

# Efficient Processing For The Edge

Ivo Bolsens | CTO

# Cloud to Edge – Different Metrics, One Platform

- | Real-time
- | Deterministic
- | Low Latency
- | Low Power

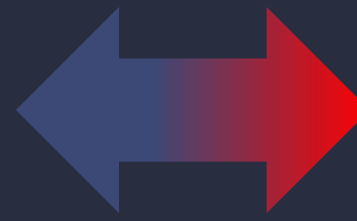
Edge



Embedded



Mobile



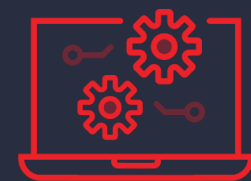
Cloud



Data Center



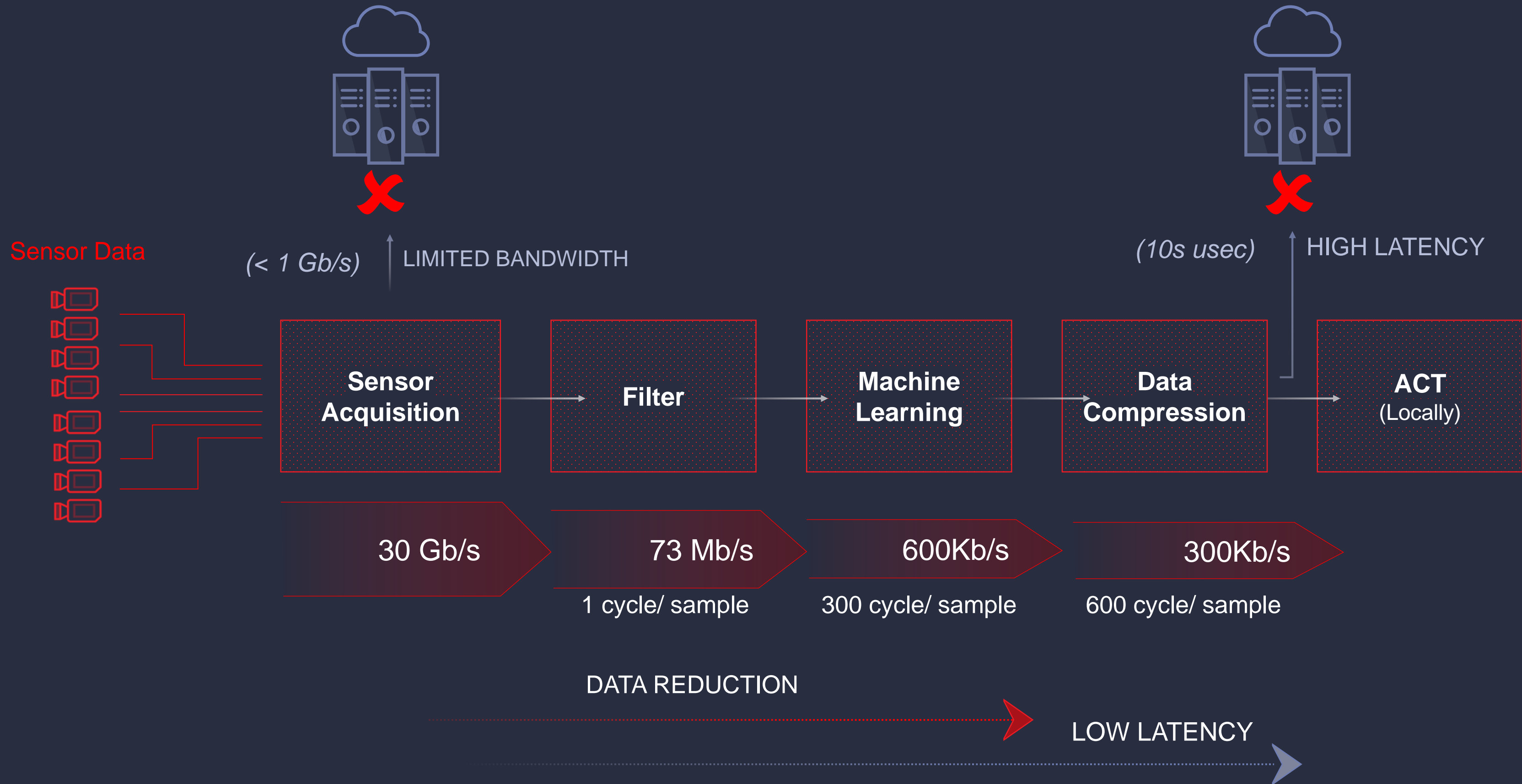
Server



Desktop

- | Transactional
- | As Fast as Possible
- | Scalable

# Edge Compute Data Flow



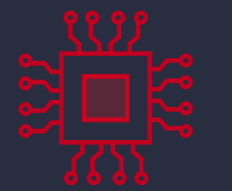


# Dr. Thomas James

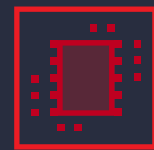
Senior Research Fellow  
CERN

# Enabling Next Level Performance & Compute Efficiency

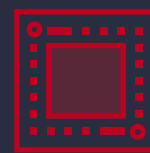
Software Programmability



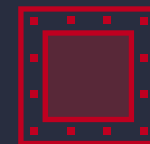
FPGA



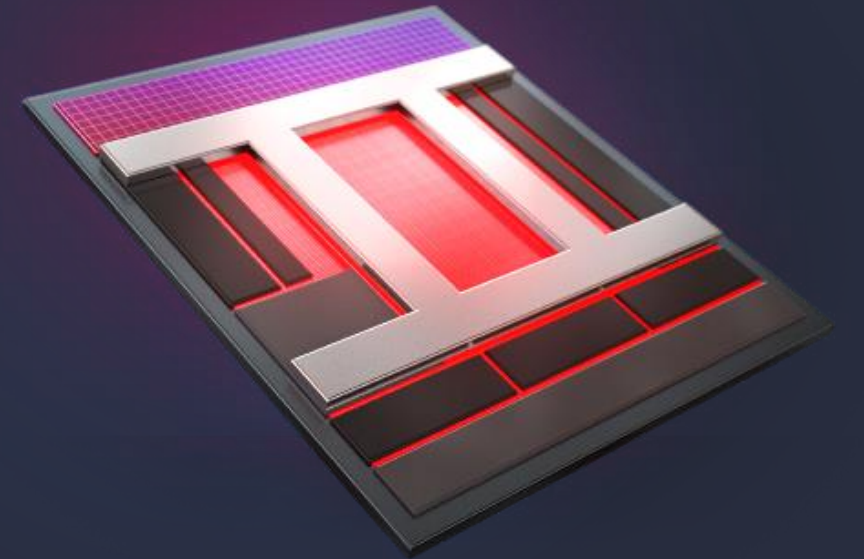
SoC



MPSoC



RFSoc



ACAP

Device Category

# Versal ACAP: A New Class of Compute Architectures

## Heterogeneous Architecture

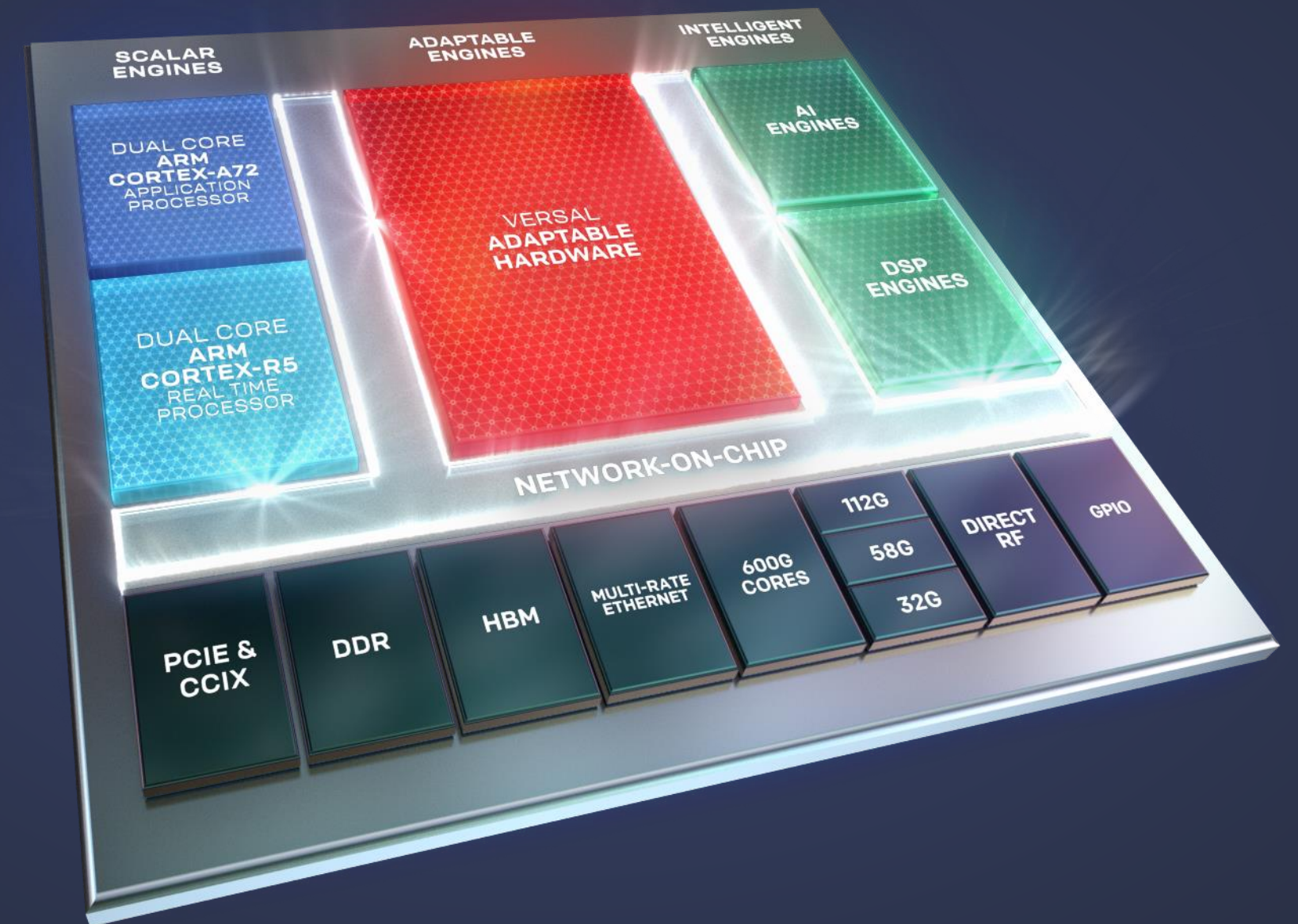
SW Programmable, HW Adaptable

## Determinism for Real-Time Systems

Predictable, Efficient Data Movement

## Spatial Compute

Concurrency, Safety, Low Latency, Low Power





# Architecture: Adaptable and Heterogeneous

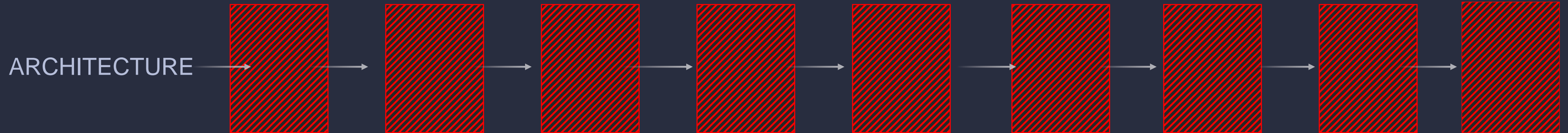
---



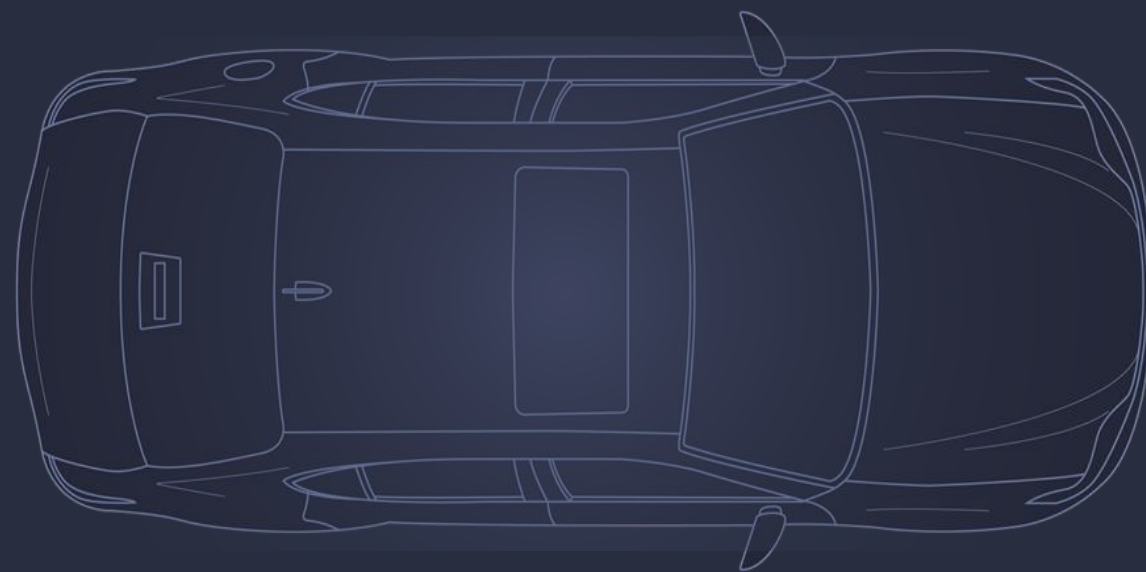
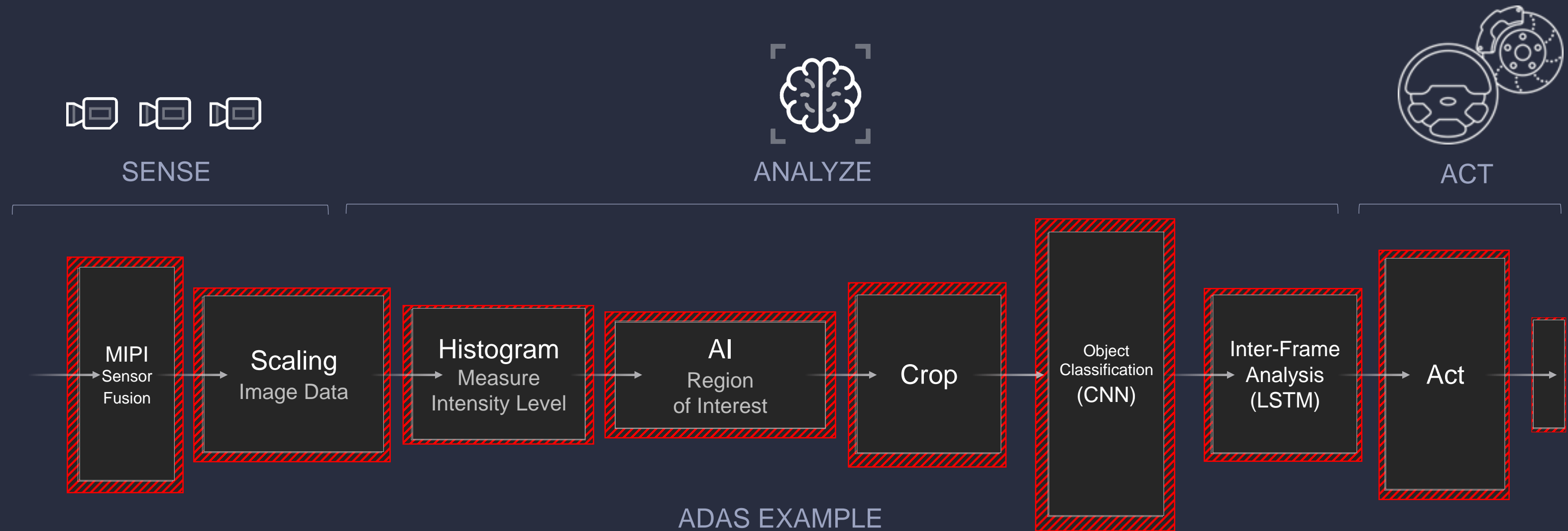
Custom Data Flow

Custom Precision

Custom Memory

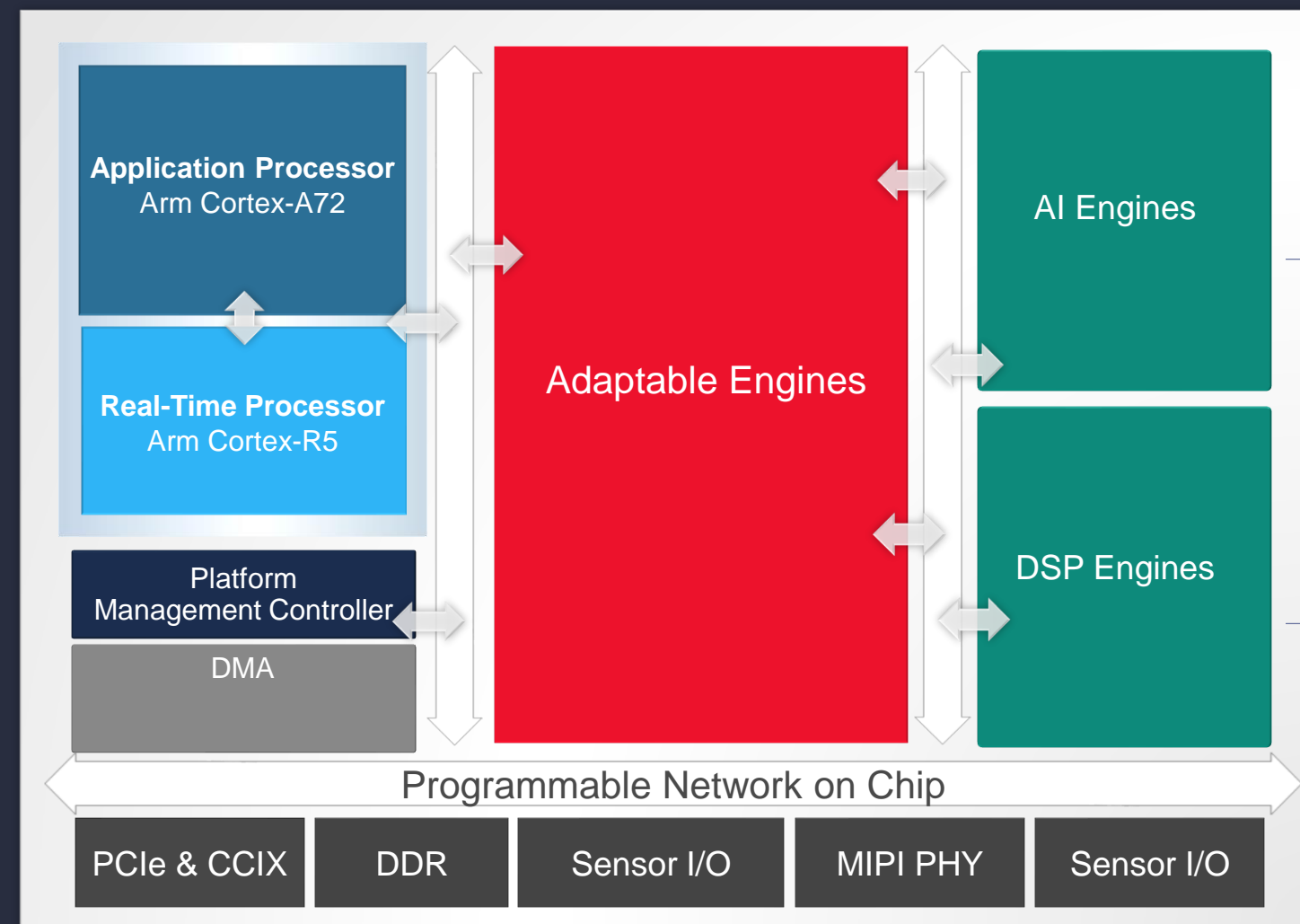
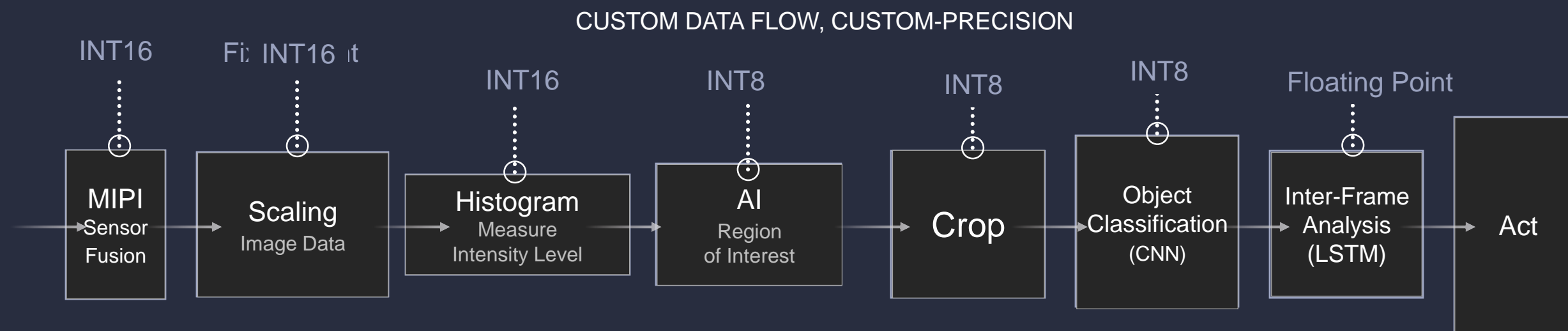


# Adapt the Architecture to the Algorithm (ADAS Example)





# Heterogeneous Programming for Custom Data Flow

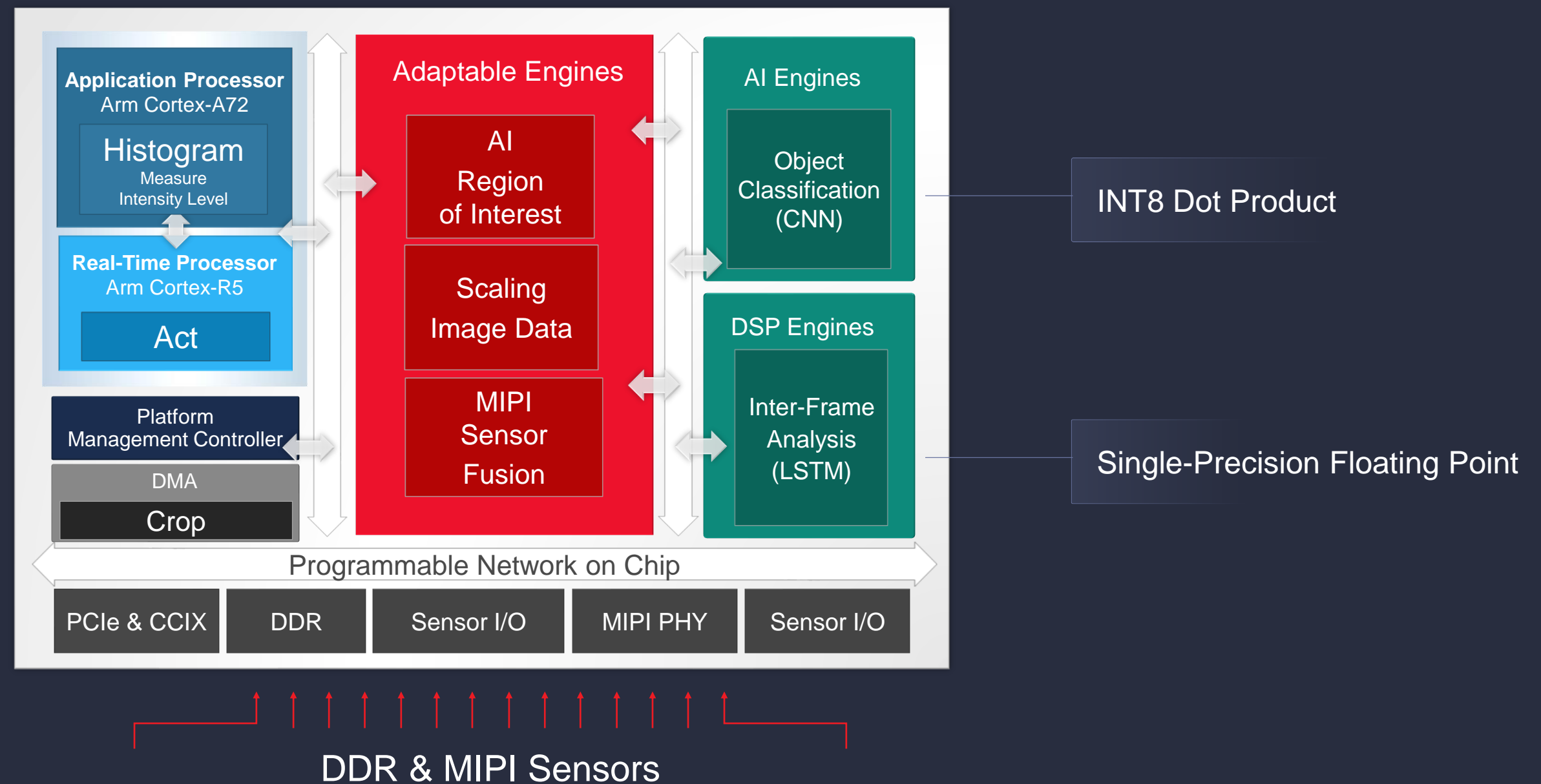
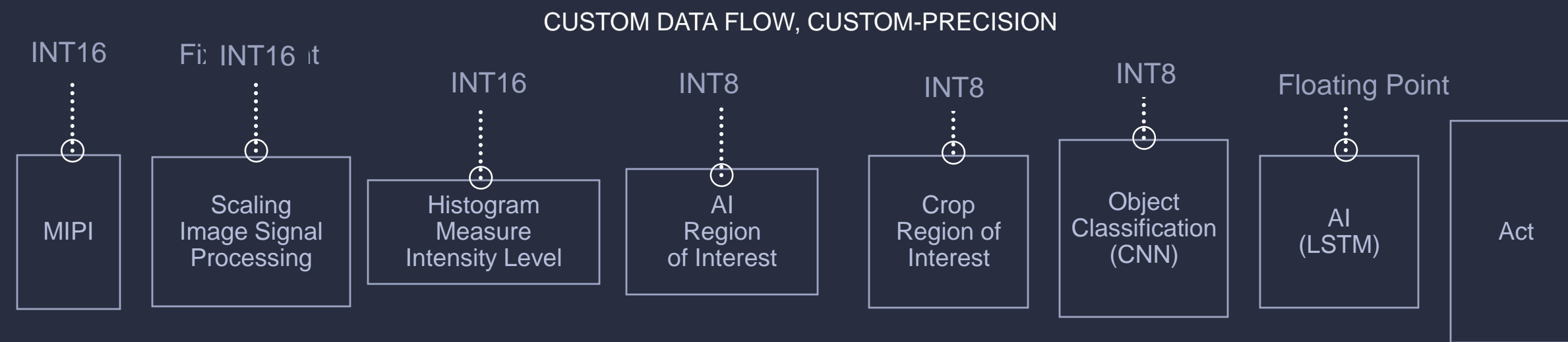


## MULTI-PRECISION ENGINES

- 8-, 16-, 32-bit Fixed Point
  - Complex Data Types (high MAC density)
  - Single Precision Floating Point
- 
- Variable Fixed-Point (incl. INT8 mode)
  - Single Precision Floating-Point
  - Complex Multiply / MACC



# Heterogeneous Programming for Custom Data Flow

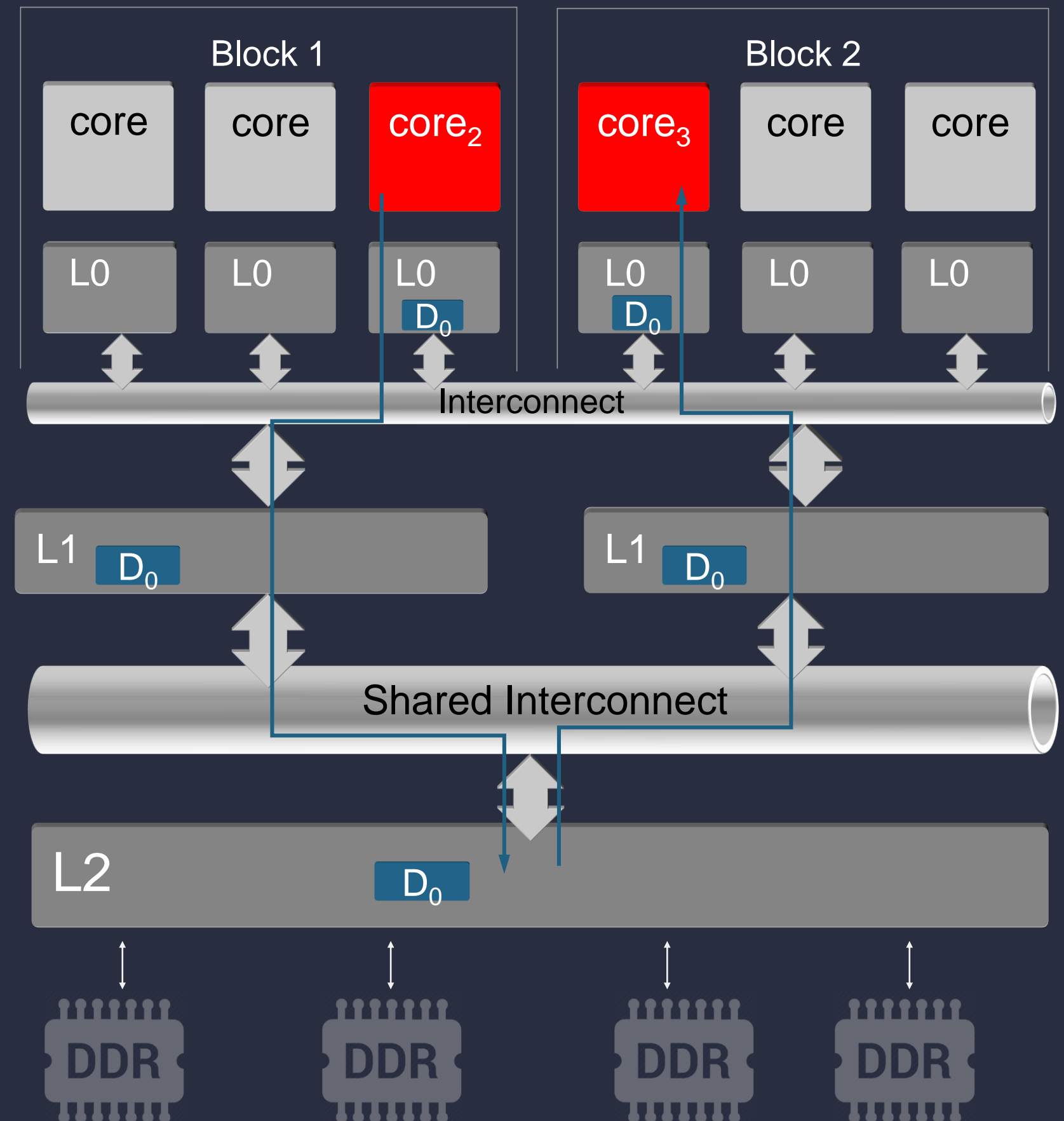


# Traditional Multi-Core

✗ Data Replication

✗ Non-deterministic

✗ High Latency

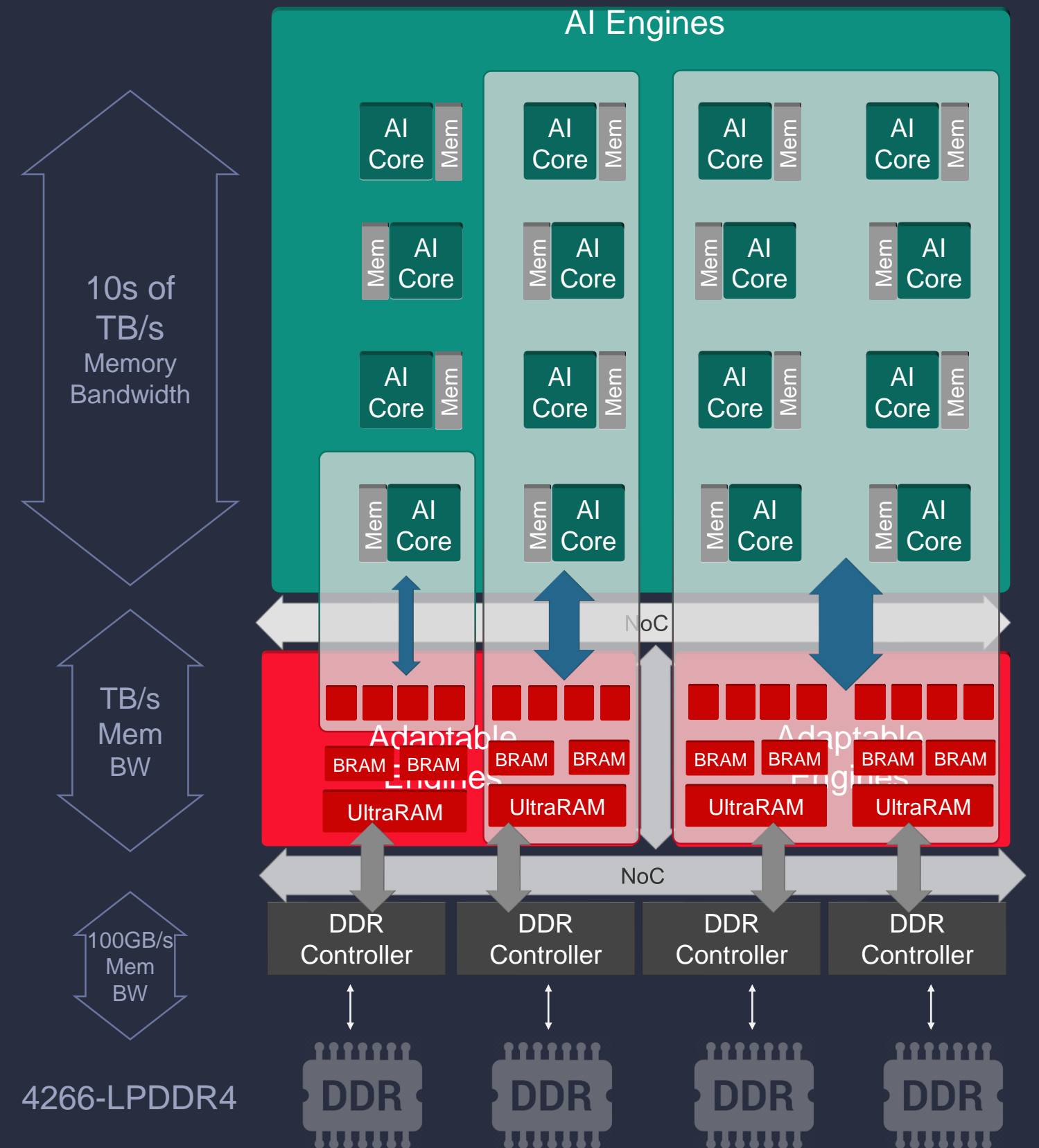


# Adaptable Memory Architecture

5X On-Chip Memory

Adaptable Memory Hierarchy

Flexible Memory & Bandwidth Allocation

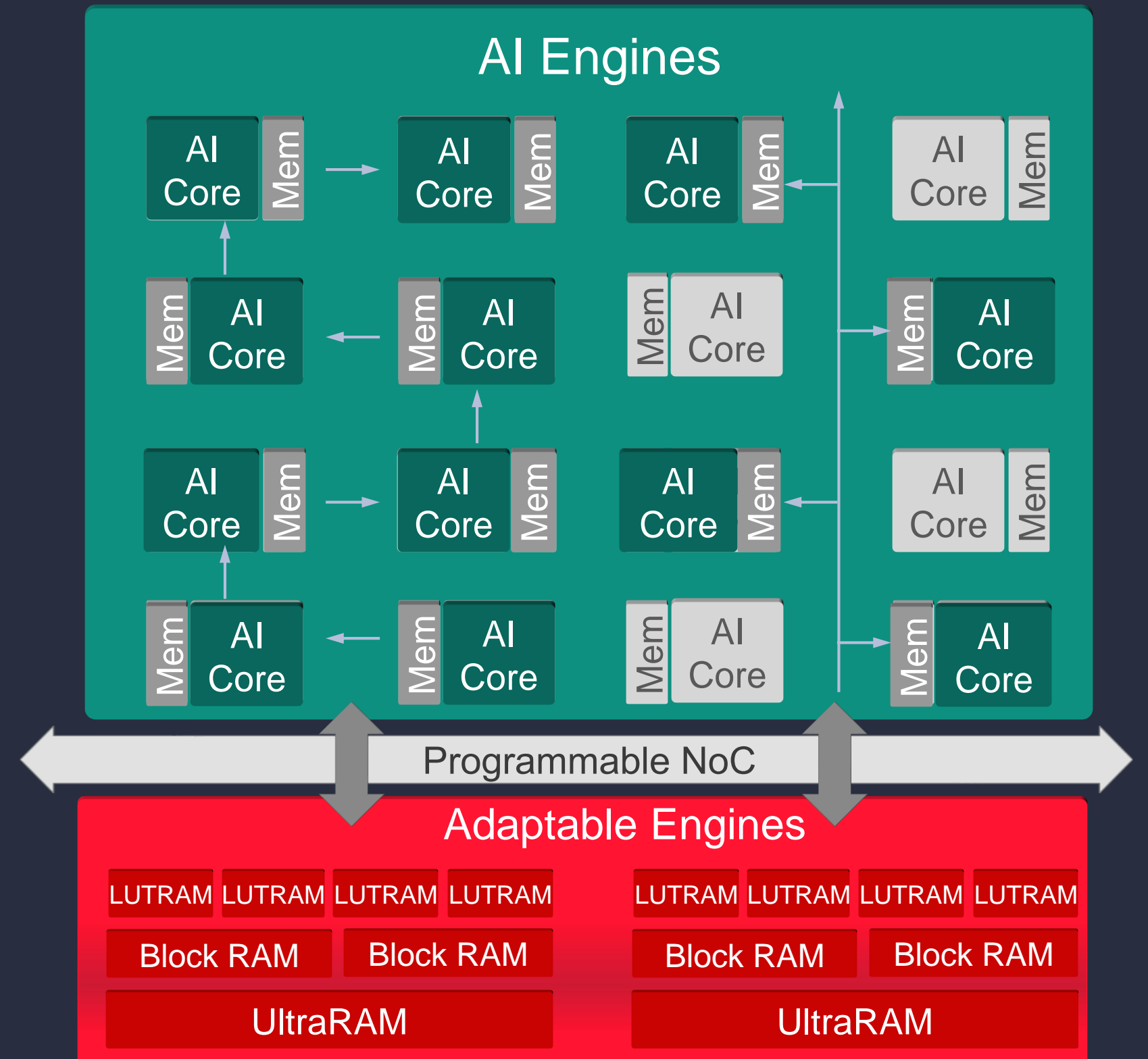


# Adaptable Dataflow Architecture

Keep data moving

Customize communication pattern

Pipeline compute and transfers



# NoC-Enabled Spatial Compute

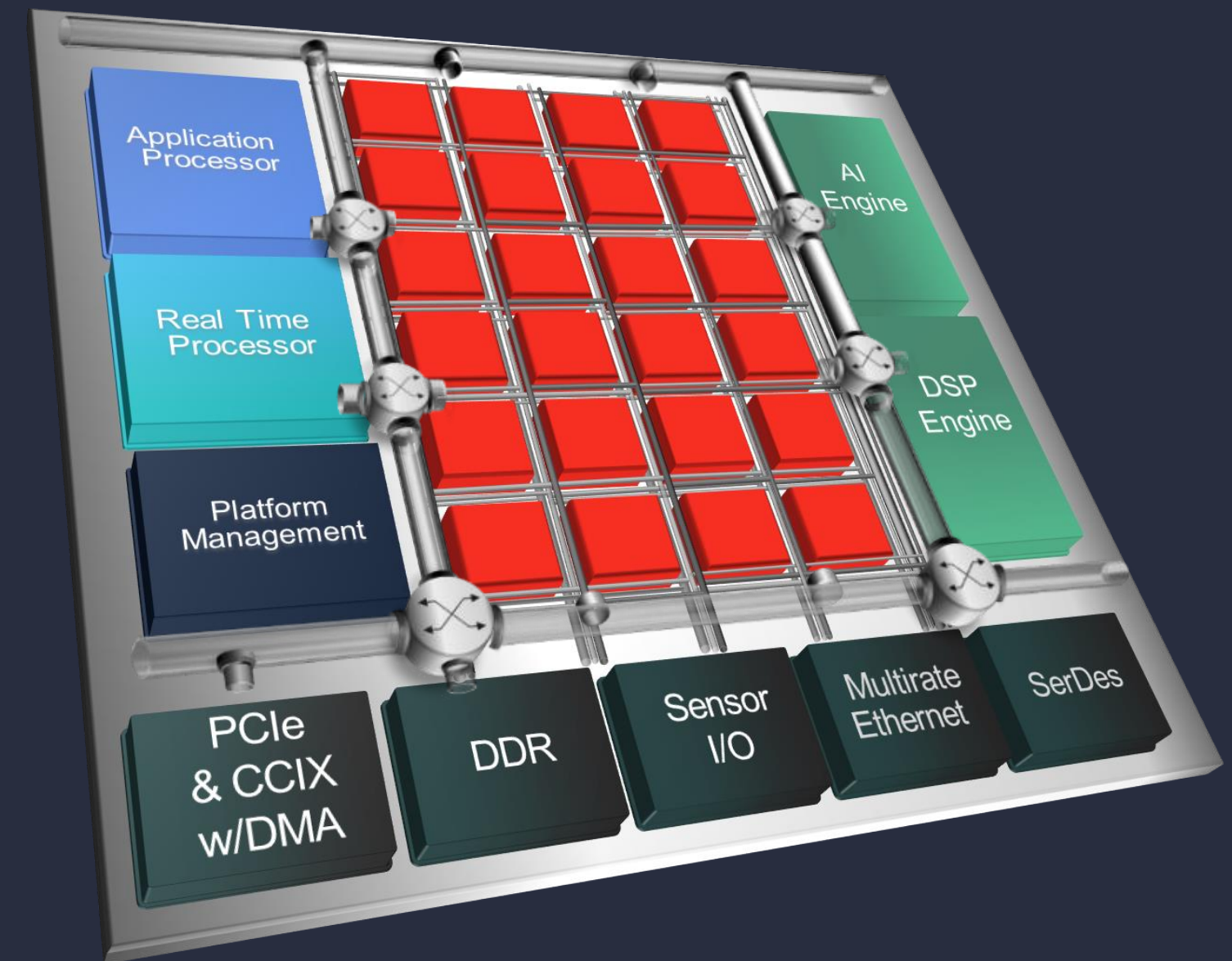
APPLICATION DOMAIN



**Modular Integration and Isolation**  
Relocatable, position-independent modules

**NoC = “Always-On” Backbone**  
Enables dynamic linking

**Dynamic Function Exchange (DFx)**  
Enables dynamic loading





# NoC-Enabled Spatial Compute

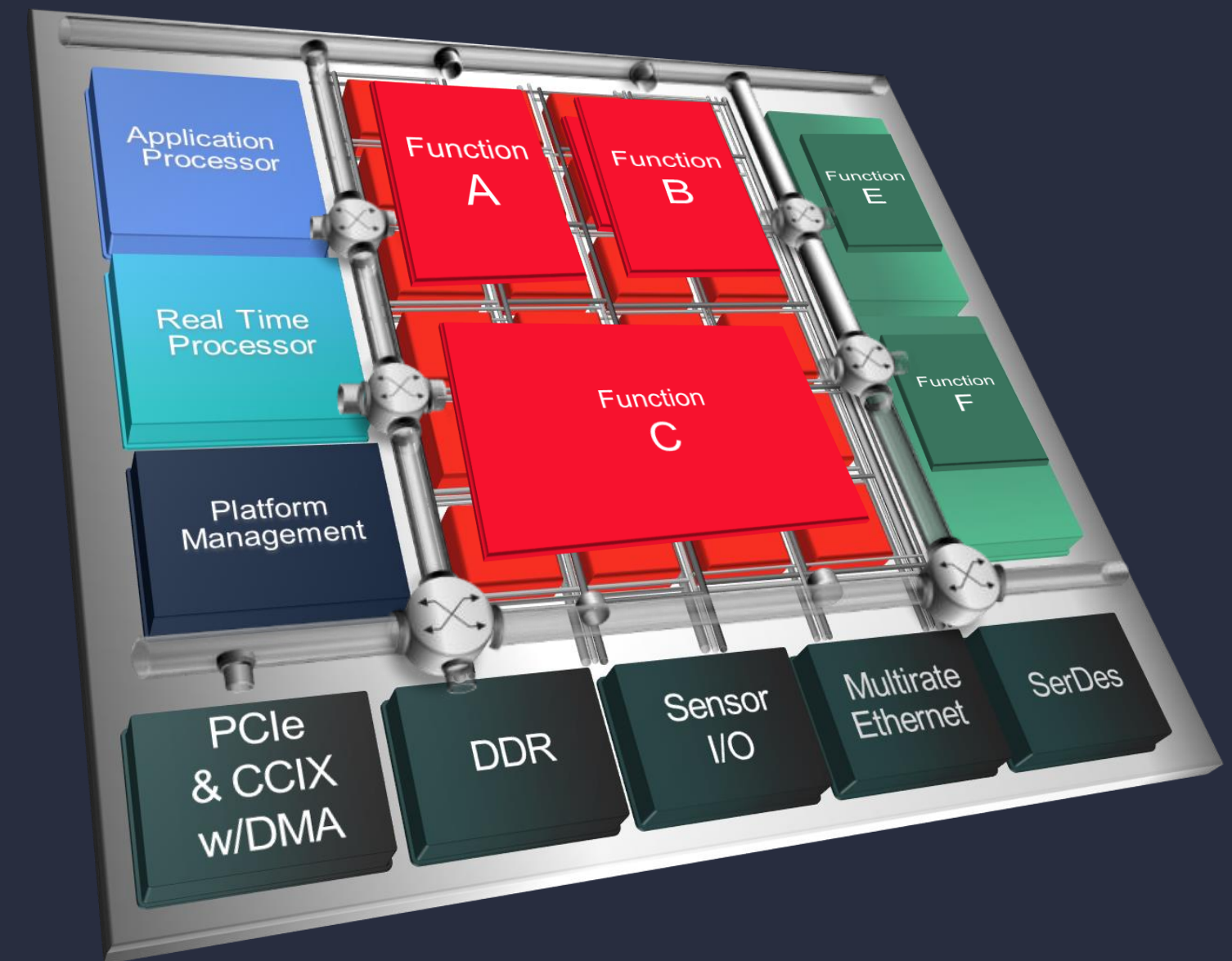
APPLICATION DOMAIN



**Modular Integration and Isolation**  
Relocatable, position-independent modules

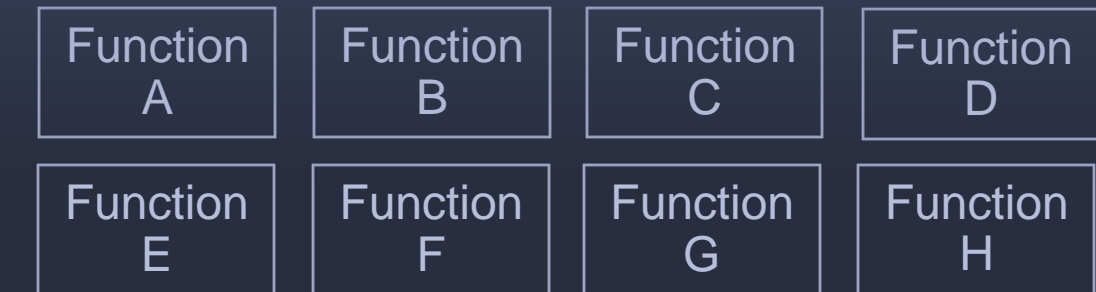
**NoC = “Always-On” Backbone**  
Enables dynamic linking

**Dynamic Function Exchange (DFx)**  
Enables dynamic loading



# NoC-Enabled Spatial Compute

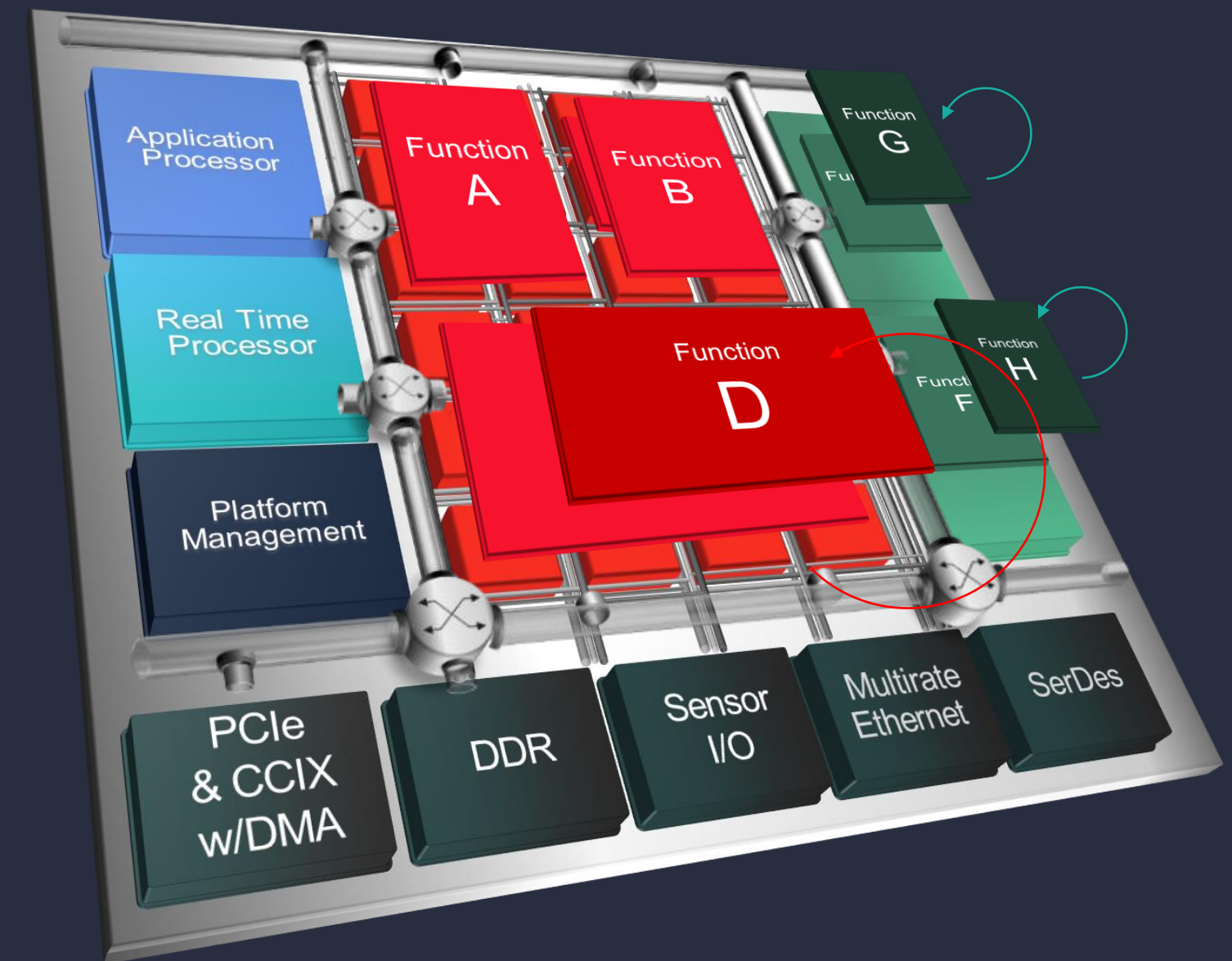
APPLICATION DOMAIN



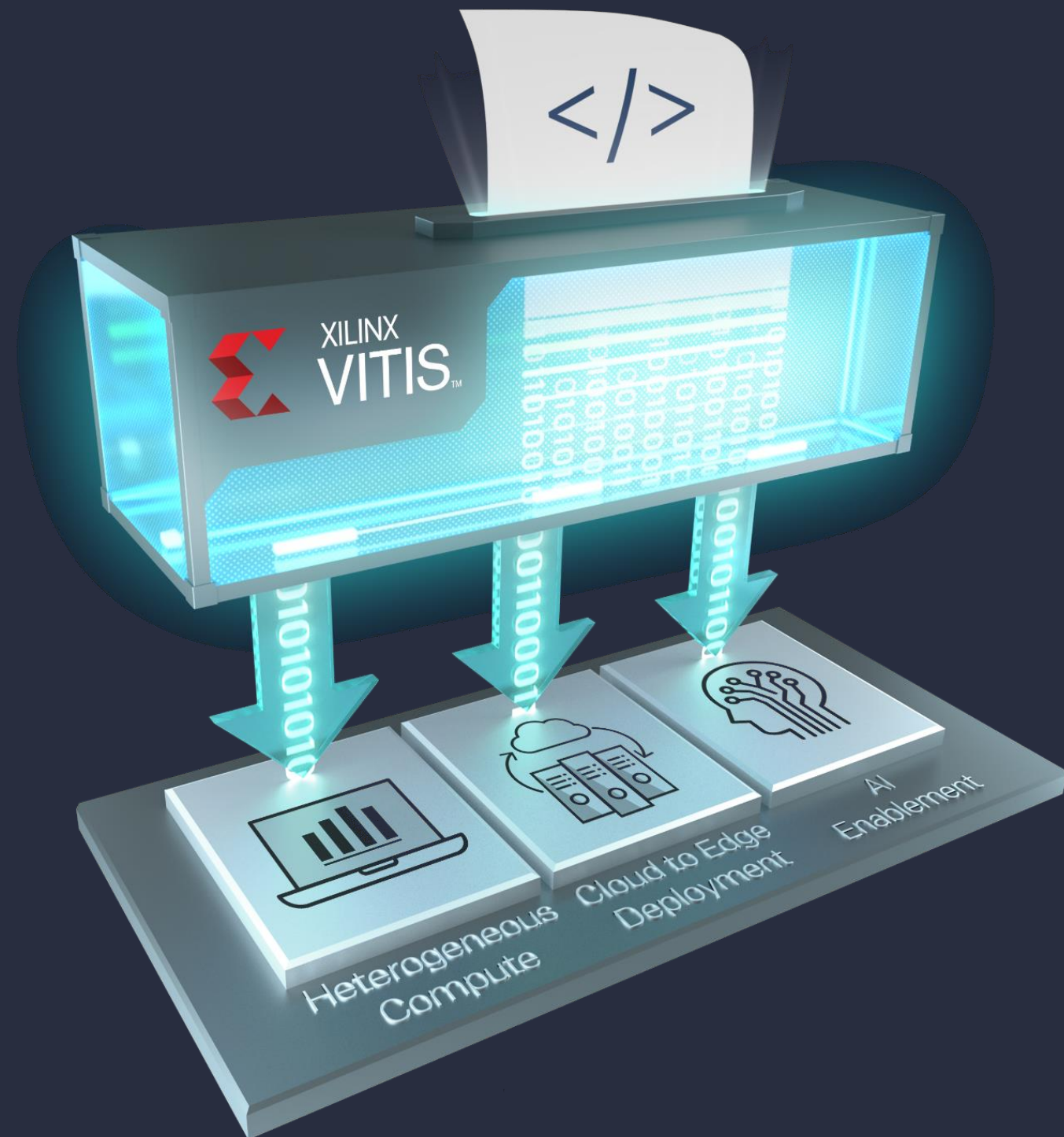
**Modular Integration and Isolation**  
Relocatable, position-independent modules

**NoC = “Always-On” Backbone**  
Enables dynamic linking

**Dynamic Function Exchange (DFx)**  
Enables dynamic loading



# Vitis Unified Software Platform



Free

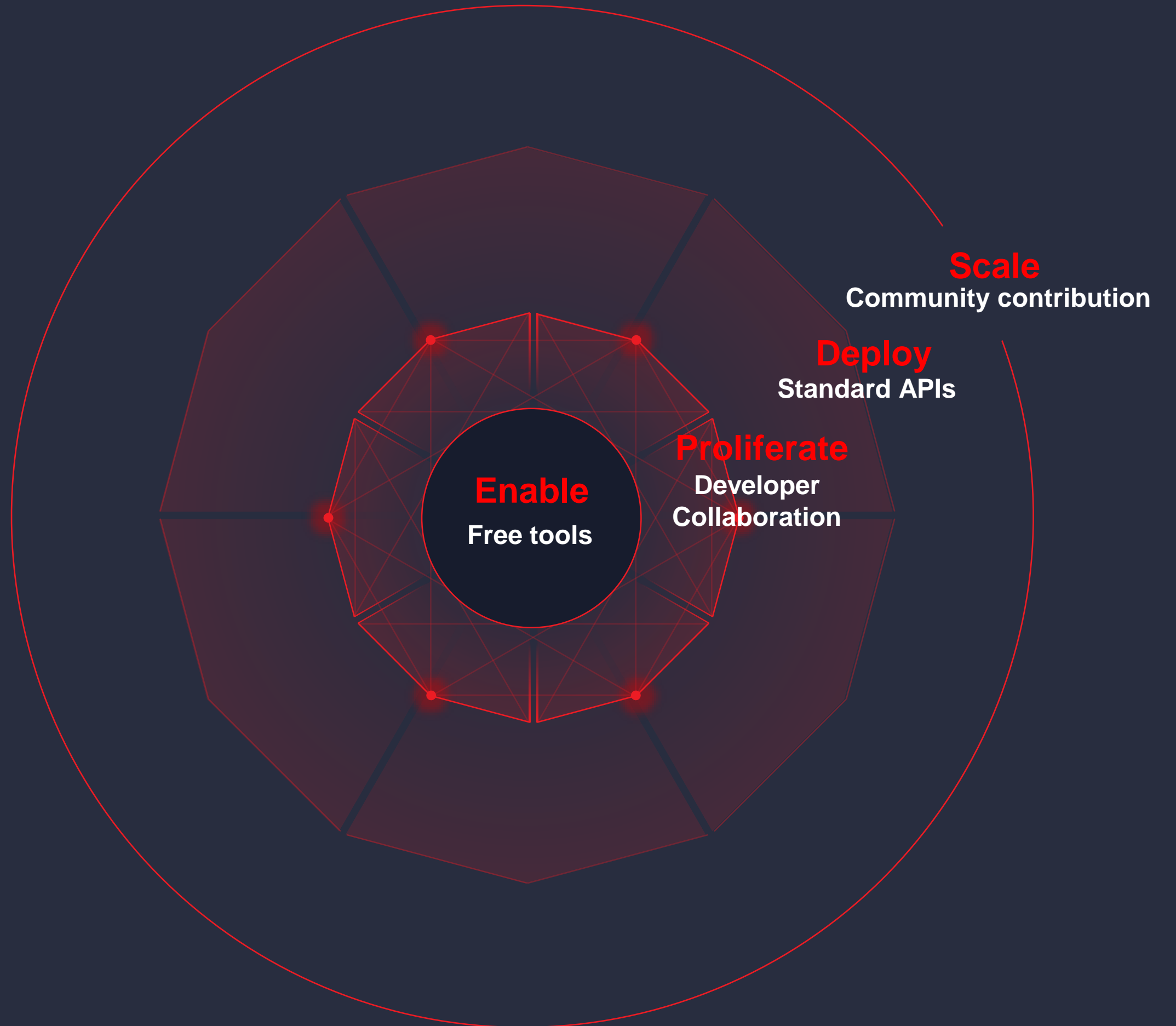
Unified

Open



# OPEN SOURCE

---



# Open Source

Today

Future



Communities

**Scale**

Available

Vitis AI Open Source

Compression Algorithms | Quantization Format | Pytorch & TF Plugins

Kubernetes & Docker Plugins



Libraries

**Deploy**

Standard

Vitis Libraries

Video | Fintech | AI Models | Math | BLAS | OpenCV

Integration with Upstream Projects



Runtime  
& System SW

**Proliferate**

Open

Vitis Runtime

Open Source

OpenAMP

Linaro Foundation

Vitis Compiler Frontends



Libraries

**Enable**

Free

Vitis Tools Free

Vivado Images in Clouds

Select Vivado modules as APT packages