Magnetic resonance spectroscopy, namely MRS, is widely used for measuring human metabolism. This technique has several advantages. For example, MRS is a noninvasive method for investigating the chemical and structural properties of molecules in vivo. It can be used to quantify the concentration of metabolites in targeted voxels. While MRS has the potential to be highly valuable in clinical practice, it poses several challenges such as low signal-to-noise ratio, overlapping metabolite signals, experimental artifacts, and long acquisition times.

For example, theoretically, the use of smaller voxels enables us to focus on more specific ROI (e.g., hippocampus). However, the use of smaller voxels requires a higher number of signal averages to be acquired to improve the SNR, and this leads to a longer scan duration ([Tognarelli et al., 2015](https://doi.org/10.1016/j.jceh.2015.10.006)). This longer scan time may cause the data accuracy to be affected due to subject motion during the in vivo MRS acquisition.

To effectively analyze spectroscopy data, we require the spectra analyzed to be robust and accurate, various considerations such as pulse sequence selection, B0 shimming, as well as preprocessing and analysis methods must be taken into account. Due to the complexity of these considerations, MRS can be a challenging technique for nonexperts to implement and oversee, hindering clinical adoption.

(100s)

The problem comes with how could we improve the data quality and enhance the overall performance. We need to review the conventional pipeline of MRS acquisition, processing and analysis. The figure here provides a schematic overview of the workflow, highlighting the human expertise involved at each step.

During data acquisition, the MR physicist or scanner operator needs to optimize the scan-configuration parameters to get the desired and optimal data output.

The processing step involves techniques that reduce the dimensionality of the data (e.g., using the Fourier transform to convert the raw data from the temporal domain to the frequency domain). Other techniques that remove spectral imperfections, improve the visual appearance of the spectra, correct the baseline or separate the macromolecule are also ([Sande et al., 2023](https://doi.org/10.1002/mrm.29793)) applied.

The MRS analysis focuses on converting the (processed) signals into meaningful and reliable metabolite concentration estimates to help the physician or surgeon make decisions.

Here, as a data scientist, from my perspective, the overall quality highly depends on the processing part, assuming that the data have already been collected properly and the raw data is available. If the data acquisition is bad, we want to fix it through some kinds of processing steps and verify it through quality assurance. If the data can not be fixed through processing, we say that the data acquisition fails and we should not use this data.

(click)

Following this idea, the deep learning approach could be involved at the very beginning of the processing part, which is reconstruction. As the spectral reconstruction lays the foundation of the spectra quality, it is very important to the following processing steps. Therefore, in this study, we aim to solve the spectroscopy inverse problems in domain transform from FIDs to spectra (e.g., Reconstruction, Spectral Denoising, etc). A unified MRS reconstruction approach is proposed in this study, as shown in this figure, which aims to reconstruct and process the FID in parallel. The framework described here is the artificial Fourier transform network, which is automatic, standard and time-saving for non-experts.

(150s all: 250s)

Here is the schematic overview of the proposed AFT-Net. The C in this figure stands for complex-valued and the R in this figure stands for real-valued. As shown in this figure, The inputs of the model are complex-valued FIDs and all the model architectures are based on complex-valued operations. The outputs of the model are real-valued spectra. We introduced two main blocks to achieve the tasks we described before. The artificial Fourier transform (or AFT) block aims to determine the mapping of the domain transformation. The residual attention complex-valued U-Net aims to extract higher features with an encoder-decoder architecture.

(40s all: 290s)

Next I want to introduce the method we used to tackle the challenges we described before. First is the idea of MRS reconstruction by domain-transform manifold learning.

The very first step of MRS processing is reconstruction, which is challenging because analytic knowledge of the exact inverse transform (e.g., Fourier transform) may not exist a prior knowledge, especially in the presence of sensor non-idealities and noise. Thus, the conventional reconstruction approach often involves approximating the inverse function in a signal-processing chain, which requires expert parameter tuning to optimize reconstruction performance. Previous work demonstrated that reconstruction could be recast as a data-driven supervised learning task that learns a reconstruction mapping between the sensor domain and the spectrum domain. As this mapping is trained, a between-manifold projection *g* from *x* (the manifold of sensor inputs) to *y* (the manifold of output spectra) is determined. A manifold mapping *ϕy* is also implicitly learned to project the spectrum from manifold *y* back to the representation in the frequency domain.

(70s all: 360s)

In this study, we used complex-valued operators to design our network, which fully leverages the richer representational capacity of complex numbers and the relationship between real parts and imaginary parts as prior knowledge. Previous studies also indicated that the complex neural network could facilitate noise-robust memory retrieval mechanisms. Another advantage of using the complex-valued neural network is pretty straightforward. A complex-valued layer has half as many parameters compared to its real-valued counterpart.

Here, we show the definition of the complex-valued operator acting on the complex number in equation 1. This form can be extended to complex matrix multiplications. In equation 2, we demonstrate some commonly used layers defined in the complex domain. Note that not all the operator is defined in this way. For example, the ReLU activation function simply applies separate ReLUs on both the real and the imaginary part of a neuron.

(60s, 420s)

The structure of the AFT-Net is displayed here. We designed the network so that it could be applied to multi-dimensional data by extending the AFT block. For the 1-dimensional MRS FID, the N is 1. Components include the complex-valued AFT block, the complex-valued residual attention UNet, the complex-valued residual block, and the complex-valued attention gate. All convolutional layers have a kernel size of 3, except those pointed out specifically. The C here stands for complex-valued. The red numbers indicate the number of channels produced by each layer. The fully connected layers approximate the between-manifold projection from the sensor domain to the spectrum domain. The convolutional layers extract high-level features from the data and force the spectra to be represented sparsely in the convolutional feature space. The architecture of the CUNet presented here is generally based on the residual attention U-Net but with all the real-valued components replaced by complex-valued components, including complex convolutional layers and complex ReLU layers. We further optimize the network for smaller batch sizes by replacing batch normalization with group normalization. One thing should be noted not all the complex-valued components are defined in the same way as we described before. For example, the complex normalization is defined in the matrix form with covariance to ensure that both real and imaginary parts have equal variance. The complex max pooling is almost the same as the real-valued version but with indices inferred from absolute values.

(120s all: 540s)

We trained our model on the MEGA-PRESS spectra from the Big GABA dataset for two reasons. First, as a proof of concept study, to guarantee the convergence of the supervised learning task, we need the dataset to be sufficient in the number of samples, good in data quality and publicly available. Thus the Big GABA dataset perfectly meets our requirements. Second, the smaller targeted signals are revealed by the subtraction of 2 spectra containing strong signals (OFF and ON) ([Mullins et al., 2012](https://doi.org/10.1016/j.neuroimage.2012.12.004)), which provide a good way to verify the performance of the proposed method by measuring the subtraction artifacts.

(40s all: 580s)

A total number of 101 subjects acquired by the Philips scanners were used in the training. For each subject, a standard GABA ON\OFF edited MRS acquisition was run, where ON editing pulses were placed at 1.9 ppm and OFF editing pulses were placed at 7.46 ppm. The acquisition number is 320 (160 ON and 160 OFF transients) per subject. The AFT-Net was trained with an input size of 2048. The ground truth of the ON/OFF spectra is derived by taking the average over 160 acquisitions ([Chen et al., 2022](https://arxiv.org/abs/2101.11442)). We denote the ground truth as noiseless signals. For the training, we combined randomly sampled acquisitions of each subject to retrieve a noisy signal. By decreasing/increasing the number of sampled acquisitions, we can generate signals with higher/lower noise. We use the reduction rate (R) to denote the level of noise, which is defined as the ratio of the total acquisition number and the number of acquisitions sampled. This quantity is very handy to assess the power of denoising methods in practical terms. Retrieving accurate denoised signals at a high R has implications for the potential reduction of total experimental time.

(80s all: 660s)

We implemented the AFT-Net unified reconstruction framework with a deep neural network feed-forward architecture composed of fully connected layers followed by a sparse convolutional autoencoder (as shown in this figure). The fully connected layers approximate the between-manifold projection from the sensor domain to the spectra domain. The convolutional layers extract high-level features from the data and force the spectra to be represented sparsely in the convolutional feature space. Our approach searches for an inverse that best represents the data in a low-dimensional feature space determined by manifold learning as well as the trained sparse convolutional filters instead of using known properties of the canonical domain transform to formulate the neural network model or performing the explicit transform before processing by a neural network.

(60s all: 720s)

The results of the AFT-Net approach and conventional numerical methods with Gaussian line broadening are illustrated in this figure. The first row shows the reconstructed spectrum from the numerical methods and the proposed AFT-Net. The second row indicates the reconstructed spectrum overlaid with the ground truth. The third row plots the difference between the reconstructed spectrum and the ground truth. Under an acceleration rate of 80, where only 2 acquisitions were used over all 160 acquisitions, the AFT-Net shows excellent performance at high acceleration factors. The AFT-Net outperforms other methods for the DIFF spectra, indicating that the AFT-Net removes the noise in the FIDs while preserving the subject-level features.

(60s all: 780s)

We used GFC to measure the similarity between the reconstructed spectra and the ground truth, as shown in the equation. The metric value increases as the reduction rate decreases, but the absolute difference between high and low acceleration rates is tiny (0.9798 for OFF spectra under an reduction rate of 10 vs. 0.9688 for OFF spectra under an reduction rate of 160). In addition, AFT-Net outperforms the DFT+GLB (Gaussian Line Broadening) method across all metrics on the right table.

(60s all: 840s)

A novel artificial Fourier transform framework is proposed that determines the mapping between FIDs and spectra as conventional DFT while having the ability to be fine-tuned/optimized with further training. The flexibility of AFT allows it to be easily incorporated into any existing deep learning network as learnable or static blocks. We then utilized AFT to design our AFT-Net, which implements complex-valued UNet to extract higher features in the temporal domain. We aim to combine reconstruction and denoising into a unified network that simultaneously enhances spectrum quality by removing artifacts directly from the FIDs. Our AFT-Net achieves competitive results and proves to be more robust to noise. One remaining limitation is that we only implement a complex-valued network with linear layers and CNNs, which are less effective than some advanced architectures. In our future work, we will aim to replace multi-layer perceptrons and CNNs with transformer-based and diffusion-based models while extending the concept of AFT to more medical spectroscopy tasks.

(120s all: 960s)

Here is a list of related works conducted by our group. We first explore the complex-valued U-Net in 2021. And we have been working for three years on the deep learning based MRS reconstruction and processing pipeline. If you are interested in the model details or mathematical derivations, you could refer to these works.

Again. I want to acknowledge the contribution of Matthieu, who has been working hard on the training and fine-tuning of the model. If you want to collaborate on any of the topic we mentioned above, or if you can provide in-vivo MRS raw data with high quality and large amount, which we are eagerly looking for the supervised learning, feel free to contact the Dr. Guo through his email which is jg3400@columbia.edu. Thanks for your listening.