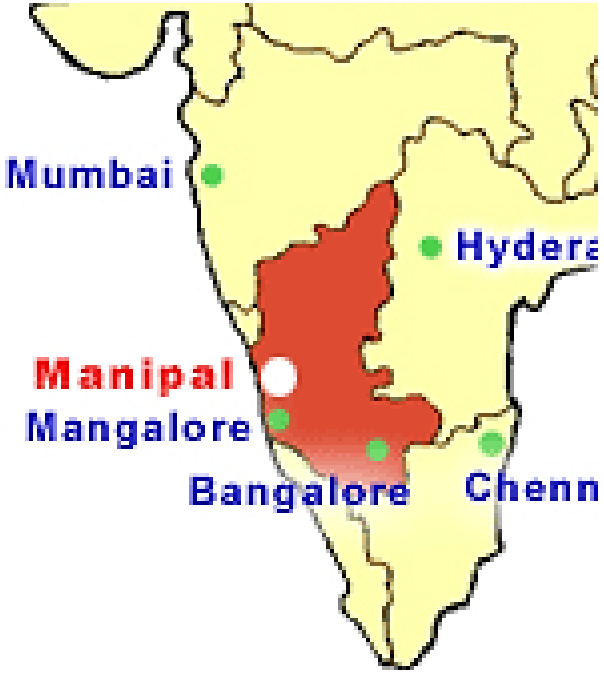
Accelerated X-Ray Imaging: Real-time Multi-plane Image Reconstruction with CUDA

Prashanth B. Bhat, Vinay Kumar, Akshaya M., Devendra Shastry, Venkata Subbarao, Niranjan U.C.

Manipal Dot Net Pvt. Ltd., India
http://www.manipal.net

Manipal Dot Net



- **Technology Outsourcing Firm in South India**
- **Founded in 2004 by Dr. Narasimha Bhat, PhD: UC Berkeley**
- **Software and Hardware Engineering Teams**
- **Areas of Expertise:**
  - Embedded Systems Software
  - Digital Design
  - High Performance Computing
  - Video and Image Processing

Agenda

- Introduction to X-Ray Fluoroscopy
- SBDX: Scanning Beam Digital X-Ray
- Image Reconstruction: Problem Definition
- Computational Aspects
- Parallel Implementation with CUDA
- Performance Results
- Summary and Future Work

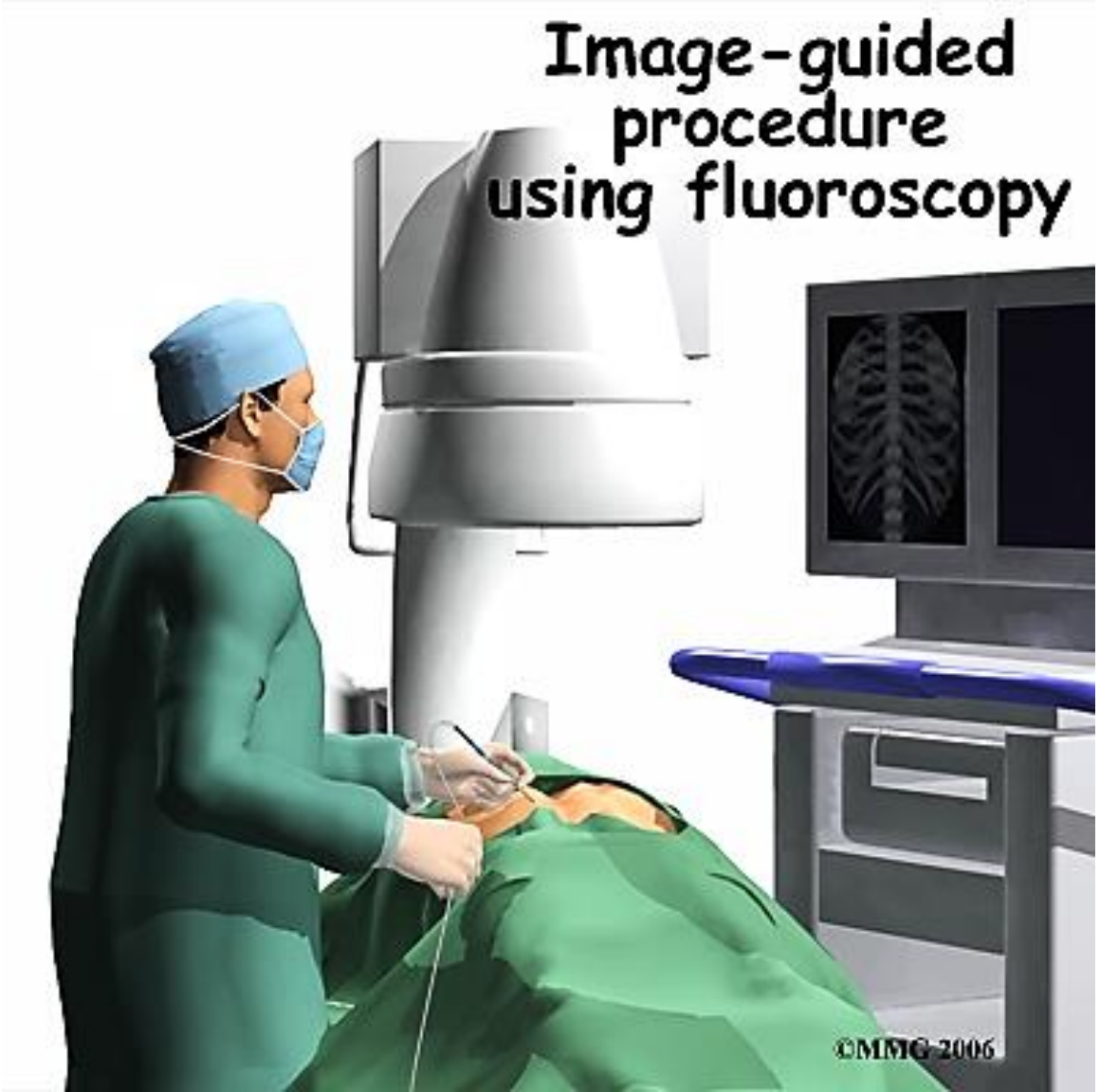
Introduction:
X-Ray Fluoroscopy

X-Ray Fluoroscopy

X-Ray Fluoroscopy

- **A Medical Imaging Technique**
  - In use for over 40 years
- **Useful during Cardiac Interventional Procedures**
  - For catheter guidance
- **Uses a Continuous X-Ray Beam**
- **Display Screen shows *Live Video***

X-Ray Fluoroscopy



X-Ray Fluoroscopy: Typical Architecture

- Single Point Source
- Large Area Detector

Concerns:

- Detector very Close to Patient
- Obstructs Access
- Scatter Degrades Images
- Radiation Hazard for Medical Personnel

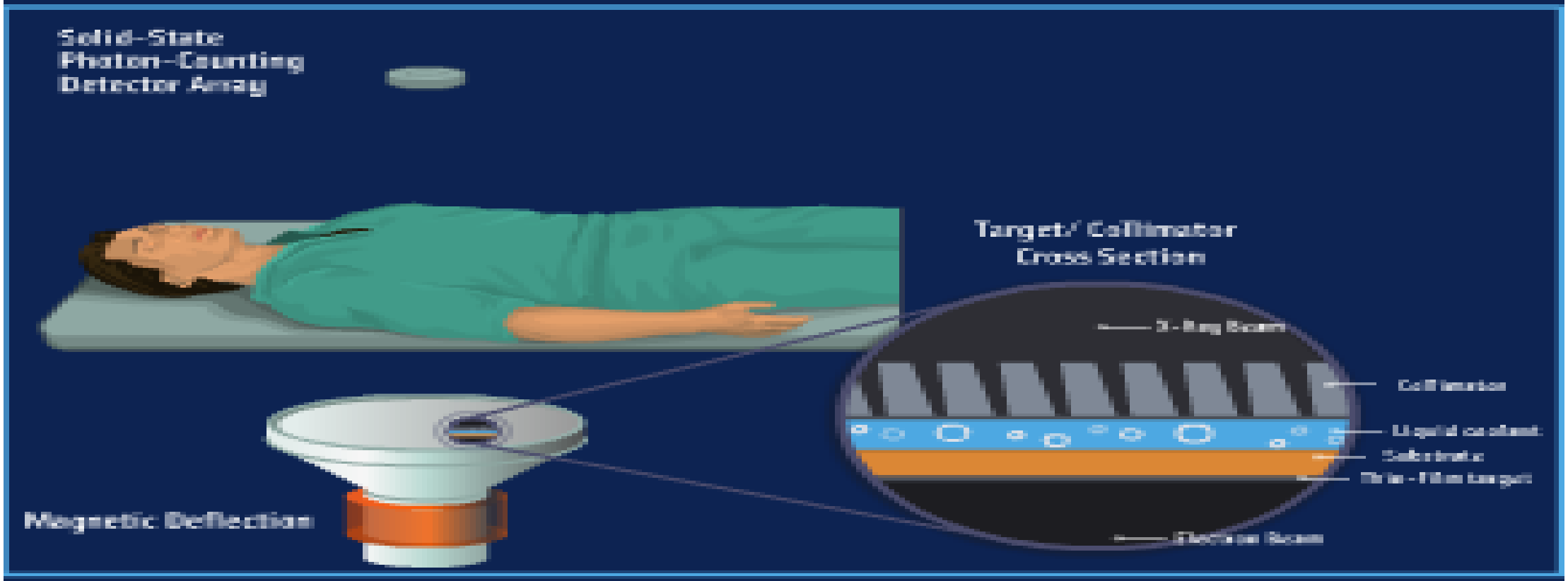
X-Ray Fluoroscopy

SBDX:
Scanning Beam Digital X-Ray

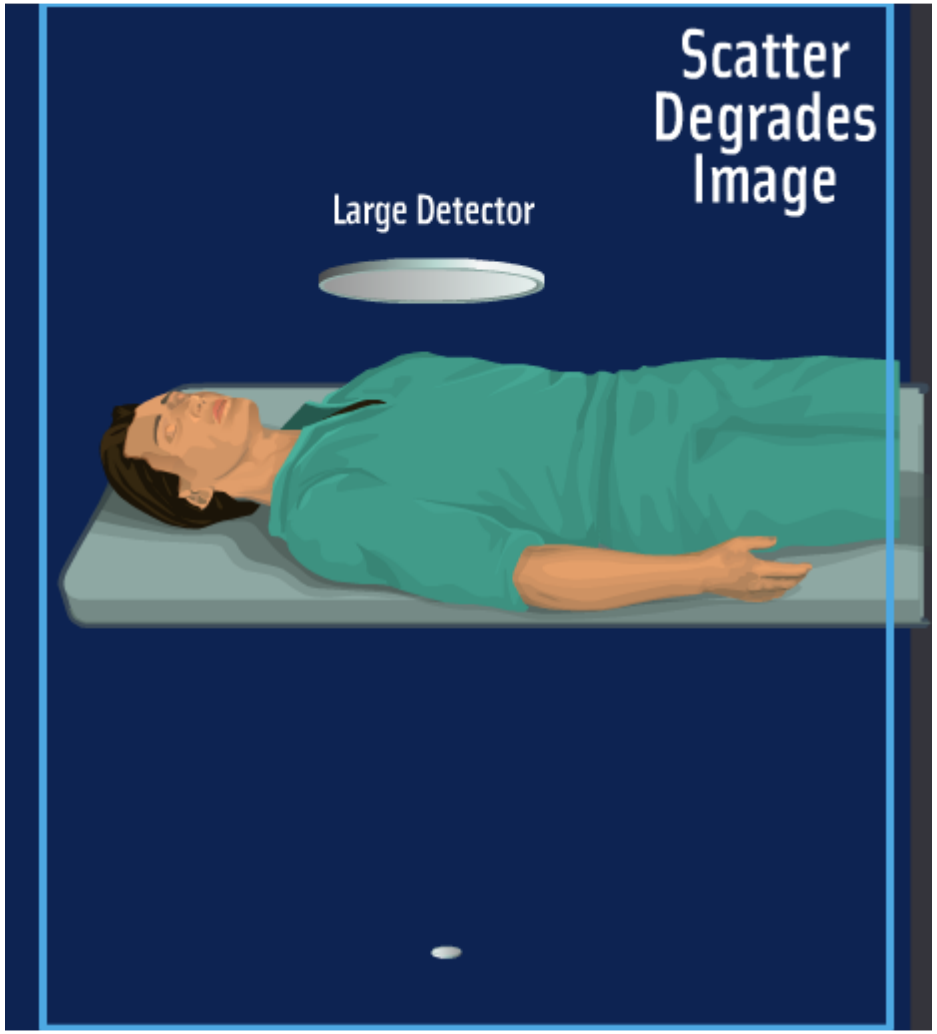
Scanning Beam Fluoroscopy

- **2D Array of X-Ray Sources**
  - Collimators
  - Activated One at a Time
  - Raster Scan Fashion
- **2D Detector Array**
  - Orthogonal Grid Topology
  - Readings Taken for Every Collimator
- **Each Collimator Covers a Small Portion of the Imaging Area**

Collimator Array



SBDX: Improved Image Quality



Point Source

Scanning Beam

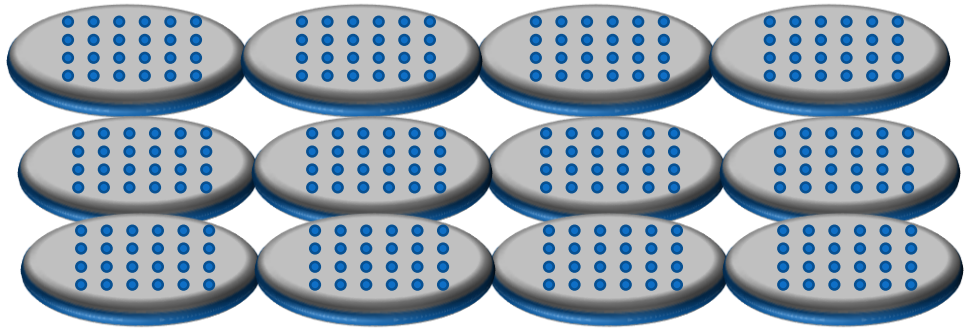
SBDX: Scanning Beam Digital X-Ray

- **Prototype System Designed at University of Wisconsin, School of Medicine**
  - *References:*
    - M.A. Speidel, B.P. Wilfley, J.M. Star-Lack, J.A. Henue, M.S. Van Lysel, 'Scanning Beam Digital X-Ray (SBDX) Technology for Interventional and Diagnostic Cardiac Angiography', Medical Physics, Vol. 33, No. 8, August 2006.
    - E.G. Solomon, B.P. Wilfley, M.S. Van Lysel, A.W. Joseph, J.A. Henue, 'Scanning Beam Digital X-Ray (SBDX) System for Cardiac Angiography', SPIE Medical Imaging Conference, February, 1999.

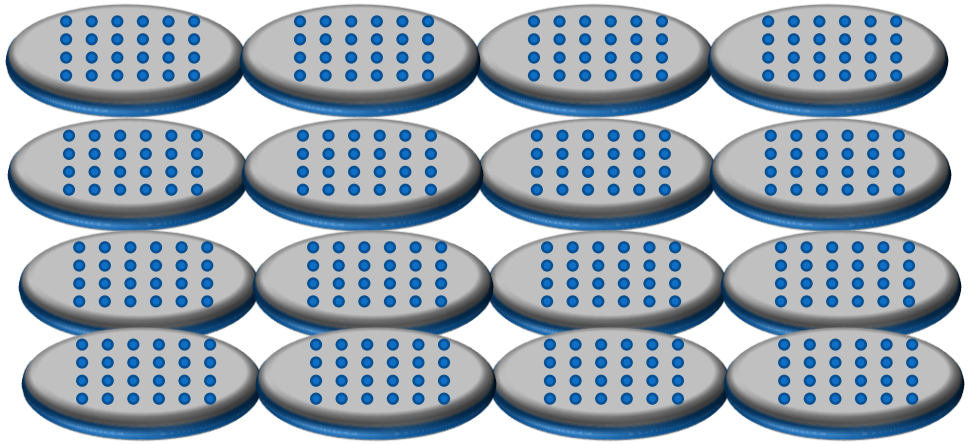
NovaRay Medical: ScanCath™

Problem Definition: Image Reconstruction

Input: Stream of Detector Data Samples



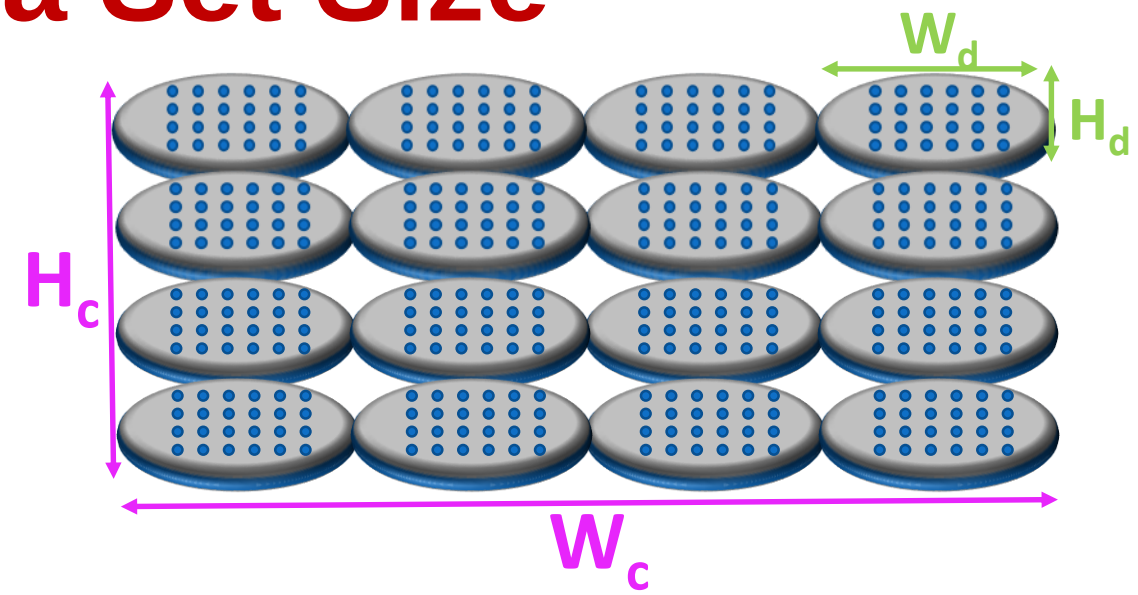
Input: Stream of Detector Data Samples



A Detector Data Set for each Collimator Hole

Each Data Set Covers a Small Portion of the Entire Imaging Area

Input Data Set Size



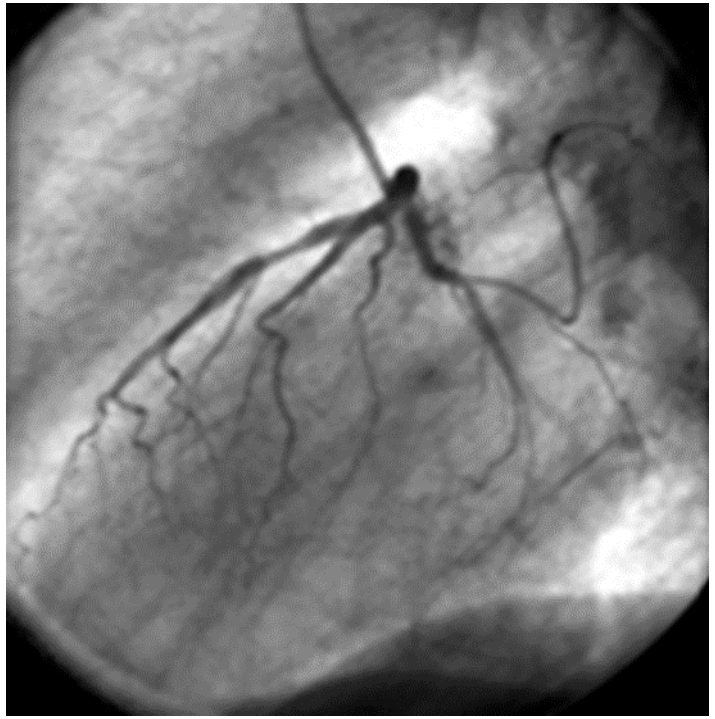
- **Detector Array: W_d x H_d Elements (160 x 80)**
- **Scanning Source: W_c x H_c Collimators (100 x 100)**
- **1 byte per sample**
- **Each data set: Approximately 128 Mbytes (1 Gbits)**
  - (160 * 80) * (100 * 100)

Output: Reconstructed Image Planes

- **32 (N) Distinct Image Planes**
  - Different Positions within Patient
  - Resolution: 1 Mega Pixel: 1000 x 1000
  - Output Data Size:
    - Approximately 128 MBytes
    - 4 Bytes per pixel

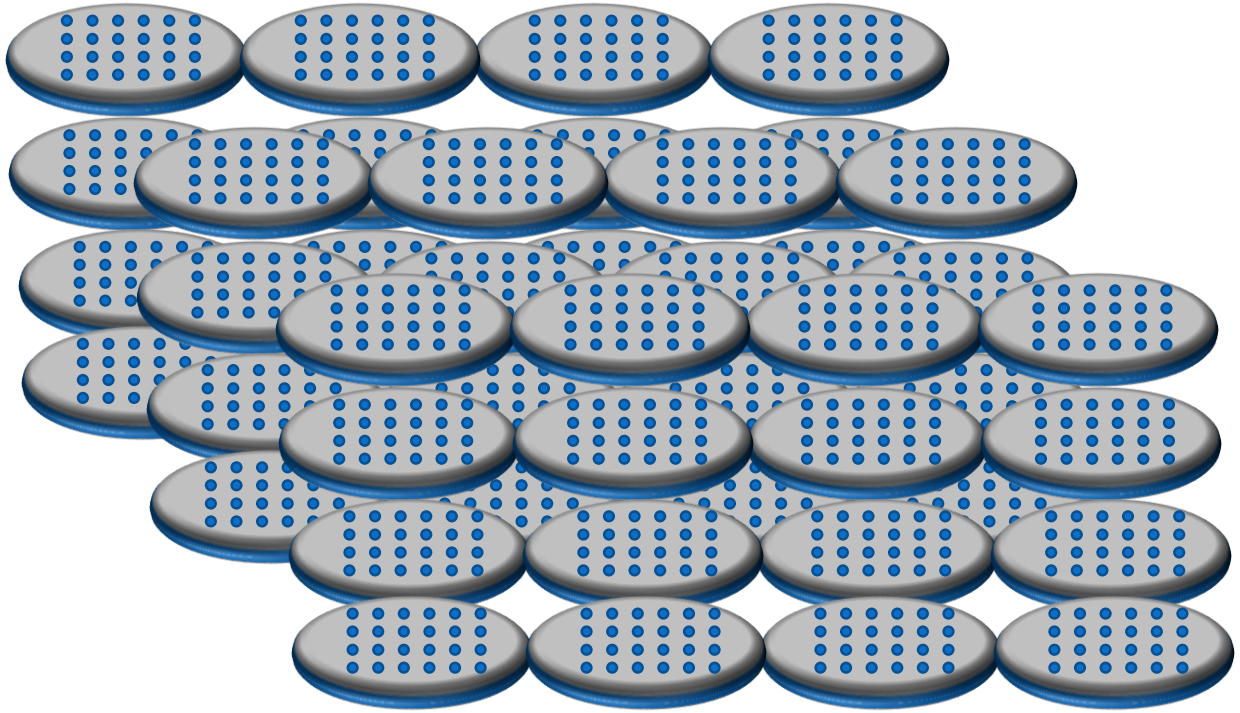
- **Each input data set must generate 32 (N) different output images**

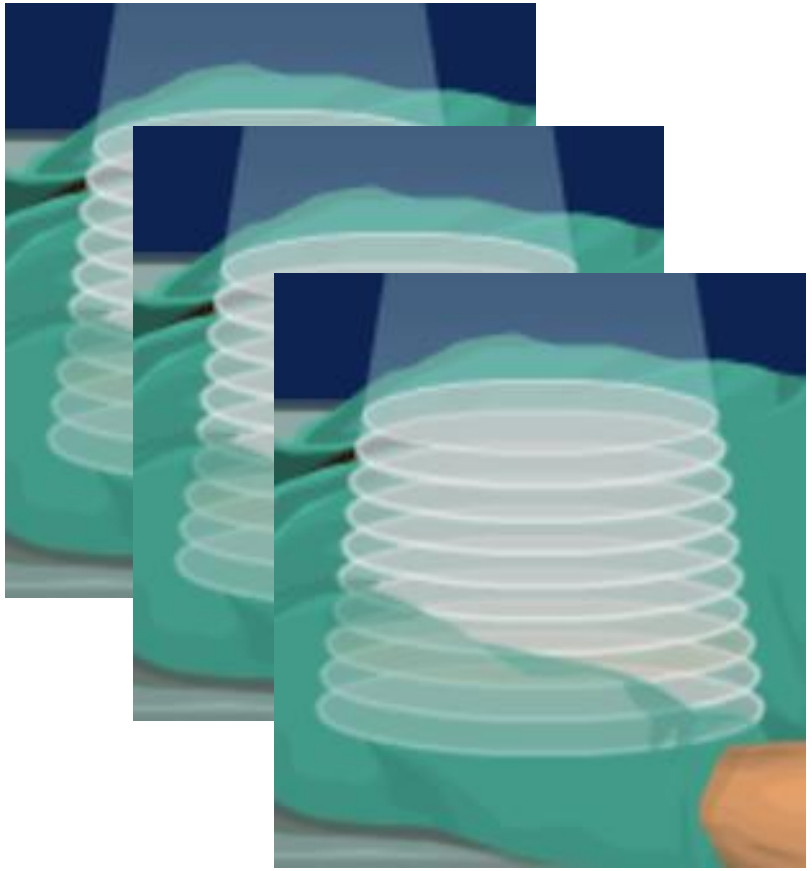
Multi-Slice Reconstruction



Each Slice Shows Some Anatomical Objects in Focus

Continuously Arriving Data Sets





Computational Requirements

- **Desired Frame Rate: 30 fps**
- **Input Data rate: 30 Gbits/sec**
  - 1 Gbits/data set * 30 frames
- **Output Data rate: 30 Gbits/sec**
  - 128 MBytes * 30
- **Computations: 1.5 Tera Flops**
  - 8 multiplications + 4 adds per pair of (detector element, collimator hole)
- **Memory Bandwidth: 245 Gbps**

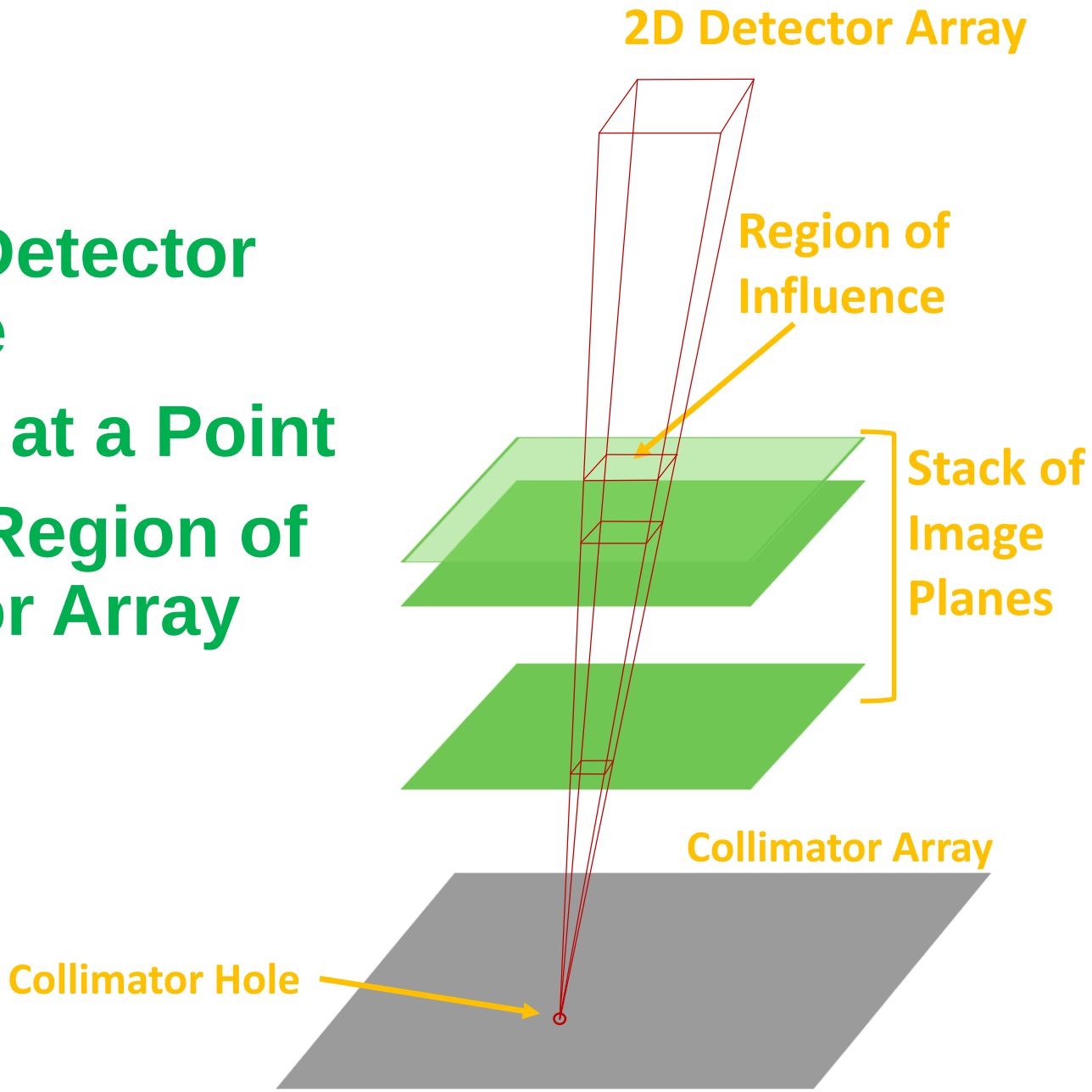
Image Reconstruction: Computational Steps

Fundamental Steps

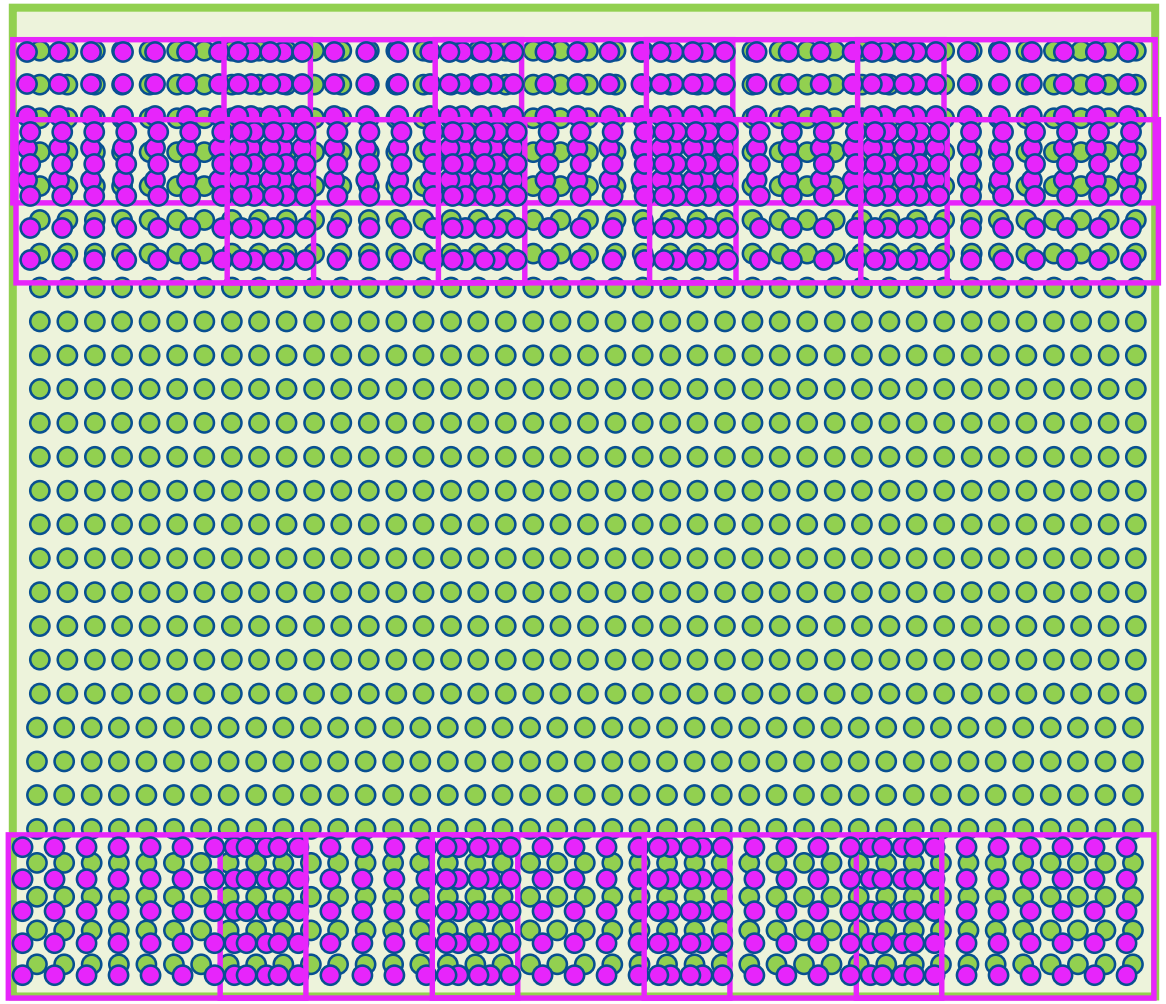
- **Step 1: Back-projection**
  - With One Collimator Hole Lit, Compute the Contribution to an Image Plane
- **Step 2: Accumulation**
  - Aggregate the Back Projection from all Collimators

Back Projection

- **A Straight Line Connects a Detector Element to a Collimator Hole**
- **Intersects each Image Plane at a Point**
- **Defines a Rectangular ROI (Region of Influence) for the 2D Detector Array**
  - Reconstruction Window
  - Reconstruction Ratio *n*



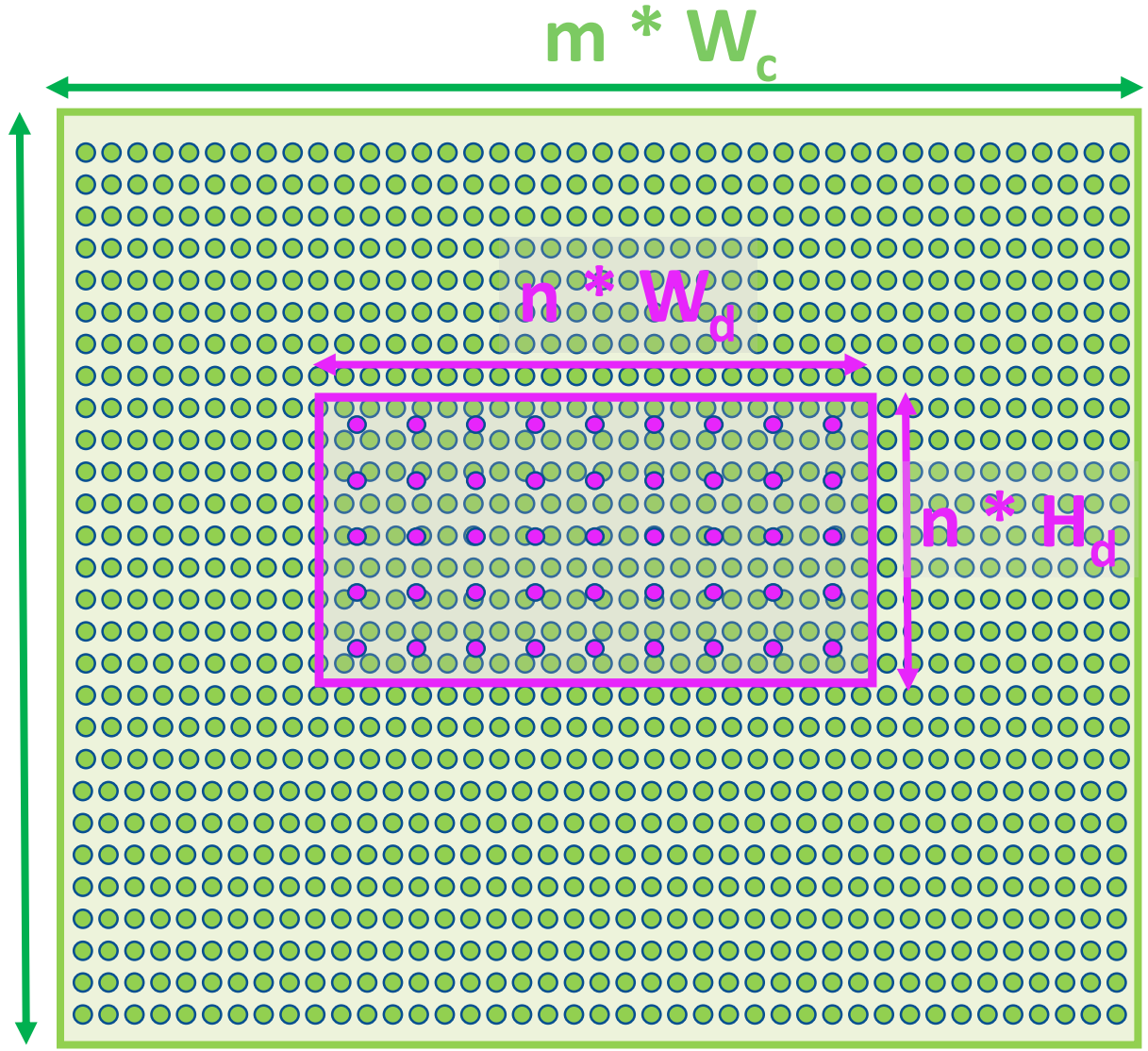
Reconstruction Windows



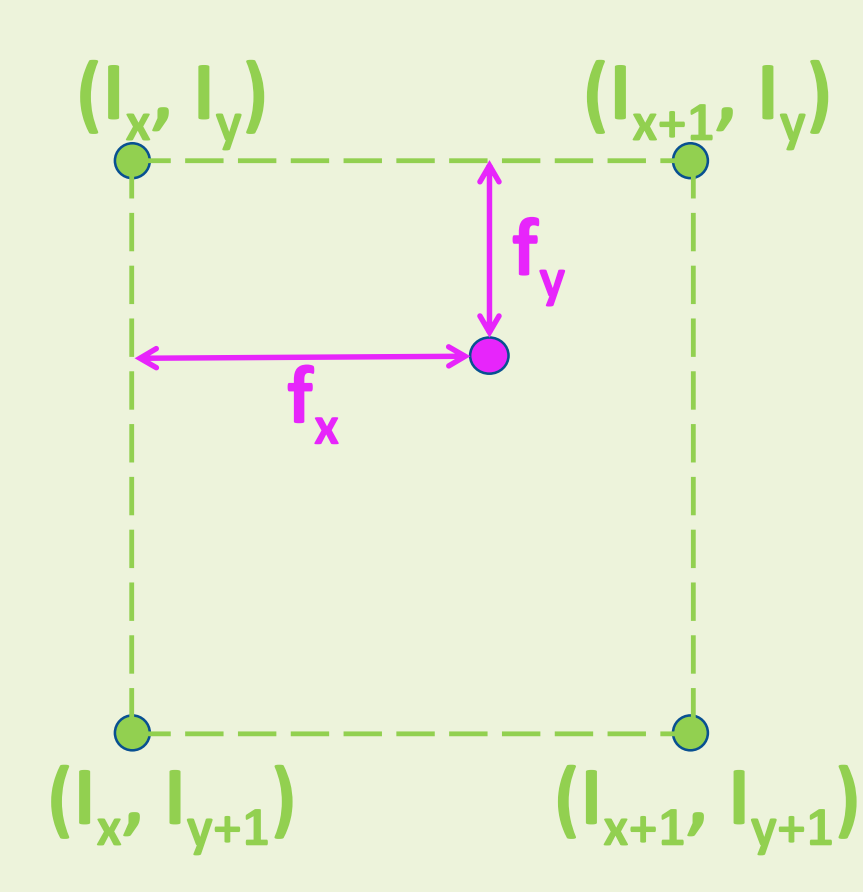
- **For a Given Image Plane**
  - Reconstruction Windows Proceed in Raster Scan Fashion
  - All Reconstruction Windows have the Same Size and Pattern
  - There is Overlap Between Reconstruction Windows
- **Across Different Image Planes**
  - Different Sizes of Reconstruction Windows
  - Size Increases with Distance from Collimator Array

Reconstruction Windows

- *n*: Reconstruction Ratio
- Constant for an Image Plane
- Increases as Image Plane moves away from Collimator Array
- *m*: Collimator Offset Constant for a Given Image Plane Resolution



Local View of Reconstruction



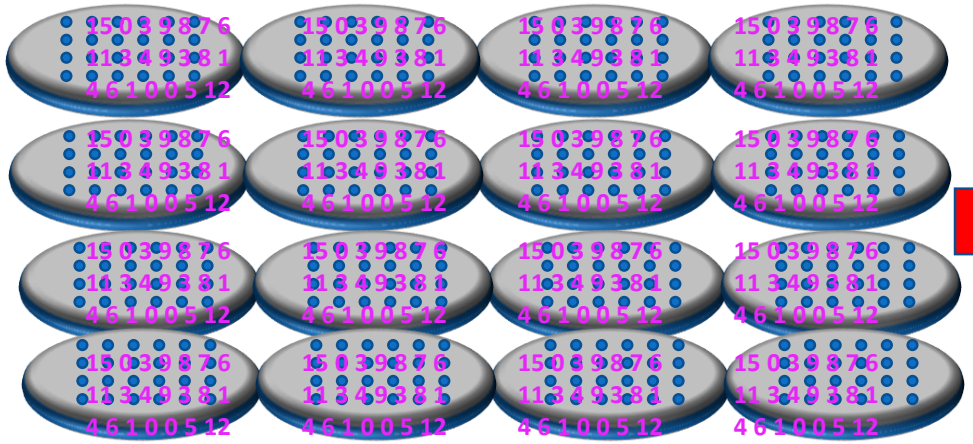
Weights (Bilinear Coefficients):

(I_x, I_y): $(1-f_x) * (1-f_y)$

(I_{x+1}, I_y): $f_x * (1-f_y)$

(I_x, I_{y+1}): $(1-f_x) * f_y$

(I_{x+1}, I_{y+1}): $(1-f_x) * (1-f_y)$



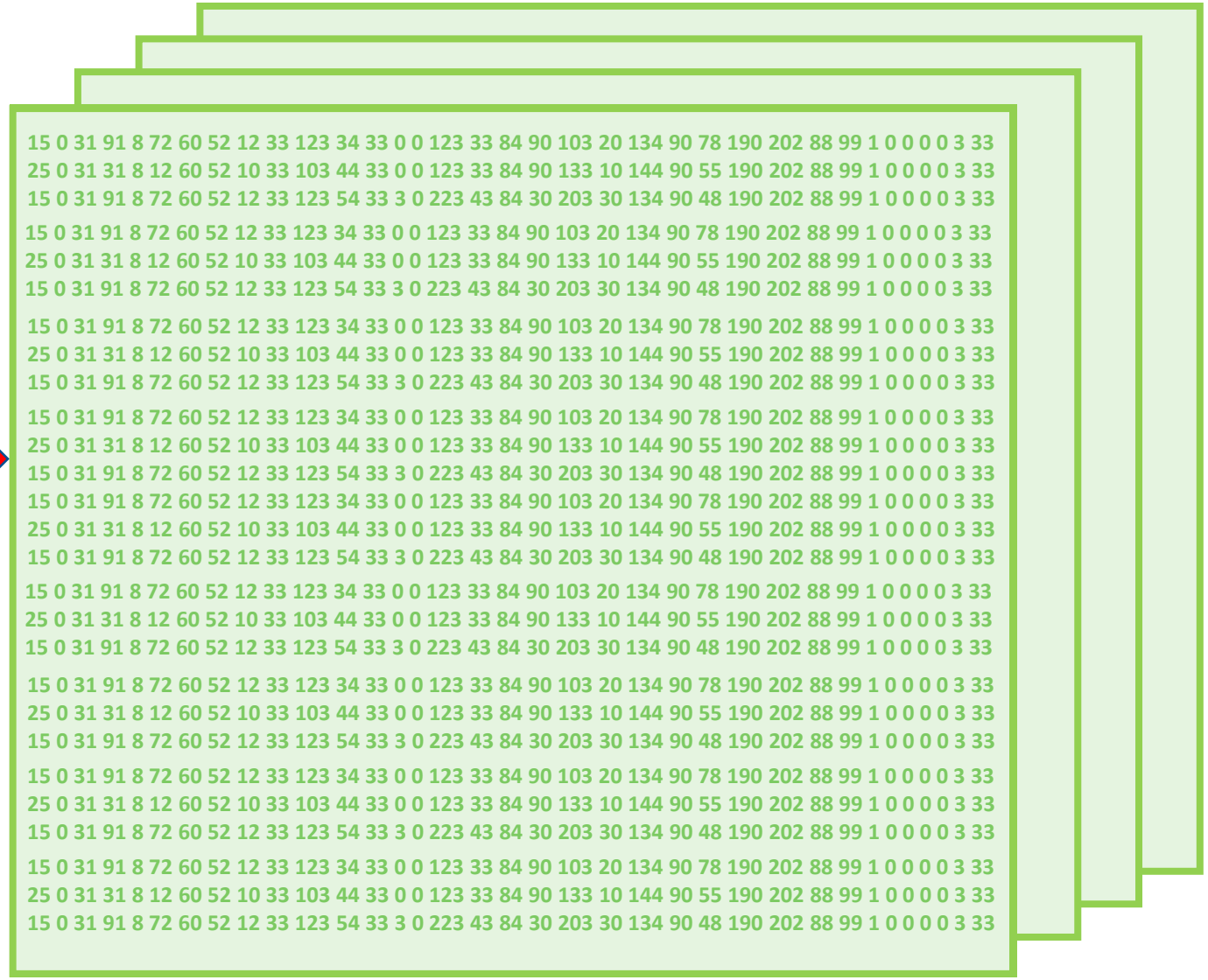


Image Reconstruction: Implementation

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

15 0 31 91 8 72 60 52 12 33 123 34 33 0 0 123 33 84 90 103 20 134 90 78 190 202 88 99 1 0 0 0 0 3 33
25 0 31 31 8 12 60 52 10 33 103 44 33 0 0 123 33 84 90 133 10 144 90 55 190 202 88 99 1 0 0 0 0 3 33
15 0 31 91 8 72 60 52 12 33 123 54 33 3 0 223 43 84 30 203 30 134 90 48 190 202 88 99 1 0 0 0 0 3 33

Image Reconstruction: Data Flow

- **Objective**
  - For each pixel, compute the grayscale value
  - Aggregate of detector contributions in neighborhood

- **All the Detector Data is Ready in Memory Before Reconstruction Begins**

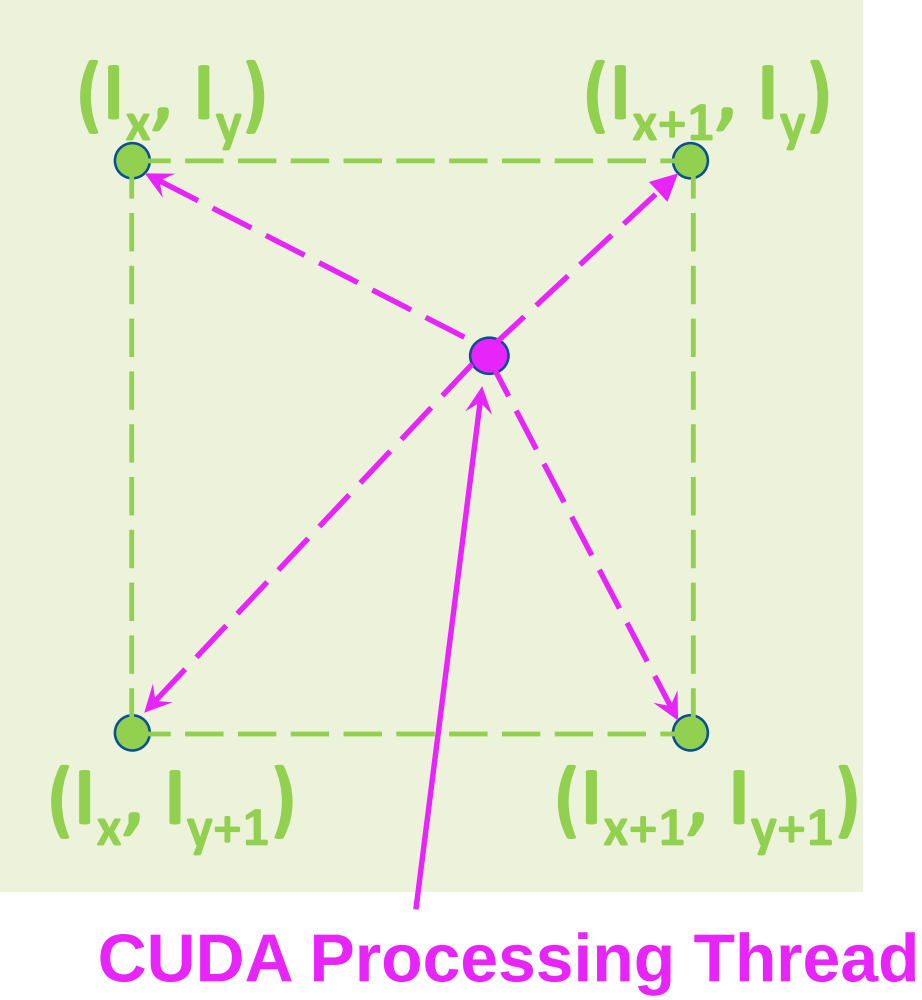
Parallel Implementation with CUDA

Scope For Parallelism

- **Parallelization Schemes**
  - Different image planes can be reconstructed concurrently
  - Within a plane, every pixel can be reconstructed concurrently
  - Every detector reading can be processed concurrently

- **Data Mapping**
  - Input Detector Data: Texture Memory
  - Output Pixel Data: Global Memory

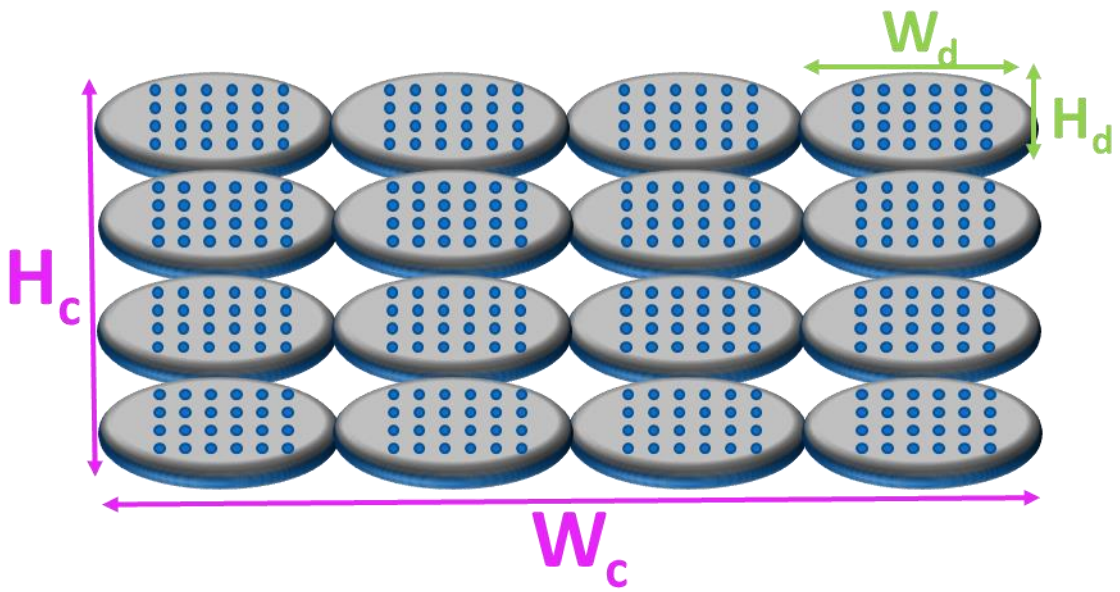
Detector Centric View



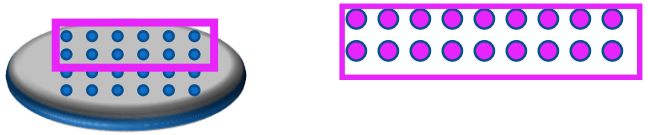
- **Each Processing Thread Handles a Single Detector Reading**
  - For all Collimator Holes
- **Main Computation: Apportion the Detector's Output to the 4 Nearest Pixels**

Data Push Model

CUDA Block and Grid Configuration

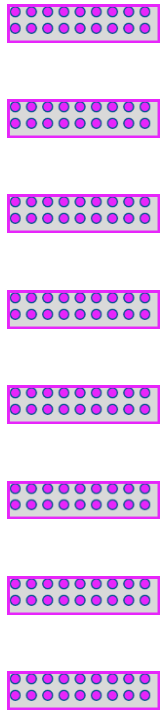


- **CUDA Blocks for a Subset of Detector Data Elements**
  - 2 D Block: W_d x H_b
  - Maximum of 1024 Threads

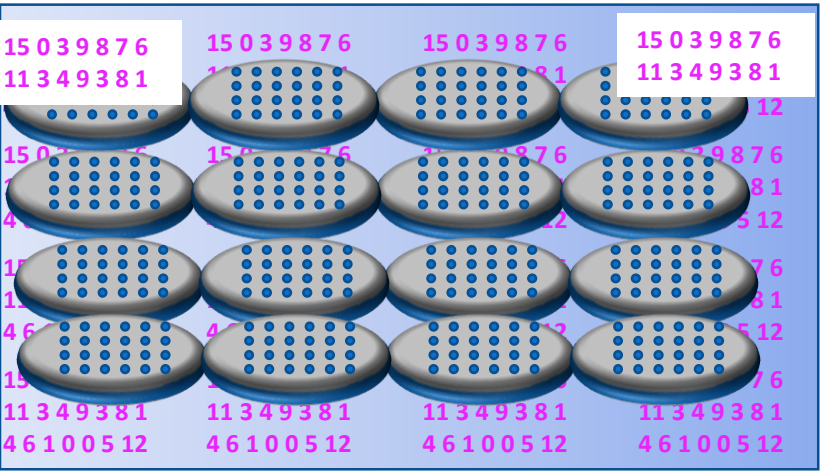


- **Blocks Organized in a 2D Grid**
  - H_d / H_b x N

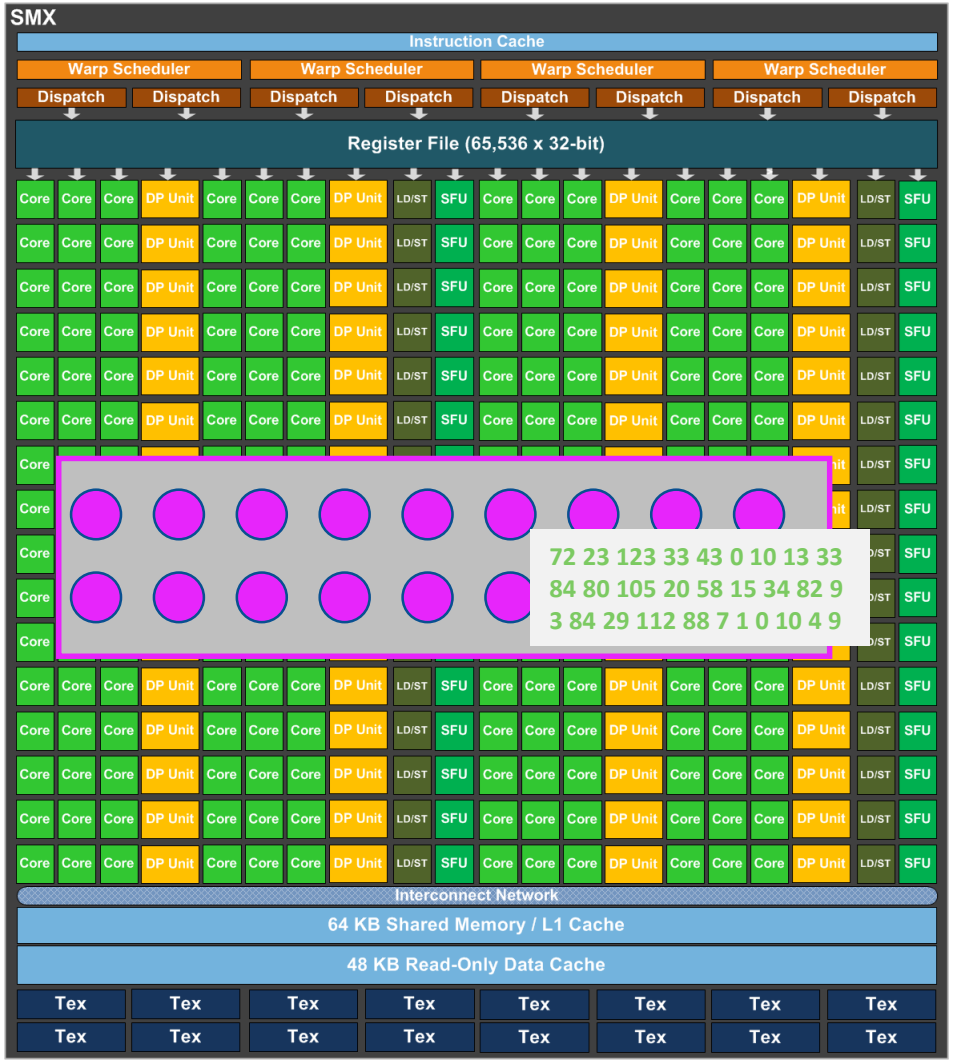
Computational Mapping

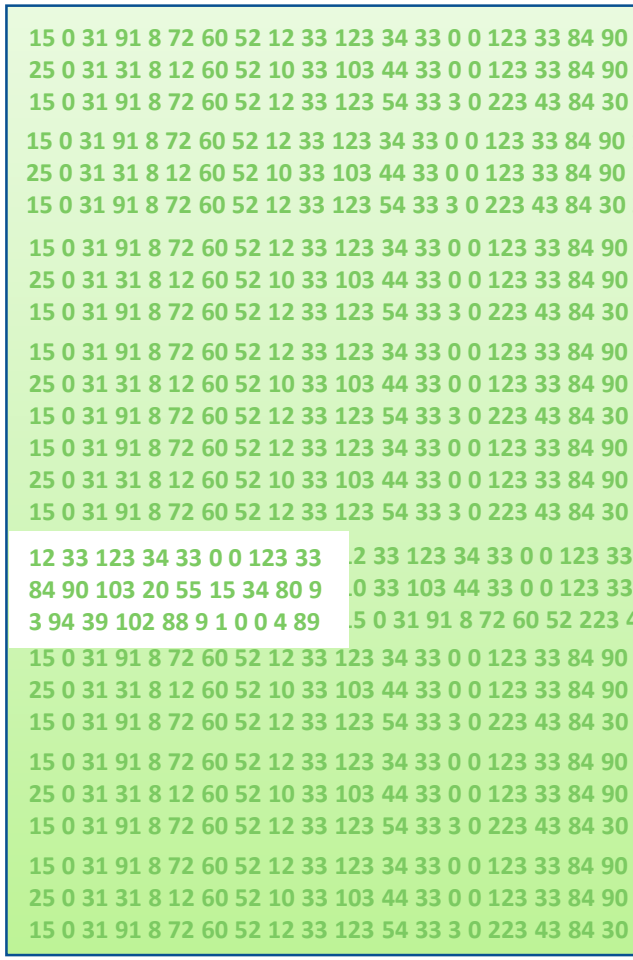


Computational Flow



Texture Memory



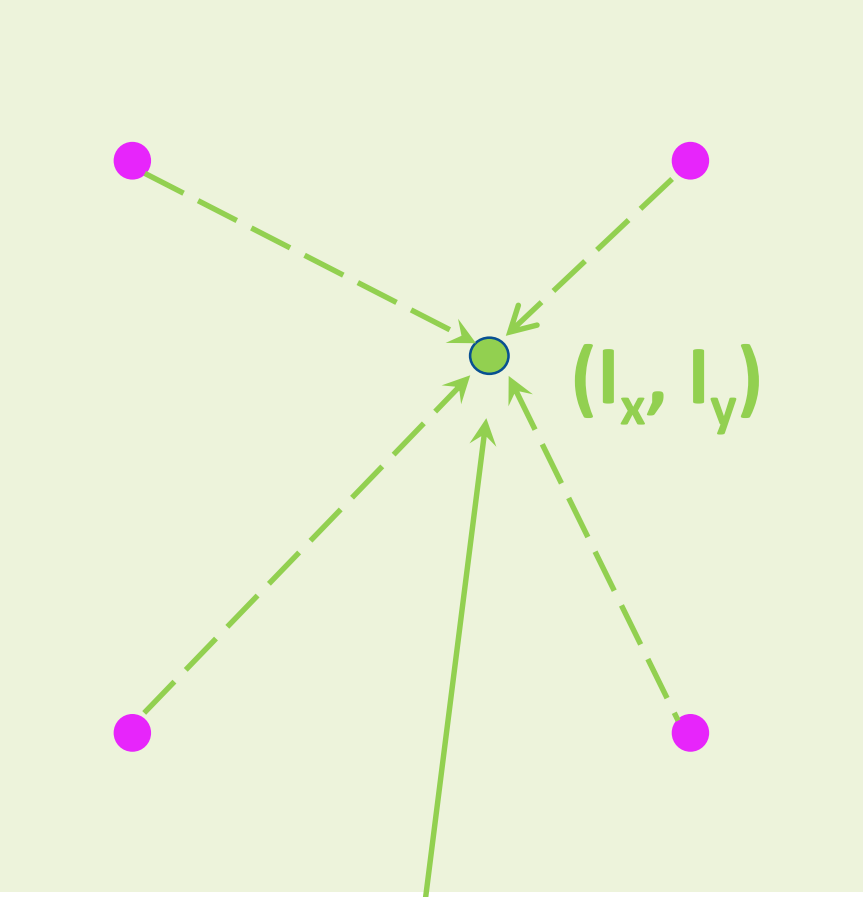


Global Memory

Detector Centric Reconstruction

- Non-Local and Asymmetric Writes
- Memory Write Contentions
- Poor Arithmetic Intensity
- Multi-GPU Scaling Cumbersome

Pixel Centric View



CUDA Processing Thread

- **Each Thread Handles the Reconstruction of a Single Pixel**
  - From all Detector Elements
- **Main Computation: Update the Pixel Value Using the Readings of Detectors Closest to it**

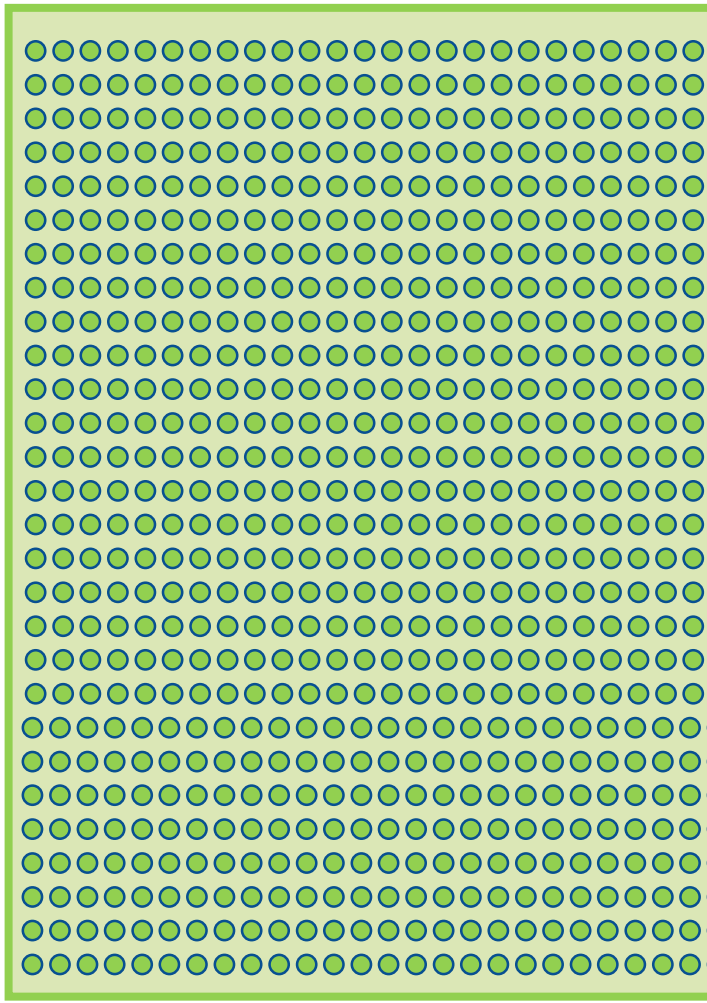
Data Pull Model

CUDA Block and Grid Configuration

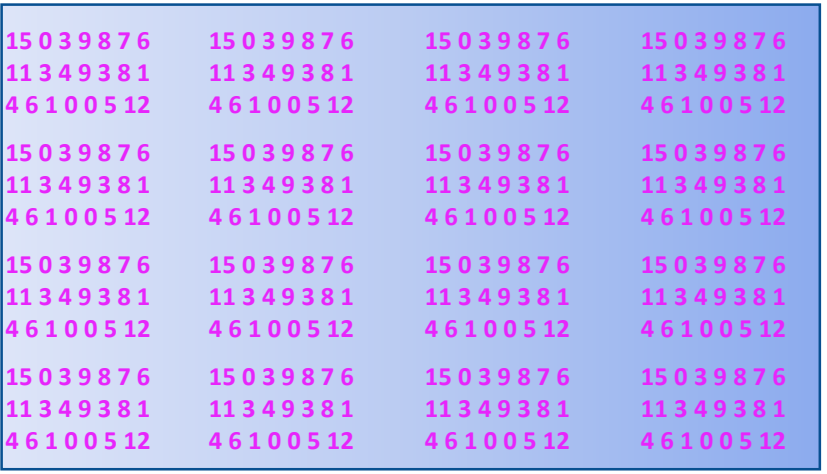
- **A CUDA Block for each Row of Pixels**
  - 1 D Block: 1000 Threads

- **Blocks Organized in a 2D Grid**
  - 1000 x N

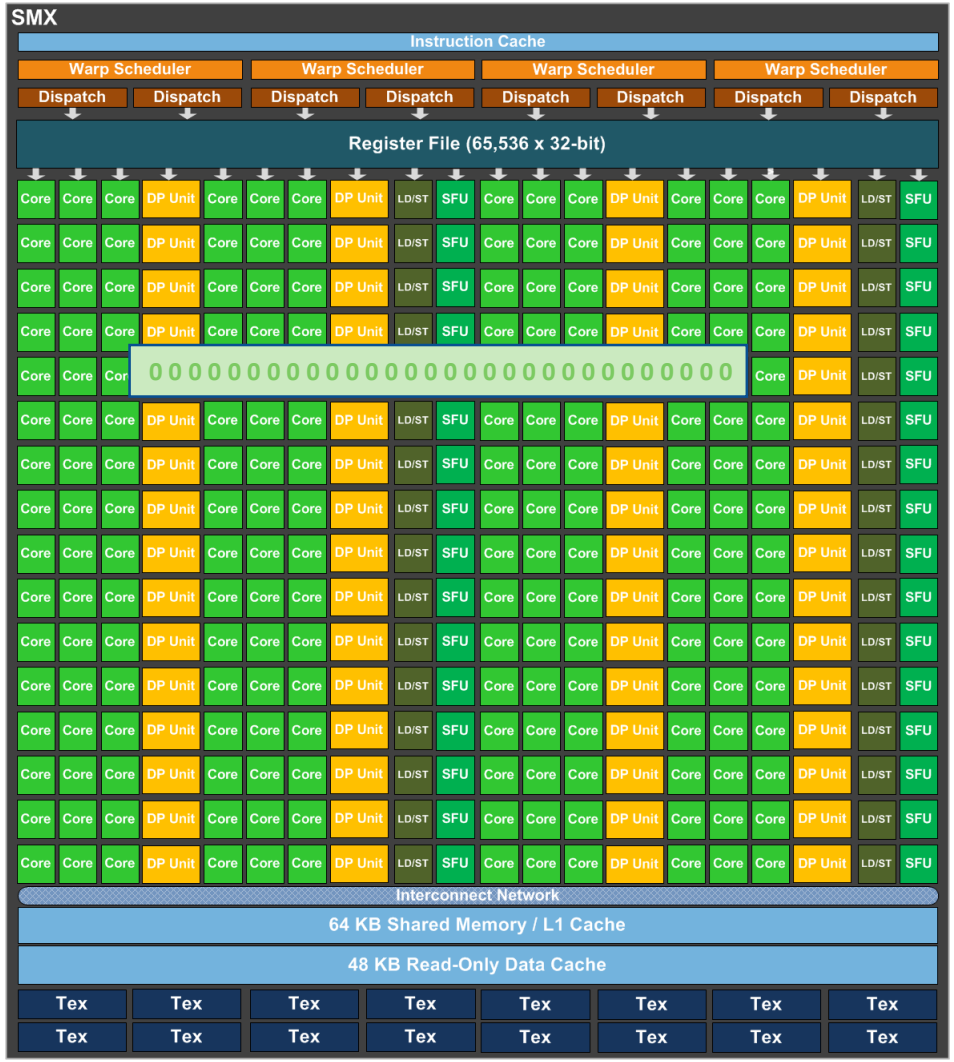
Computational Mapping

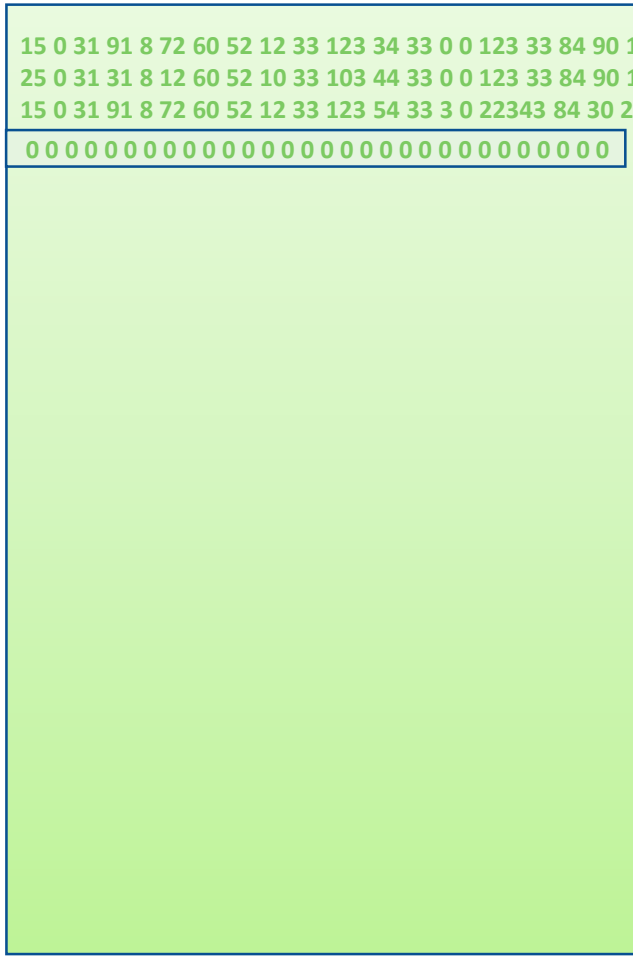


Computational Flow



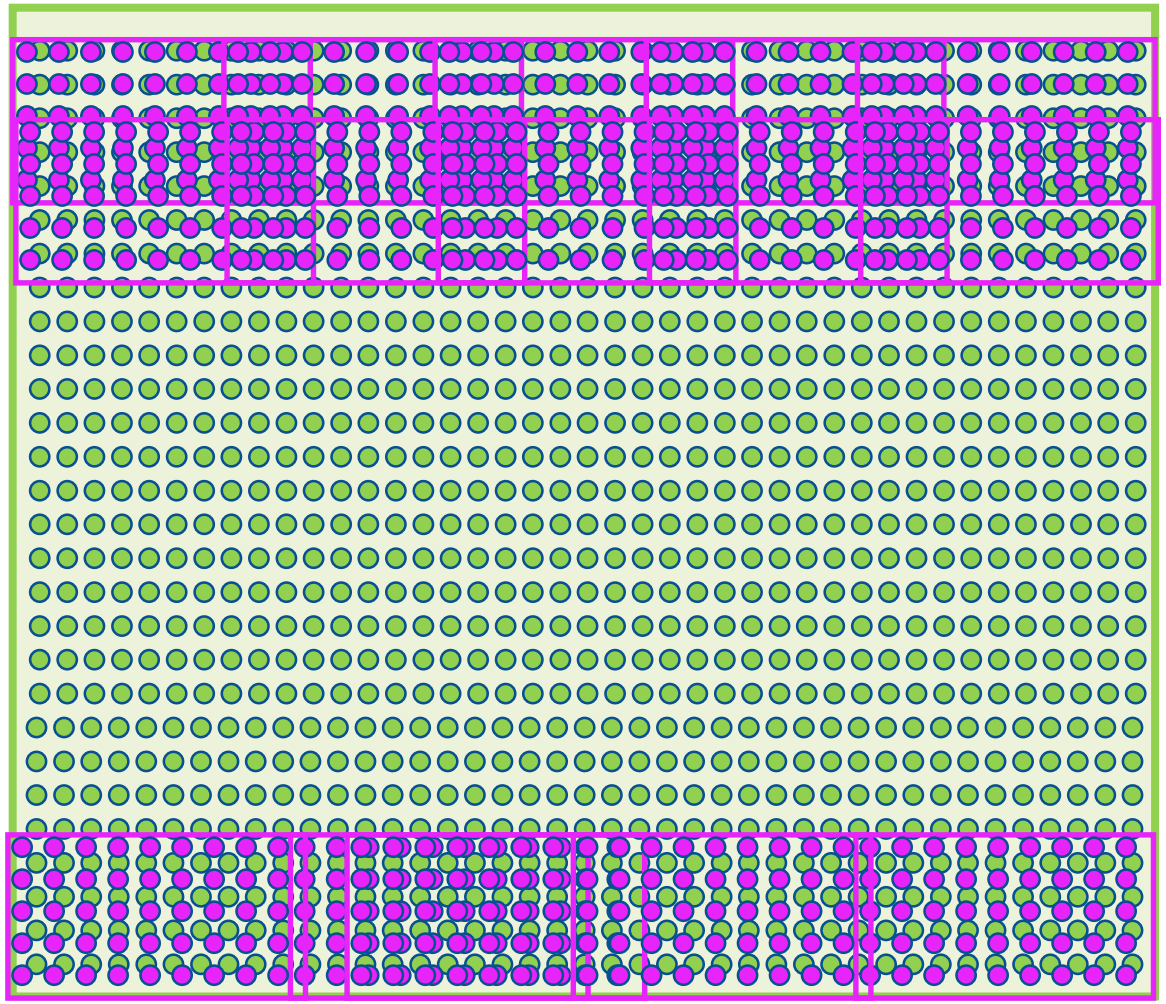
Texture Memory





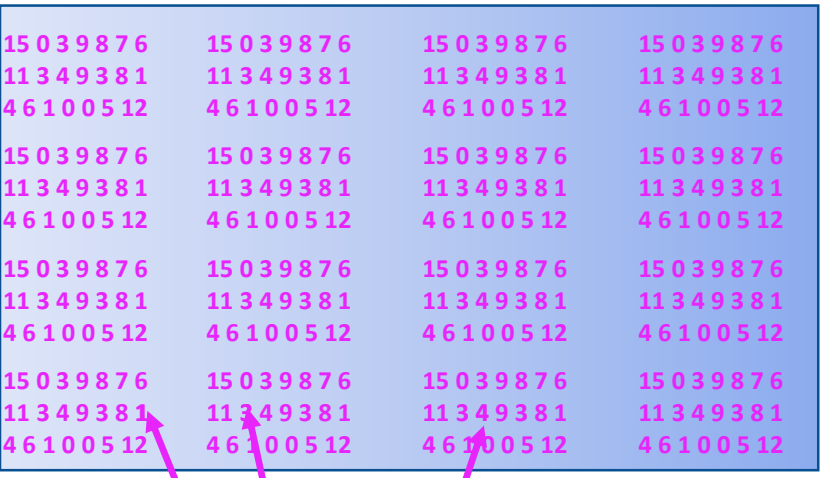
Global Memory

Pixel Centric Reconstruction: Observations



- Each Pixel is Affected by a Range of Collimators
- Across Collimators, the Affected Pixels Have the Same Pattern
- Use Look Up Tables (LUTs) to Remember these, Avoiding Repeated Computations
- Save LUTs in Shared Memory

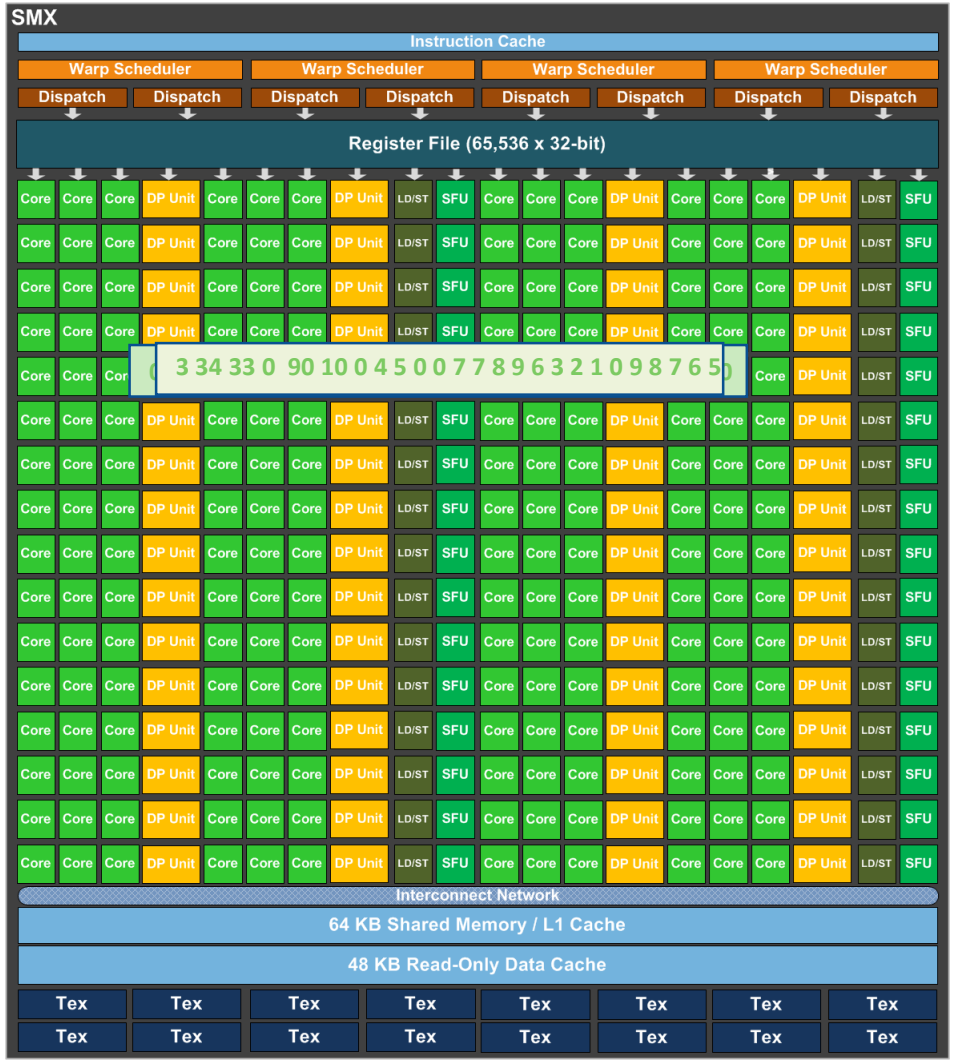
Computational Flow

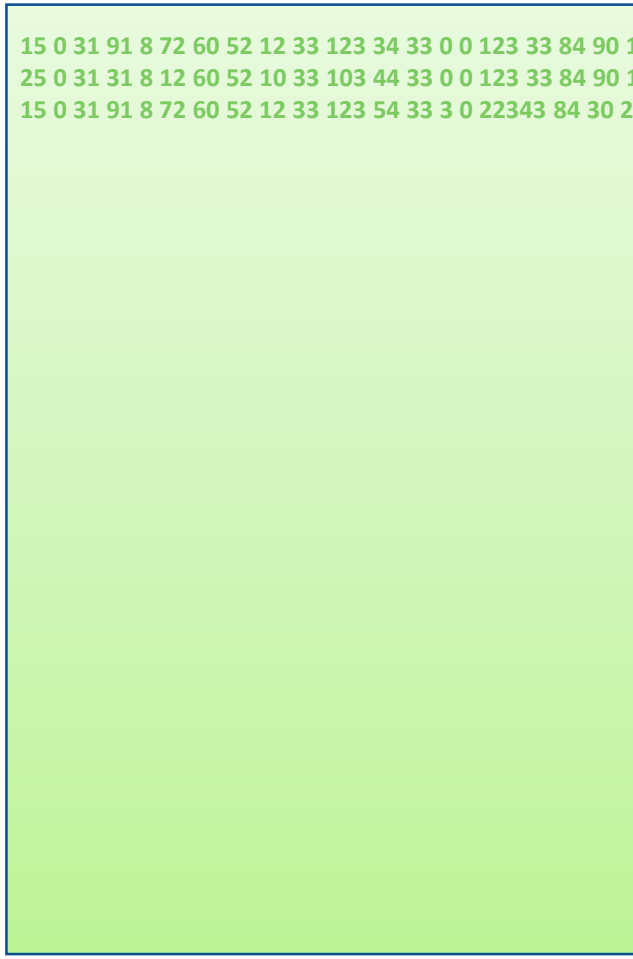


Texture Memory

Look Up Tables

Shared Memory





Global Memory

Pixel Centric Reconstruction

- Redundant Memory Writes Avoided
- Four Memory Write Contentions/pixel
- Better Arithmetic Intensity
- Easier Multi-GPU Scaling

Performance Results

Kepler GPU Specifications

	Tesla K10 1 card 2 GPU	Tesla K20 1 card 1 GPU	GeForce GTX 690 1 card 2 GPU
No. of cores	2 x 1536 = 3072	2496	2 x 1536 = 3072
GPU Clock rate	745MHz (0.75GHz)	704MHz (0.71 GHz)	1020MHz (1.02GHz)
Global Memory Size	2 x 4 = 8 GB	5 GB	2 x 2 = 4 GB
Single Precision	4.58 T Flops	3.52 T Flops	4.58 T Flops
Double Precision	190 G Flops	1.17 T Flops	190 G Flops
Memory Clock rate	2500 MHz	2600 MHz	3004 MHz
Max. No. of threads / MP	2048	2048	2048
Max.No.of threads / block	1024	1024	1024
Shared Memory / block	49152 bytes	49152 bytes	49152 bytes

Performance Experiments

- **Measure Execution Time for Image Reconstruction**
- **3 Kepler GPU Models:**
  - Tesla K10, Tesla K20, GeForce GTX 690
- **N = 32 planes, 1000 x 1000 pixels each**
- **Detector Array Size: 160 x 80**
- **Collimator Array: 100 x 100**
- **Detector Inputs: Random numbers in the range 0 - 15**

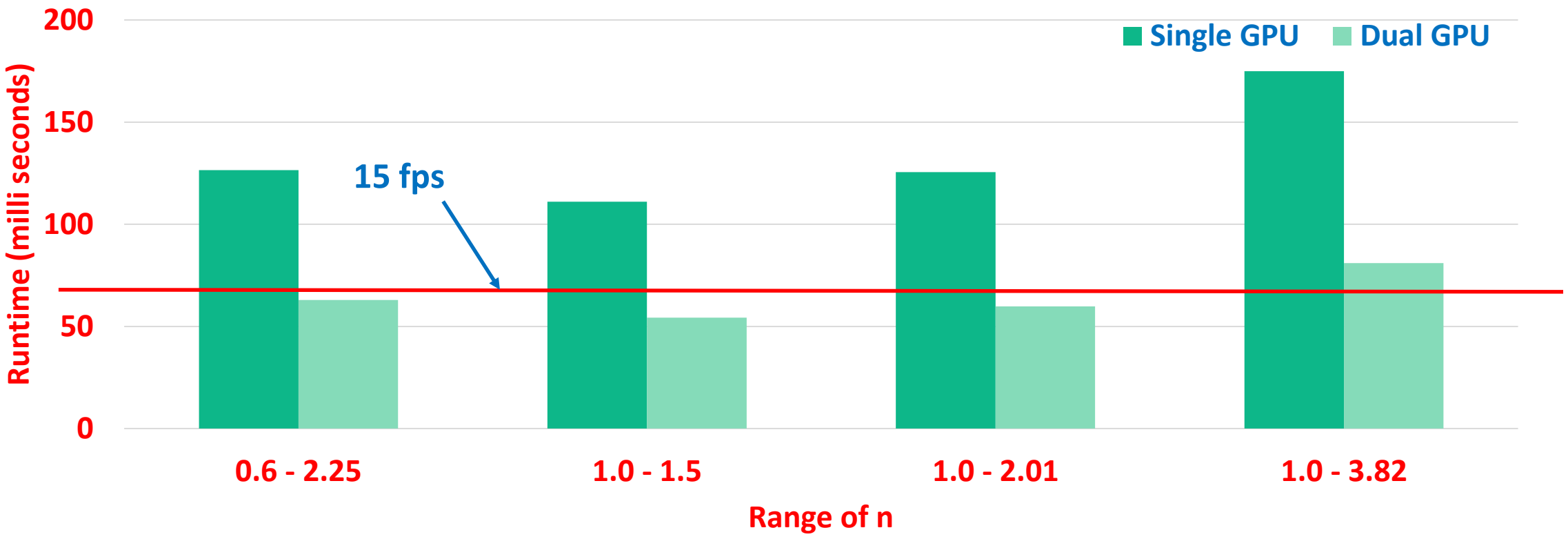
Performance: Detector Centric

Range of n	Run time on 1 GPU of GTX690 (msec)	Run Time On Q8400 CPU (sec)	Speed Up
0.60 – 2.25	1029.64	394.85	383.483
1.00 – 1.50	1039.86	366.90	352.83
1.00 – 2.01	1003.45	387.36	386.02
1.00 – 3.82	966.57	443.21	458.53

Performance: Pixel Centric

Range of n	Run Time on GPUs (msec)					
	Tesla K10		Tesla K20		GeForce GTX 690	
	Single GPU	Dual GPU	Single GPU	Dual GPU	Single GPU	Dual GPU
0.60 – 2.25	155.16	77.11	152.26	74.89	126.55	62.98
1.00 – 1.50	147.10	73.18	145.33	71.60	111.02	54.32
1.00 – 2.01	165.70	82.32	162.46	79.90	125.50	59.76
1.00 – 3.82	227.84	95.66	223.60	109.23	174.94	80.94

Performance: GeForce GTX 690



Linear Speedup: Dual GPU Runtime = 50% of Single GPU Runtime

Observations

- **GTX 690 performed slightly better than Tesla K10**
- **Tesla K10 performed better than Tesla K20**
  - Better single precision Flops rating
- **Linear scaling with 2 GPUs**
- **15 fps with one GPU card (2 GPUs) of GTX 690**
- **When 2 GPUs were used, the image was divided by rows – top half and bottom half**

Summary and Future Work

Summary

- **Application**
  - Scanning Beam X-Ray Fluoroscopy
  - Image Reconstruction
- **CUDA Implementation**
  - Kepler GPUs
  - Single Precision Floating Point
- **Performance**
  - Real time Performance at 15 fps
  - Scalable Parallel Implementation

Future Work: Plane Reduction

- Combine Several Planes into One

Future Work

- **Plane Reduction**
- **Generalize to More Configurations**
  - Non-uniform Detector Arrays
- **Use GPU-Direct for Data Movement**

Thanks to …

- **NVIDIA GPU Test Drive**
  - Tesla K10, Tesla K20
- **NovaRay Medical, Inc.**
  - Introductory Slides with Animations
- **Triple Ring Technologies**
  - Problem Definition

Questions?

- **Email: prashanth.bhat@manipal.net**
- **Web page: http://www.manipal.net**