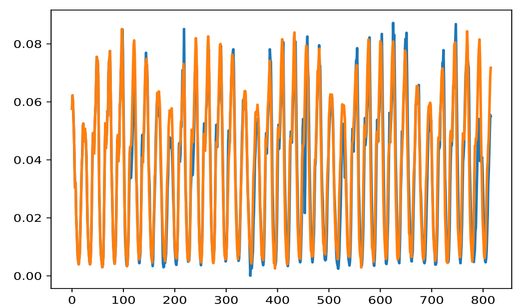
(a) Traffic I/O = 96/96



(b) Traffic I/O = 96/192

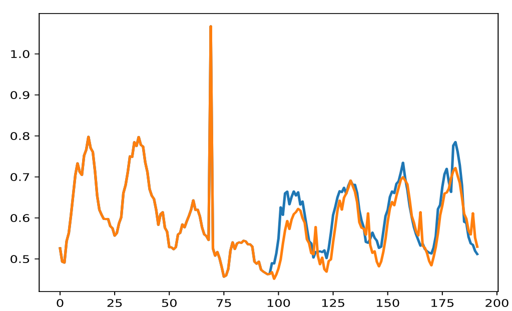


(c) Traffic I/O = 96/336

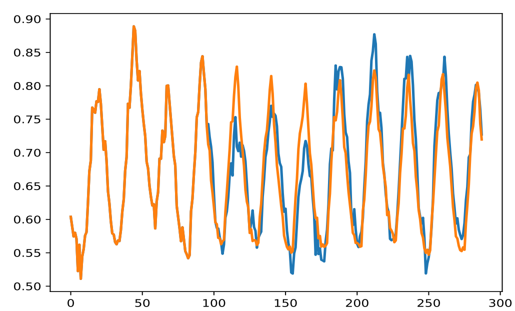


(d) Traffic I/O = 96/720

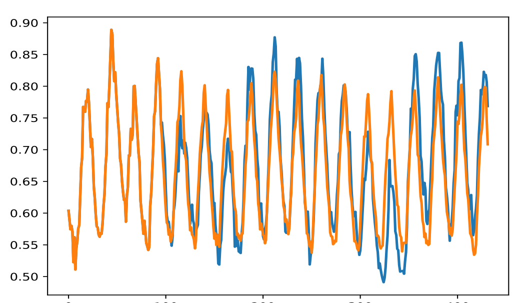
Figure 12: Ground Truth vs. Predictions for Different I/O Settings (Traffic Dataset).



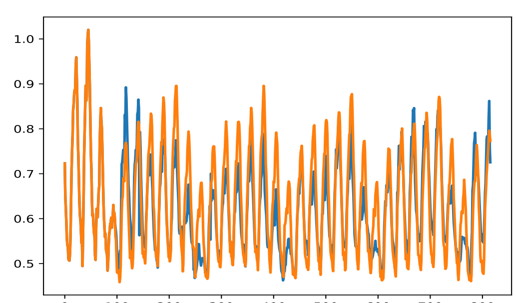
(a) Electricity I/O = 96/96



(b) Electricity I/O = 96/192



(c) Electricity I/O = 96/336



(d) Electricity I/O = 96/720

Figure 13: Ground Truth vs. Predictions for Different I/O Settings (Electricity Dataset).