Research Statement

Xiao Wang

1 My Research Topic and Significance

My research interest is to apply high performance computing, machine learning, image processing, and medical physics to image reconstruction applications, such as computed tomography (CT), single emission positron emission tomography (SPECT) and magnetic resonance imaging (MRI). Among different imaging applications, my primary focus at the moment is CT reconstruction. More specifically, **my research focuses on increasing reconstruction spatial resolution and reducing artifacts as well as image noise all at the same time without sacrificing diagnostic values. In addition, my research develops high performance and parallel algorithms to achieve real-time reconstructions for time-sensitive CT imaging.**

2 Background and Summary of My Contributions

To reconstruct from the CT projections to clinical images, Filtered Back-Projection (FBP) is the dominant approach by performing Fourier and inverse Fourier transform and interpolates discretely sampled projections back to a reconstruction in the continuous spatial domain. However, an exact FBP implementation cannot fully model the complexity of the state-of-the-art CT scanner geometry, such as the dual-source and flying focal spot Siemens Force scanner. To simplify geometry and save computations, the existing clinical FBP methods are all approximate techniques that rebin the helical projections into parallel-beams for reconstruction. A clear advantage of the approximate FBP methods is that they enable a short reconstruction time. The disadvantage is that these approximate FBP methods may produce undesirable cone-beam artifacts when the cone angle is large [4, 1]. In addition, both the exact and the approximate FBP methods are originally derived under the assumption of continuously sampled projections, although the actual detector projections are all discretely sampled. Therefore data interpolation is heavily used in the FBP methods. When the projections are not sufficiently many, such as with a high pitch, data interpolation may limit the achievable spatial resolution and cause undersampling artifacts [4].

In contrast, statistical iterative reconstruction, also known as model-based iterative reconstruction (MBIR), is based on linear algebra and Bayesian estimation, and formulates each projection as a weighted linear summation of voxel intensities along the path of the X-rays. A unique advantage of MBIR is that it requires no data interpolation or rebinning and operates directly on discrete measurements [5]. In addition, MBIR has great flexibility to incorporate precise scanner hardware characteristics into forward geometry model [4]. Therefore, MBIR has the potential to be more faithful to the true acquisition physics and geometry than FBP. Because of these distinct advantages, MBIR often produces clearer image details with fewer artifacts than FBP, especially when the radiation doses are low and the number of projections is limited [4].

MBIR, however, has several key technical issues. (1) The number of operations for MBIR is several magnitudes more than those for FBP [5, 1]. Therefore, MBIR has a slow reconstruction time and is unpopular for clinical practice. Furthermore, MBIR exhibits poor cache locality as the data access layout for each voxel requires an array of memory to be accessed in sinusoidal patterns, and thus leads to a poor per-core efficiency. (2) the existing implementation for MBIR is only applicable to single source CT and there is no implementation for advanced scanner geometry, such as dual-source flying focal spot CT of Siemens. (3) The realistic but complex forward geometry model of MBIR has a very large memory requirement to store its physics-geometry information, whose size grows linearly with image size. Limited by its memory footprint, MBIR often cannot fully show its super-resolution capability. Any naive way to shrink or distribute the forward model among nodes corrupts the physics-geometry information.

To address the above issues, my research work made the following contributions:

1. Introduced the idea of super-voxel (SV) to significantly increase cache locality, prefetching, and completely regularize data access for improved SIMD operations [5, 2] (Project 1).
2. Proposed a parallel SV algorithm for fully 3D reconstruction, and achieves massive parallelism while achieving a convergence rate faster than other numerical algorithms [6] (Project 2).
3. Proposed an Asynchronous Consensus MBIR (AC-MBIR) that significantly reduces memory footprint and increases image super-resolution while keeping all the physics-geometry information intact [7] (Project 3).
4. Developed a Consensus Equilibrium framework that allows MBIR to use advanced but non-differentiable denoising filters, such as Non-Local Mean (NLM) and Block-Matching 3D Filtering (BM3D), as a prior model. With these advanced prior models, MBIR can better denoise at low-frequency with an image texture similar to that of the FBP methods [3] (Project 4).
5. Developed the first physics based statistical iterative reconstruction algorithm for Dual-Source Flying Focal Spot CT, such as Siemens Force Scanner. The proposed solution takes advantage of complete knowledge of the CT geometry and X-ray acquisition physics, and outperforms the state-of-the-art clinical method in terms of spatial resolution and artifacts elimination [8] (Project 5).

On the practical side, my collaborators and I have developed an open-source github project for the work in Projects 1 & 2 (`https://github.com/cabouman/svmbir`) and our users include scientists from Los Alamo, Lawrence Berkeley and Oak Ridge National Labs. We have also implemented a cloud-based MBIR solution, Inversion Engine, that allow users to run high performance MBIR over the internet. As to the MBIR clinical solution for dual-source CT from project 5, the image quality has brought attention from major CT vendors, such as Siemens, to investigate how to port the MBIR solution to their next-generation commercial CT products. For each contribution and related project mentioned above, here is a brief introduction for each:

Project 1: High Performance and Cache-Friendly MBIR [5, 2].
MBIR produces higher quality images and allows for the use of more general CT scanner geometries than is possible with more commonly used methods. The high computational cost of MBIR, however, often makes it impractical in applications for which it would otherwise be ideal. This project describes a new MBIR implementation that significantly reduces the computational cost of MBIR while retaining its benefits. It proposes a novel organization of the scanner data into super-voxels (SV) that, combined with a super-voxel buffer (SVB), dramatically increase cache locality, prefetching, SIMD operations, and enable parallelism across SVs with an average speedup of 187 times on 20 CPU cores compared to the conventional MBIR.

A separate endeavor of this project develops the first GPU algorithm for Iterative Coordinate Descent based MBIR. The algorithm leverages the SV design, and efficiently exploits three levels of parallelism available in MBIR to better utilize the GPU hardware resources. This project also transforms the CT signal data layout to obtain more coalesced accesses with several GPU-specific optimizations to boost performance. Across 3200 testing datasets, our GPU implementation obtains a mean speedup of 4.43 times over the iso-power CPU implementation.

Project 2: Fast Converging and Massively Parallel MBIR [6]. Despite of its high performance, project 1 is a single node implementation and restricts its parallelism to shared-memory CPU or GPU cores, and cannot scale to a large cluster with distributed memory. Furthermore, implementations for MBIR have a trade-off among parallelism, cache locality and algorithmic convergence. This project proposes a massively parallel MBIR that ensures a fast convergence through non-homogeneous updates, but also keeps the high cache locality benefits from the SV design. The massively parallel MBIR is compared with two state-of-the-art MBIR implementations on a 69632-core distributed system for both synchrotron and security datasets. Results indicate that the massively parallel MBIR has an average speedup of 1665 compared to the fastest state-of-the-art implementations. The huge speedup achieved in this project has received attention from ACM award committee and this work was selected as a *finalist for 2017 ACM Gordon Bell Prize.*

Project 3: Super-Resolution MBIR with Minimal Memory Footprint [7]. MBIR has a high memory requirement that limits the achievable image resolution. Because of the memory requirement, MBIR often has a

limited super-resolution capability. This project proposes Asynchronous Consensus MBIR (AC-MBIR) that uses Consensus Equilibrium (CE) to provide a super-resolution algorithm with a small memory footprint, low communication overhead and a high scalability. Super-resolution experiments on large synchrotron datasets show that AC-MBIR has a 6.8 times smaller memory footprint and 16 times more scalability, compared with project 2's massively parallel MBIR, and maintains a 100% strong scaling efficiency at 146880 cores. In addition, AC-MBIR achieves an average bandwidth of 3.5 petabytes per second at 587520 cores.

Project 4: Plug-and-Play MBIR with Non-Differentiable Prior Model [3]. The conventional MBIR prior model estimates a voxel's value based on its neighboring voxels. The local-neighbor prior model, however, is a low-pass filter and does not denoise effectively at low frequency and significantly changes the noise texture. In contrast, advanced denoisers such as Non-Local Mean (NLM) or Block Matching 3D Filtering (BM3D) not only learn from local neighborhoods, but also from remote voxels, and with much better low frequency denoising capability. Unfortunately, these denoisers are non-differentiable and cannot be easily fused together with the forward model and solved as an optimization problem. This project uses a Consensus Equilibrium (CE) Method to fuse non-differnetiable denoiser as a prior model for MBIR. Experiments on security screening datasets show that MBIR with a BM3D prior model improves image spatial resolution, with less noise and fewer artifacts.

Project 5: Physics Based Statistical Reconstruction for Dual-Source and Flying-Focal Spot CT [8]. The newest CT geometry, such as Siemens Force, has dual-source and flying-focal spot (DS-FFS CT) and features a much shorter patient scan time and a much higher sampling rate. With the DS-FFS CT design, the temporal resolution for cardiac imaging is much higher and sedating patients can possibly be avoided. Unfortunately, DS-FFS CT scanners have no solution for statistical iterative reconstruction currently. This project aims to fill in this gap by proposing novel physics models to reconstruct from the native cone-beam geometry and interleaved dual source helical trajectory of a DS-FFS CT. To do so, we construct a noise model to represent data acquisition noise and a prior image model to represent image noise and texture. In addition, we design forward geometry physics models to compute the locations for deflected focal spots, the dimension and sensitivity for voxels and detector units, as well as the length of intersection between X-rays and voxels. The forward geometry models further represent the coordinated movement between the dual sources by computing their X-ray coverage gaps and overlaps at an arbitrary helical pitch. Preliminary data on phantom and thoracic scans show that the proposed physics-based reconstruction has a much higher spatial resolution than the state-of-the-art clinical methods while having fewer image undersampling artifacts.

3 My Vision for Future Work

The 2019 novel coronavirus (SARS-cov-2) is a highly contagious illness with a wide range of symptoms and clinical evidence shows that 94% discharged patients demonstrate lasting residual lung injury from the disease. Two kinds of the most common lung injury from discharged COVID-19 patients are pulmonary embolism and lung consolidation, with 23% of all COVID-19 patients diagnosed with pulmonary embolism and nearly all discharged patients showing lung consolidation. The detection and monitoring of embolism and lung consolidation, however, have two issues. First, the detection of pulmonary embolism is challenging, with a high rate of delayed diagnosis and misdiagnosis because emboli are difficult to see on CT due to their small sizes and limited contrast. Second, the CT lung images from current clinical practice poorly quantify fluid accumulation and air in the lungs for lung consolidation, as the mass attenuation coefficient for water fluid is highly similar to that for the surrounding soft tissues. Without precise fluid density, radiologists rely on qualitative findings instead of quantitative measures to estimate patients' recovery progress on lung consolidation. In my vision for future work, I will (1) develop a physics-based MBIR reconstruction method for dual-energy Computed Tomography Angiography (CTA) emboli detection with enhanced spatial resolution and reduced image artifacts. The new image reconstruction will facilitate the localization of pulmonary emboli that arise from COVID-19 associated intravascular coagulation. (2) I will enable for the first time an effective *multi-material* decomposition method to quantify blood, fluid and air in lungs for COVID patients. This new decomposition capability will offer a quantitative biomarker for blood, fluid and air to characterize lung consolidation and ground glass opacities, and thereby provide a means of monitoring COVID

therapy response.

1: Model dual-energy geometry and X-ray acquisition to improve resolution and reduce image streaking artifacts. Project 5 from Sec. 2 does not model the X-ray acquisitions at different energies. Therefore, this aim proposes to develop and evaluate a dual-energy physics model for CTA imaging that represents the true acquisition physics of dual-energy. To do so, this physics model accounts for the different tube current, peak voltage, noise floor bowtie filtering and energy spectra between the two energy sources, so that reconstructed images can have optimal streaking and blurring artifact reduction. The *working hypothesis* is that the dual-energy physics based reconstruction from this aim eliminate image blurring and streaking artifacts, thereby improving emboli detection rate.

2: Design a multi-material decomposition method that quantifies the distribution and concentration of fluid, soft tissues, bones, iodine contrast and air. The clinical CT biomarker characterizes ground glass opacities and lung consolidations by performing material decomposition for at most three materials, but poorly quantifies their concentrations. This aim will develop a multi-material decomposition capability and uses the physics model from Aim 1 to distinguish and quantify material densities of fluid, bones, iodine, soft tissues, and air. The *working hypothesis* is that the new multi-material decomposition algorithm can accurately decompose and measure the concentration and distribution for the five materials, thereby providing a means for monitoring COVID therapy response.

3: Adaptive deep learning priors for dual-energy MBIR decomposition. An issue with multi-material decomposition is that the decomposed images can sometimes have significant amount of image noise and cross-talk artifacts, where one material appears in the image of another material. To achieve better image noising and cross-talk artifact reduction, we propose to integrate dual-energy MBIR decomposition with deep learning prior models in a Bayesian estimation framework. The forward model of this Bayesian estimation framework is the physics-based decomposition work from Aims 1 and 2. The prior models are a set of deep learning convolution neural network (CNN) denoisers. This set of CNN deep learning prior models are trained with different noise variances and each CNN prior model is weighted based on the similarity between its training noise variance and the actual noise in the decomposed images. When image noise in the decomposed images is significant, more prior model weights are given to CNN priors trained with large noise variance and fewer weights are given to priors trained with small noise variance. Vice versa with little image noise, more weights are given to CNN prior models trained with small noise variance. With adaptive choices of CNN prior models, image noise and cross-talk artifacts can be decimated, while over-denoising and loss of image details can possibly be prevented too.

References

[1] J. Hsieh, B. Nett, Z. Yu, K. Sauer, J.-B. Thibault, and C. A. Bouman. Recent advances in ct image reconstruction. *Current Radiology Reports*, 1(1):39–51, 2013.

[2] A. Sabne, X. Wang, S. J. Kisner, C. A. Bouman, A. Raghunathan, and S. P. Midkiff. Model-based iterative ct image reconstruction on gpus. In *Proceedings of the 22Nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, PPoPP '17, pages 207–220, New York, NY, USA, 2017. ACM.

[3] V. Sridhar, X.Wang, G. Buzzard, and C. A. Bouman. Distributed memory framework for fast iterative ct reconstruction using multi-agent consensus equilibrium. Under review at *IEEE Transaction on Computational Imaging*. Preprint availabe at ArXiv, November 2019.

[4] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical ct. *Medical Physics*, 34(11), 2007.

[5] X. Wang, A. Sabne, S. J. Kisner, A. Raghunathan, C. A. Bouman, and S. P. Midkiff. High performance model based image reconstruction. In *Proceedings of the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, PPoPP '16, pages 2:1–2:12, New York, NY, USA, 2016. ACM.

[6] X. Wang, A. Sabne, P. Sakdhnagool, S. J. Kisner, C. A. Bouman, and S. P. Midkiff. Massively parallel 3d image reconstruction. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '17, pages 3:1–3:12, New York, NY, USA, 2017. ACM.

[7] X. Wang, V. Sridhar, Z.Ronaghi, R. Thomas, J. Deslippe, D. Dilworth, G. Buzzard, S. Midkiff, C. Bouman, and S. Warfield. Consensus equilibrium framework for super-resolution and extreme-scale ct reconstruction. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '19, New York, NY, USA, 2019. ACM.

[8] X.Wang, R. MacDougall, C. A. Bouman, and S. K. Warfield. Model-based iterative reconstruction for dual-source and flying-focal spot computed tomography. Under review at *AAPM Medical Physics Journal*. Preprint availabe at ArXiv, November 2019.