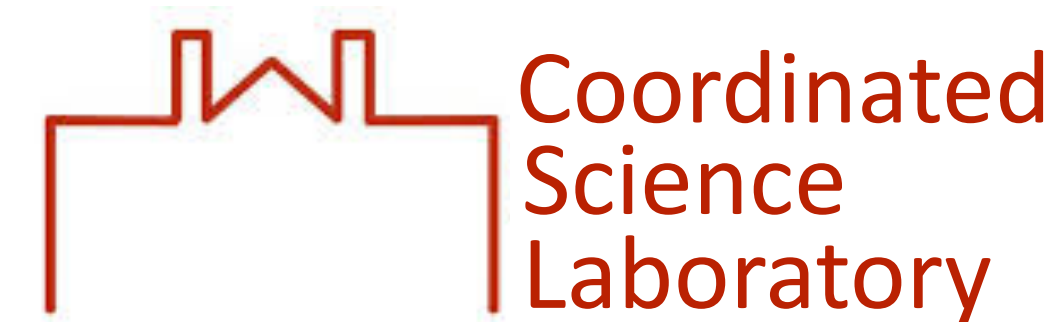
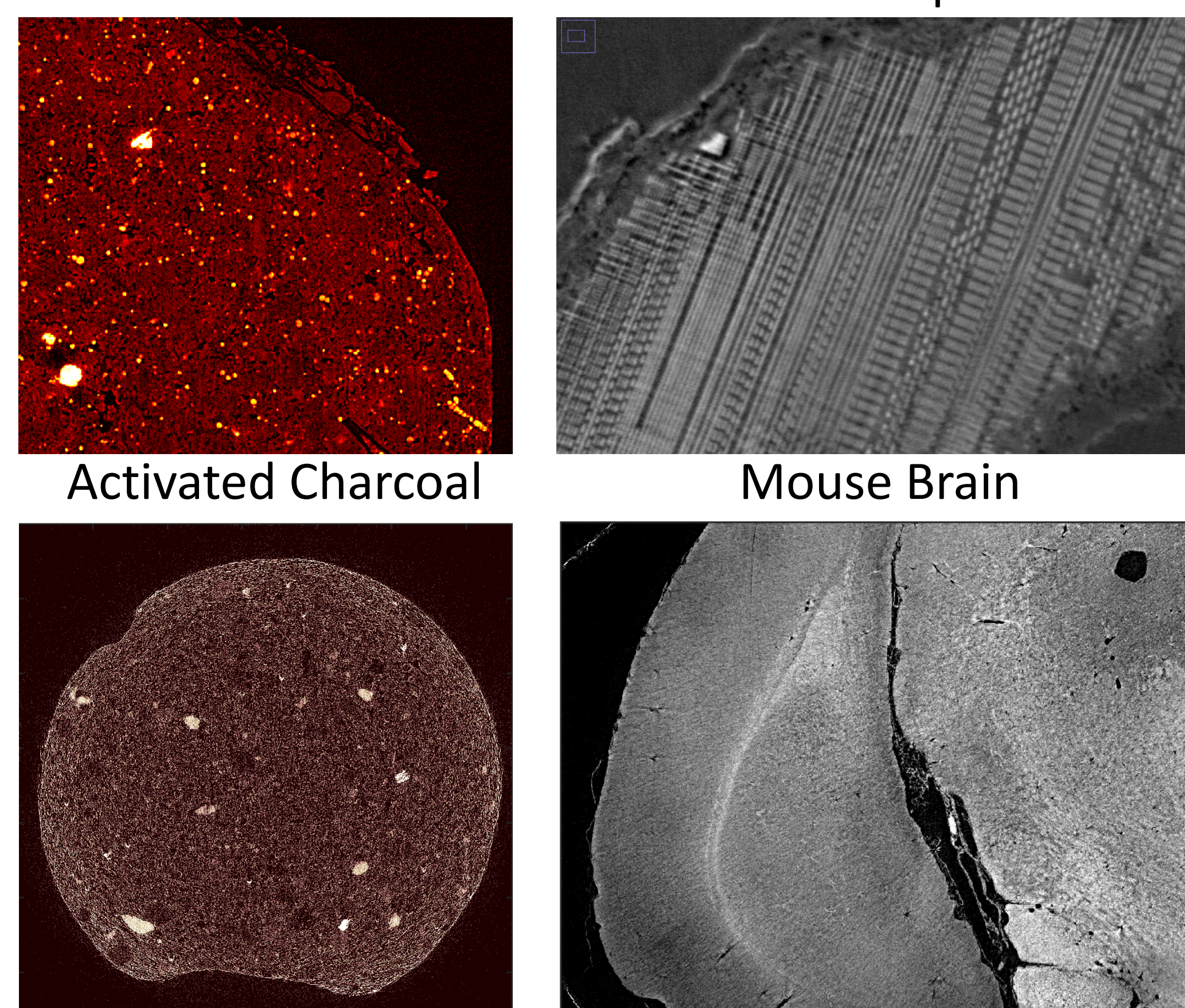


# Memory-Centric 3D Image Reconstruction with Hierarchical Communications on Multi-GPU Node Architecture

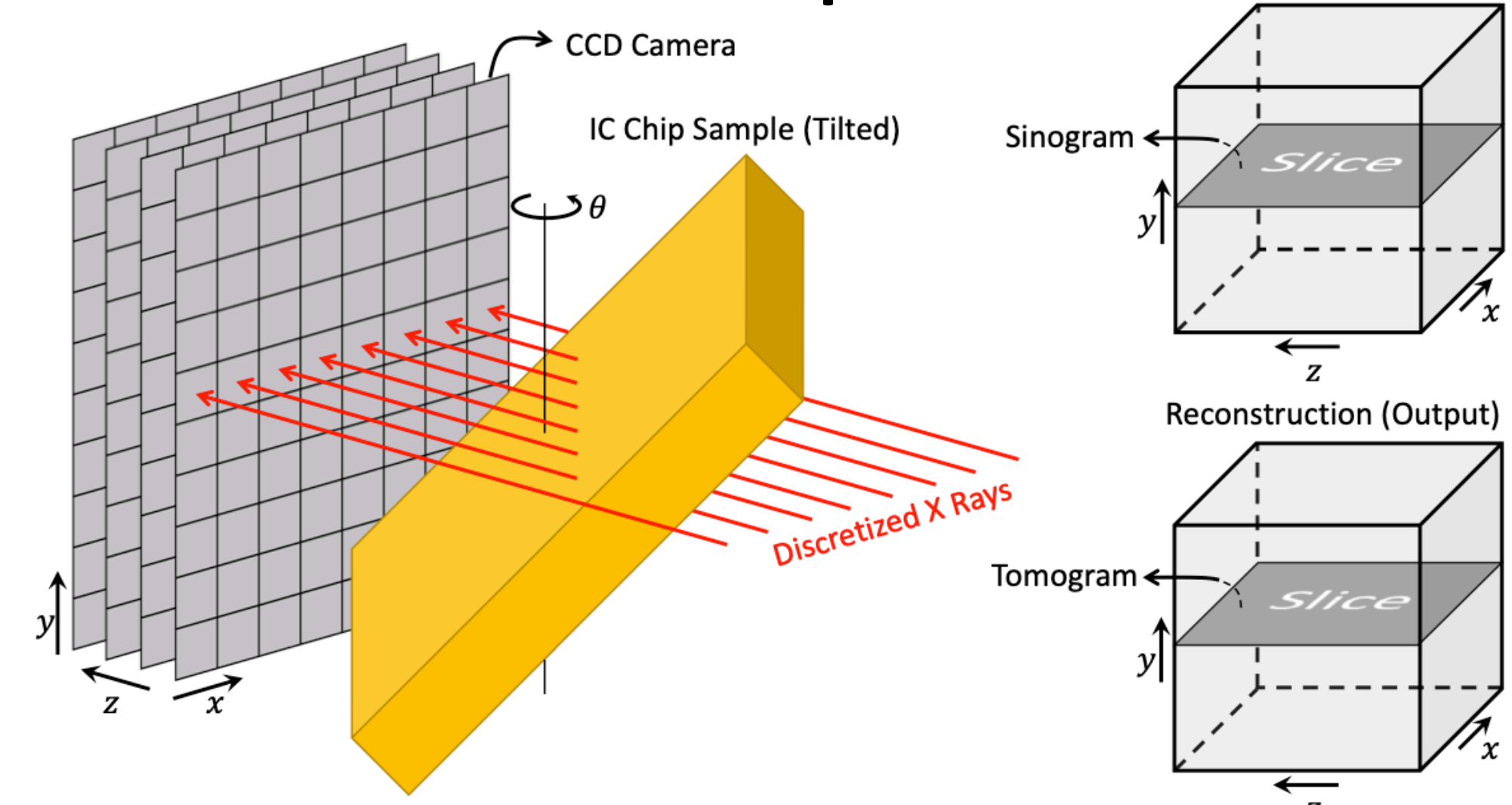
Mert Hidayetoglu and Wen-Mei Hwu (Advisor)  
University of Illinois at Urbana-Champaign



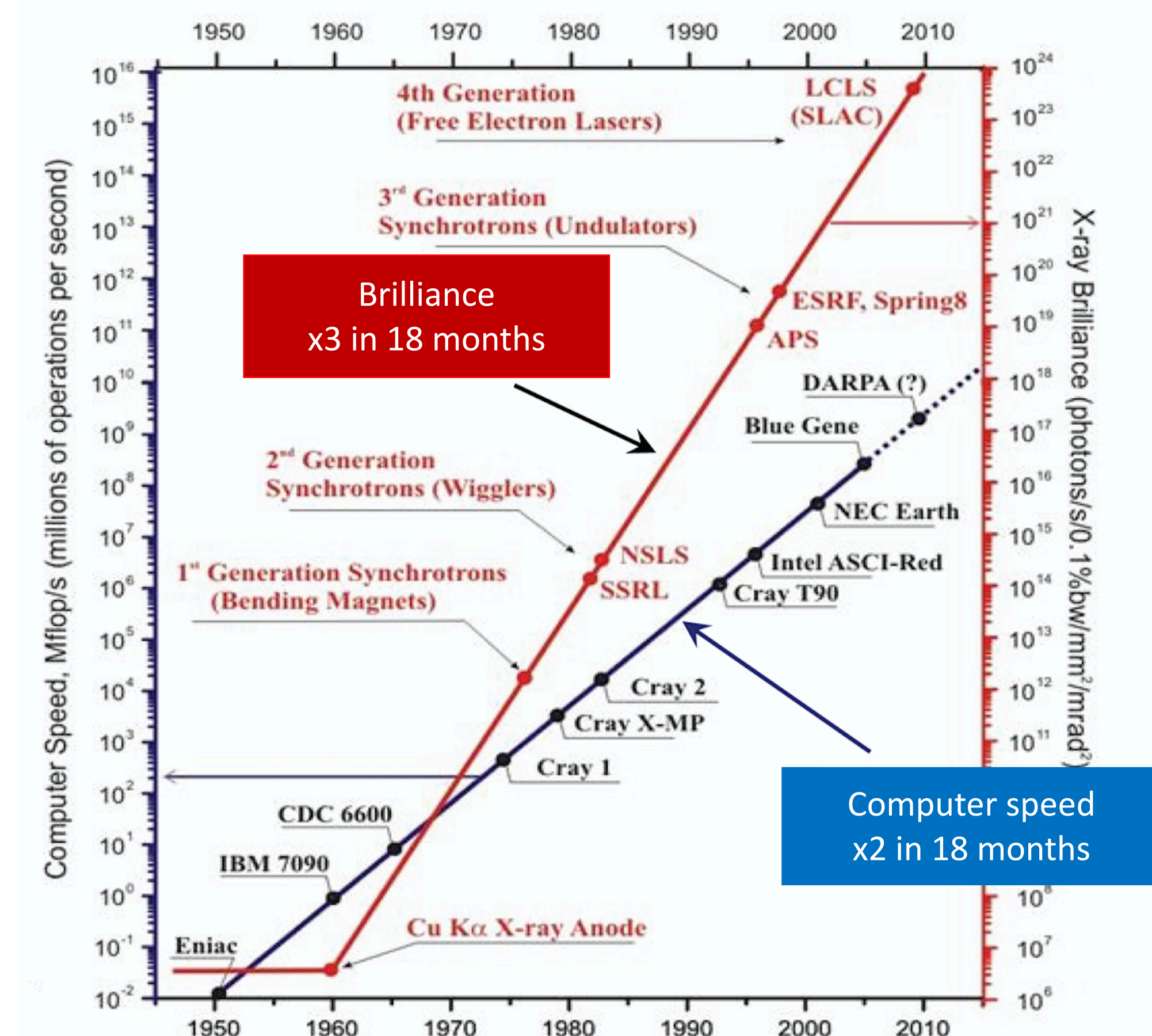
## 3D X-Ray Imaging Applications



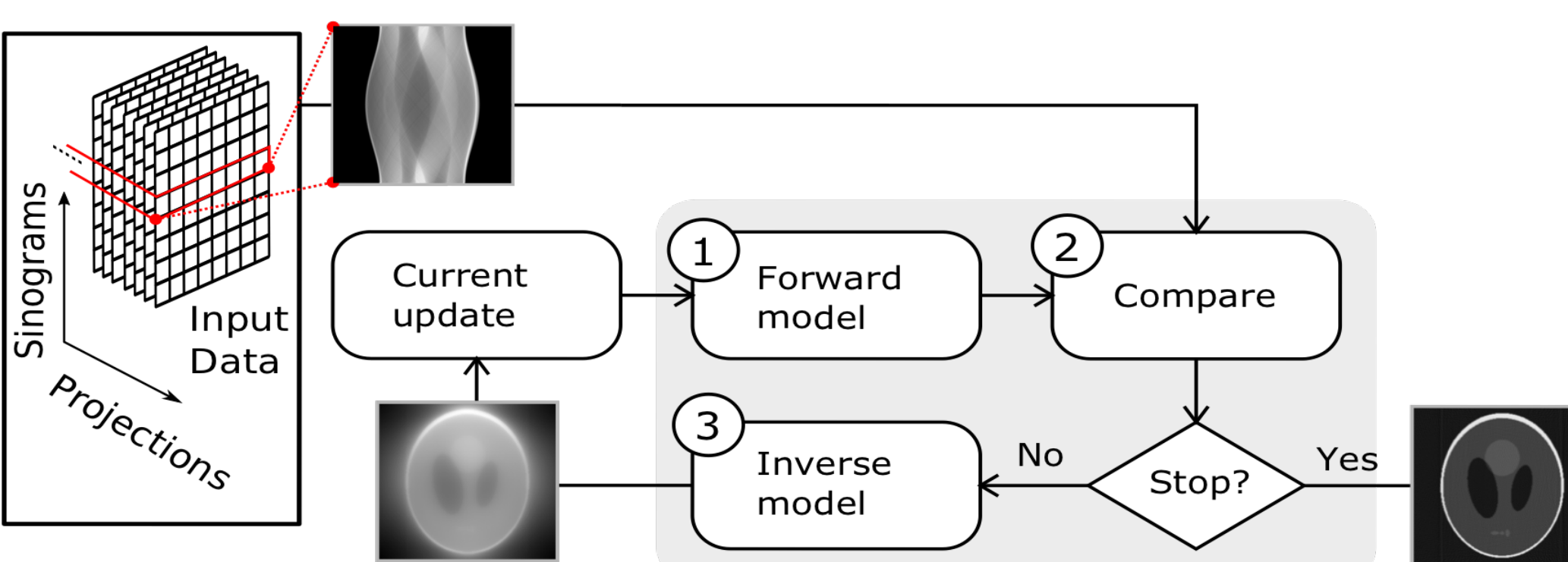
### Data Acquisition



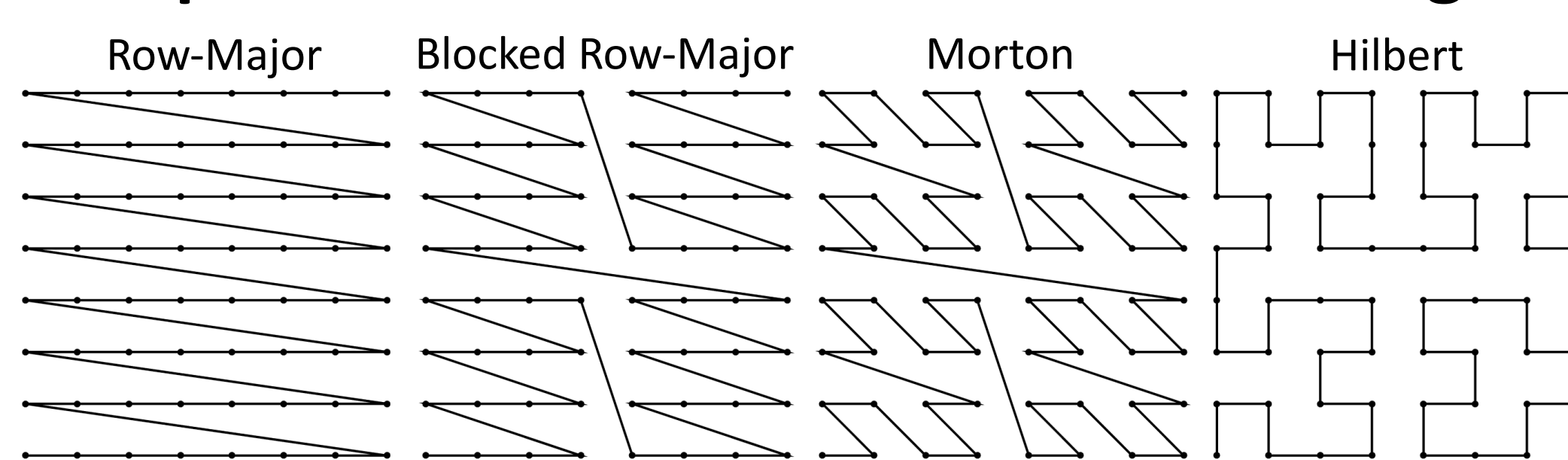
### Advancement Trends



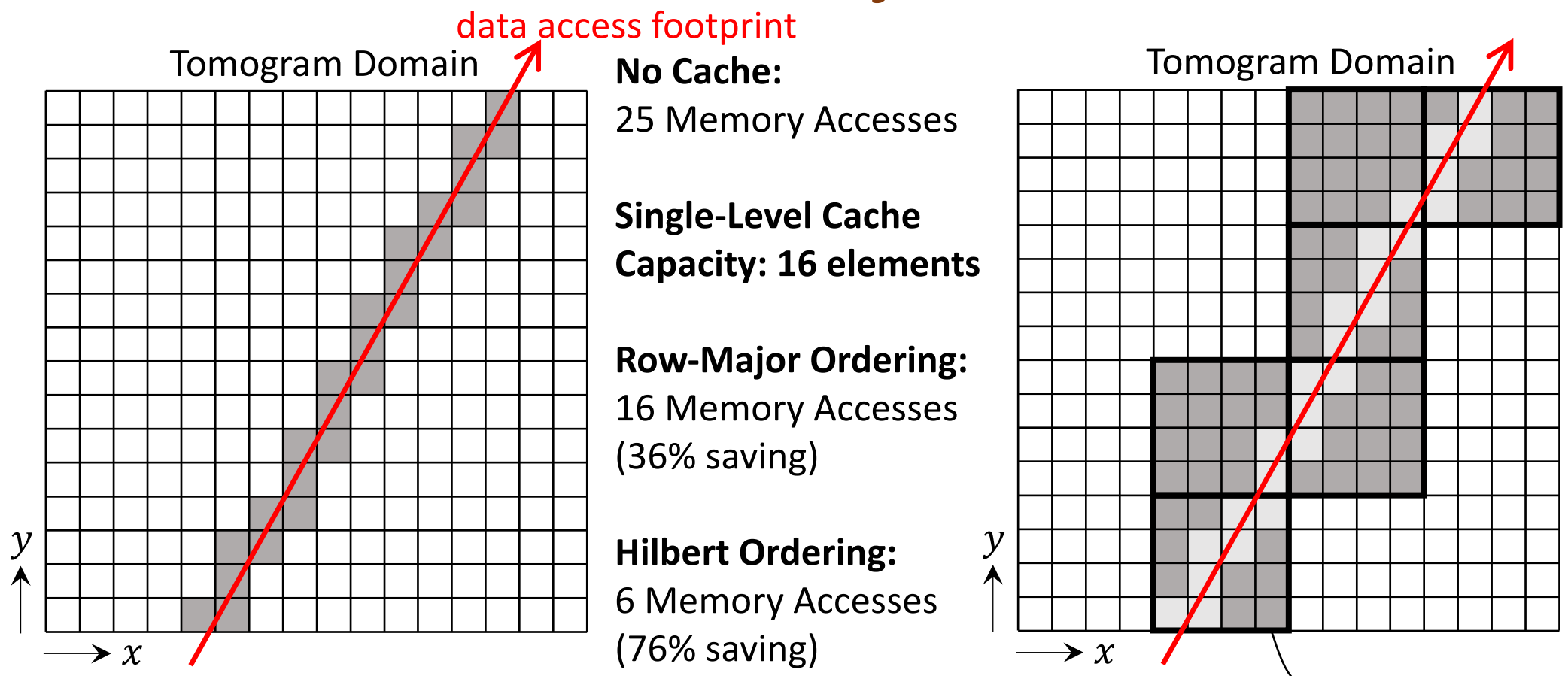
### Iterative Reconstruction



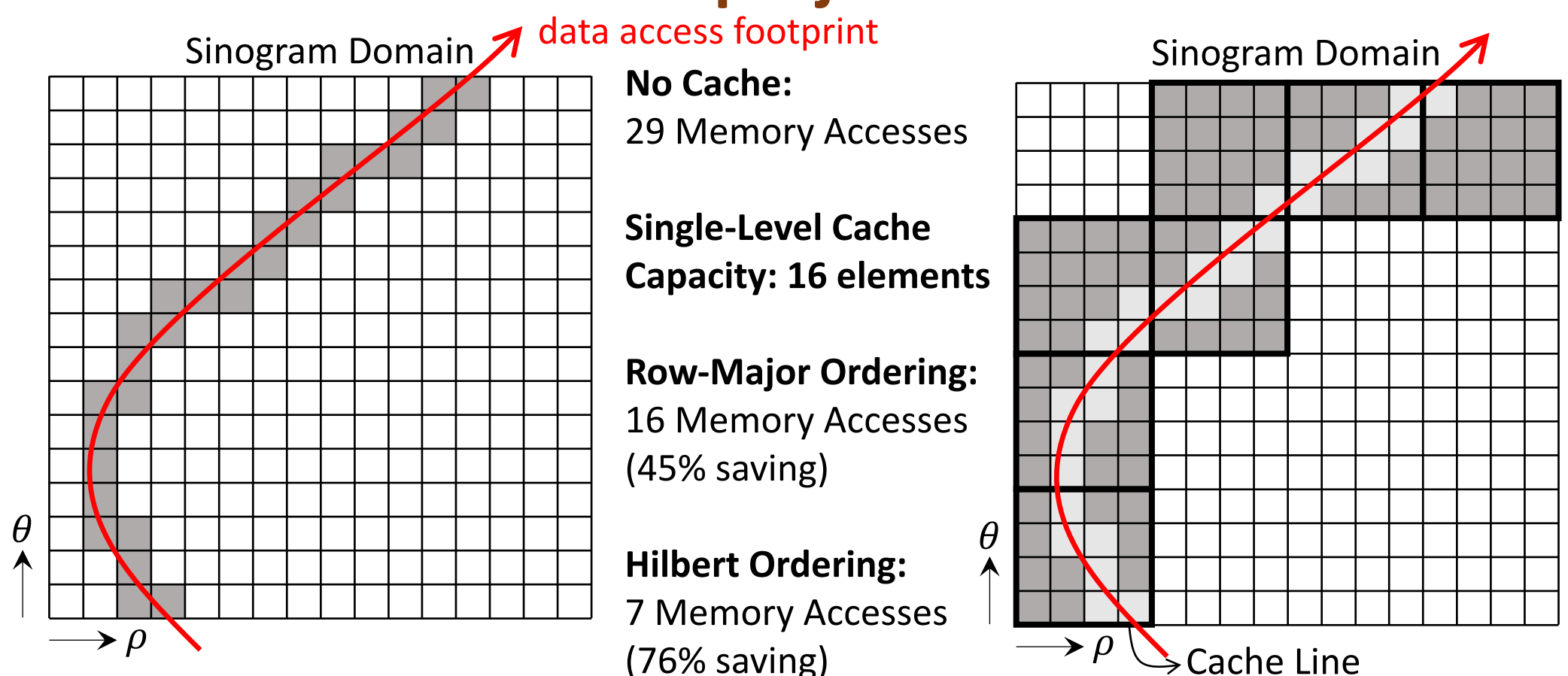
## Optimization 1: Pseudo-Hilbert Ordering



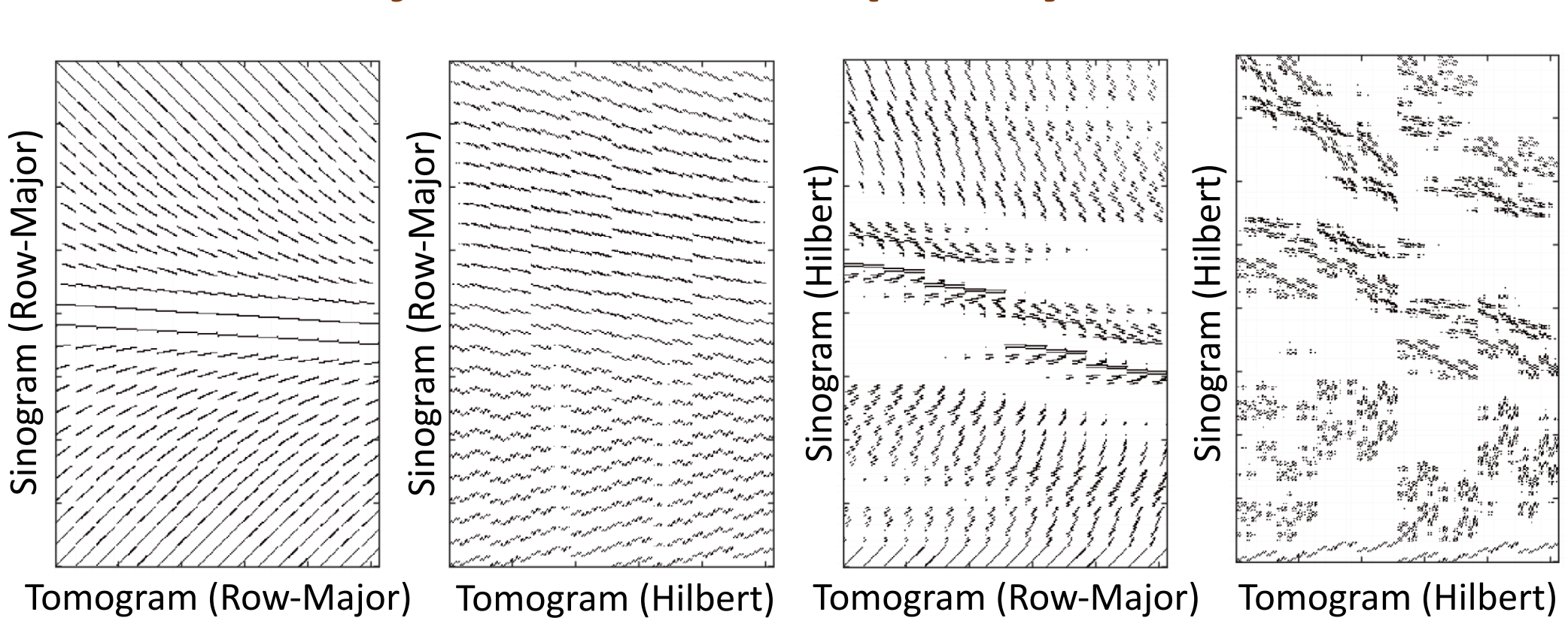
### Forward Projection



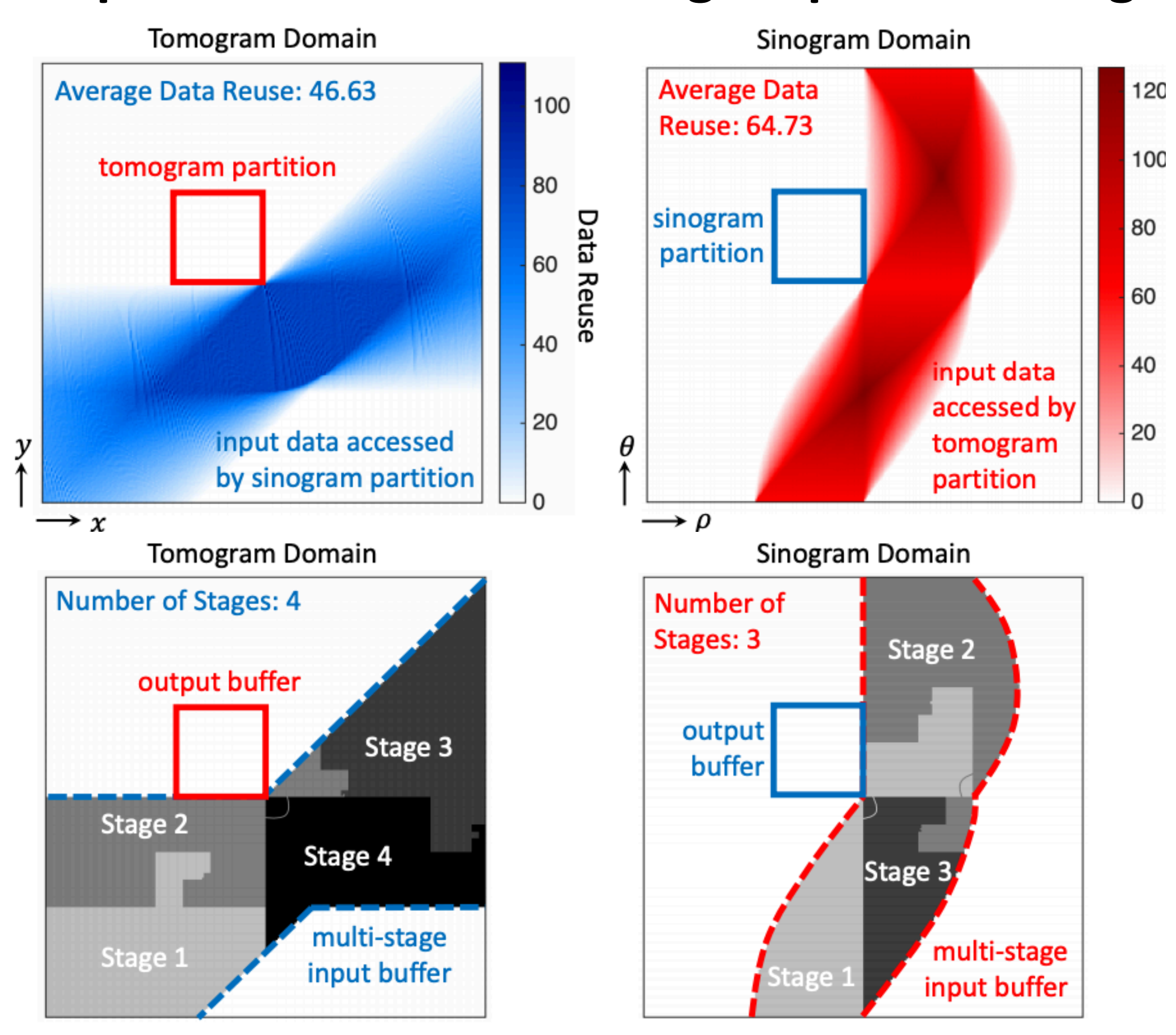
### Backprojection



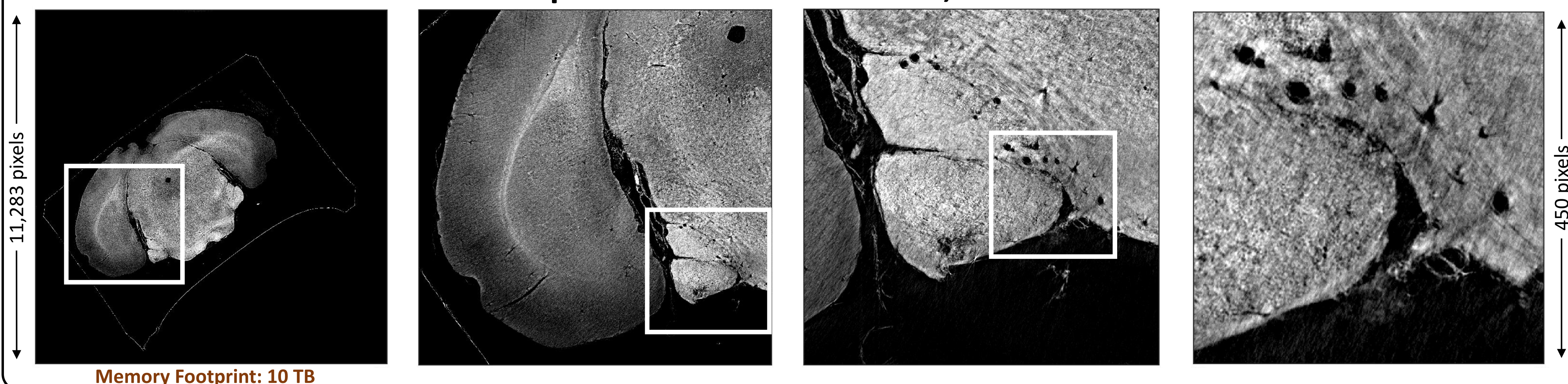
### Projection Matrix Sparsity Patterns



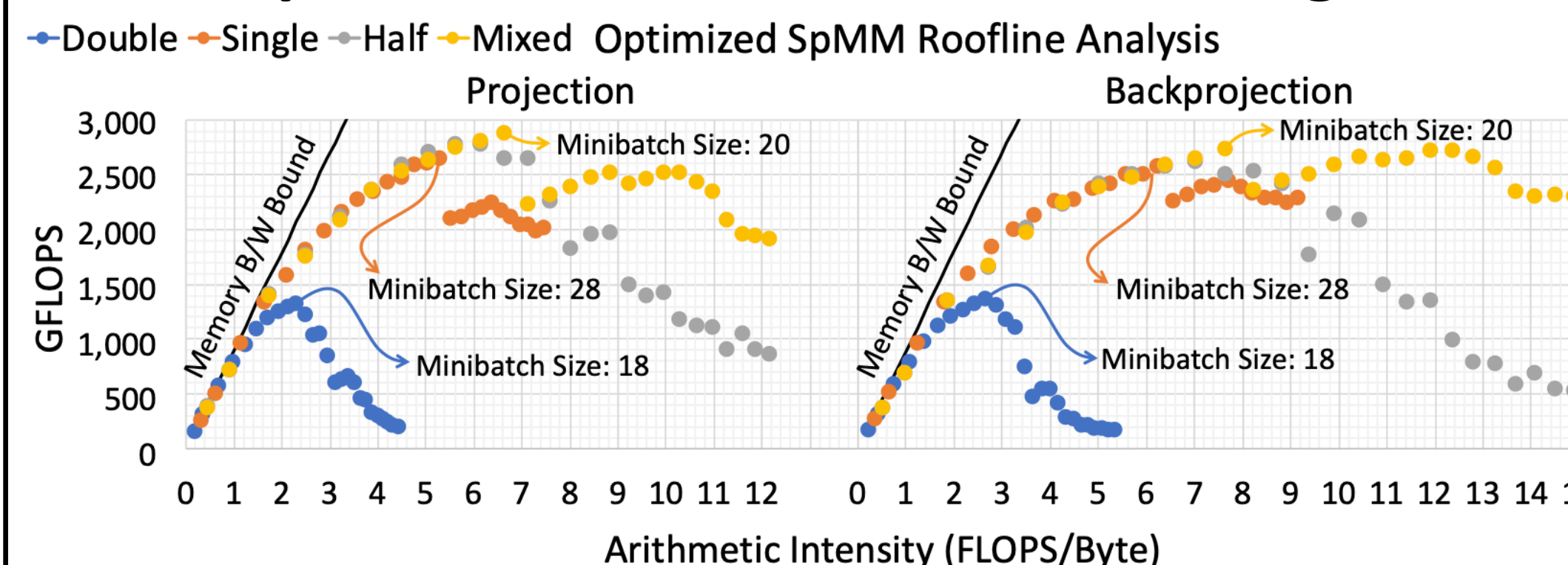
## Optimization 2: Multi-Stage Input Buffering



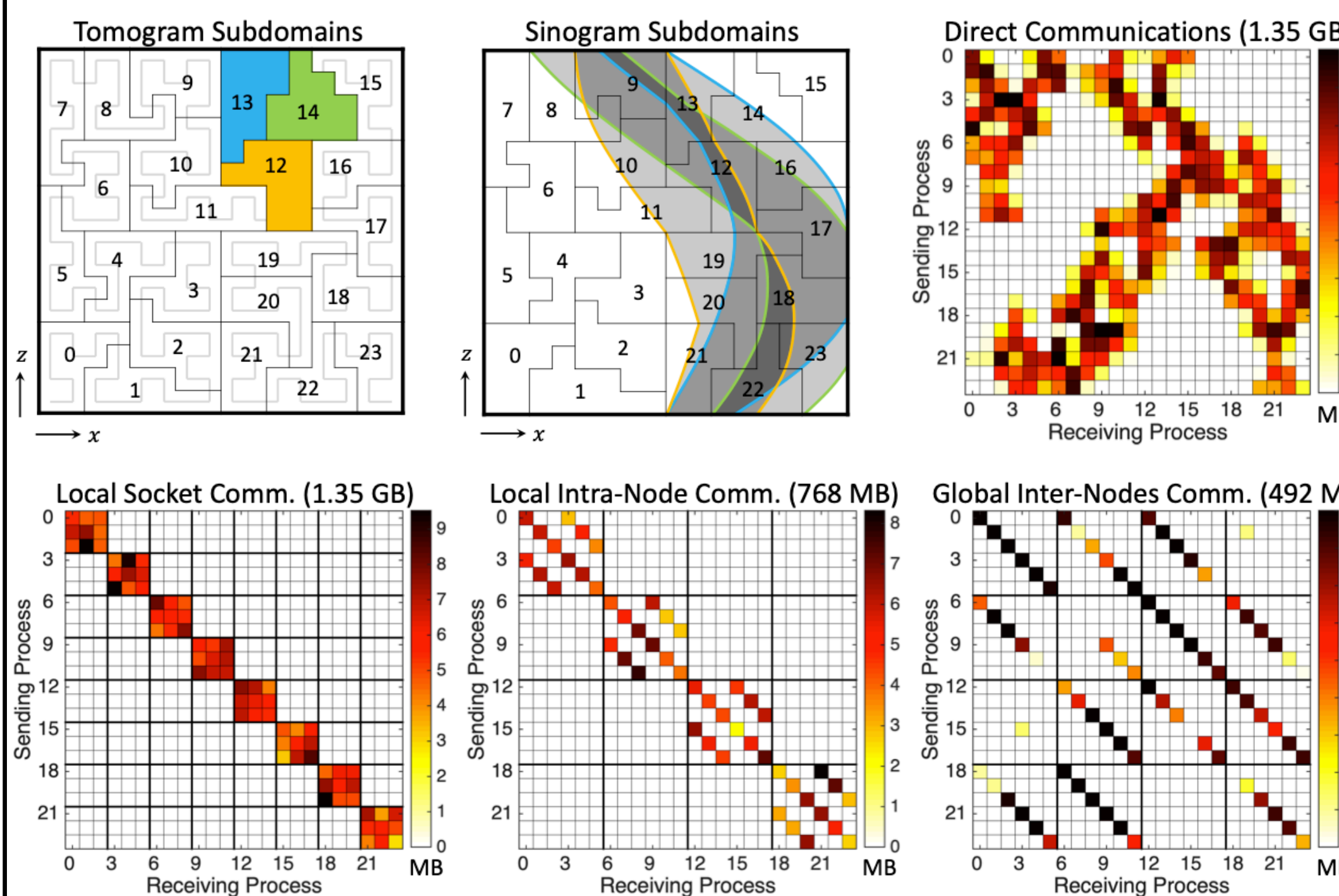
## Brain Sample Reconstruction on 24,576 V100 GPUs



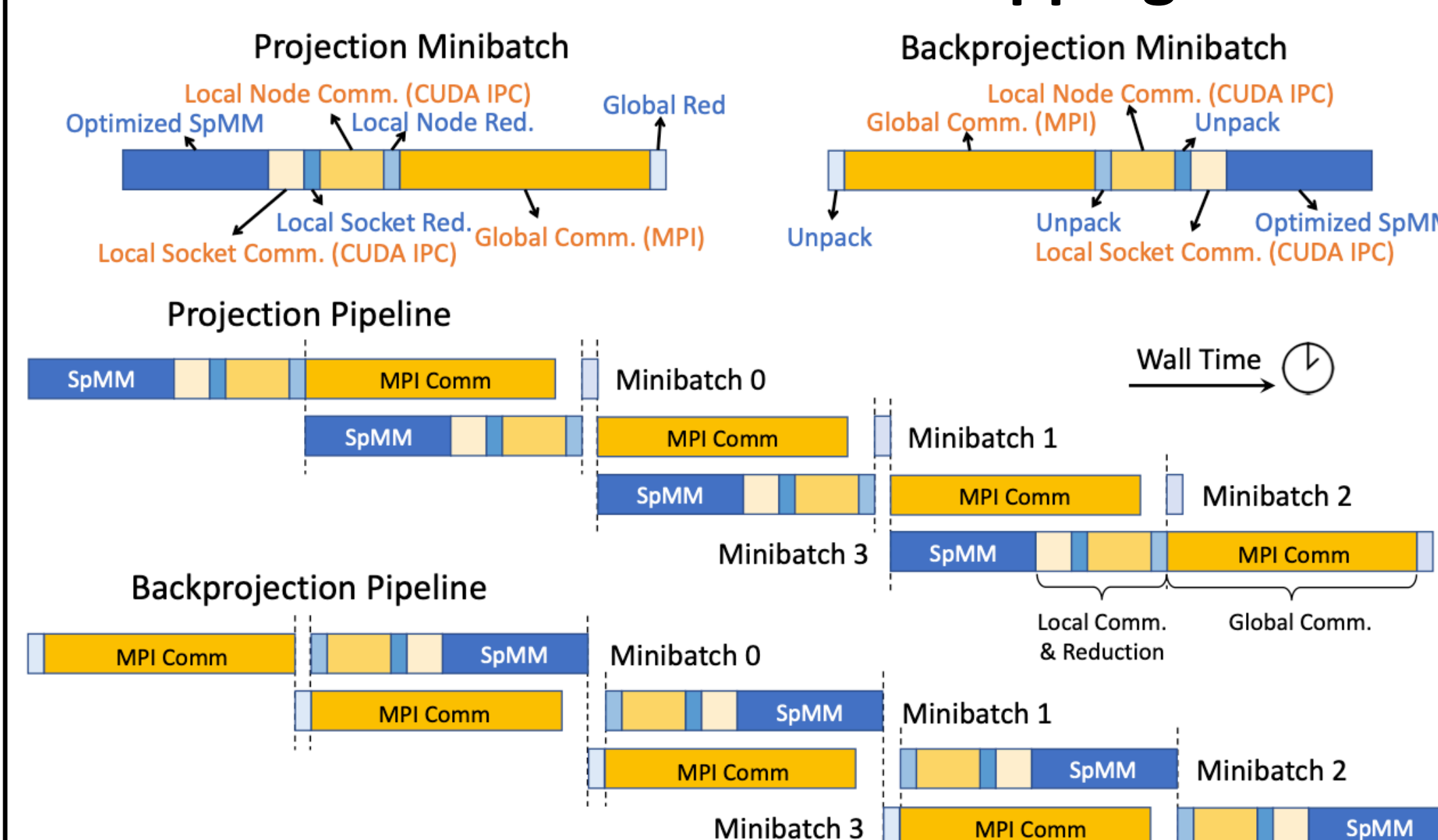
## Optimization 3: GPU Minibatching



## Hierarchical Communication & Data Reduction



## Communication Overlapping



## Sample Datasets

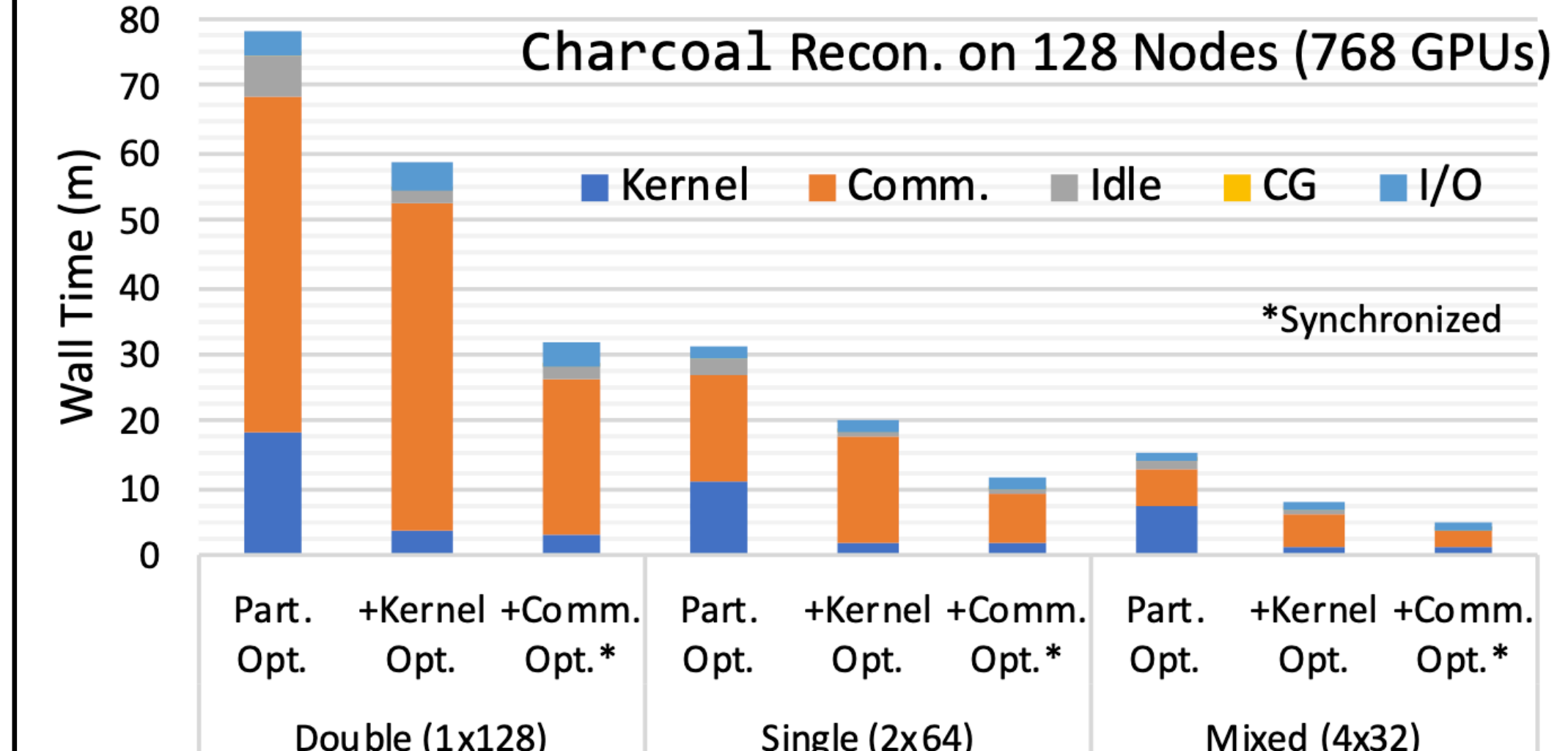
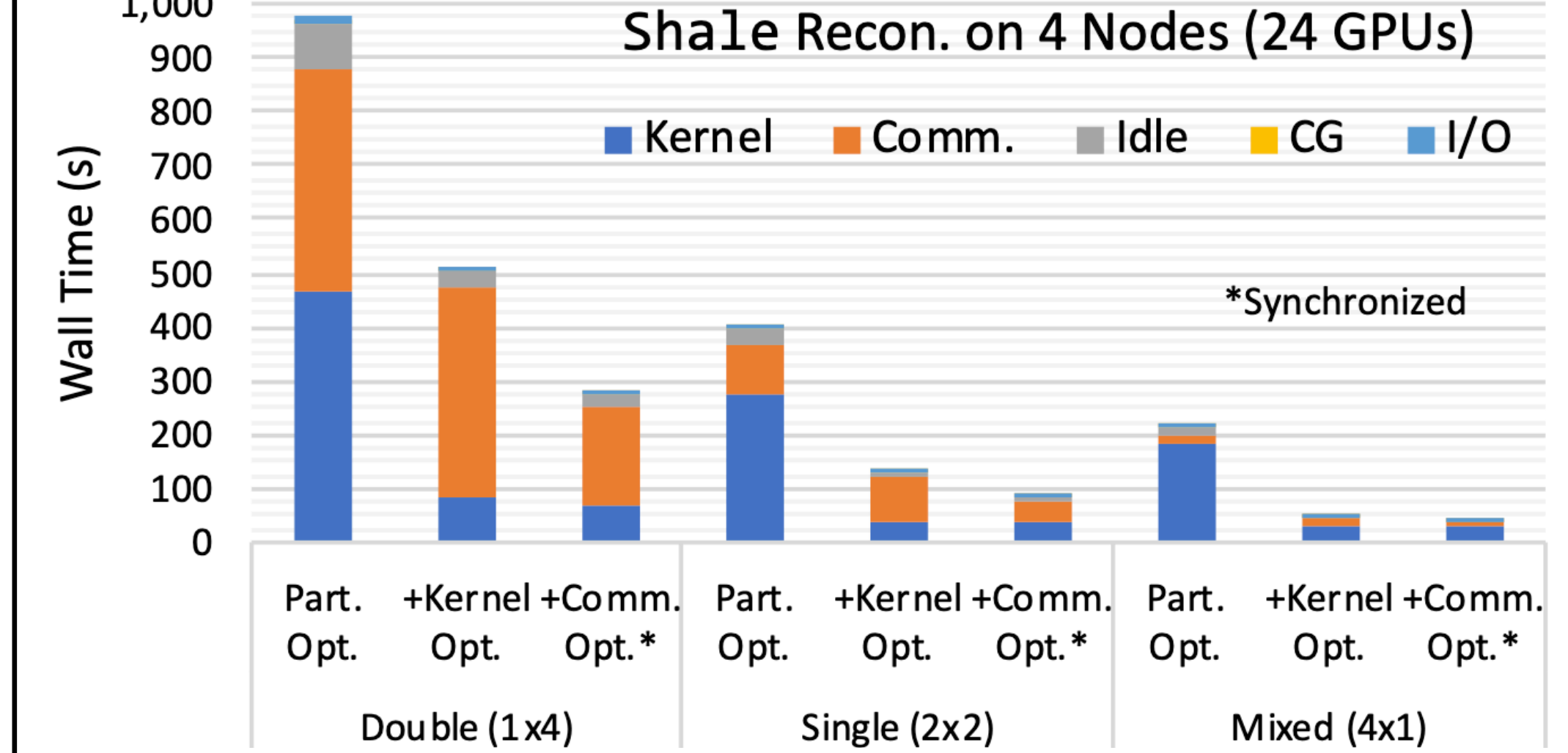
Sample	Measurement Data Cube ( $K \times M \times N$ )	I/O Data Footprint	Memory Footprint
Shale Rock	1501 × 1792 × 2048	52.1 GB	120 GB
IC Chip	1210 × 1024 × 2448	36.7 GB	139 GB
Activated Charcoal	4500 × 4198 × 6613	1.23 TB	2.82 TB
Mouse Brain	4501 × 9209 × 11 283	6.56 TB	10.9 TB

## Overall Reconstruction Speedup

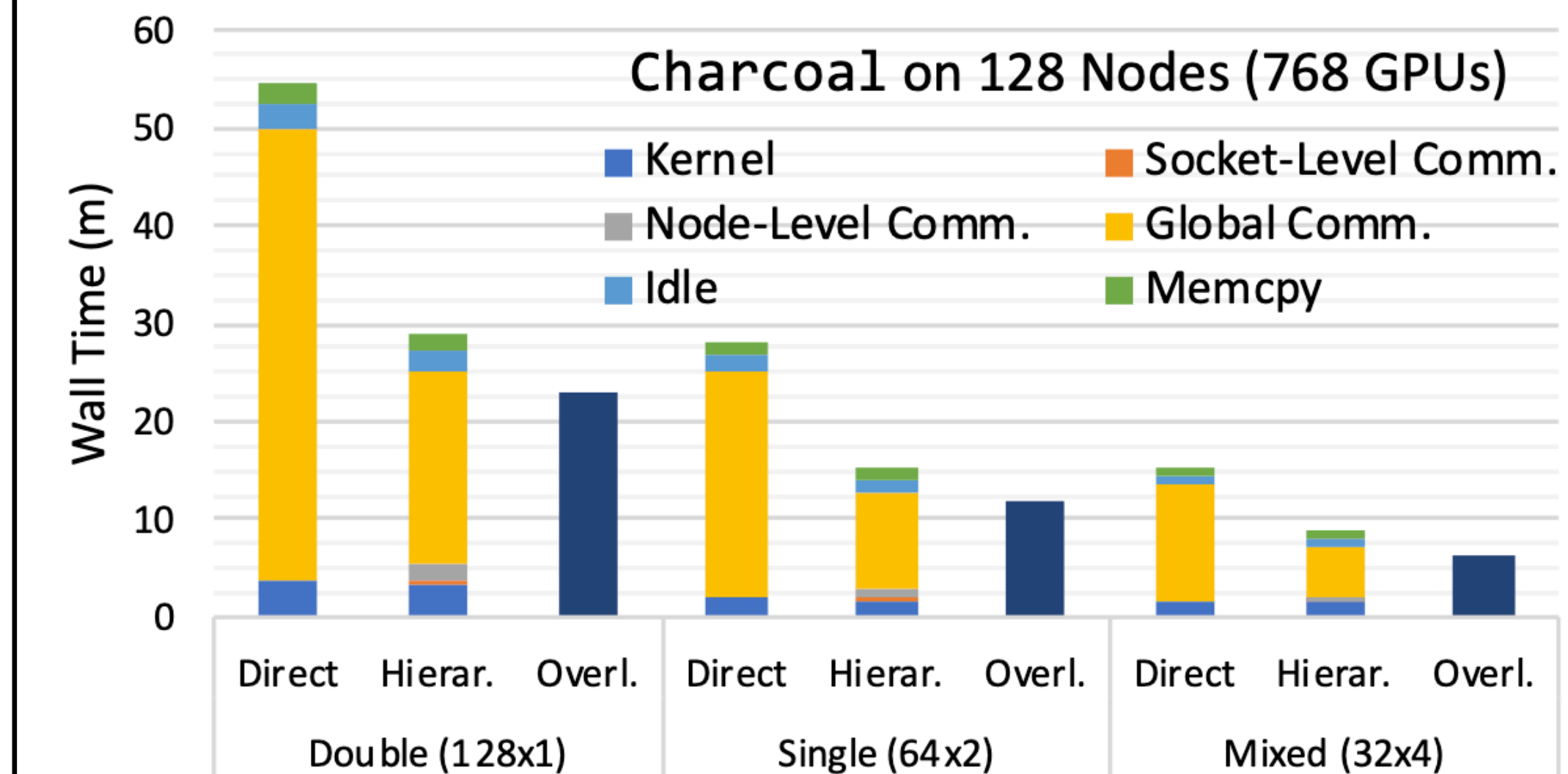
Part. Opt.	Shale on Four Nodes			Charcoal on 128 Nodes		
	Part.*	Recon.	Speed.	Part.*	Recon.	Speed.
Double	1 × (4 × 6)	979 s	1 ×	1 × (128 × 6)	78.4 m	1 ×
	2 × (2 × 6)	405 s	2.42 ×	2 × (64 × 6)	31.3 m	2.51 ×
	4 × (1 × 6)	215 s	4.56 ×	4 × (32 × 6)	15.1 m	5.20 ×
Part. + Kernel Opt.	1 × (4 × 6)	513 s	1.91 ×	1 × (128 × 6)	58.4 m	1.34 ×
	2 × (2 × 6)	134 s	7.30 ×	2 × (64 × 6)	20.4 m	3.85 ×
	4 × (1 × 6)	51.1 s	19.2 ×	4 × (32 × 6)	8.0 m	9.78 ×
Part. + Kernel + Comm. Opt.	1 × (4 × 6)	218 s	4.49 ×	1 × (128 × 6)	27.0 m	3.00 ×
	2 × (2 × 6)	76.5 s	12.79 ×	2 × (64 × 6)	10.0 m	7.87 ×
	4 × (1 × 6)	42.2 s	23.19 ×	4 × (32 × 6)	4.30 m	18.19 ×

\*Total Number of Partitions = Batch Nodes × (Data Nodes × Partitions per node). Data partitions per node is set to six because each node consists of six GPUs.

## Breakdown of End-to-End Recon. Times



## Breakdown of Communication Times



## Communicated Data and System B/W

	Prec.	Socket-Level Comm. Data B/W	Node-Level Comm. Data B/W	Global Comm. Data B/W	Memcpy B/W
Direct	Double	N/A	N/A	36.6 TB	1.61 TB/s
	Single	N/A	N/A	18.3 TB	1.61 TB/s
	Mixed	N/A	N/A	9.16 TB	1.59 TB/s
Hierar.	Double	36.6 TB	174 TB/s	21.4 TB	21.3 TB/s
	Single	18.3 TB	170 TB/s	10.7 TB	22.8 TB/s
	Mixed	9.16 TB	164 TB/s	5.35 TB	23.5 TB/s

\*Per projection (and backprojection). Fig. 11 involves 30 projections and 31 backprojections.

## Strong and Weak Scaling on Summit

