A 219mW Image Signal Processor for Portable MRI Systems

Yueh-Feng Tsai and Chia-Hsiang Yang
Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan

***Abstract*—This work presents the *first* dedicated image signal processor for portable MRI systems. The proposed design is jointly optimized across algorithm and architecture based on the CSR-PERT algorithm. At the algorithm level, convergence-guided early termination avoids redundant iterations in fast iterative shrinkage thresholding (FIST) and adaptive grid selection (AGS), reducing the computational complexity by 73%. At the architecture level, a sample fetcher supports pointer-based and movement-aware memory access, reducing memory access by 99%. A non-uniform fast Fourier transform (NUFFT) engine with parallel multiply-accumulate (MAC) units is optimized to improve the utilization by 23%. A regularization engine that includes parallel coordinate rotation digital computers (CORDICs) enables to reduce the latency by 50%. Fabricated in 40nm CMOS, the proposed processor dissipates 219 mW at a clock frequency of 206 MHz. The chip achieves a throughput of 3.1 frame/s in constructing 128×128 MRI images. Compared with a high-end CPU, this work achieves an 8.2×10^2 higher throughput with 1.6×10^5 higher energy efficiency.**

I. Introduction

Magnetic resonance imaging (MRI) is a widely used technology for medical diagnosis, particularly effective for soft-tissue imaging. As illustrated in Fig. 1, in an MRI system, hydrogen atoms (^{1}H) are aligned by a time-varying external magnetic field. These atoms are then activated by a radio-frequency (RF) pulse during the slice-selection phase, and the induced signals from activated ^{1}H are sampled multiple times during the frequency-encoding phase. Sequential samples form a trajectory in the spatial-frequency domain (k-space), and multiple trajectories are collected by repeating the RF pulses. These trajectories constitute a spectral map, and that spectral map is utilized to construct an image. Interest in portable MRI systems has surged, as their compact form factor enhances diagnostic accessibility [1], [2]. Such technology provides essential imaging for critical patients and remote populations.

For portable MRI with a low-field magnet, signal quality usually drops [3], [4] and prolonged sampling is required. Also, a lower magnetic-field slew rate (due to limited power supply) leads to trajectory deviations [5], thereby causing artifacts. Compressive sensing (CS)-based MRI algorithms [6]–[8] can address these issues, as illustrated in Fig. 2. These algorithms construct images by leveraging compressive sensing to reduce the sampling time. Meanwhile, they eliminate the effect of trajectory deviations by deviation estimation. However, their computational complexity is excessively high. It takes over an hour to construct a 128×128 image with the best quality (evaluated by peak signal-to-noise ratio (PSNR)) using the CSR-PERT algorithm on a high-end CPU [6]. This

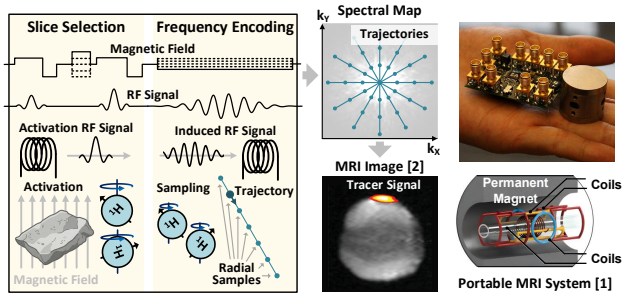


Fig. 1: Illustration of MRI and portable MRI system.

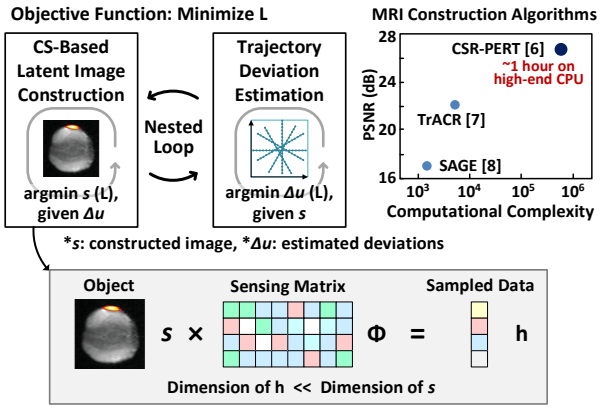


Fig. 2: Overview of the CS-based MRI algorithms.

work presents a dedicated image signal processor that achieves 3.1 frame/s with 219mW for CSR-PERT image construction. An MRI processing system with power below 1W can be achieved by integrating the frontend shown in [1], [2].

II. Algorithm-Architecture Co-optimization

Fig. 3 illustrates the CSR-PERT algorithm [6], which employs radial-pattern-based compressive sensing to reduce sampling time and deviation estimation to correct trajectory deviations. The discrete wavelet transform (DWT) is utilized to suppress high-frequency noise, and the finite-difference operator (FDO) is utilized to enhance image smoothness. As shown in Fig. 3(a), an MRI image is constructed by minimizing the objective function, including one l_2 term (associated with the non-uniform fast Fourier transform (NUFFT)) and two l_1 regularization terms (with respect to the costs associated with

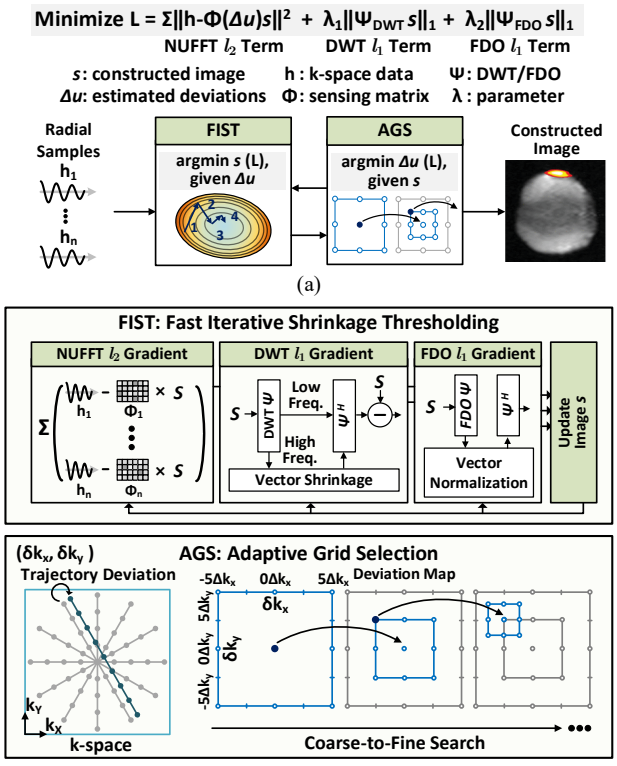


Fig. 3: CSR-PERT algorithm: (a) workflow and (b) details of FIST and AGS.

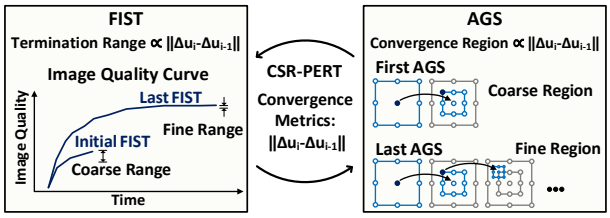


Fig. 4: Convergence-guided early termination.

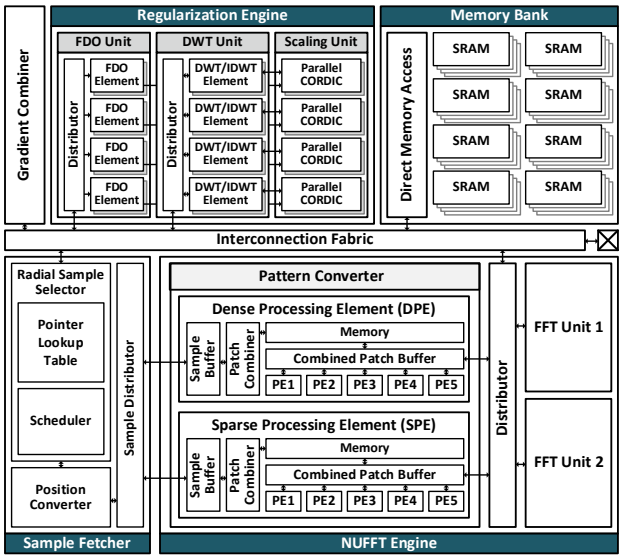


Fig. 5: System architecture.

DWT and FDO). During the minimization process, fast iterative shrinkage thresholding (FIST) and adaptive grid selection (AGS) are performed alternately. Fig. 3(b) shows the details of FIST and AGS. FIST constructs a latent image by performing iterative gradient descent on the objective function given fixed deviations; whereas AGS estimates trajectory deviations by performing iterative coarse-to-fine search on the objective function given a fixed latent image. Forward NUFFT that transforms a latent image into a radial pattern spectral map is included for both FIST and AGS. A latent image in the image space is first converted into a spectral map in the k-space via FFT. In the k-space, the spectral map with a grid pattern is then converted into the one with a radial pattern via interpolation. Backward NUFFT is the inverse function of forward NUFFT, which is required for FIST.

Fig. 4 illustrates algorithmic optimization for reducing computational complexity. In the early iterations of the CSR-PERT, the trajectory deviations are usually inaccurately estimated (based on a noisy image), and a latent image is therefore poorly constructed (due to the inaccurate deviations). Thus, gradient descent for FIST and coarse-to-fine search for AGS can be terminated early. Additionally, different iteration counts can be assigned for distinct convergence behaviors. In this work, convergence-guided early termination is proposed to save computations: an adaptive termination range is set for performing gradient descent, and an adaptive convergence region is determined for estimating trajectory deviations. The termination range and convergence region are narrowed according to the convergence of the CSR-PERT, reducing the total computational complexity by 73%.

III. System Architecture

Fig. 5 shows the system architecture of the proposed MRI image signal processor, including a sample fetcher, an NUFFT engine, and a regularization engine. The sample fetcher facilitates memory access to radial samples for the NUFFT operation. The NUFFT engine comprises a pattern converter, including dense processing element (DPE) and sparse processing element (SPE), and two FFT units. The pattern converter handles conversions between the spectral map with the grid pattern and the one with the radial pattern. The two FFT units perform operations for both FFT and inverse FFT. The regularization engine calculates the values of the two regularization terms and generates the corresponding gradients simultaneously.

A. Sample Fetcher

Fig. 6 shows the design of the sample fetcher that utilizes the radial-sample access scheme. Massive data (with up to 6GB/image) for generating spectral maps with a radial pattern

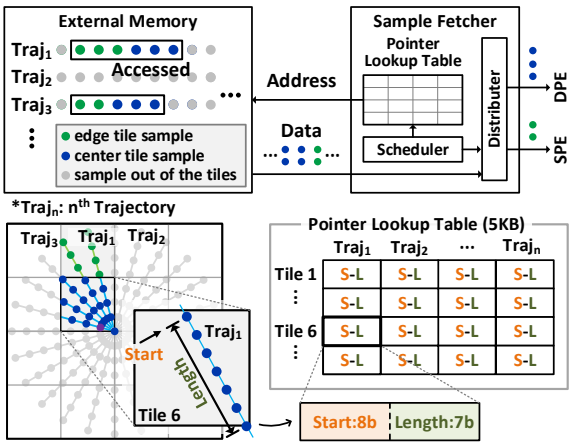


Fig. 6: Design of the sample fetcher.

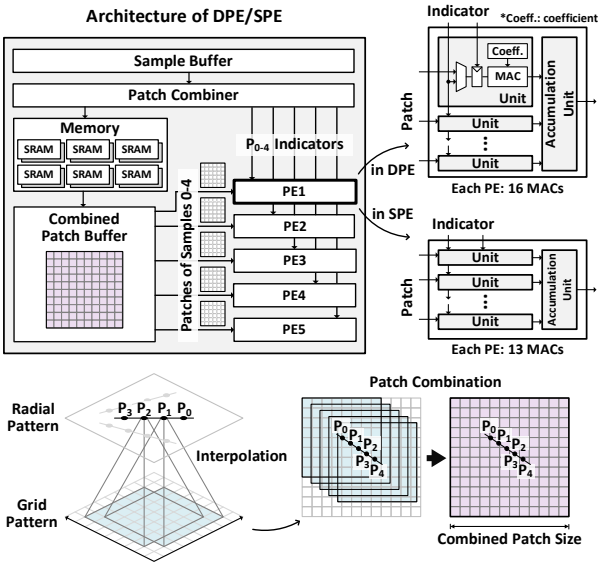


Fig. 7: NUFFT engine: pattern converter.

needs to be processed, so the radial samples are divided into tiles, and the pattern conversions are performed in a tile-by-tile manner. Since the radial samples across trajectories are not continuously stored, the radial samples outside a tile are accessed frequently, leading to redundant memory access. A pointer-based scheme is used to access radial samples efficiently. Instead of the position, the start point and length of a segment are stored as a pointer in a lookup table. This reduces the memory access by 94%. For pattern conversion, the value of the partial sum associated with radial samples with the same positions remains unchanged and only moving radial samples need to be fetched. Such movement-aware access further reduces memory access by 85%. Overall, memory access for fetching radial samples is reduced by 99%.

B. NUFFT Engine

Fig. 7 shows the critical pattern converter in the NUFFT engine. The pattern converter performs conversion between the latent image and the radial pattern spectral map for both dense and sparse samples by utilizing DPE and SPE, respectively. Since the sample density in the k-space center region is much higher than that in the edge region, processing the center region becomes the bottleneck in parallel processing. Therefore, the DPE, equipped with higher computational resources, processes the center region, and the SPE processes the edge region. For both DPE and SPE, 5 processing elements (PEs) are deployed since 5 interpolations need to be performed in parallel for a combined patch. Multiple multiply-accumulate (MAC) units are included to perform the required interpolations. Starting from the design with equal MAC units, 80 MAC units and 65 MAC units are deployed in the DPE and SPE, respectively, increasing the utilization by 23%.

In the DPE and the SPE, multiple interpolations are performed in parallel to support a high throughput. Interpolation is applied to an 8×8 patch, and parallel interpolation operations on multiple patches lead to high SRAM bandwidth demand. The spatial overlap among neighboring patches allows combining multiple smaller patches into a larger one. A larger combined patch size enhances sample reuse in the overlap regions, but it also includes more samples outside the original smaller patches. An optimized 12×12 patch is selected to minimize memory access by 55% compared to the design with an 8×8 patch.

C. Regularization Engine

Fig. 8 shows the regularization engine, which is equipped with an FDO unit, a DWT unit, and a scaling unit. The FDO unit includes 8 FDO elements, each with 2 convolution elements; the DWT unit includes 8 DWT elements, each with 4 convolution elements. The scaling unit is implemented by coordinate rotation digital computer (CORDIC) in an area-efficient manner. Conventionally, a CORDIC architecture is operated in the vectoring mode for multiple cycles to retrieve the angle of a vector, and in the rotation mode to obtain the unit vector at that angle. The proposed parallel CORDIC architecture is operated in both modes simultaneously, rotating the vector in rotation mode by a partial angle whenever the partial angle is obtained in vectoring mode. This design reduces the processing latency by 50% compared to the original CORDIC architecture.

IV. Chip Implementation

Fig. 9 shows the chip implementation results. Fabricated in 40nm CMOS, the proposed MRI signal processor integrates 8.3M logic gates and 169 KB SRAM in a core area of 5.47 mm^2. The chip dissipates 31-to-219 mW at a clock frequency of 70-to-206 MHz from a 0.6-to-1.0 voltage supply. It achieves a maximum throughput of 3.1 frame/s, with an energy efficiency of 14.1 frame/J, for constructing MRI images with 128×128 pixels, as shown in Fig. 10. The performance of the chip is evaluated on the MRI empirical data [7] and synthetic data (with a normalized deviation of 0.003 rad), as shown in Fig. 11. Each dataset consists of 201 trajectories,

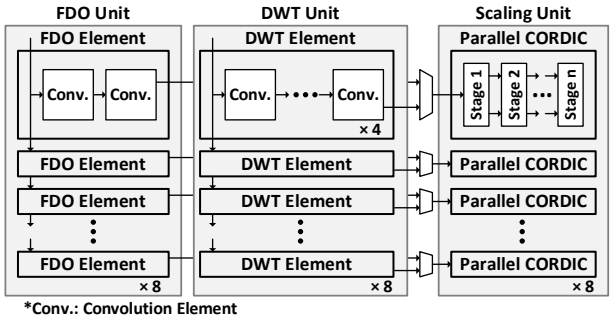
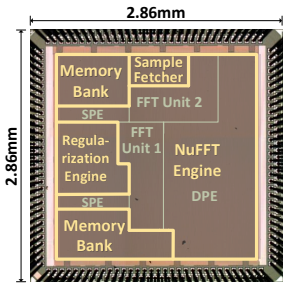


Fig. 8: Hardware design of the regularization engine.



Technology		40 nm CMOS
Area	Die	2.86 x 2.86 mm^2
	Core	2.34 x 2.34 mm^2
On-Chip SRAM (KB)		169
Frequency (MHz)		70-206
Voltage (V)		0.6-1.0
Power (mW)		31-219
Maximum Throughput (frame/s)		3.1
Energy Efficiency (frame/J)		14.1-33.8

Fig. 9: Chip micrograph and summary.

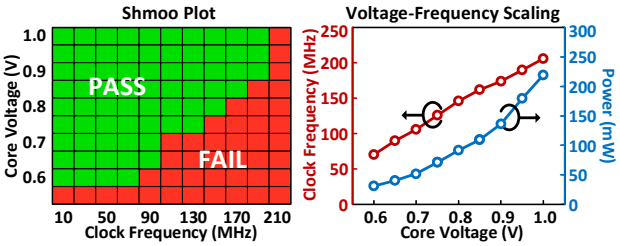


Fig. 10: Measurement results.

and each trajectory contains 256 samples. An undersampling ratio (UR) is set to accelerate sampling. The results show that the PSNR degradation for 1, 4, 16 URs is less than 0.01 dB compared to the ground truth. This work achieves 1.6 $\times$ 10^5 and 8.2 $\times$ 10^2 higher energy efficiency and throughput, respectively, than a high-end CPU, as shown in Fig. 12.

V. Conclusion

This work presents the first image signal processor designed for portable MRI systems. The computational complexity is significantly reduced by leveraging the convergence behaviors of FIST and AGS through convergence-guided early termination. For hardware implementation, the sample fetcher reduces memory access by efficiently fetching non-continuous samples and skipping unchanged samples. In the NUFFT engine, the MAC units are properly allocated according to the workloads, thus improving the hardware utilization, and the overlapped patches are combined to minimize memory access. In the regularization engine, the CORDIC architecture is operated

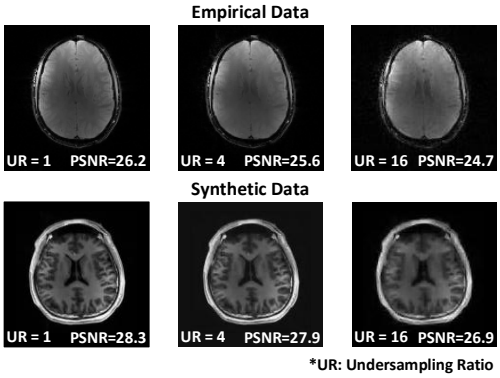


Fig. 11: Performance evaluation.

	Intel Core i7	This Work
Technology	7 nm	40 nm
Die Area (mm^2)	257	8.18
Frequency (MHz)	3900	206
Power (mW)	43000	219
Latency (s)	260	0.32
Throughput (frame/s)	3.8x10^{-3}	3.1
Energy Efficiency (frame/J)	8.8x10^{-5}	14.1

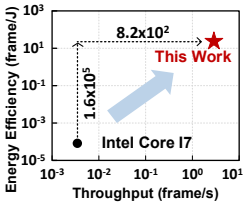


Fig. 12: Comparison with a high-end CPU.

in both the vectoring mode and rotation mode simultaneously to reduce the latency. The chip achieves orders-of-magnitude improvements in both energy efficiency and throughput over a high-end CPU. This work provides a promising low-power, real-time image construction solution for portable MRI systems.

Acknowledgments

This work is supported by National Science and Technology Council (NSTC) of Taiwan. The authors also thank Taiwan Semiconductor Research Institute (TSRI) for technical support on chip design and fabrication.

References

[1] D. Krüger *et al.*, "A Portable CMOS-Based Spin Resonance System for High-Resolution Spectroscopy and Imaging," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 7, pp. 1838–1849, May 2023.

[2] S. Fan *et al.*, "A Miniature Multinuclei NMR/MRI Platform with a High-Voltage SOI ASIC," *IEEE Journal of Solid-State Circuits*, vol. 60, no. 6, pp. 2013–2024, Nov. 2024.

[3] J. P. Marques *et al.*, "Low-Field MRI: an MR Physics Perspective," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 6, pp. 1528–1542, June 2019.

[4] T. C. Arnold *et al.*, "Low-Field MRI: Clinical Promise and Challenges," *Journal of Magnetic Resonance Imaging*, vol. 57, no. 1, pp. 25–44, Jan. 2023.

[5] N. O. Addy *et al.*, "Simple Method for MR Gradient System Characterization and K-Space Trajectory Estimation," *Magnetic Resonance in Medicine*, vol. 68, no. 1, pp. 120–129, July 2012.

[6] A. Vaish *et al.*, "CSR-PERT: Joint Framework for MRI and HARDI Data Reconstruction using Perturbed Radial Trajectory Estimated from Compressively Sensed Measurements," *Computers in Biology and Medicine*, vol. 150, p. 106117, Nov. 2022.

[7] J. D. Ianni *et al.*, "Trajectory Auto-Corrected Image Reconstruction," *Magnetic Resonance in Medicine*, vol. 76, no. 3, pp. 757–768, Sep. 2016.

[8] A. M. Stokes *et al.*, "A Simplified Spin and Gradient Echo Approach for Brain Tumor Perfusion Imaging," *Magnetic Resonance in Medicine*, vol. 75, no. 1, pp. 356–362, Jan. 2016.