MathWorks
AUTOMOTIVE CONFERENCE 2023
Europe

Targeting GPUs for Automotive Applications

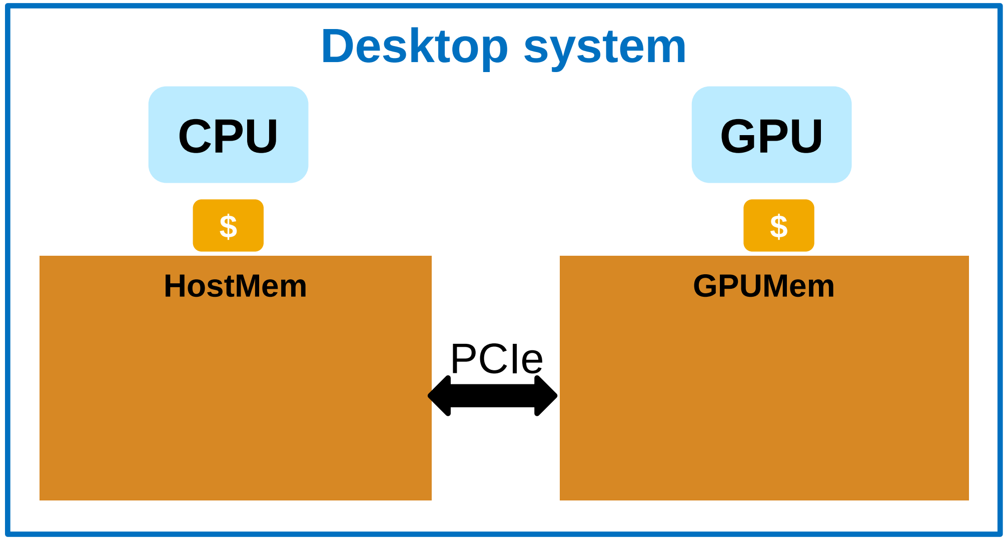
Christoph Stockhammer, MathWorks

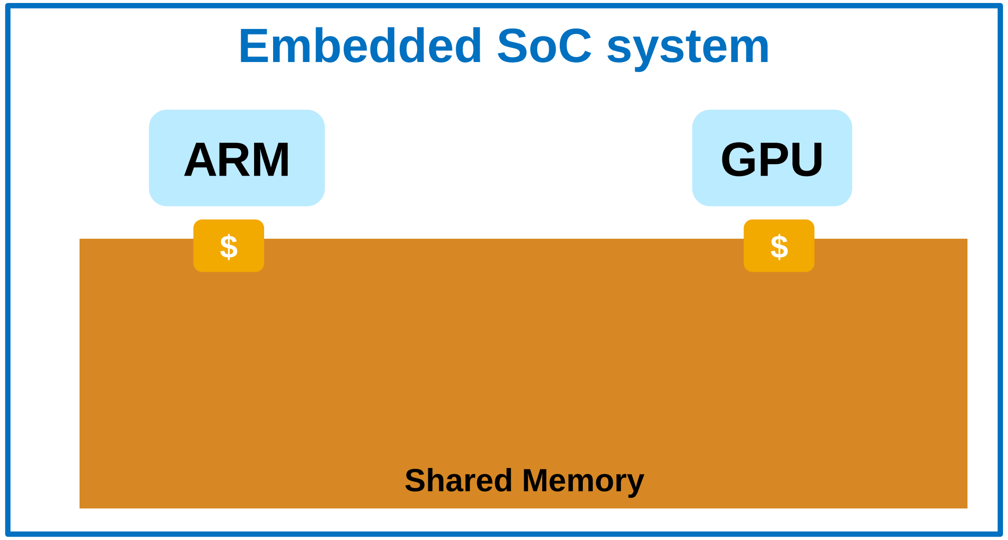
MathWorks

Key Takeaways

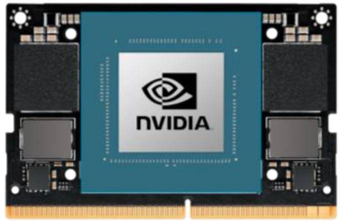
- GPU Coder generates CUDA code from MATLAB & Simulink
- Accelerate MATLAB & Simulink simulations
- Deploy algorithms (signal/deep learning,…) to embedded GPUs

Types of GPUs





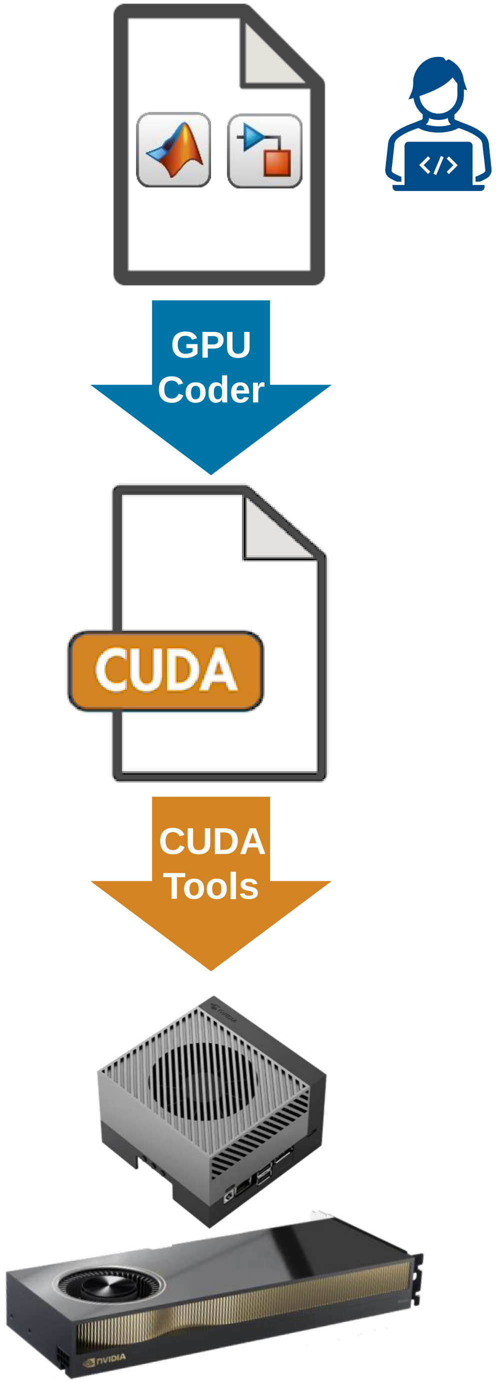
Desktop GPUs
(and Cloud GPUs)



Embedded GPUs

CUDA code generation

- Generate optimized CUDA code from MATLAB and Simulink for deep learning, embedded vision, and autonomous systems
- Generated CUDA is <u>portable</u> across NVIDIA desktop GPUs
- Prototype algorithms on modern GPUs including the A100/ V100 and Jetson AGX Orin
- Accelerate computationally intensive portions of your MATLAB code and Simulink models using generated CUDA code



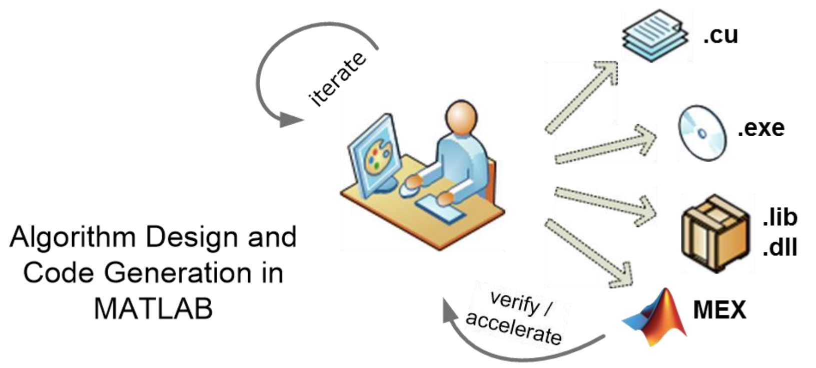
Why Use CUDA code generation ?

Pains: Hand code

- **Cannot code in CUDA**
- Time consuming
- Manual Coding Errors
- Multiple implementations
- Expensive

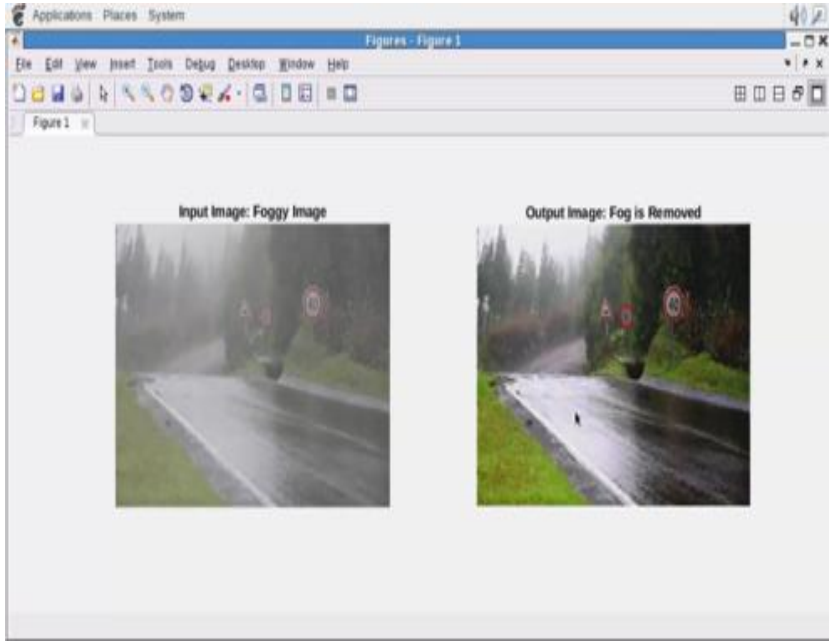
Solution: GPU Coder

- Automatically convert to CUDA
- Get to CUDA faster
- Eliminate manual coding errors
- Maintain Single “Truth”
- Stay within MATLAB/Simulink at a higher level

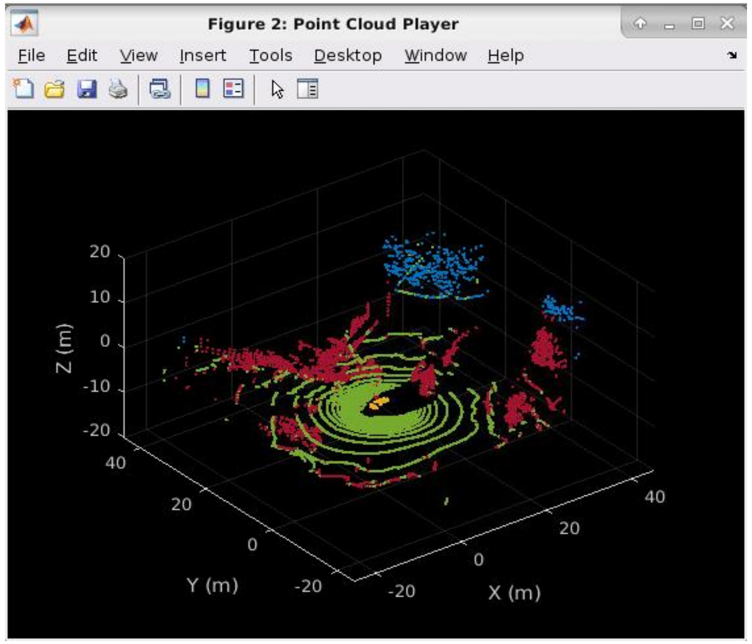


Two Application examples today

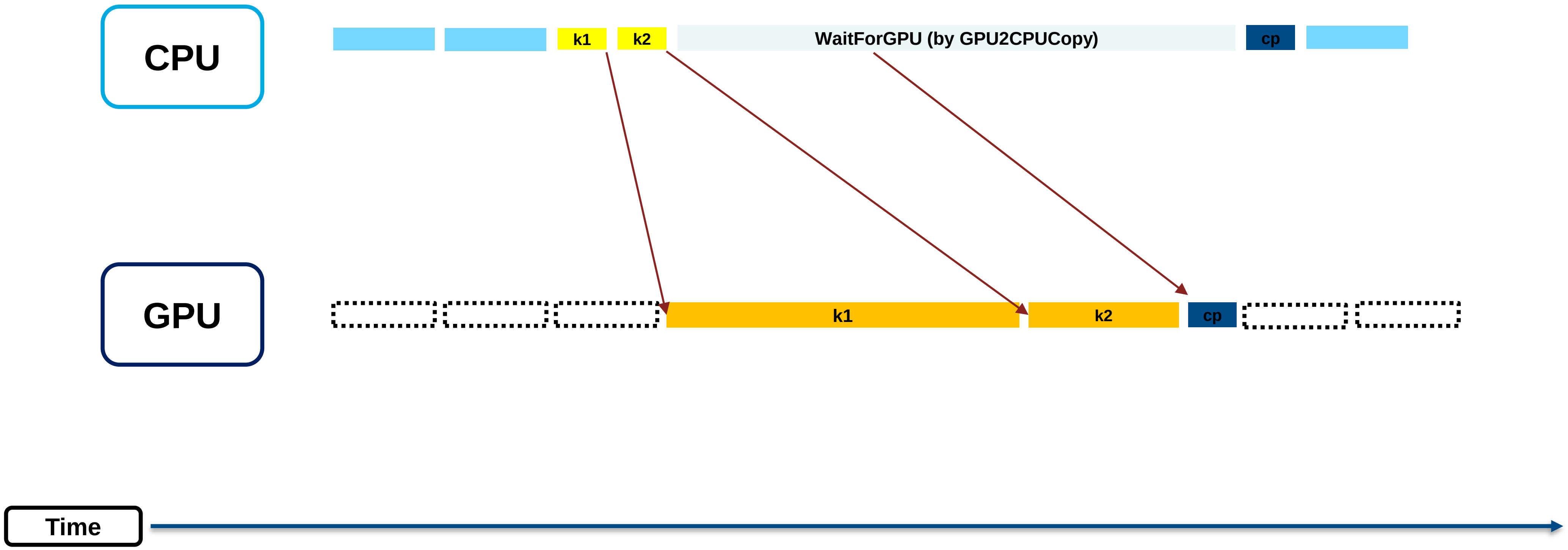
1) Fog Rectification



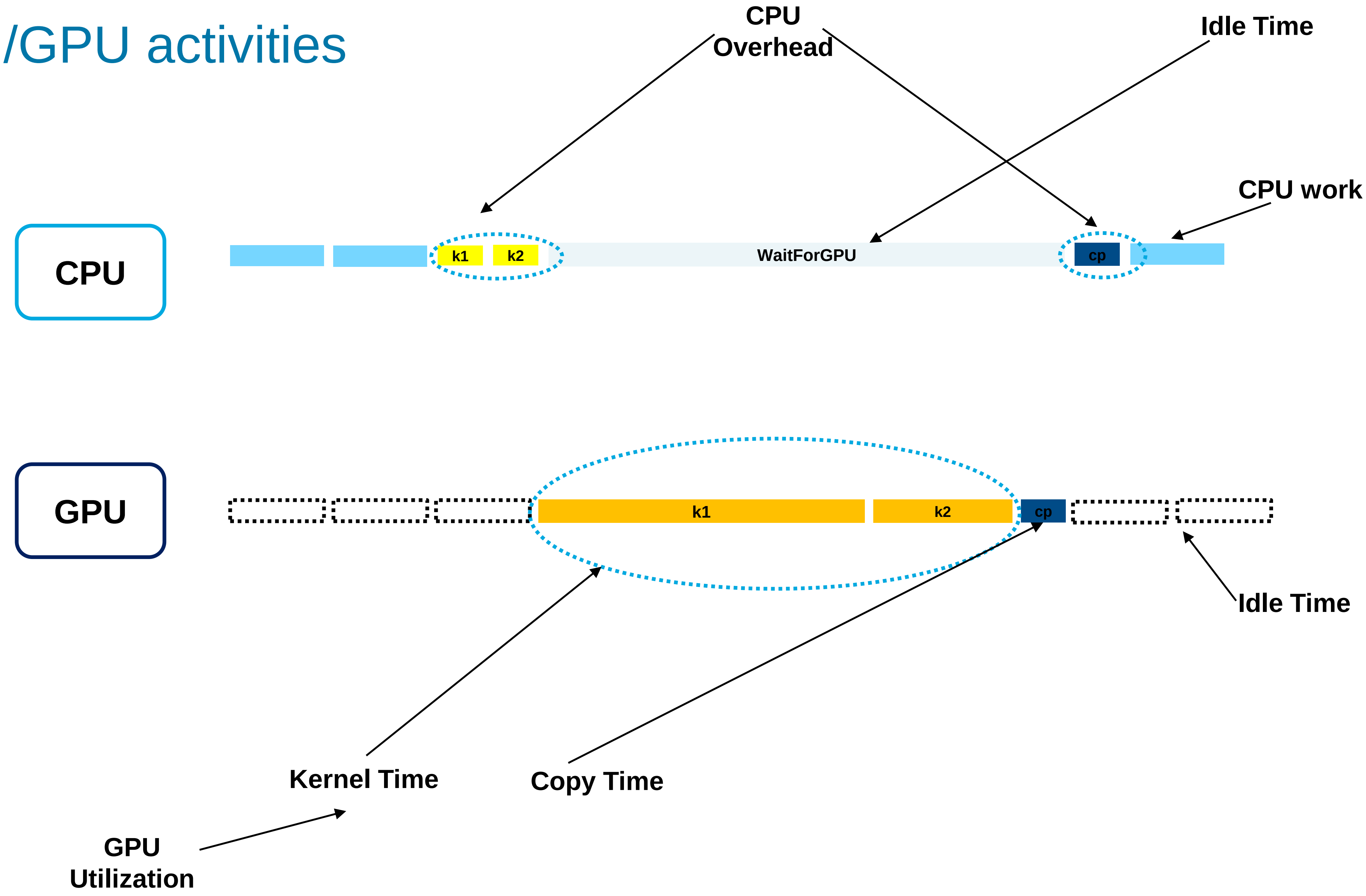
2) Lidar point cloud segmentation



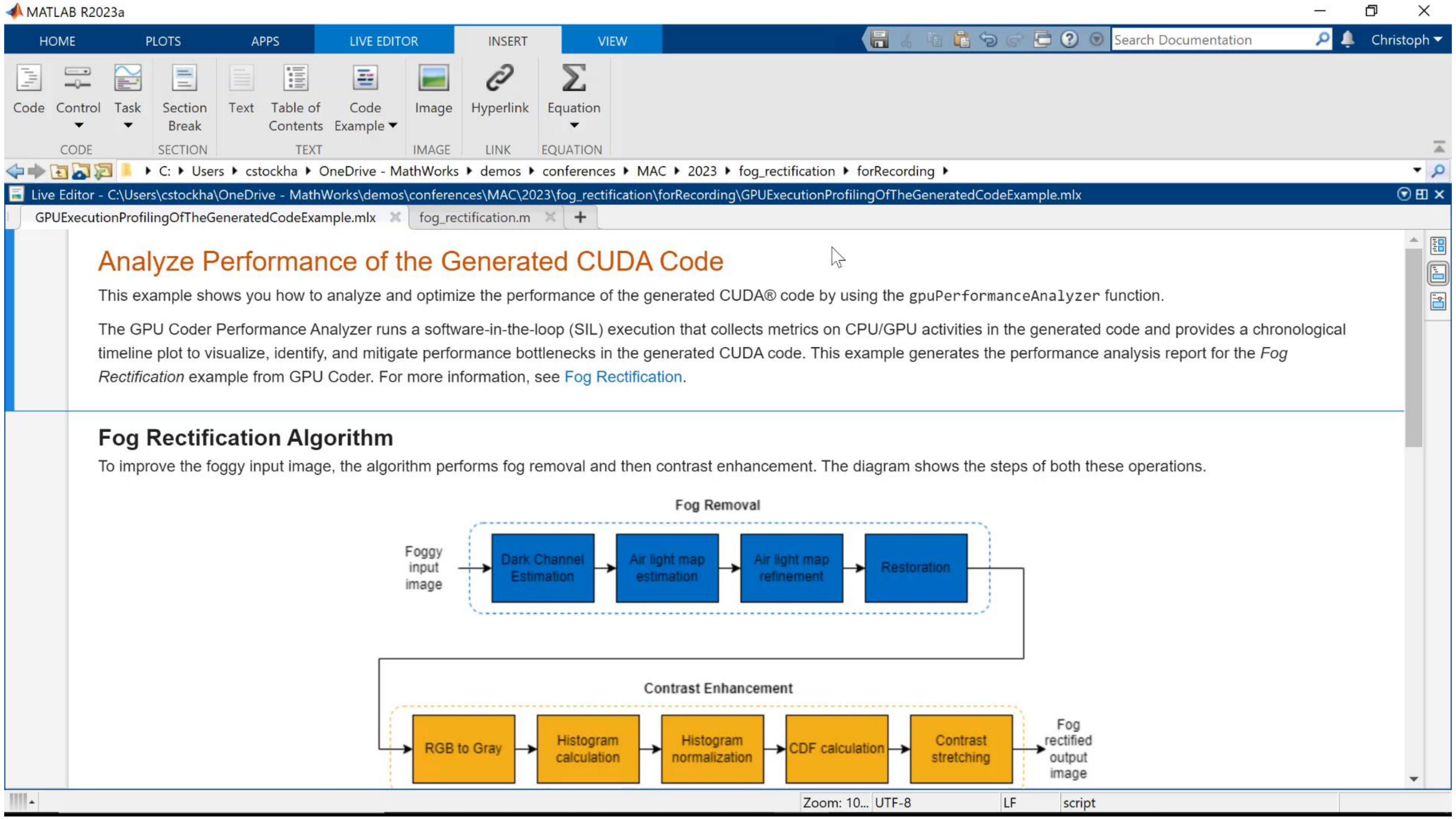
Analyze CPU/GPU interaction for performance improvements



CPU/GPU activities



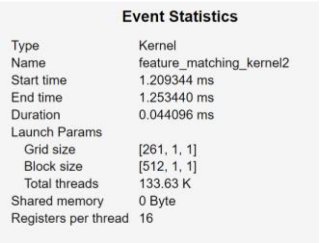
Example 1: Fog rectification + GPU Performance Analyzer

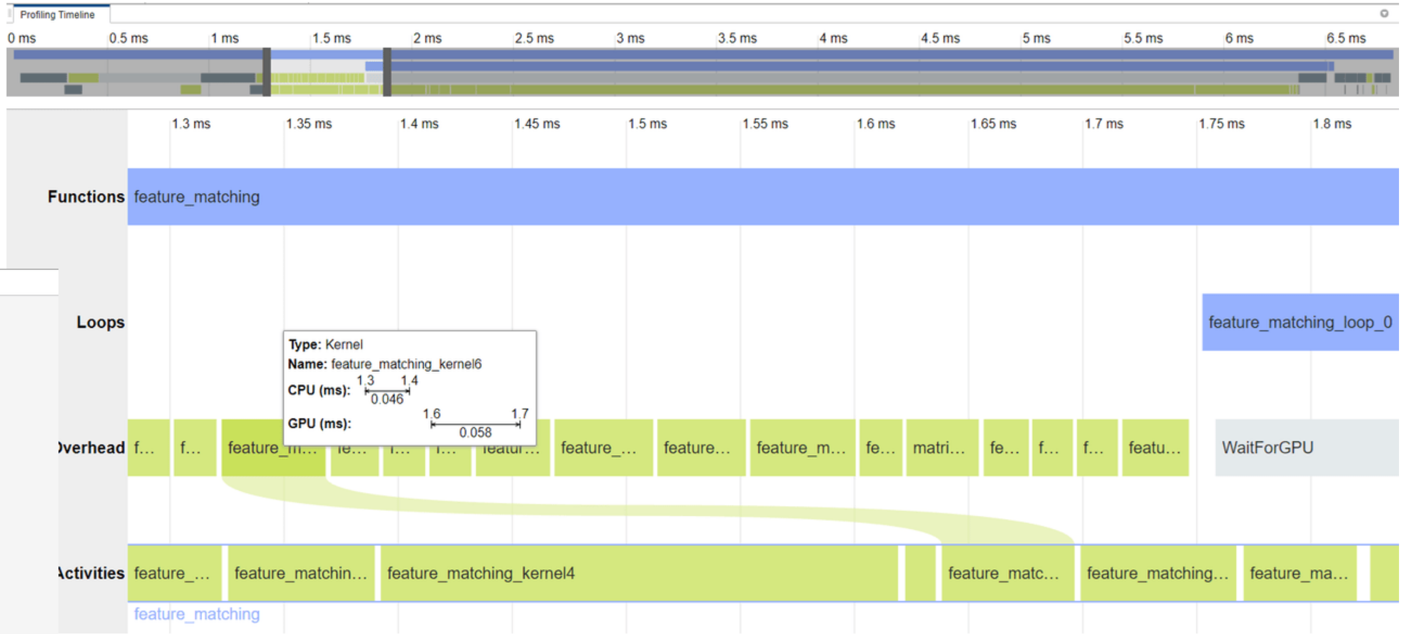


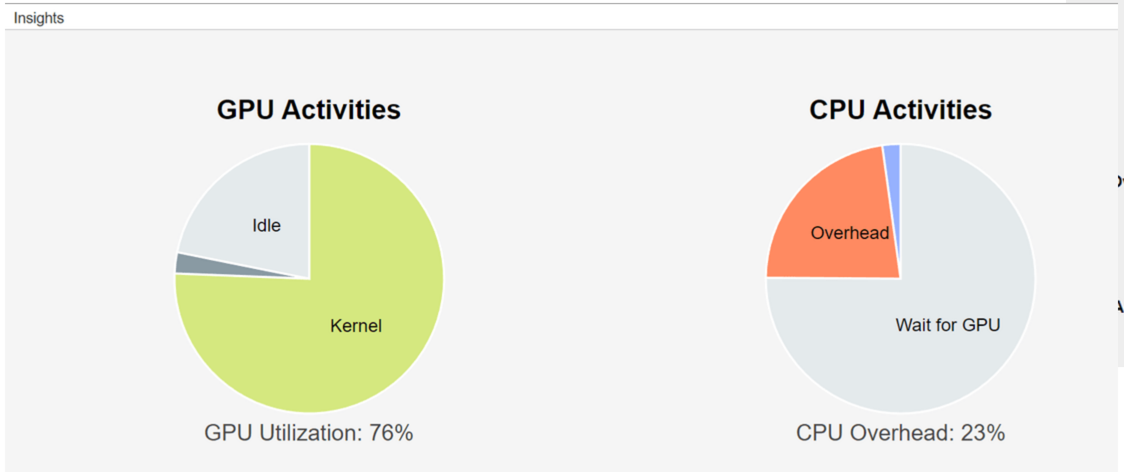
R2023a

Code Profiling using GPU Performance Analyzer

- Profile and understand GPU and CPU activities, events, and performance metrics in a chronological timeline plot
- Use the profiling info to analyze and optimize the performance of the generated CUDA
- Visualize code metrics and identify optimization and tuning opportunities

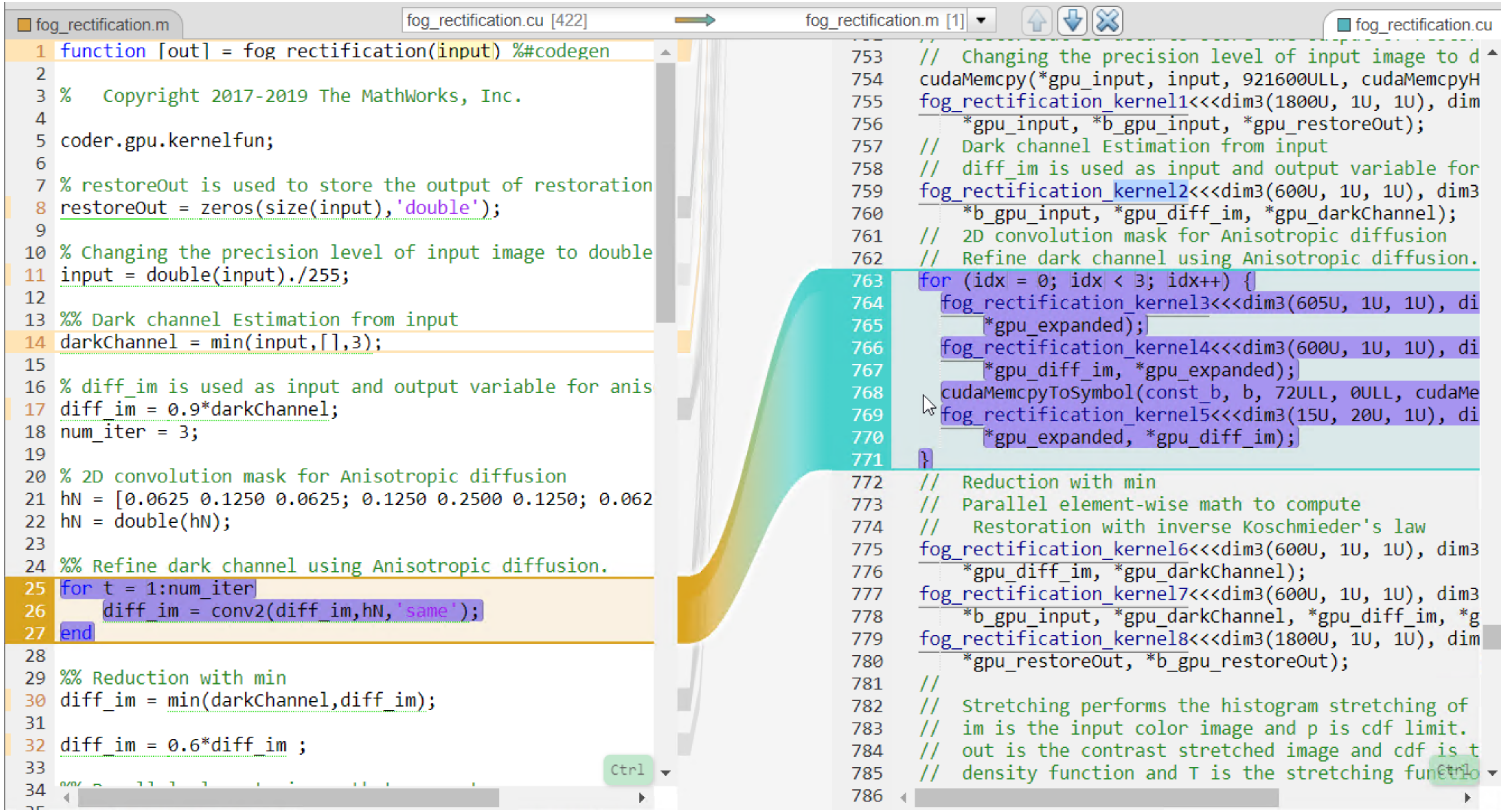






Bidirectional Traceability

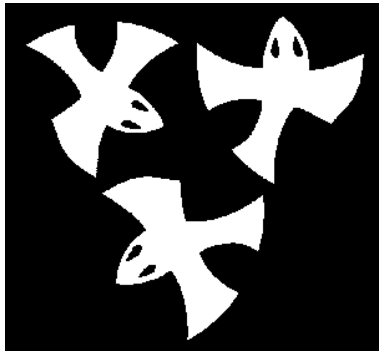
Understand how GPU Coder maps the MATLAB algorithm to CUDA kernels



GPU Coder for Image Processing and Computer Vision

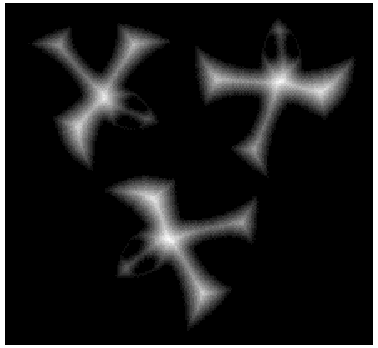
Fog removal

4x speedup



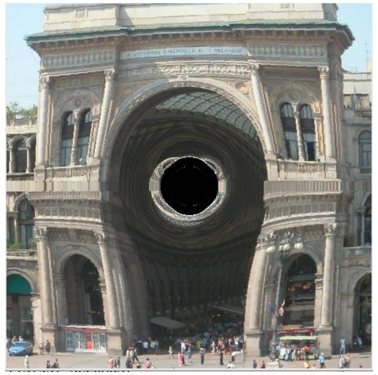
Distance transform

8x speedup



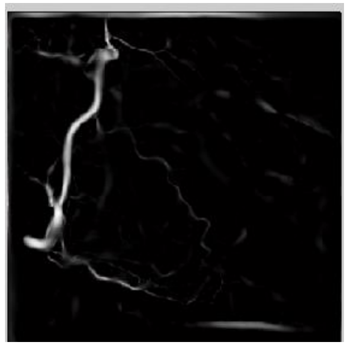
Ray tracing

18x speedup



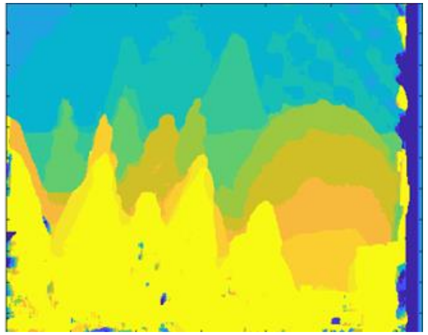
Frangi filter

3x speedup



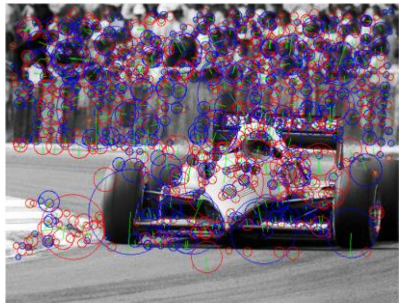
Stereo disparity

50x speedup



SURF feature extraction

700x speedup

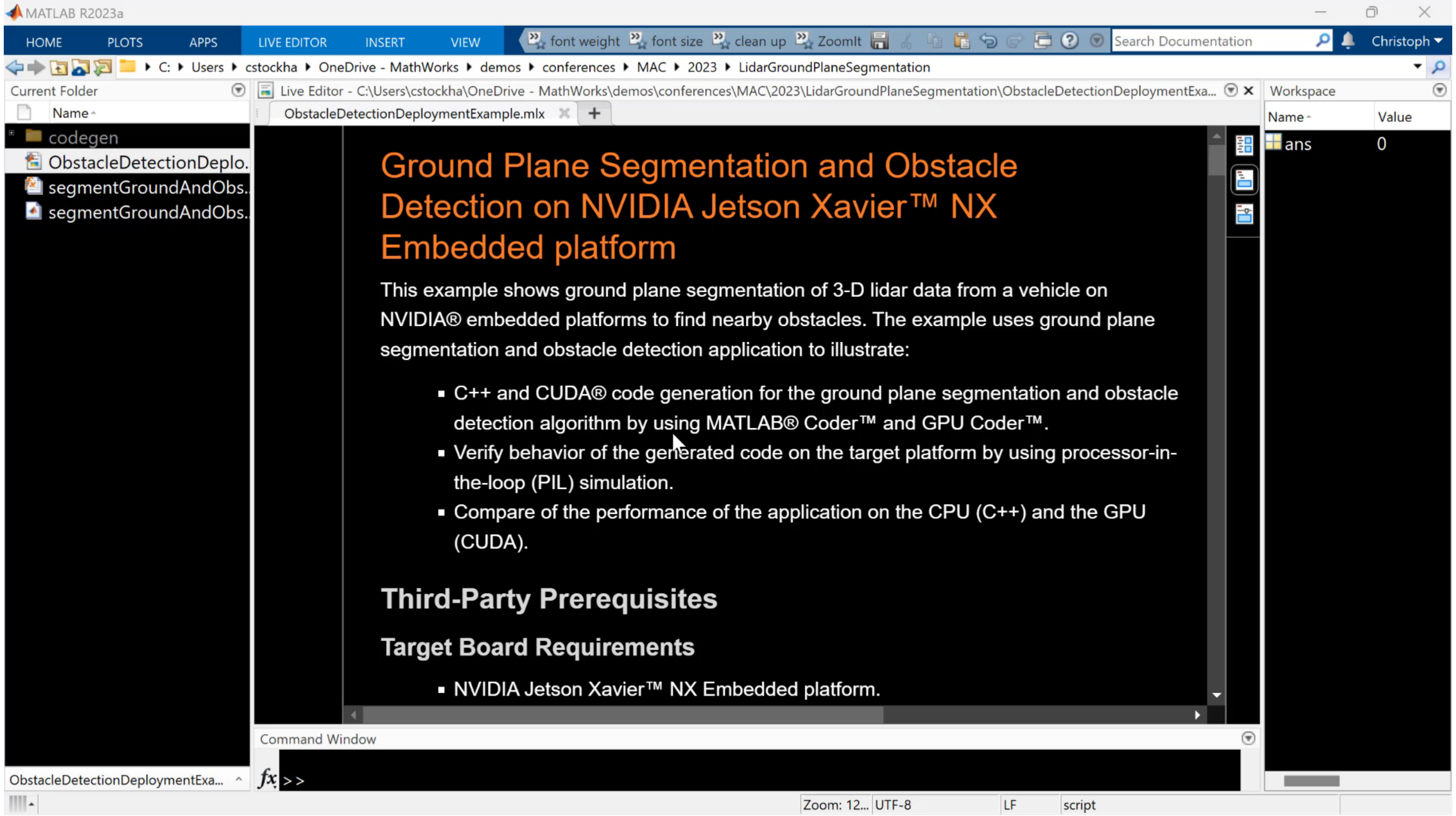


Optimizations for Generated CUDA Code

- Accelerated library support
  - cuFFT, cuBLAS, cuSolver, Thrust, cuDNN, & TensorRT

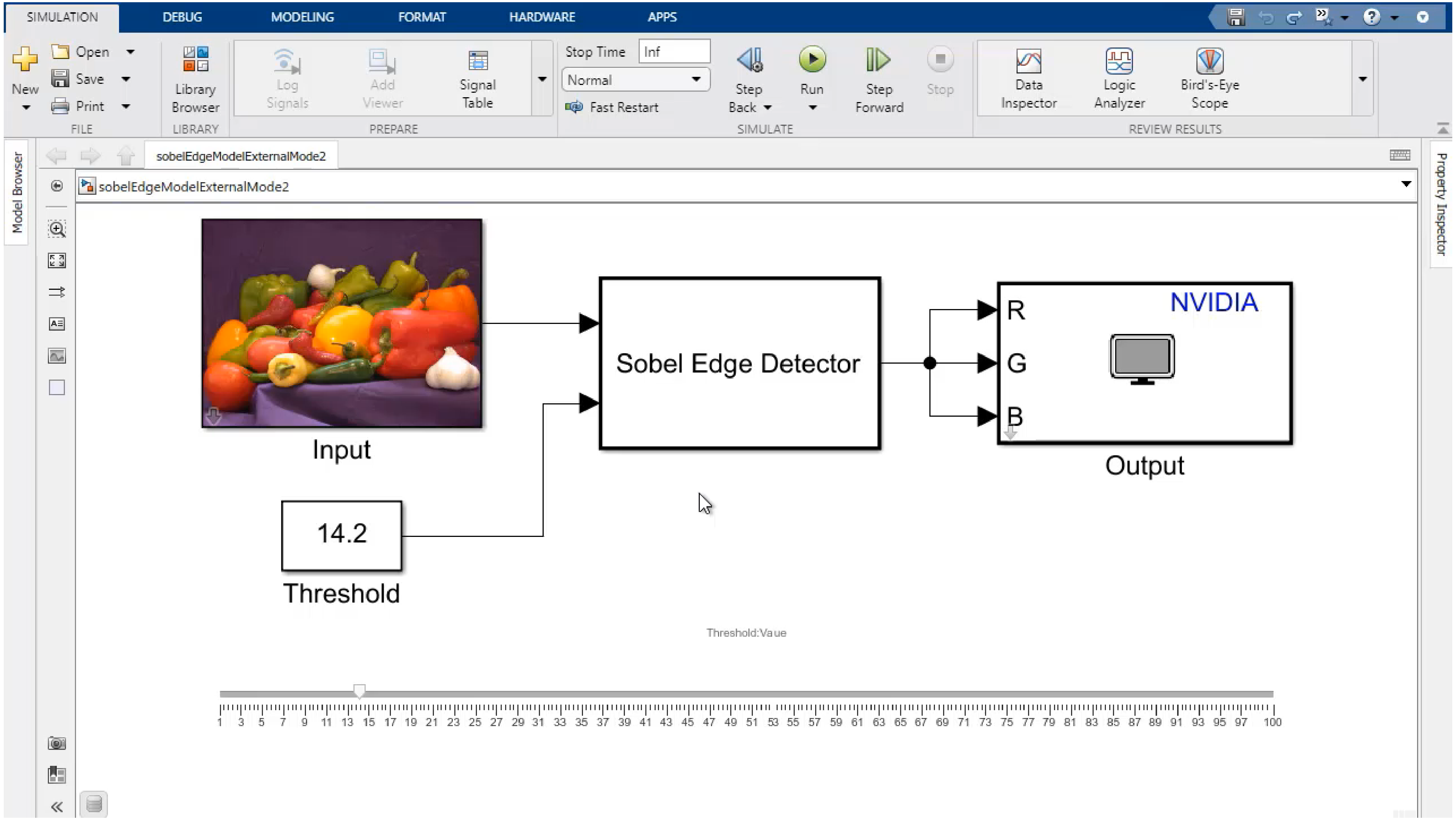
- Data Transfer Minimization
  - Analyzes data dependency between the CPU and GPU partitions to determine minimum set of locations where data must be copied between CPU and GPU using `cudaMemcpy`

Example 2: Lidar point cloud segmentation

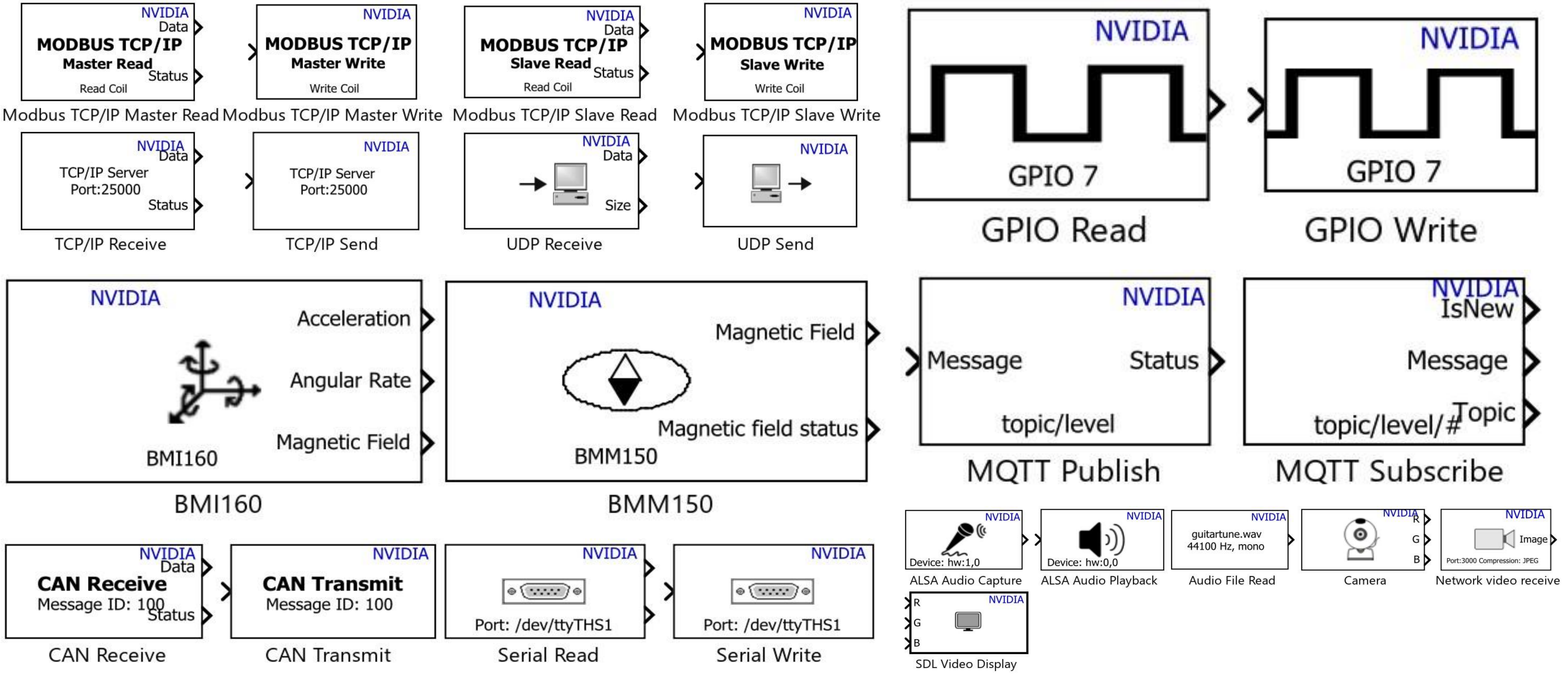


OK, but what about Simulink?

Tuning Parameters using External Mode

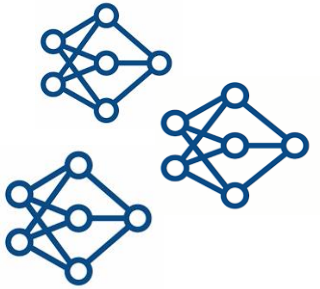


NVIDIA Peripheral Support – block library



OK, but what about AI?

Deploy Complete Deep Learning Application

Pre-processing - - - ->  - - - -> **Post-processing**

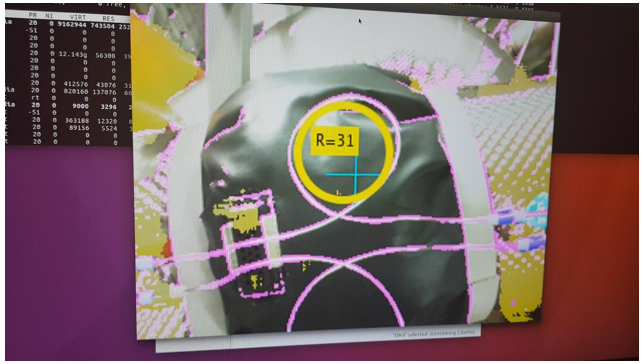
↓

GPU Coder

↓

NVIDIA

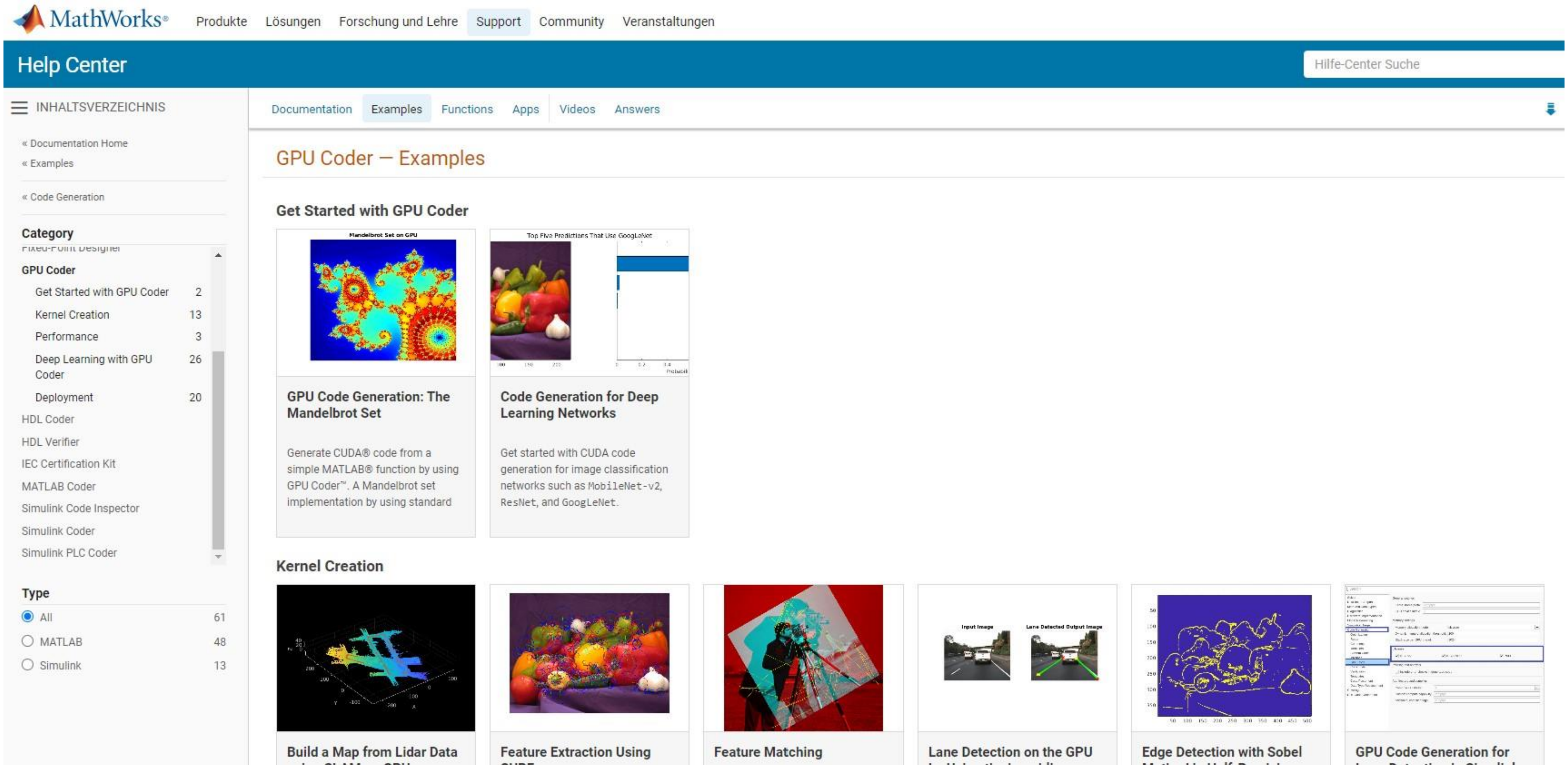
Automatic Defect detection at Airbus



Deep Learning-based Visual inspection at Musashi

Shipping Examples

https://www.mathworks.com/help/gpucoder/examples.html



Key Takeaways

- GPU Coder generates CUDA code from MATLAB & Simulink
- Accelerate MATLAB & Simulink simulations
- Deploy algorithms (signal/deep learning,…) to embedded GPUs

MathWorks
AUTOMOTIVE CONFERENCE 2023
Europe

Thank you