DISTRIBUTED MEMORY-EFFICIENT PHYSICS-GUIDED DEEP LEARNING RECONSTRUCTION FOR LARGE-SCALE 3D NON-CARTESIAN MRI

Chi Zhang[1,2]*, Davide Piccini*[3,4]*, Omer Burak Demirel*[1,2]*, Gabriele Bonanno*[4]*, Burhaneddin Yaman*[1,2]*, Matthias Stuber*[3,5]*, Steen Moeller*[2]*, Mehmet Akçakaya*[1,2]

[1] Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, United States
[2]Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN, United States
[3]Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[4]Advanced Clinical Imaging Technology, Siemens Healthcare AG, Bern, Switzerland.
[5]Center for Biomedical Imaging, Lausanne, Switzerland.

ABSTRACT

Physics-guided deep learning (PG-DL) reconstruction has emerged as a powerful strategy for accelerated MRI. However, adopting PG-DL on 3D non-Cartesian MRI remains a challenge due to GPU hardware limitations. In this paper, we utilize multiple memory-efficient techniques to accomplish PG-DL on large-scale 3D kooshball coronary MRI. We first leverage a recently proposed approach to keep only one unrolled step on GPUs. We then utilize a Toeplitz approach to represent the multi-coil encoding operator. Subsequently, we distribute the most memory-consuming data consistency operations into multiple GPUs, enabling conjugate gradient iterations without necessitating coil compression. Finally, we employ mixed-precision training to further reduce memory consumption.The combination of these methods enable training of high-quality PG-DL reconstruction for 3D kooshball trajectories, and our results show reconstruction improvement compared to existing strategies.

***Index Terms*—** accelerated imaging, non-Cartesian MRI, deep learning, GPU, implementation

1. INTRODUCTION

Non-Cartesian MRI trajectories offer more efficient coverage of k-space, higher motion robustness, and less visually apparent aliasing artifacts when compared to Cartesian sampling [1]. In particular, 3D non-Cartesian trajectories, including kooshball [2] and 3D cones [3], can be used for efficient volumetric coverage, and have found use in multiple applications [4, 5]. Such acquisitions are conventionally reconstructed using parallel imaging [6] or compressed sensing [4, 7, 8].

Physics-guided deep learning (PG-DL) reconstruction has emerged as a powerful alternative for accelerated MRI [9–12]. In PG-DL, the multi-coil inverse problem is solved by unrolling a conventional optimization algorithm, alternating between a regularizer and data consistency (DC) term, for a fixed number of iterations [13]. PG-DL has been shown to have excellent performance and improved robustness compared to data-driven DL approaches [9–11]. The performance of PG-DL depends on several architectural components, including the depth of the neural network that implicitly performs the regularization [13], a sufficient number of unrolled iterations, and the implementation of the linear DC unit, for instance via unrolling of conjugate gradient (CG) itself [10]. These in turn translate to higher memory requirements, and insufficient GPU memory ends up being the major limitation for training of PG-DL methods [14].

This issue becomes even more challenging in the case of non-Cartesian MRI, since the non-Cartesian encoding operator require memory-consuming gridding operations on an over-sampled grid, with a factor up to 2-fold along each dimension [15]. This is reflected in the existing DL reconstruction methods for non-Cartesian MRI, where majority of studies concentrate on data-driven networks [16, 17], or PG-DL at low-resolutions [18] or with simulated 2D single coil data [19]. In particular, PG-DL reconstruction of high-resolution 3D non-Cartesian datasets has remained elusive.

In this study, we build on recent developments to enable distributed memory-efficient learning (MEL) for high-resolution 3D non-Cartesian acquisitions. We combine the MEL idea proposed in [14] with the Toeplitz method for non-Cartesian trajectories [20, 21] that allows the solution of the DC problem without memory-consuming gridding-regridding operations. This enables us to distribute the learning task over multiple GPUs to accommodate a DC unit using all coils and sufficient iterations in CG, as well as a deep CNN regularizer. Finally, to further extend the network depth, we employ mixed-precision training for lowering memory usage without a substantial loss in accuracy [22]. The proposed method is trained and tested on 3D kooshball coronary MRI at different undersampling rates, and compared to conventional reconstructions. Our results suggest PG-DL is a potentially powerful tool for accelerating 3D non-Cartesian MRI.

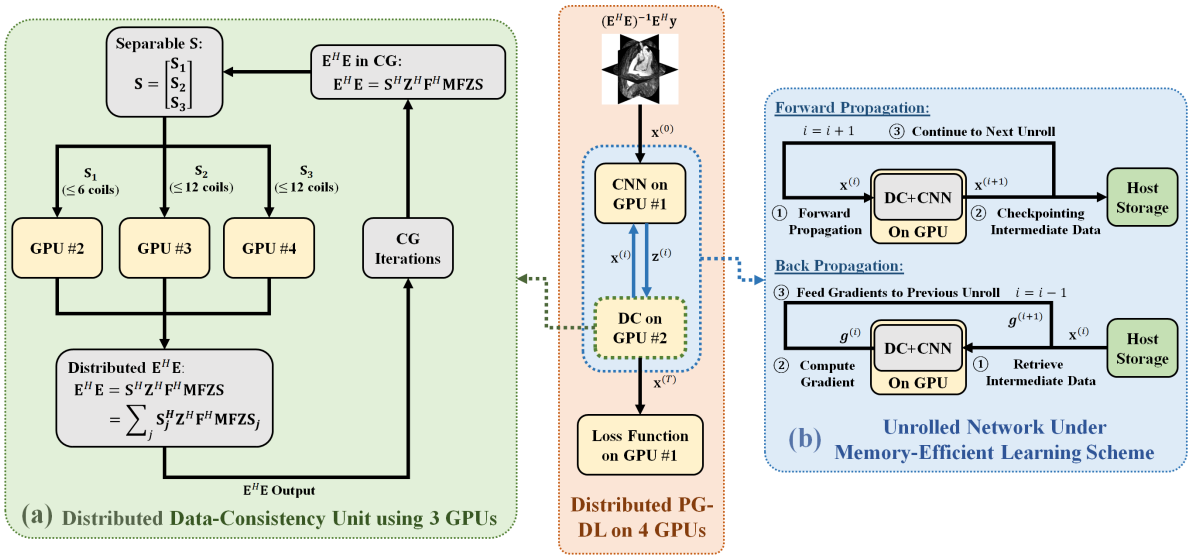

Fig. 1: (a) Distributed MEL using 4 GPUs. The data consistency (DC) unit is distributed using 3 GPUs, while the CNN regularizer is processed with an individual GPU. (b) MEL for PG-DL [14]. A single unrolled step will be allocated on device. Intermediate data are saved (checkpoingting) on the host memory during forward propagation. Backpropagation gradients are computed using preserved intermediate data, and accumulated through unrolled steps.

2. METHODS

2.1. 3D Kooshball Coronary MRI Data

Nine 3D non-Cartesian coronary MRI datasets were acquired on a Siemens Magnetom Aera 1.5T scanner using an ECG triggered T2-prepared, fatsaturated, navigator-gated prototype bSSFP sequence, with relevant parameters: resolution = $1.15 \times 1.15 \times 1.15$ mm^3, matrix size = $384 \times 384 \times 384$, FOV = $440 \times 440 \times 440$mm^3 with 2-fold readout oversampling. A total of 12320 radial projections (sub-Nyquist rate of 5) were acquired in 385 heartbeats with the spiral phyllotaxis pattern [23] with one interleaf of 32 projections per heartbeat. Number of coils in these datasets were between 20 and 30.

2.2. Physics-Guided Deep-Learning Reconstruction

PG-DL solves the objective function of a regularized multi-coil reconstruction model

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{E}\mathbf{x}||_2^2 + \mathcal{R}(\mathbf{x}), \tag{1}$$

where $\mathbf{E}$ is the multi-coil encoding operator, $\mathbf{x}$ is the image to be reconstructed, $\mathbf{y}$ is the acquired k-space data, and $\mathcal{R}(\cdot)$ is a regularization term. Optimization techniques such as variable splitting with quadratic penalty [9, 10, 12] can be used to solve this objective function via

$$\mathbf{z}^{(i)} = \arg\min_{\mathbf{z}} \ \mu||\mathbf{x}^{(i-1)} - \mathbf{z}||_2^2 + \mathcal{R}(\mathbf{z}) \tag{2}$$

$$\begin{aligned}\mathbf{x}^{(i)} &= \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{E}\mathbf{x}||_2^2 + \mu||\mathbf{x} - \mathbf{z}^{(i)}||_2^2 \\ &= (\mathbf{E}^H\mathbf{E} + \mu\mathbf{I})^{-1}(\mathbf{E}^H\mathbf{y} + \mu\mathbf{z}^{(i)})\end{aligned} \tag{3}$$

where μ is a trainable scalar, (2) is implicitly solved with a neural network, and the DC term in (3) is solved using the conjugate gradient (CG) method [10].

2.3. Proposed Efficient Learning Strategies for Large-scale 3D Non-Cartesian MRI

2.3.1. Toeplitz Method for the DC Term

Since the DC term in (3) requires the calculation of $\mathbf{E}^H\mathbf{y}$ only once, but necessitates repeated application of $\mathbf{E}^H\mathbf{E}$ throughout iterations, it is desirable to avoid using gridding-regridding operations to implement the latter. To this end, a Toeplitz method can be used to efficiently represent $\mathbf{E}^H\mathbf{E}$ as [20, 21]:

$$\mathbf{E}^H\mathbf{E} = \mathbf{S}^H\mathbf{Z}^H\mathbf{F}^H\mathbf{M}\mathbf{F}\mathbf{Z}\mathbf{S} \tag{4}$$

where $\mathbf{S}$ denotes the coil sensitivities, $\mathbf{Z}$ denotes zero-padding in image domain into double the matrix size along each dimension, $\mathbf{Z}^H$ is cropping to the original FOV, $\mathbf{F}$ is the Cartesian FFT over this enlarged FOV. The trajectory-dependent operator $\mathbf{M}$ is obtained by running an impulse through $\mathbf{E}^H\mathbf{E}$ with or without appropriate density compensation weights for the given trajectory [21]. Note that this Toeplitz method performs non-Cartesian gridding-regridding operations by point-wise multiplications and other fast operators instead, which is extremely beneficial to reduce memory consumption during backpropagation.

2.3.2. *Memory-Efficient Learning*

In the MEL scheme for PG-DL [14], intermediate outputs from each unrolled step are preserved on the host memory during forward propagation (referred to as checkpointing in [14]), and backpropagation gradients are computed one-by-one through all the unrolled steps, using the preserved intermediate data and gradients from the previous unroll (Fig. 1b). Conceptually, this strategy supports an unlimited amount of unrolled steps. The only drawback is that it requires additional data transferring between devices and the host, leading to trade-offs between GPU memory and processing speed.

2.3.3. *Distributed Processing*

Note MEL [14] assumes a single unrolled step is able to be fit into memory. However, this may still be difficult in the case of multi-coil 3D non-Cartesian MRI, especially for the DC units. Instead of compromising on data quality or network depth, we propose to use distributed learning (Fig. 1a). Noting that (4) is separable across the coil sensitivity profiles $\mathbf{S}$, and that $\mathbf{M}$ depends only on the trajectory, it is feasible to distribute $\mathbf{E}^H\mathbf{E}$ into different devices, where each device handles part of $\mathbf{E}^H\mathbf{E}$ corresponding to a subset of $\mathbf{S}$. In this work we utilize three GPUs for the DC unit. One among the three is chosen as the base device, which is responsible for the overall CG steps, as well as part of the $\mathbf{E}^H\mathbf{E}$ operation with ≤ 6 coils. The rest of the coils are evenly distributed on the other two GPUs, which only handle $\mathbf{E}^H\mathbf{E}$ operation, and transfer the results to the base device. To maximize the CNN regularizer depth, an individual GPU is allocated to process this CNN unit. Both forward and backpropagations are distributed in the same manner.

2.3.4. *Mixed Precision Processing*

Although distributed MEL further reduces memory occupation on a single device, our experiments suggest that these efforts are still not enough to support a 3D CNN of sufficient depth. Thus we further exploit mixed-precision processing in CNN training, where majority of the computations in CNN, such as convolutions are processed in half-precision (float16 and complex32). This leads to approximately 50% less memory consumption without a noticeable loss on accuracy [22].

3. EXPERIMENTS AND RESULTS

3.1. Implementation Details

Prior to any processing, 40% of oversampling was removed to reduce the matrix size to $224 \times 224 \times 224$. The datasets were retrospectively further undersampled by rates (R) of 4, 5 and 6. For all undersampling patterns, $\mathbf{M}$ is obtained by gridding density compensation weights [24] using Kaiser-Bessel NUFFT [25] with an oversampling ratio 2. $\mathbf{E}^H\mathbf{E}$ operations are density-compensated for maximum convergence rate in CG [21]. Training labels were generated using the CG-SENSE reconstruction of the acquired fully-sampled data. Coil maps were estimated from central $20 \times 20 \times 20$ Nyquist rate-sampled region of k-space. 6 datasets were used for training and 3 for testing. All implementations used Pytorch 1.9, with APEX 0.1.0 for mixed-precision processing, on 4 NVIDIA A100 GPUs (40GB memory each). 10 unrolled steps were used in the PG-DL network. Linear data-consistency was solved using 9 CG iterations. The ResNet in [12] is employed as the CNN regularizer, but with 3D convolutions accordingly for 3D kooshball data. Adam optimizer with learning rate of $3 \cdot 10^{-3}$ was used for network training. Training with 100 epochs took around 28 hours.

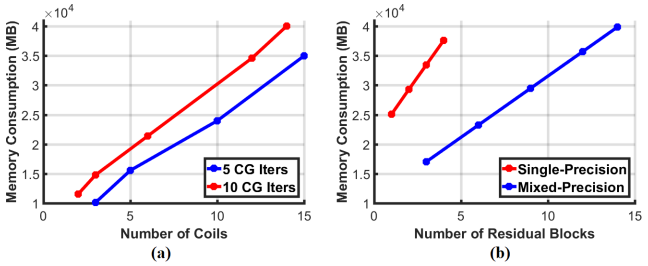

Fig. 2: Memory consumption of (a) data consistency (DC) unit handling different number of coils under 5 and 10 CG iterations; (b) CNN (ResNet) regularizer, using different number of residual blocks under single and mixed-precision.

3.2. Memory Consumption

Fig. 2a depicts the single GPU memory occupation of a data-consistency unit with different amounts of coils and CG iterations. In PG-DL, the CG iteration is commonly chosen between 5 to 10. Fig.2b presents the single GPU memory occupation of a CNN (ResNet) regularizer using different number of residual blocks. Using single-precision, an A100 GPU supports up to 4 residual blocks in 3D ResNet, which is far from sufficient depth. By using mixed-precision, this limit has been pushed to up to 14 residual blocks. These results suggest that even using highly compressed coils, MEL alone still cannot support a reasonable PG-DL for such large-scaled kooshball data. Using distributed learning and mixed-precision techniques enable PG-DL to have sufficient number of CG iterations and CNN depth.

3.3. PG-DL Reconstruction Results

Fig. 3 and 4 show reconstruction results for retrospective undersampling rate of 6 (30-fold sub-Nyquist acceleration), using CG-SENSE, Tikhonov-regularized CG-SENSE, compressed sensing using ℓ_1 regularization of Daubechies4 wavelets, and PG-DL, along with the fully sampled reference images. For all rates, CG-SENSE suffers from noise amplification and artifacts, which are reduced with Tikhonov

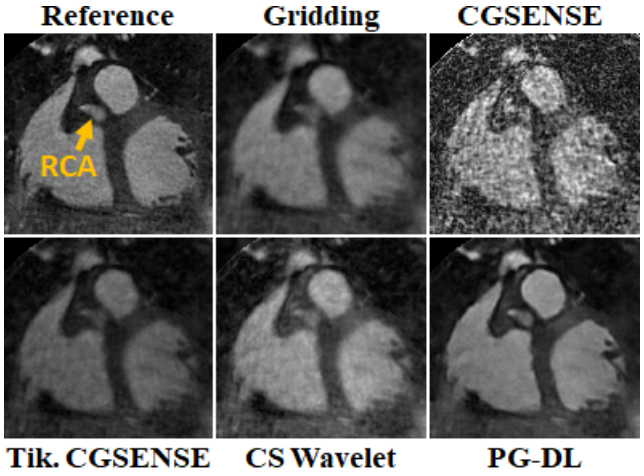

Fig. 3: Representative coronal slices of 3D kooshball coronary MRI at a retrospective acceleration rate of 6 (sub-Nyquist rate of 30). CG-SENSE suffers from noise amplification, which is reduced by Tikhonov regularization. Compressed sensing further improves image quality. PG-DL visibly outperforms other approaches in terms of noise and artifacts removal.

regularization. Compressed sensing further improves image quality, but still shows residual streaking artifacts. PG-DL outperforms other approaches in terms of noise reduction and artifact removal, with recovery of fine structures, such as the coronary arteries.

4. CONCLUSIONS

In this work, we enabled PG-DL reconstruction for high-resolution 3D large-scale non-Cartesian coronary MRI. This is achieved by utilizing multiple memory-efficient techniques. Our results suggest PG-DL can offer promising image quality in 3D coronary MRI, compared to conventional reconstruction approaches, which are susceptible to noise and residual artifacts. In turn, this may facilitate the use of highly-undersampled 3D non-Cartesian acquisitions in different applications.

5. ACKNOWLEDGMENTS

This work was partially supported by NIH R01HL153146, NIH P41EB027061, NIH R21EB028369, NSF CAREER CCF-1651825.

6. COMPLIANCE WITH ETHICAL STANDARDS

The research study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Institutional Review Board of University of Lausanne.

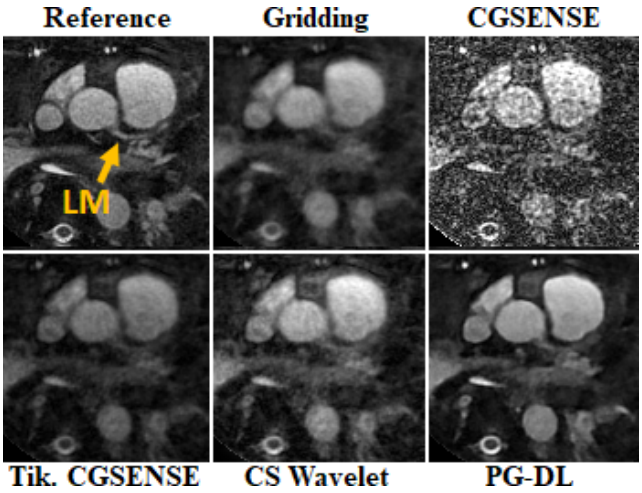

Fig. 4: Representative axial slices of 3D kooshball coronary MRI at R = 6. PG-DL visibly outperforms alternative methods, providing a clear delineation of the left coronary arteries.

References

[1] K. L. Wright, J. I. Hamilton, et al., "Non-Cartesian parallel imaging reconstruction," *J Magn Reson Imaging*, vol. 40, no. 5, pp. 1022–1040, Nov 2014.

[2] C. Stehning, P. Börnert, et al., "Fast isotropic volumetric coronary MR angiography using free-breathing 3D radial balanced FFE acquisition," *Magn Reson Med*, vol. 52, no. 1, pp. 197–203, Jul 2004.

[3] P. T. Gurney, B. A. Hargreaves, and D. G. Nishimura, "Design and analysis of a practical 3D cones trajectory," *Magn Reson Med*, vol. 55, pp. 575–582, Mar 2006.

[4] L. Feng, R. Grimm, et al., "Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magn Reson Med*, vol. 72, no. 3, pp. 707–717, Sep 2014.

[5] A. V. Barger, W. F. Block, et al., "Time-resolved contrast-enhanced imaging with isotropic resolution and broad coverage using an undersampled 3D projection trajectory," *Magn Reson Med*, vol. 48, no. 2, pp. 297–305, Aug 2002.

[6] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn Reson Med*, vol. 46, no. 4, pp. 638–651, Oct 2001.

[7] S. Nam, M. Akçakaya, et al., "Compressed sensing reconstruction for whole-heart imaging with 3D radial trajectories: a graphics processing unit implementation," *Magn Reson Med*, vol. 69, no. 1, pp. 91–102, Jan 2013.

[8] M. Akçakaya, S. Nam, et al., "An augmented Lagrangian based compressed sensing reconstruction for non-Cartesian magnetic resonance imaging without gridding and regridding at every iteration," *PLoS One*,

vol. 9, no. 9, pp. e107107, 2014.

[9] K. Hammernik, T. Klatzer, et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magn Reson Med*, vol. 79, pp. 3055–3071, 2018.

[10] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans Med Imag*, vol. 38, no. 2, pp. 394–405, 02 2019.

[11] S. A. H. Hosseini, B. Yaman, et al., "Dense recurrent neural networks for accelerated MRI: history-cognizant unrolling of optimization algorithms," *IEEE J Sel Top Sig Proc*, vol. 14, no. 6, pp. 1280–1291, Oct 2020.

[12] B. Yaman, S. A. H. Hosseini, et al., "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magn Reson Med*, vol. 84, no. 6, pp. 3172–3191, 12 2020.

[13] F. Knoll, K. Hammernik, et al., "Deep-learning methods for parallel magnetic resonance imaging reconstruction: a survey of the current approaches, trends, and issues," *IEEE Sig Proc Mag*, vol. 37, pp. 128–140, Jan 2020.

[14] M. Kellman, K. Zhang, et al., "Memory-efficient learning for large-scale computational imaging," *IEEE IEEE Trans Comp Imag*, vol. 6, pp. 1403–1414, 2020.

[15] P. J. Beatty, D. G. Nishimura, and J. M. Pauly, "Rapid gridding reconstruction with a minimal oversampling ratio," *IEEE Trans Med Imag*, vol. 24, no. 6, pp. 799–808, Jun 2005.

[16] H. El-Rewaidy, A. S. Fahmy, et al., "Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI," *Magn Reson Med*, vol. 85, no. 3, pp. 1195–1208, 03 2021.

[17] L. Fan, D. Shen, et al., "Rapid dealiasing of undersampled, non-Cartesian cardiac perfusion images using U-net," *NMR Biomed*, vol. 33, no. 5, pp. e4239, 05 2020.

[18] M. O. Malavé, C. A. Baron, et al., "Reconstruction of undersampled 3D non-Cartesian image-based navigators for coronary MRA using an unrolled deep learning model," *Magn Reson Med*, vol. 84, pp. 800–812, 2020.

[19] Z. Ramzi, J. Starck, and P. Ciuciu, "Density compensated unrolled networks for non-cartesian MRI reconstruction," in *Proc IEEE ISBI*, 2021, pp. 1443–1447.

[20] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans Med Imag*, vol. 30, no. 3, pp. 694–706, Mar 2011.

[21] C. A. Baron, N. Dwork, J. M. Pauly, and D. G. Nishimura, "Rapid compressed sensing reconstruction of 3D non-Cartesian MRI," *Magn Reson Med*, vol. 79, no. 5, pp. 2685–2692, 05 2018.

[22] P. Micikevicius, S. Narang, et al., "Mixed precision training," *arXiv preprint arXiv:1710.03740*, 2017.

[23] D. Piccini, A. Littmann, S. Nielles-Vallespin, and M. O. Zenge, "Spiral phyllotaxis: the natural way to construct a 3D radial trajectory in MRI," *Magn Reson Med*, vol. 66, no. 4, pp. 1049–1056, Oct 2011.

[24] J. G. Pipe and P. Menon, "Sampling density compensation in MRI: rationale and an iterative numerical solution," *Magn Reson Med*, vol. 41, pp. 179–186, 1999.

[25] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans Sig Proc*, vol. 51, no. 2, pp. 560–574, 2003.