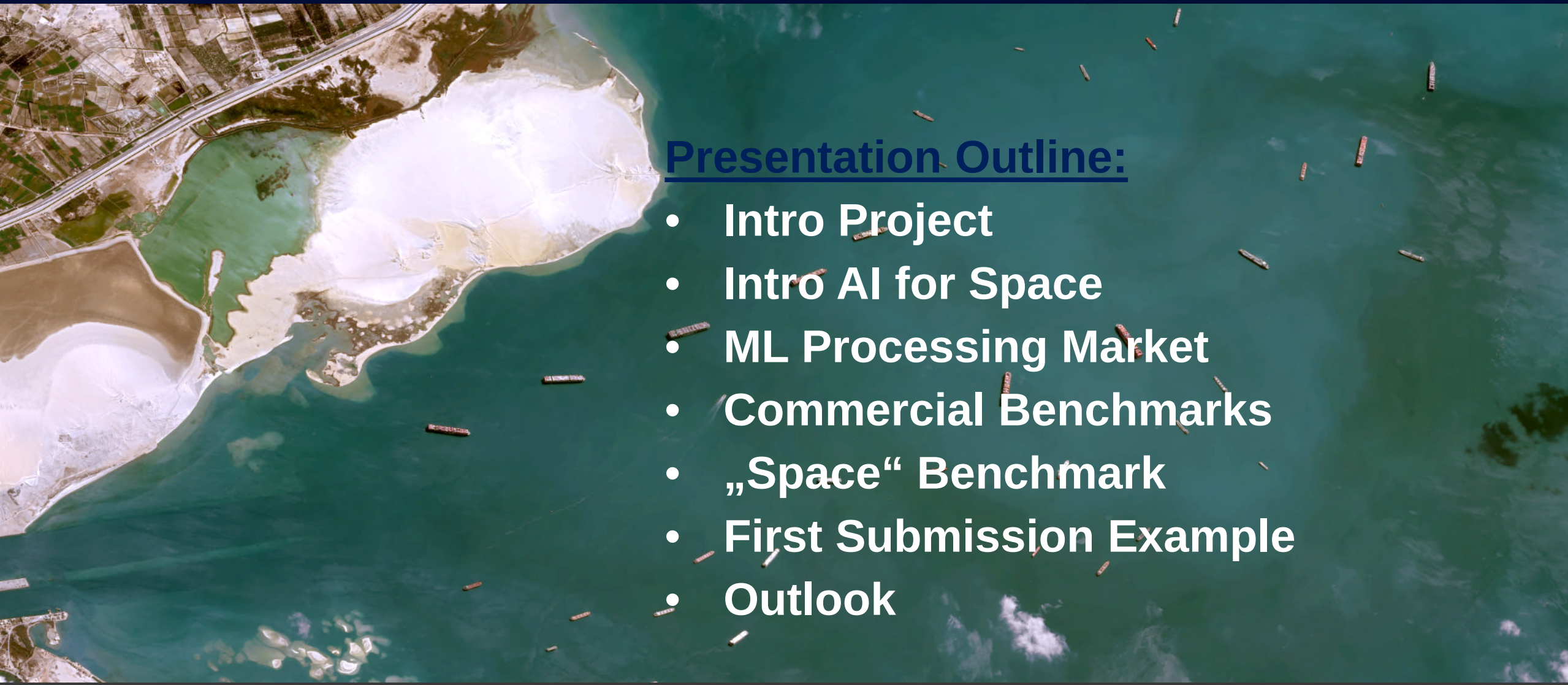
OBDP2021
Machine Learning Application Benchmark
TSTCG – Space Systems
Telecom & Navigation Processing Germany

DEFENCE AND SPACE

- Max Ghiglione
- Airbus Defence and Space

AIRBUS

MLAB – Machine Learning Application Benchmark



Presentation Outline:

- Intro Project
- Intro AI for Space
- ML Processing Market
- Commercial Benchmarks
- „Space“ Benchmark
- First Submission Example
- Outlook

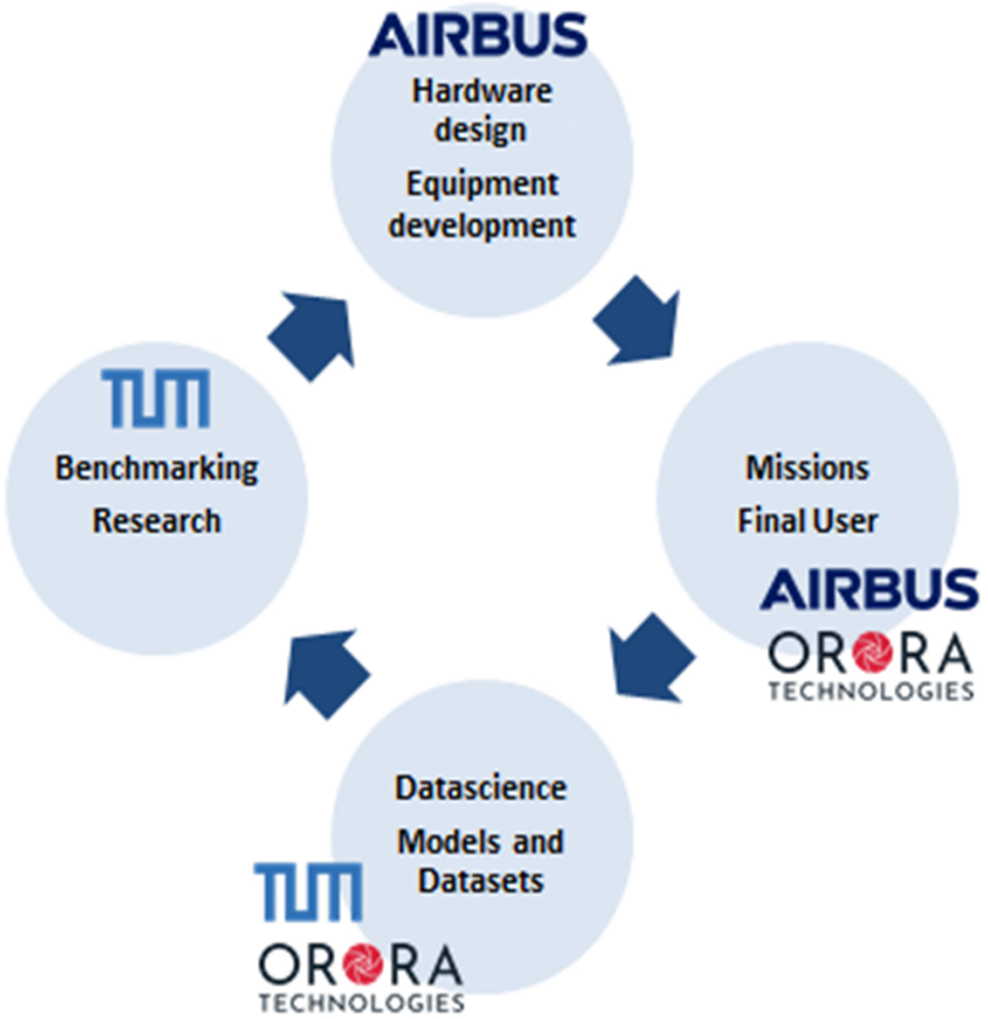
MLAB – Machine Learning Application Benchmark

Project Details:

- GSTP – AO10370
- Design of a **benchmark** for ML applications in Space
- Trade-off of: **Versal, KU060, Myriad, Zynq US+**
- First Phase started, **requirement definition**
- Conclusion planned **June 2021**

Teaming:

- **ESA Technical Officer:** Gianluca Furano
- **Airbus:** Max Ghiglione, Vittorio Serra, Richard Wiest
- **TUM CAPS:** Amir Raoofy, Carsten Trinitis, Martin Schulz
- **TUM BDGM:** Gabriel Dax, Martin Werner
- **OroraTech:** Martin Langer



Goals

Step 1:

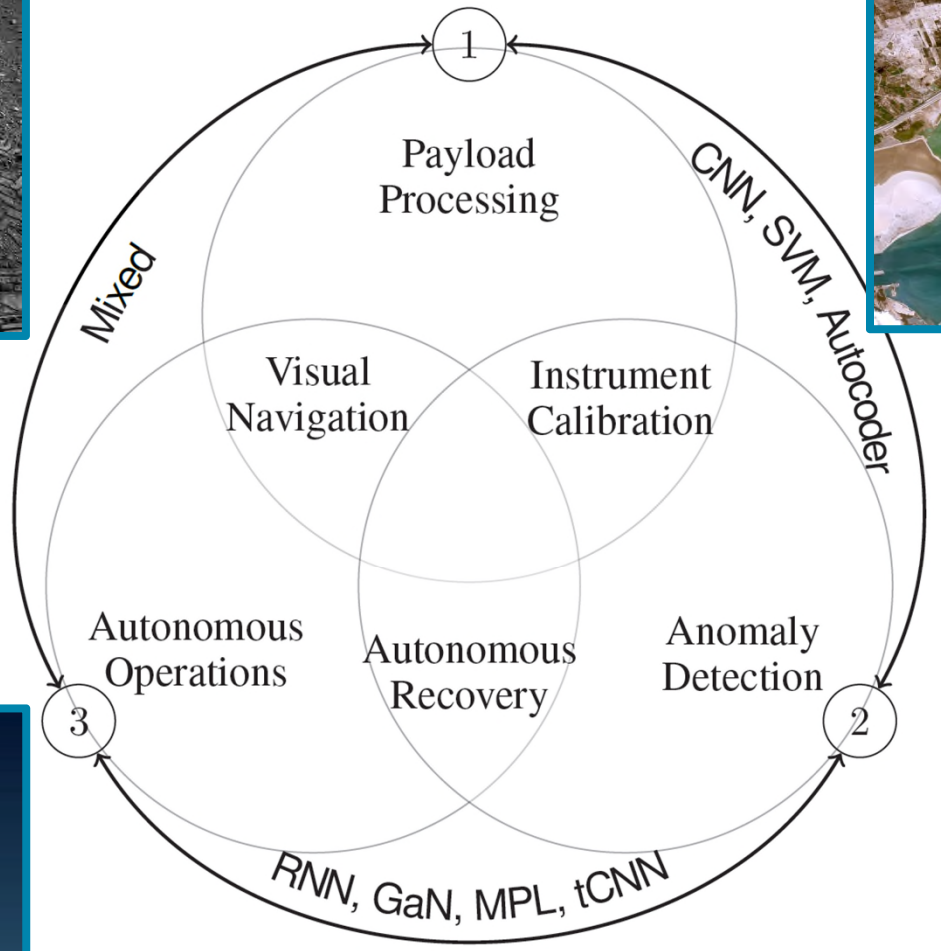
Design and implement a **benchmark for Machine Learning** applications for the space market. The benchmark shall:

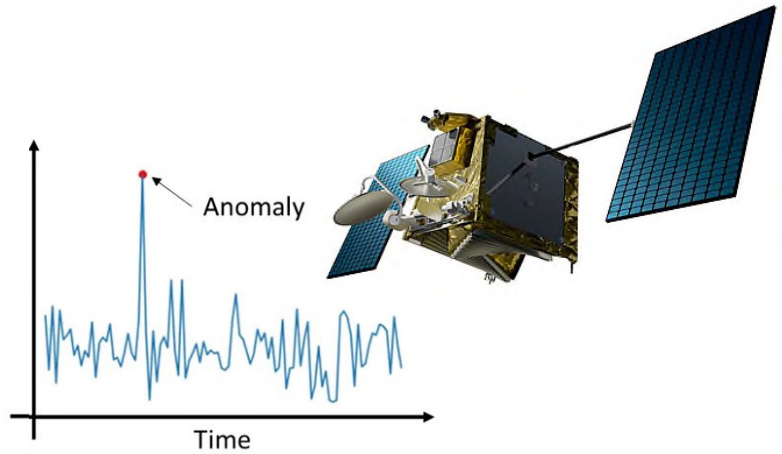
- **Enable the industry to compare and replicate** Machine Learning implementations
- Drive to **develop ML applications** and **processing platforms for space**

Step 2:

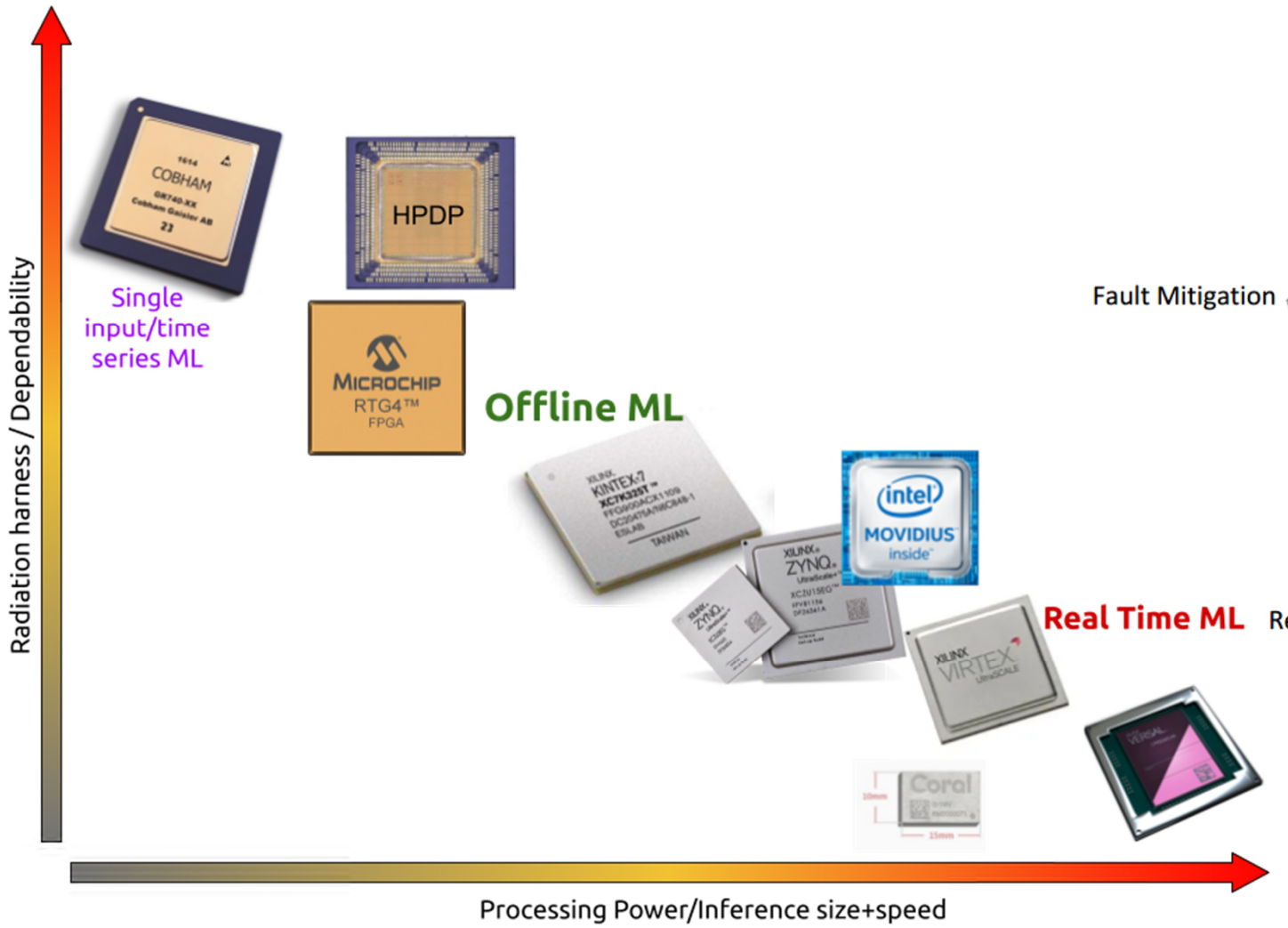
Demonstrate the feasibility of such **algorithms on a demonstrator breadboard** which shall be representative of future co-processing computers

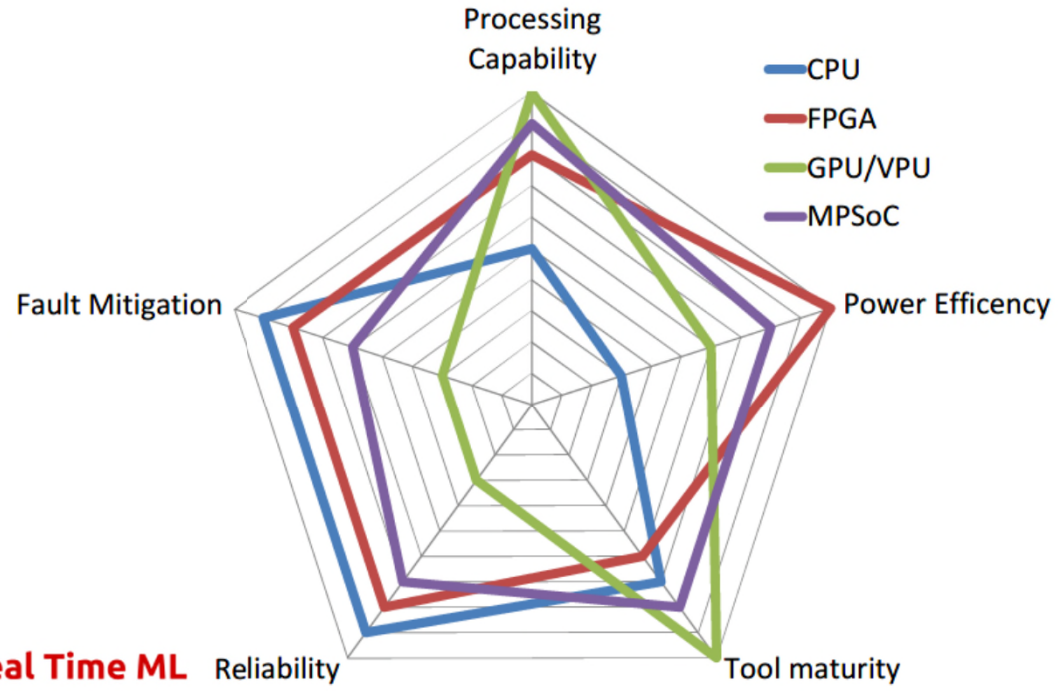
State of Art – AI Onboard Data Processing



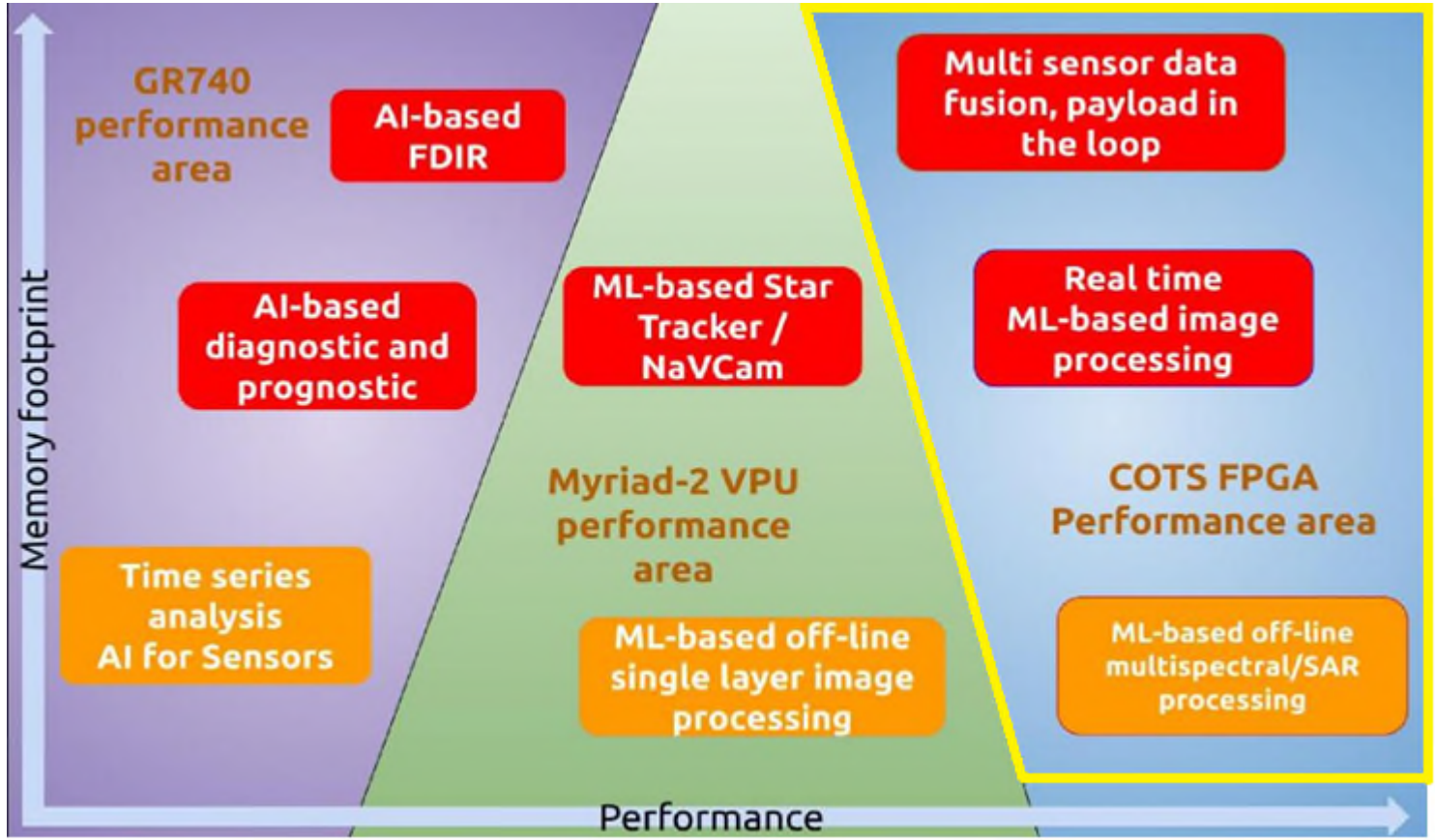


Market options for Machine Learning



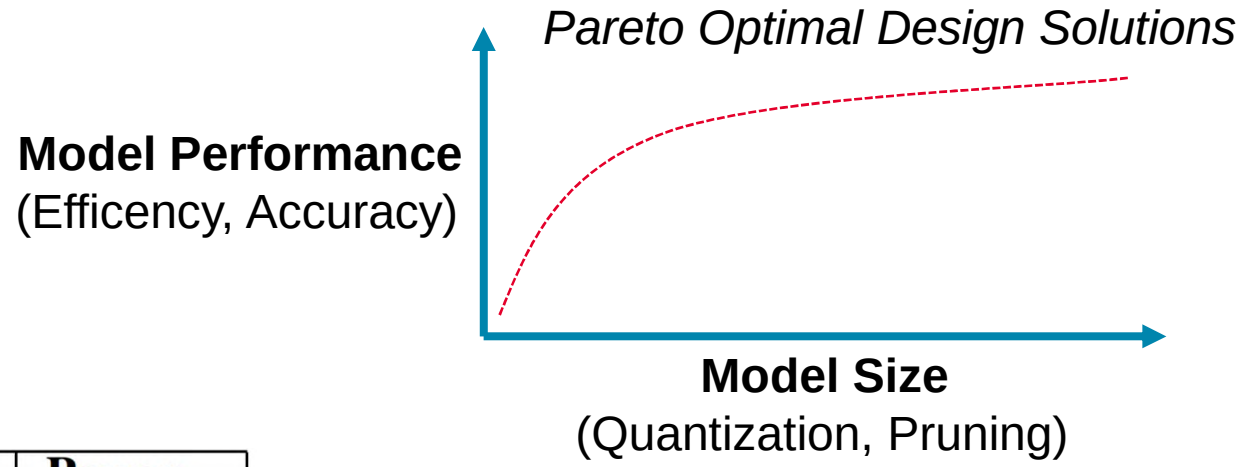


Processing areas – COTS FPGA



Performance Requirements

- Power?
- Troughput?
- Accuracy?



Use Case	Accuracy	Throughput	Power
Image Class. (Heavy)	*Quality Metric*	5 FPS	10W
Image Class. (Light)	90%	10 FPS	*Quality Metric*
Obj. Detection (Heavy)	85%	*Quality Metric*	10W
Obj.Detection (Light)	80%	10 FPS	*Quality Metric*

- Memory?
- Threading?
- Quantization?

Commercial Inference Benchmarks - Models

Area	Task	Model	Dataset	QSL Size	Quality	Server latency constraint
Vision	Image classification	Resnet50-v1.5	ImageNet (224x224)	1024	99% of FP32 (76.46%)	15 ms
Vision	Object detection (large)	SSD-ResNet34	COCO (1200x1200)	64	99% of FP32 (0.20 mAP)	100 ms
Vision	Medical image segmentation	3D UNET	BraTS 2019 (224x224x160)	16	99% of FP32 and 99.9% of FP32 (0.85300 mean DICE score)	N/A
Speech	Speech-to-text	RNNT	Librispeech dev-clean (samples < 15 seconds)	2513	99% of FP32 (1 - WER, where WER=7.452253714852645%)	1000 ms
Language	Language processing	BERT	SQuAD v1.1 (max_seq_len=384)	10833	99% of FP32 and 99.9% of FP32 (f1_score=90.874%)	130 ms
Commerce	Recommendation	DLRM	1TB Click Logs	204800	99% of FP32 and 99.9% of FP32 (AUC=80.25%)	30 ms

V. J. Reddi *et al*., "MLPerf Inference Benchmark," *2020*
https://mlcommons.org/en/

Commercial Inference Benchmarks – Scenarios

Scenario	Query Generation	Duration	Samples/query	Latency Constraint	Tail Latency	Performance Metric
Single stream	LoadGen sends next query as soon as SUT completes the previous query	1024 queries and 60 seconds	1	None	90%	90%-ile measured latency
Multiple stream	LoadGen sends a new query every *latency constraint* if the SUT has completed the prior query, otherwise the new query is dropped and is counted as one overtime query	270,336 queries and 60 seconds	Variable, see metric	Benchmark specific	99%	Maximum number of inferences per query supported
Server	LoadGen sends new queries to the SUT according to a Poisson distribution	270,336 queries and 60 seconds	1	Benchmark specific	99%	Maximum Poisson throughput parameter supported
Offline	LoadGen sends all queries to the SUT at start	1 query and 60 seconds	At least 24,576	None	N/A	Measured throughput

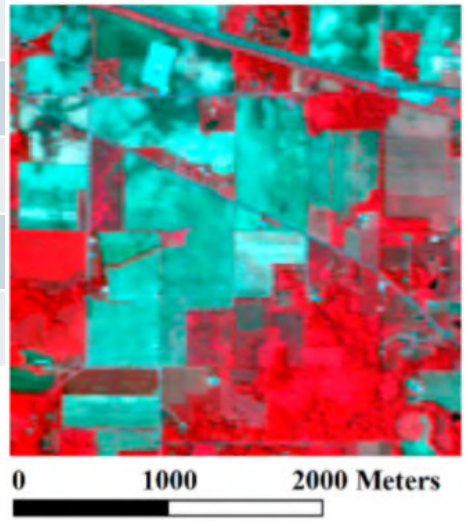
V. J. Reddi *et al*., "MLPerf Inference Benchmark," *2020*
https://mlcommons.org/en/

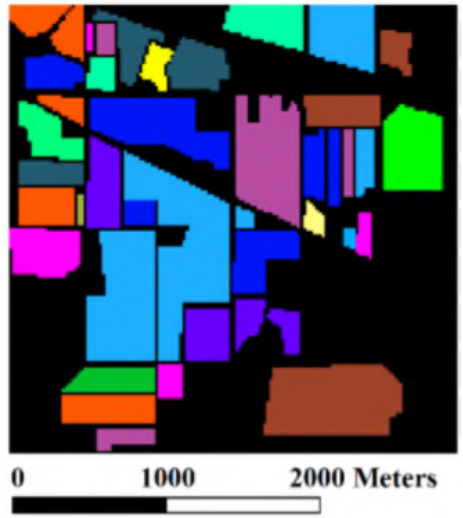
“Space” Benchmark

Benchmark design:

- Use Cases
- Datasets
- Models
- Performance Metrics

Area	Payload Processing		Payload Processing	
Dataset	Airbus Ship Detection		Indian Pine	
Task	Object Detection		HyperSpectral	
Architecture	Unet-MobileNet-v2	Unet-ResNet-50	3D Unet	…
Size	224x224	…		
Oracle acc.	90%	…		
Quantiz. acc.	85%	…		
Resources	DPU*	…		
Troughput	15 FPS	…		





Submission Types:

- **Open** (optimized)
- **Closed** (fixed parameters to compare hardware)
- **Hardware evaluation** (Fixed Model and Quantization)
- In the future could be extended for **Model evaluation** (Fixed accuracy or power)

MLAB Benchmark – First proposal for the model set

Use Case	Model	Dataset
Image Classification (Heavy)	Resnet50 Multilabel	HyRank $\geq$384x384
Image Classification (Light)	MobileNet-v1 Singlelabel	BigEarthNet 224x224
Object Detection (Heavy)	UNet ResNet50	Airbus Ship $\geq$384x384
Object Detection (Light)	UNet Tiny YOLOv3	Airbus Ship 224x224
Hyperspectral ($220 \geq f \geq 13$)	3d UNet	Indian Pine fx224x224
Anomaly Detection (Heavy)	tCNN,Wavenet*	multi-input timeseries*
Anomaly Detection (Light)	GAN,RNN*	single-input timeseries*

MLAB Benchmark – Application examples

Future mission use case:

"**Payload Processing** performed in parallel to regular operation"

- Not mission critical
- Low Reliability and availability req.

Use Case	Model	Dataset
Image Classification (Heavy)	Resnet50 Multilabel	HyRank ≥384x384
Image Classification (Light)	MobileNet-v1 Singlelabel	BigEarthNet 224x224
Object Detection (Heavy)	UNet ResNet50	Airbus Ship ≥384x384
Object Detection (Light)	UNet Tiny YOLOv3	Airbus Ship 224x224
Hyperspectral ($220 \geq f \geq 13$)	3d UNet	Indian Pine fx224x224
Anomaly Detection (Heavy)	tCNN,Wavenet*	multi-input timeseries*
Anomaly Detection (Light)	GAN,RNN*	single-input timeseries*

COTS Processing: ***Zynq Ultrascale+***

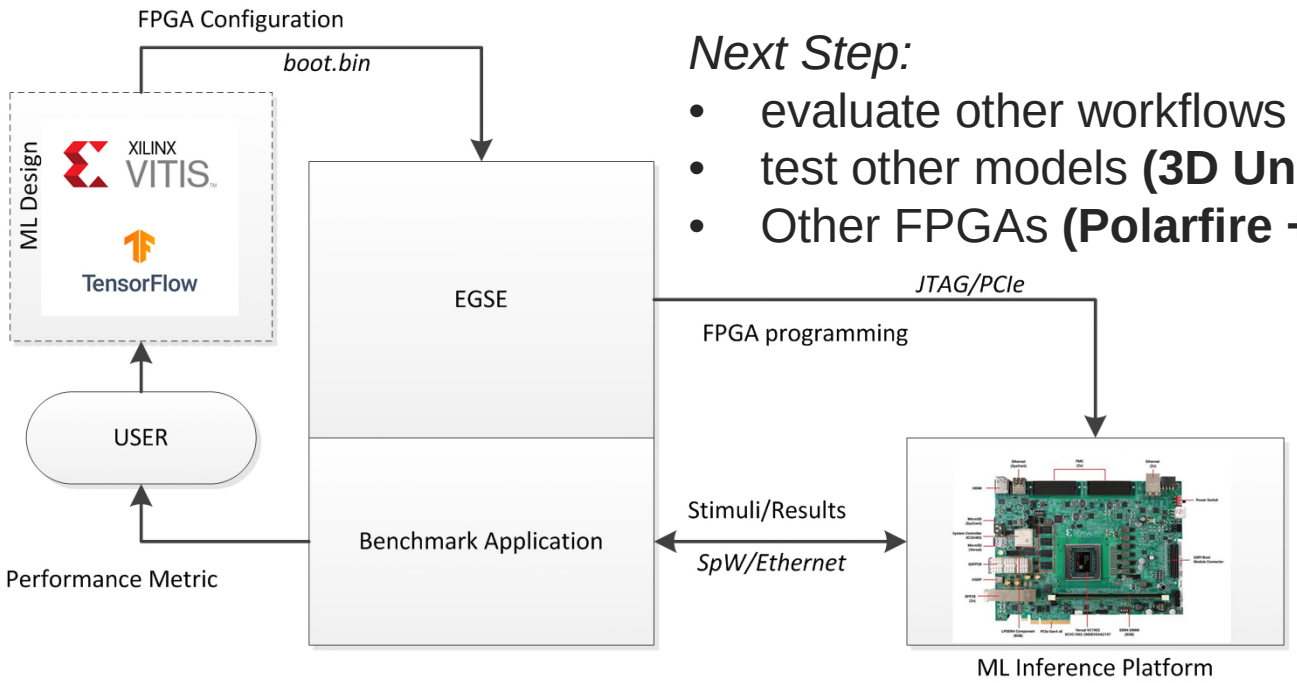
Demonstration of a Benchmark Submission

Use Case:
- **EO Payload Processing**

Hardware:
- **Zynq Ultrascale+**

Workflow:
- **Vitis-AI**

Models:
- **Unet**
- **Unet + ResNet50**
- **Unet + MobileNetv2**



Results:
- Image sizes up to 224x224 tested
- +10 to 20 FPS reached in single threading
- Shows first **feasibility for future missions**
- Inputs used to determine benchmark metrics

Next Step:
- evaluate other workflows **(FINN, Matlab)**
- test other models **(3D Unet, YOLO)**
- Other FPGAs **(Polarfire + VectorBlox)**

Thank you

AIRBUS