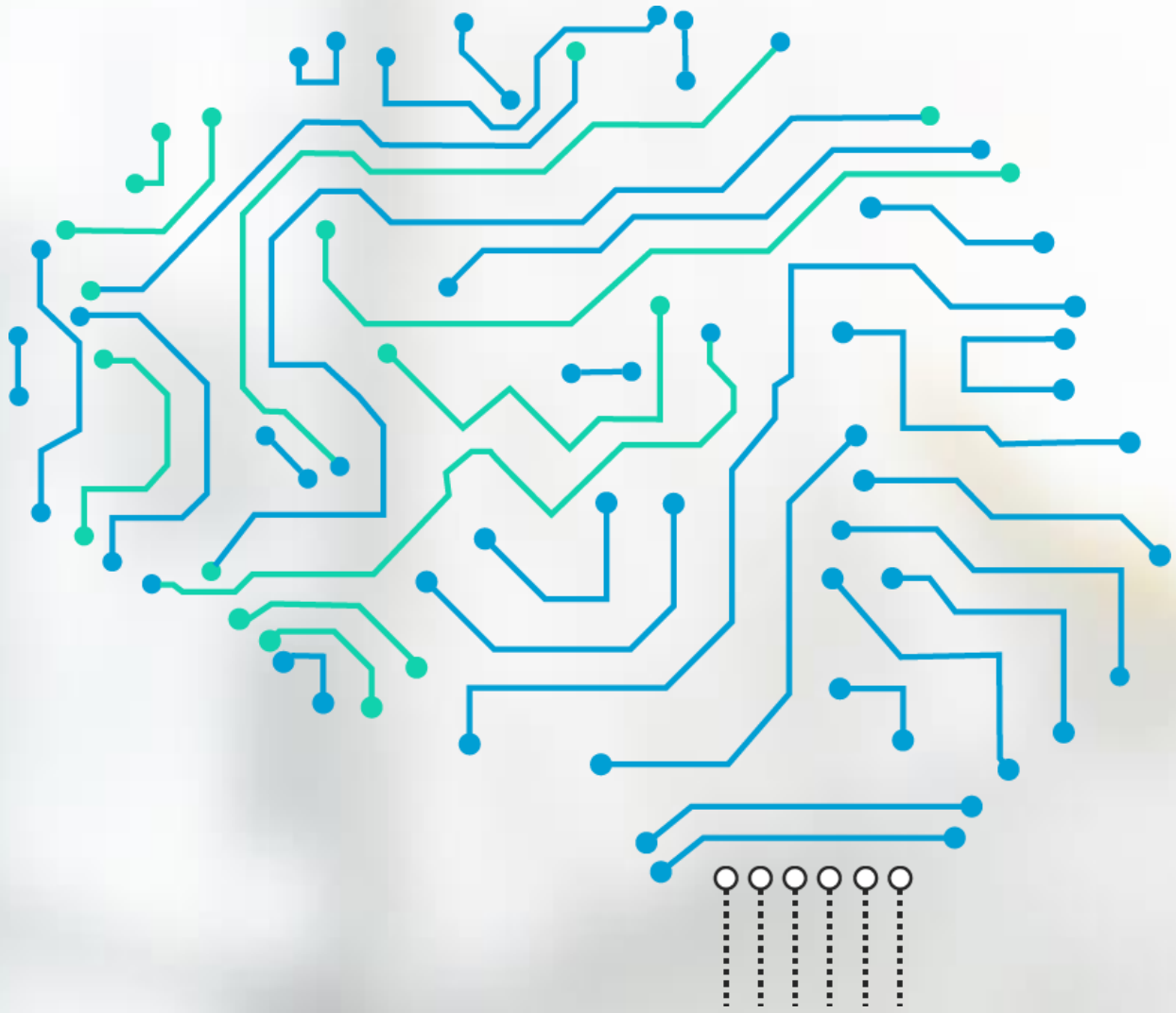
MPSoC 2017

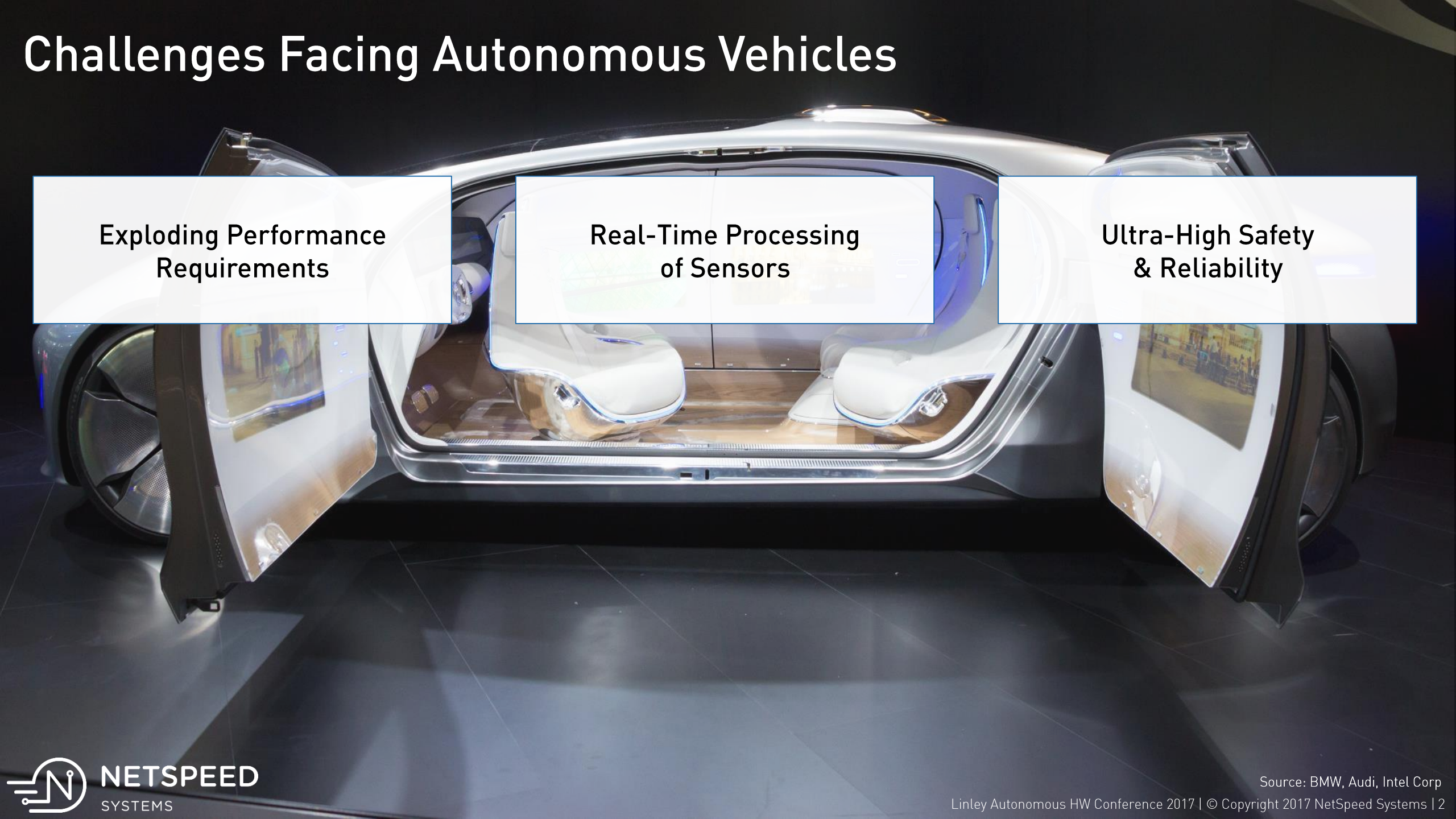
Machine learning Based Intelligent Interconnect for Next Generation Autonomous Vehicle SoCs

Rocco Jonack

NETSPEED SYSTEMS



Challenges Facing Autonomous Vehicles



Exploding Performance Requirements

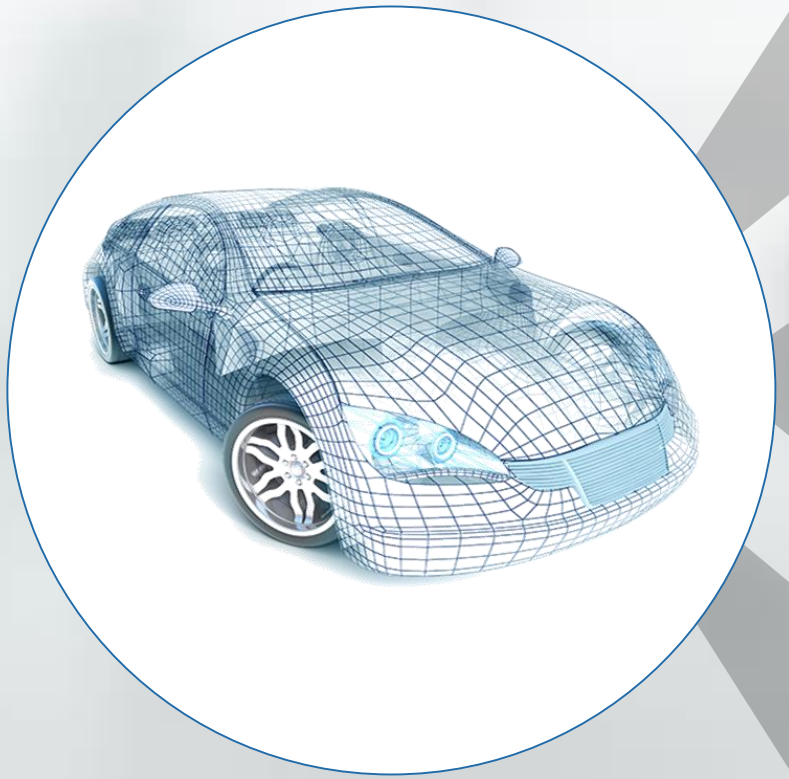
Real-Time Processing of Sensors

Ultra-High Safety & Reliability

NETSPEED SYSTEMS

Source: BMW, Audi, Intel Corp

Translating System-Level Requirements → SoC Level



Exploding Performance Requirements

- Rise of heterogeneous architectures & right-sized compute
- Cache coherency & End-to-end QoS of critical importance

Real-Time Sensor Processing

- Different IPs with differing requirements
- Ensuring communication happens without any deadlocks

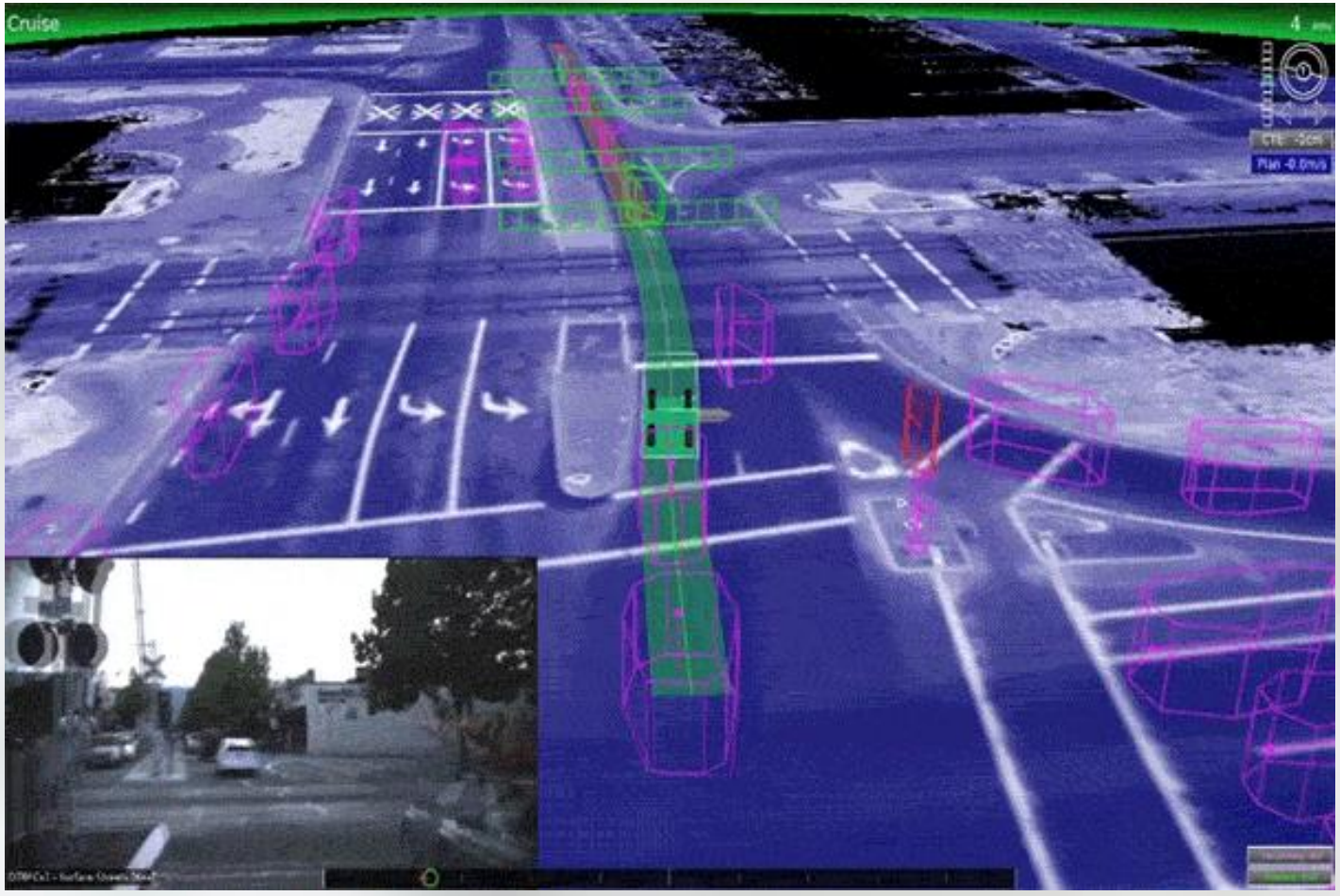
Ultra-High Safety & Reliability

- Pressure to comply to industry standards – ISO 26262
- Functional Safety – Performance – Area Tradeoffs

NETSPEED SYSTEMS

Autonomous Driving: Use Case Compute Flow

Image Acquisition	Noise removal
	Pixel processing
	Image pyramids
Feature Extraction	Optical flow
	Edge detection
	Gradient detection
Feature Processing	Segmentation & filtering
	Object tracking
	Object detection
Pattern Recognition	Feature reduction
	Feature classification
	Augmentation
Feedback and Action	Computation & processing
	Feedback loop
	Avoidance signalling



NETSPEED SYSTEMS

Source: Extreme Tech, Google, ARM

Need For Heterogeneous Computing

Image Acquisition	Noise removal
	Pixel processing
	Image pyramids
Feature Extraction	Optical flow
	Edge detection
	Gradient detection
Feature Processing	Segmentation & filtering
	Object tracking
	Object detection
Pattern Recognition	Feature reduction
	Feature classification
	Augmentation
Feedback and Action	Computation & processing
	Feedback loop
	Avoidance signalling

- Smaller amounts of data
- Highly structured data
- Complex computation/item

- Lots of data
- Simple computation/item
- Massive parallelism

CPU

DSP, Accel

GPU, ISP

Feedback loop

Noise removal

Noise removal

Segmentation & filtering

Pixel processing

Feature reduction

Image pyramids

Feature classification

Edge detection

Computation & processing

Object tracking

Gradient detection

Object detection

Avoidance signalling

Augmentation

Optical flow

Source: Extreme Tech, Google, ARM

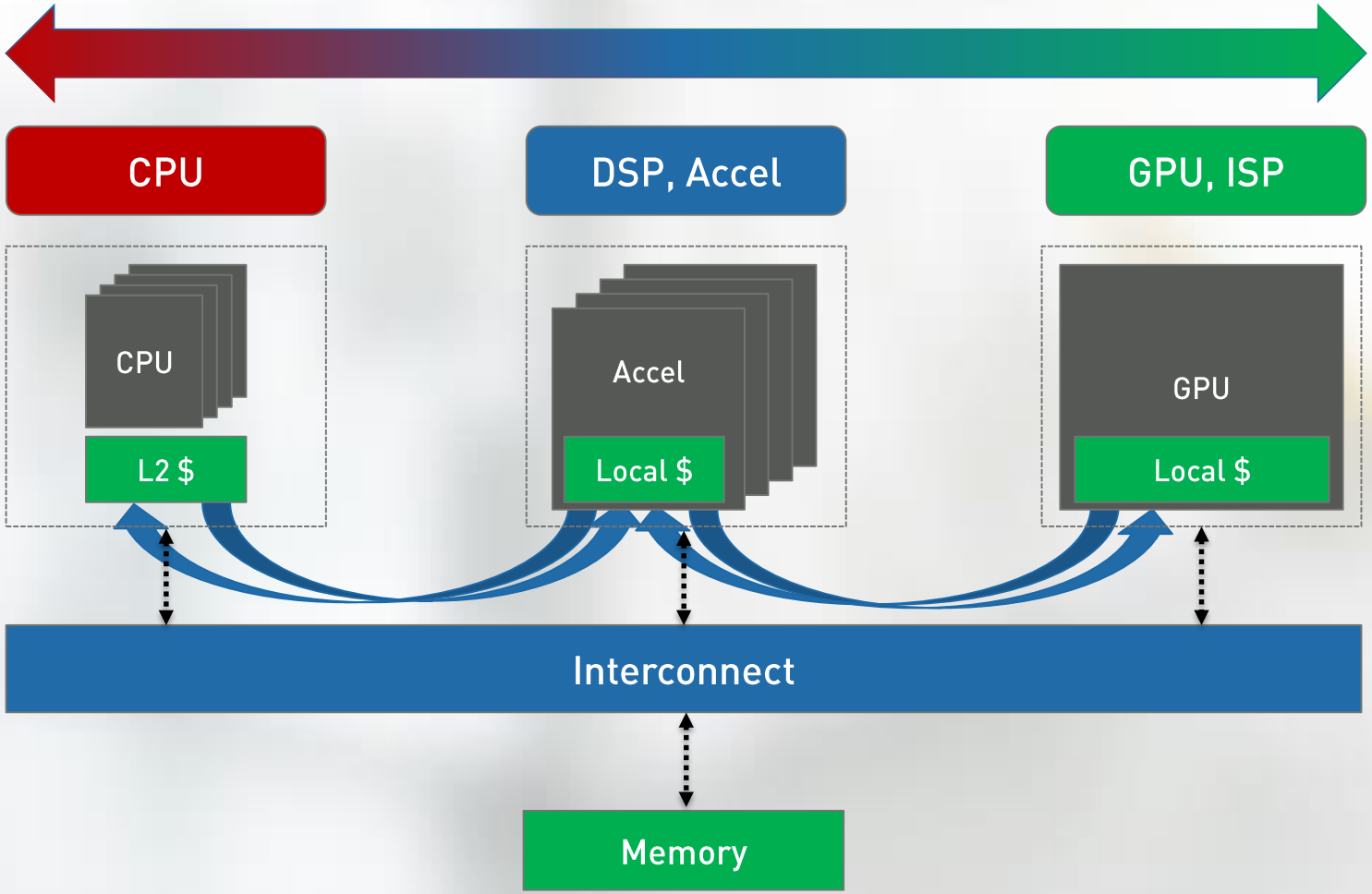
NETSPEED SYSTEMS

Challenges With Heterogeneous Computing

Cache Coherency	Ensure every compute engine has uniform view of system memory
System-Level QoS	BW-Hogs, Latency-sensitive & Real-time engines need seamless access to memory
End-to-End Functional Safety	Underlying architecture should be resilient & tolerant to random & systematic errors

- Smaller amounts of data
- Highly structured data
- Complex computation/item

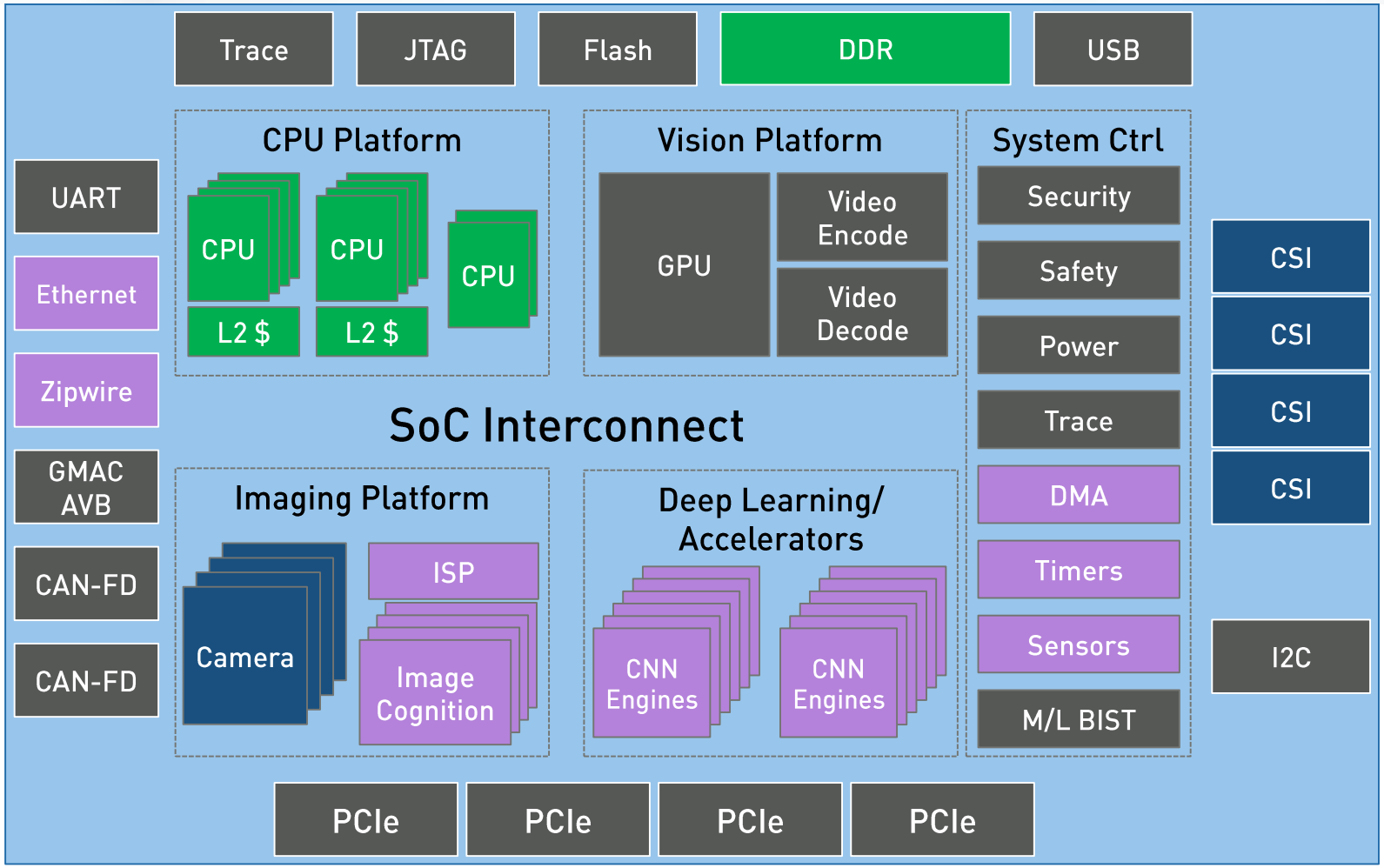
- Lots of data
- Simple computation/item
- Massive parallelism



Source: Extreme Tech, Google, ARM

NETSPEED SYSTEMS

Next-Gen Autonomous SoC Architecture Details



RIGHT SIZED COMPUTING

- Single & multi-threaded compute engines
- Differing access patterns, spatial/temporal locality and performance requirements

SEAMLESS CACHE COHERENCY

- Uniform shared view of system memory
- Interprocessor communications lead to network and protocol level deadlocks

ROBUST ARCHITECTURE

- Dynamically changing workloads
- Handle changing use cases/SW needs
- Highest level of fault tolerance

High priority Dynamic Real time/Isochronous Low Priority

NETSPEED SYSTEMS

Existing Approaches Fall Short

Hard, Fixed Point Designs

- Fixed topology with limited configurability
- Coherency through "tiled" structures and regular connection patterns

Manual, Hand-optimized

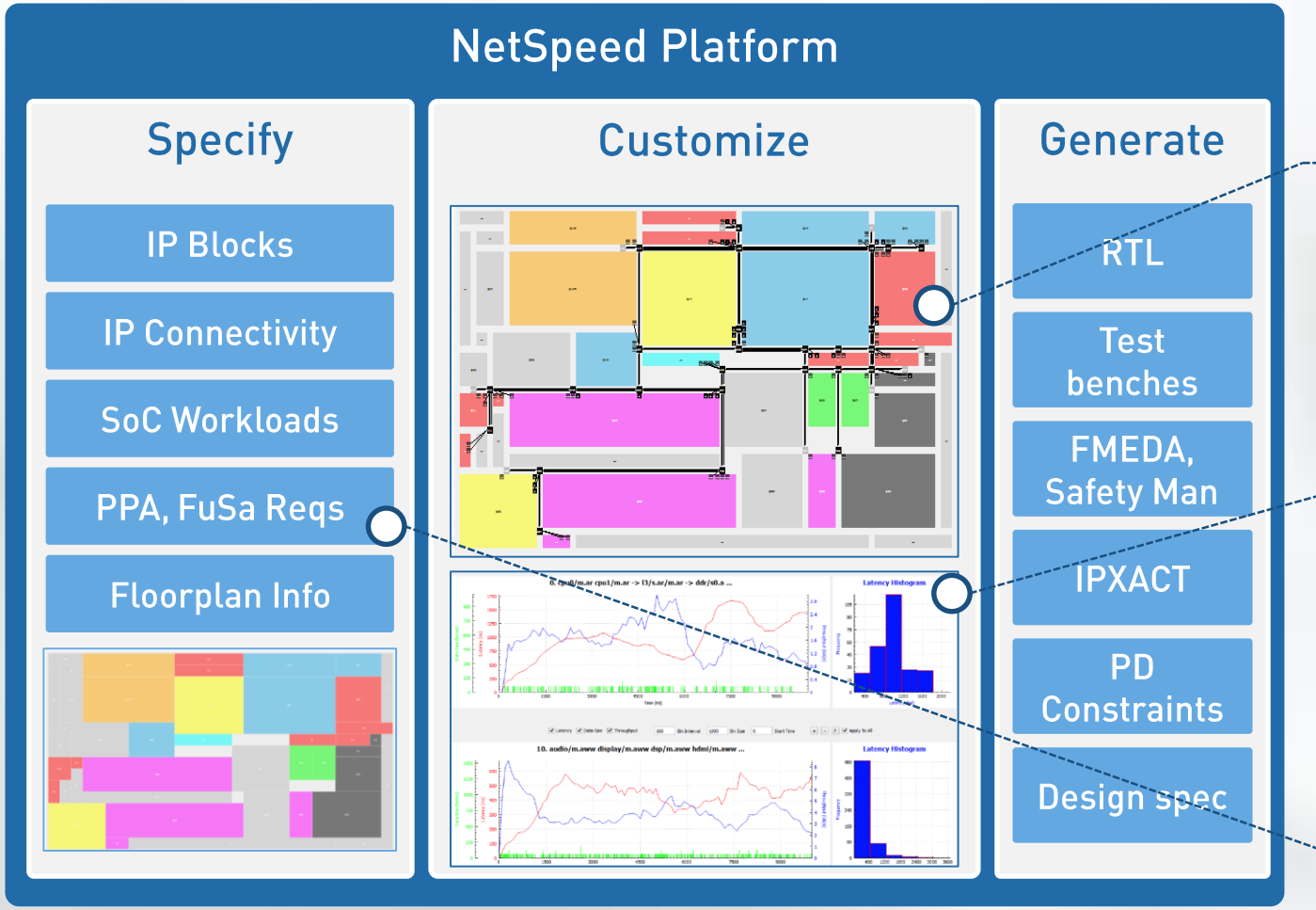
- Hand-optimized sub systems
- Divide and conquer approach: Separate coherent, non-coherent
- Deadlock prone designs

Key Elements Are Afterthoughts

- QoS schemes are patched, built on top of existing infrastructure
- Functional safety features is "added-on" instead of being architected in the solution

NETSPEED SYSTEMS

NetSpeed Technology: Intelligent Interconnect Architecture



Machine Learning Interconnect Construction

Scalable Coherency & SoC-Level QoS

Design Cockpit for PPA & FuSa Tradeoffs

NETSPEED SYSTEMS

Machine Learning Based Interconnect Construction

NetSpeed Platform

SoC Specifications

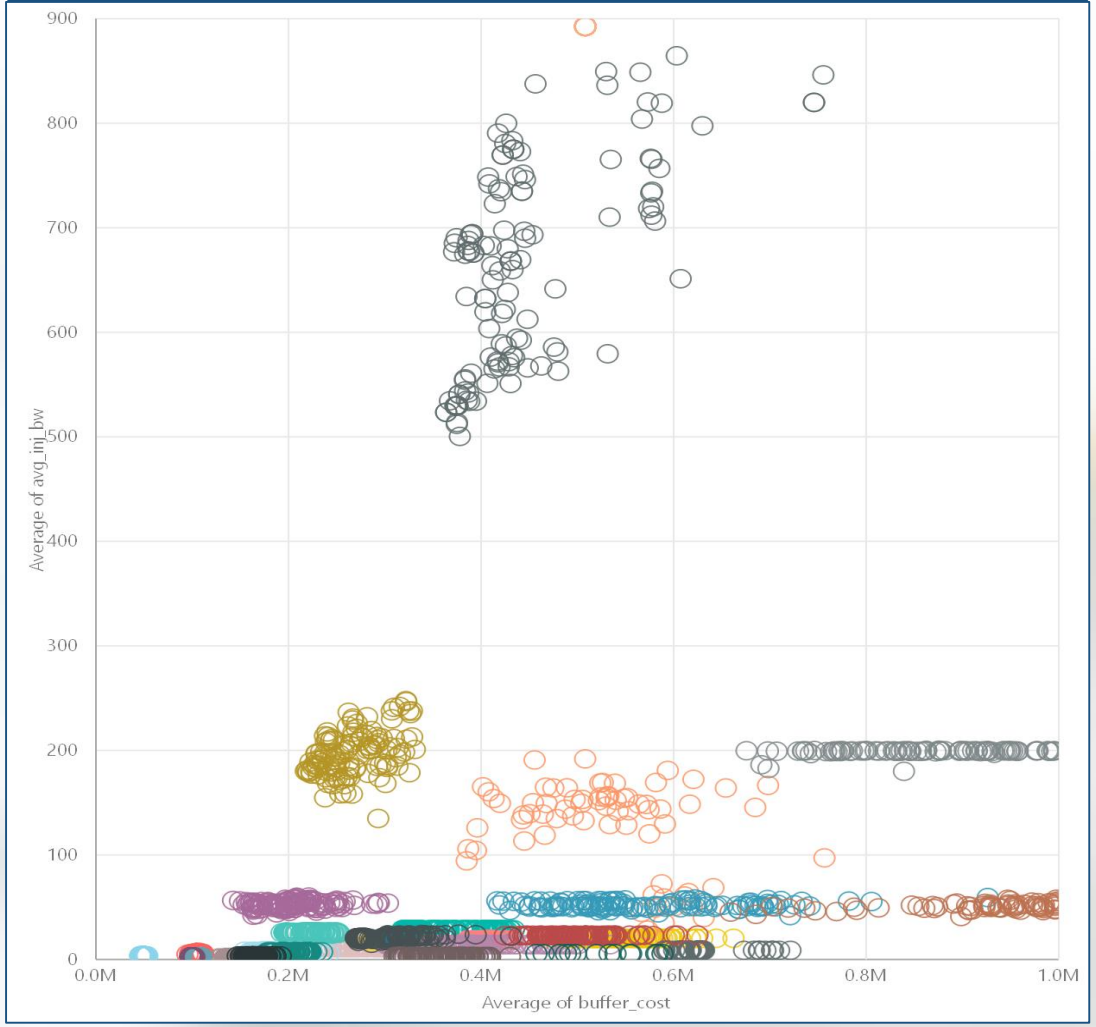
- Performance, Area, FuSa
- Workloads, SoC Use Cases
- IP Blocks, Compute Engines

Turing

NetSpeed Machine Learning Engine ⟷ Predictive Models

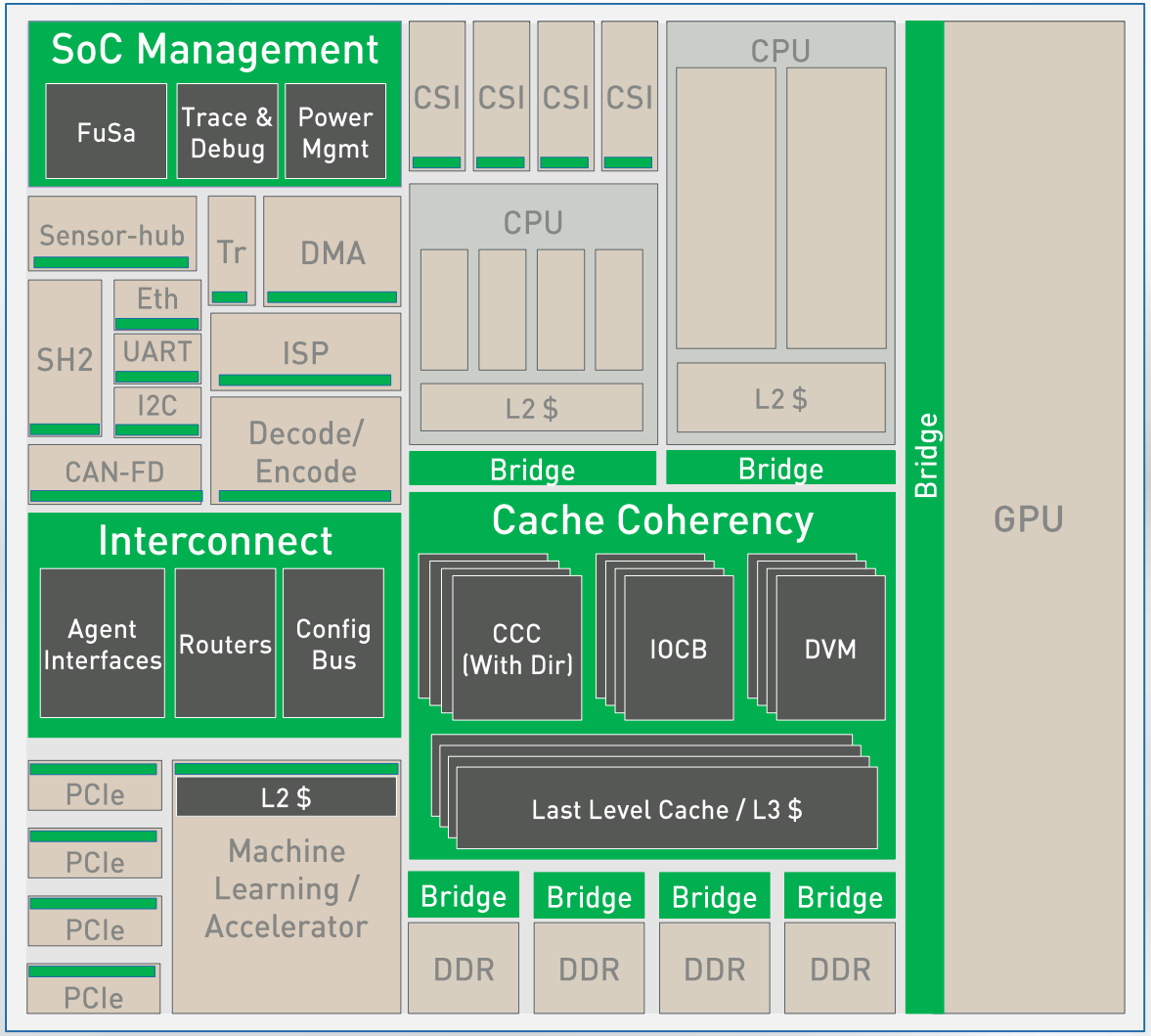
- **Orion** — Physically-Aware Interconnect
- **Gemini** — Cache-Coherent Interconnect
- **Crux** — High Performance Interconnect
- **Pegasus:** Last-Level Cache

Results



NETSPEED SYSTEMS

Scalable Cache Coherent Interconnect



NetSpeed Gemini Components

Configurable Cache Coherency

- Scalable coherency solution: Modular & programmable
- In-built directory support
- Specialized IO-Coherency accelerator

Multi-Level Caching Options

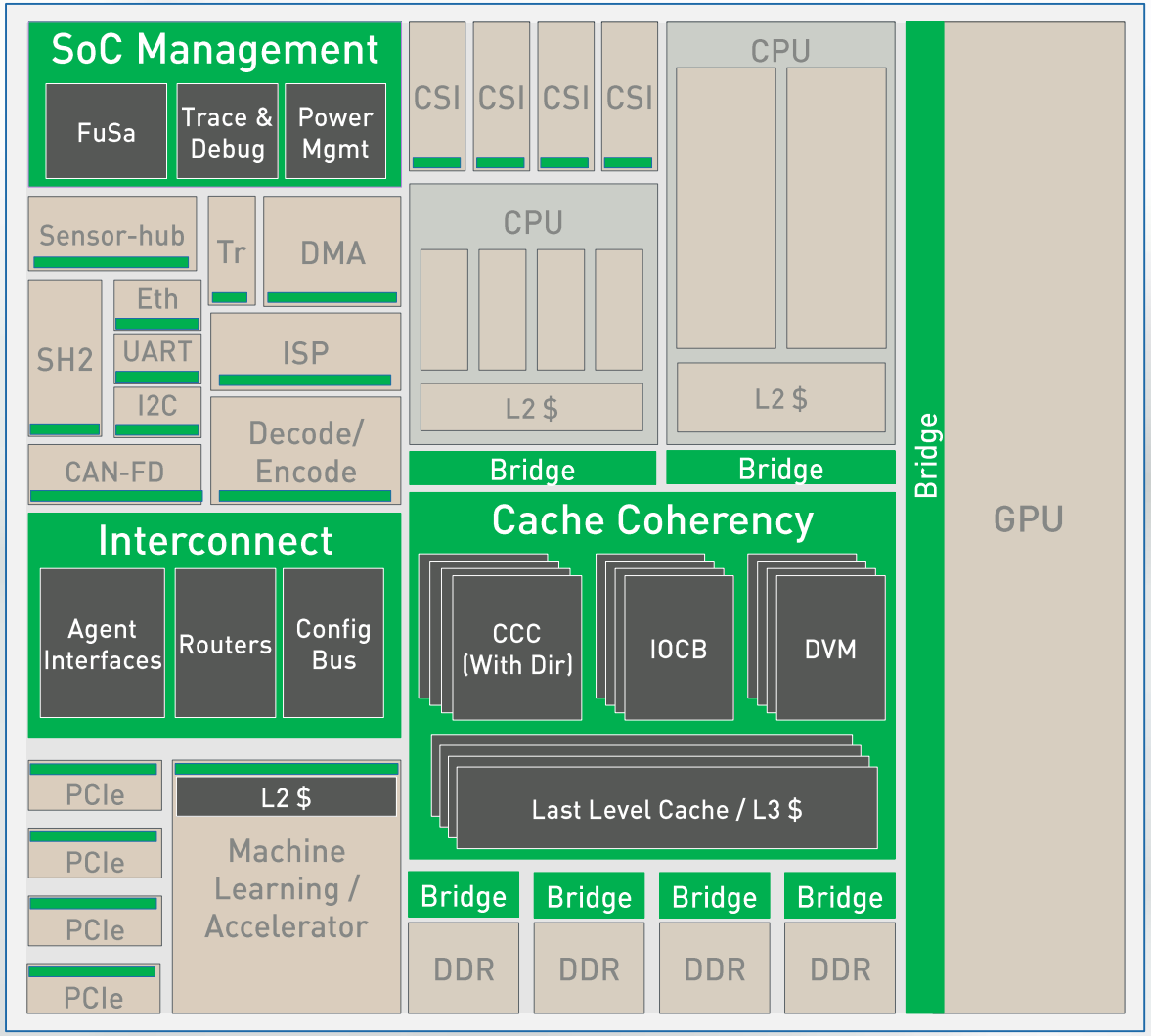
- Last-level cache, L3 $ → Higher bandwidth
- L2 $ → Reduce critical latency
- Programmable allocation policies

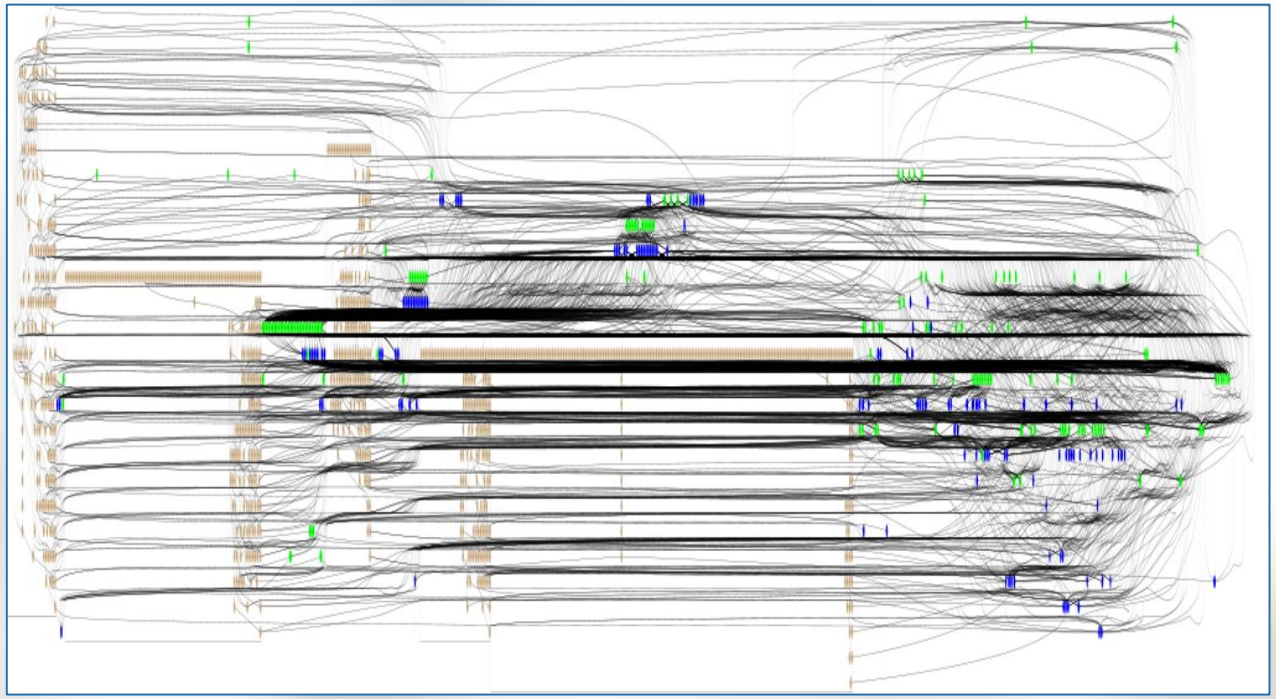
Scalable Solution

- 64 cache coherent cluster; 250 IO coherent IPs
- Seamless connection with DDR, HBM memories
- In-built deadlock detection and avoidance

NETSPEED SYSTEMS

Scalable Cache Coherent Interconnect: Built-In Deadlock Avoidance





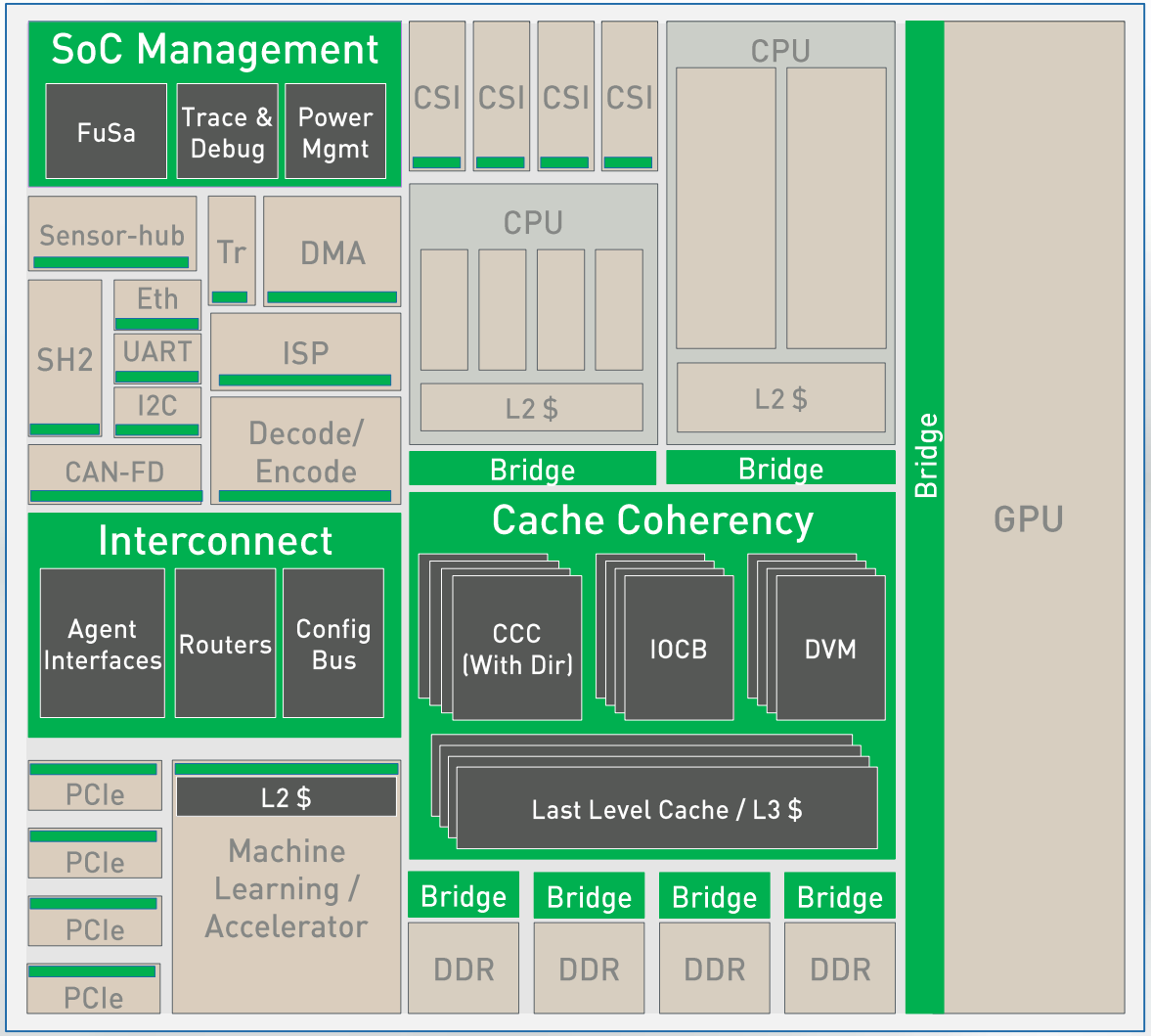
Built-In Deadlock Detection & Avoidance

- Formal methods and graph theory algorithms
- User-driven traffic dependencies
- Handles complex topologies and routing

NetSpeed Gemini Components

NETSPEED SYSTEMS

Advanced SoC-Level QoS Schemes



NetSpeed Gemini Components

Dynamic QoS Control

- 16 Traffic Classes & 64 Virtual Channels with Dynamic priority
- Low-latency QoS control for Isochronous traffic flows

End-to-End QoS

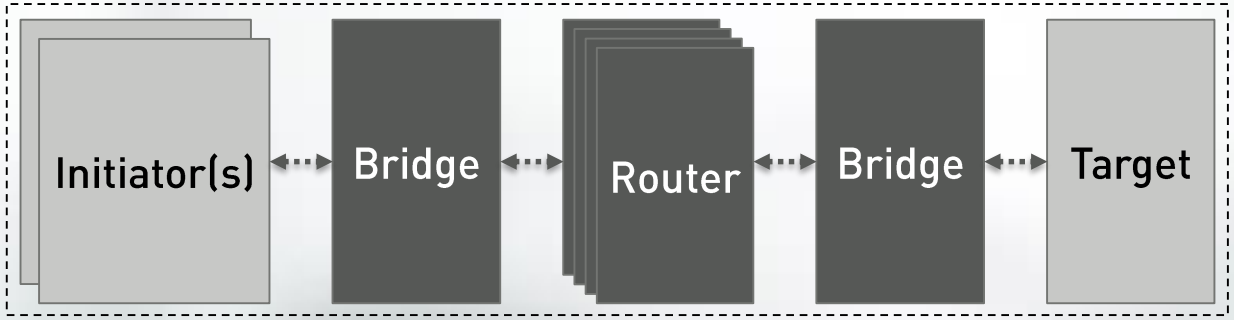
- Improved, lower-latency flow-control with memory scheduler
- Non head-of-line blocking schemes with guaranteed delivery

Runtime Programmability

- Runtime programmable weighted BW allocation
- Adaptive control to DVFS modes without software intervention

NETSPEED SYSTEMS

Functional Safety: ISO 26262 ASIL-D



End-to-end User ECC/Parity

ECC/Parity | Interconnect ECC/Parity | ECC/Parity

Hop-to-Hop ECC/Parity

Port Check | Logic Duplication, Redundant Routes

Initiator Timeout | Interconnect Timeouts | Target Timeouts

Error Detection & Correction

Fault Tolerance & Resilience

Timeouts

FuSa Architected In With Top-Down Approach

- FuSa features considered first class citizens from Day #1
- Interplays cleanly with coherency and ISO 26262 standard
- End-to-end protection, logic redundancy & timeouts

Unprecedented Configurability

- Fine grained FuSa feature control for low area overhead
- Design cockpit for Performance vs. FuSa vs. area tradeoffs
- Rapid analysis and convergence

ASIL-D Ready

- First & Only Coherent Interconnect IP
- Detailed FMEDA analysis & reporting for any configuration
- Comprehensive Safety Manual & Safety Report

SGS TÜV SAAR ASIL D READY Functional Safety www.sgs-tuev-saar.com

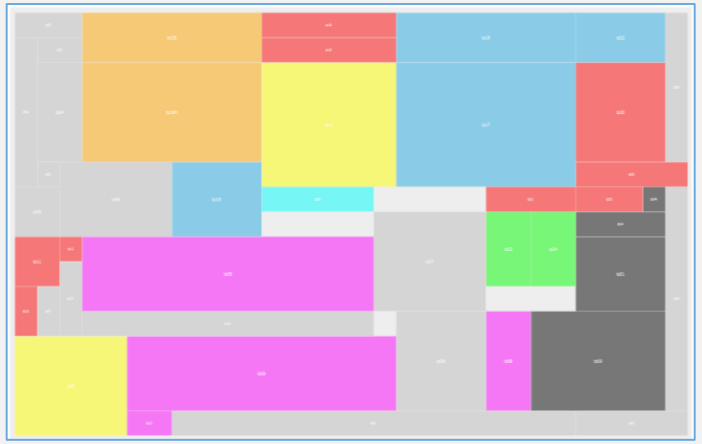
NETSPEED SYSTEMS

Design Cockpit: Balancing Performance vs. Area vs. FuSa

NetSpeed Design Cockpit

Step 1: Specify

- IP Blocks & Connectivity
- PPA Requirements
- Functional Safety Goals
- SoC Workloads



Step 2: Customize

Design Exploration: PPA, FuSa Tradeoffs

Step 3: Generate

Design
Synthesizable, Partitionable RTL

Verification
Checkers, Monitors, Scoreboards

Physical Design
DEF, SDC, Clk Skew Mgmt, PD Scripts

SoC Integration
IPXACT, CPF/UPF, Arch Manual

Functional Safety
FMEDA, Safety Manual

NETSPEED SYSTEMS

Customer Case Study

Customer

- Tier #1 Automotive manufacturer and Tier #1 ADAS company
- Architecture and frontend design by customer
- Back-end by 3rd party ASIC vendor

Challenge

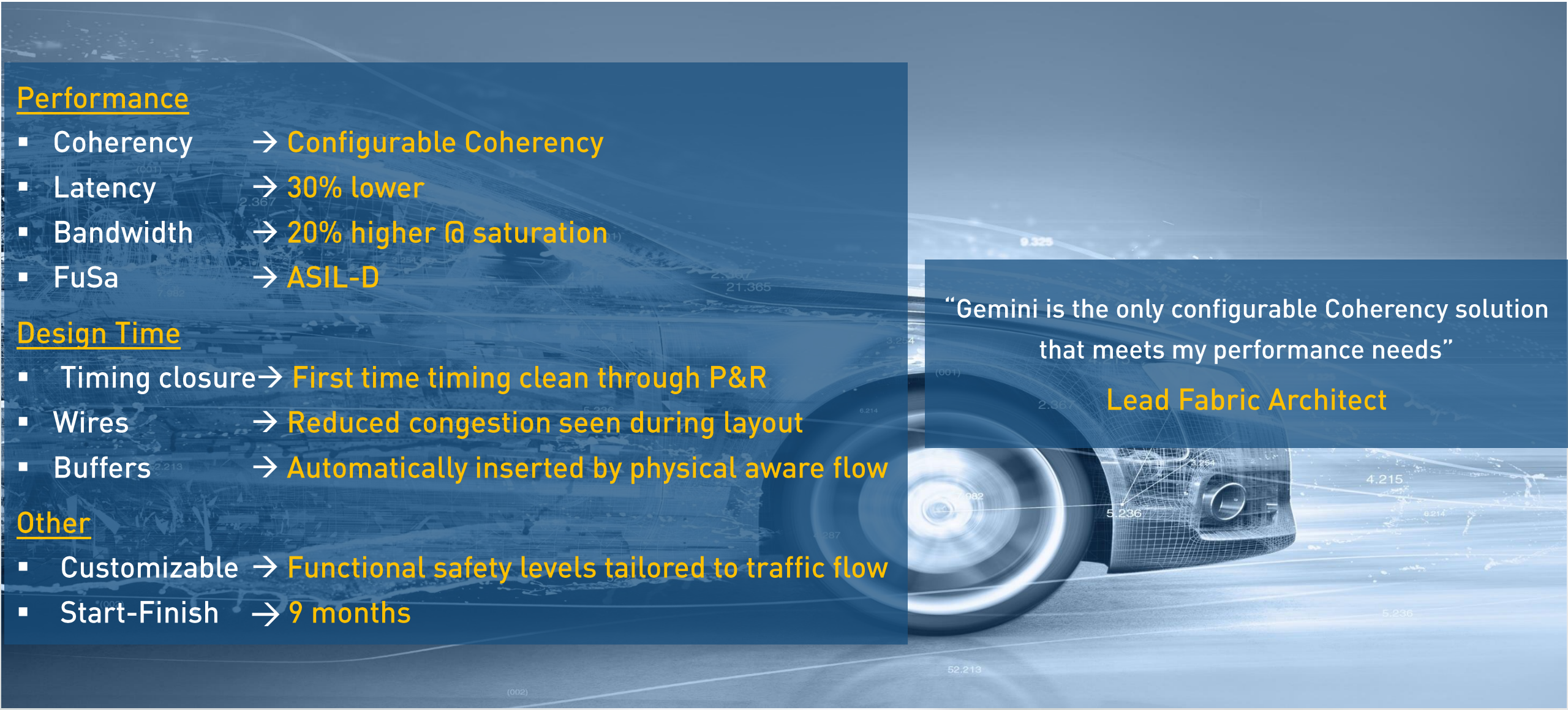
- Many clusters of CPUs and proprietary accelerators requiring high-bandwidth, low latency, distributed coherent interconnect.
- No other solution in the market has Robust Real-time QoS support
- Reliability and safety features for ADAS market

Solution

- Scalable Gemini performance far exceeds competing interconnect capabilities
- User-controlled automation with integrated performance analysis allows architect to tune Bandwidth/latency/area tradeoffs for specific traffic flows
- Functional safety architected into interconnect, configured on a per-traffic-flow granularity

NETSPEED SYSTEMS

Customer Case Study



Performance

- Coherency → Configurable Coherency
- Latency → 30% lower
- Bandwidth → 20% higher @ saturation
- FuSa → ASIL-D

Design Time

- Timing closure → First time timing clean through P&R
- Wires → Reduced congestion seen during layout
- Buffers → Automatically inserted by physical aware flow

Other

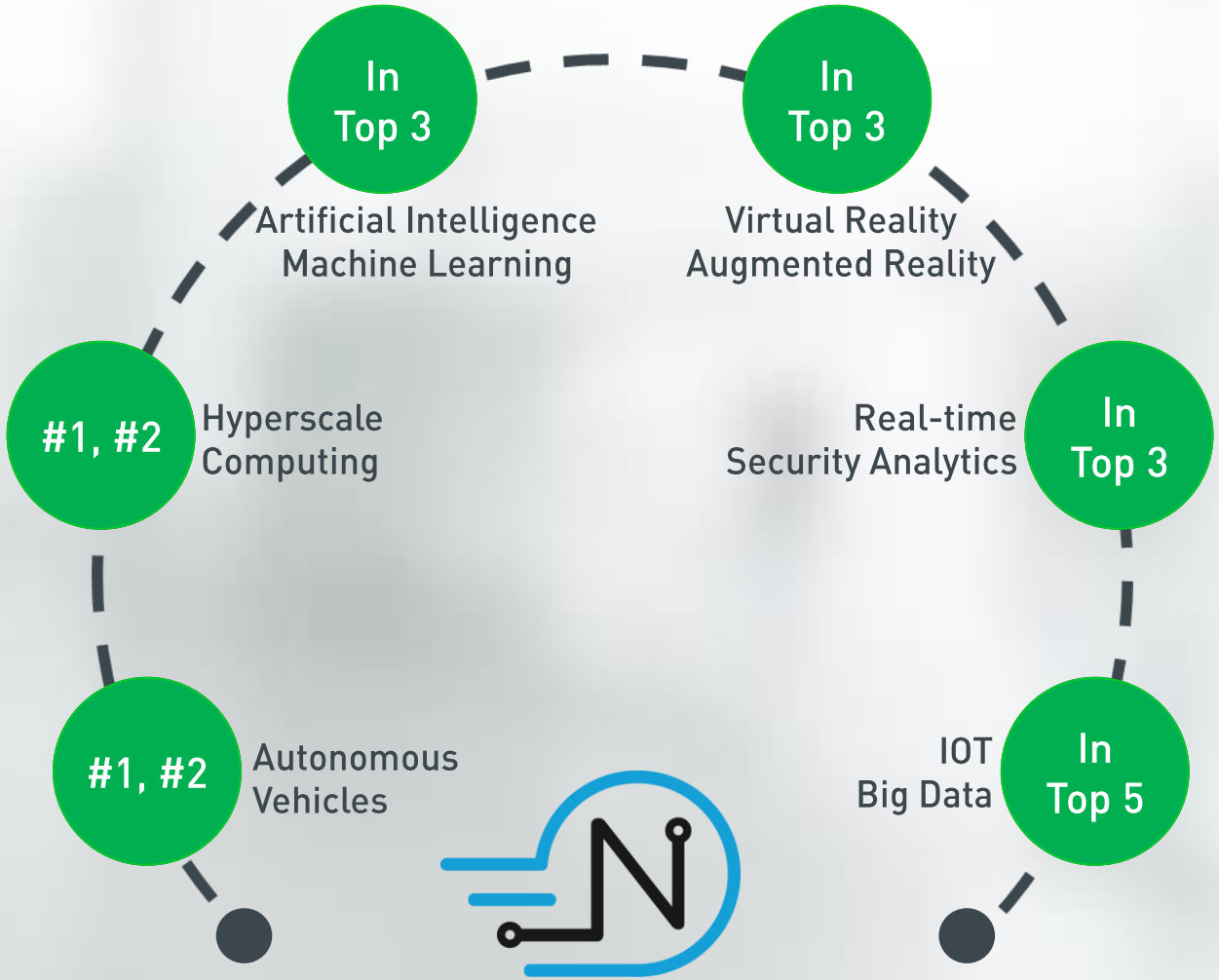
- Customizable → Functional safety levels tailored to traffic flow
- Start-Finish → 9 months

"Gemini is the only configurable Coherency solution that meets my performance needs"

Lead Fabric Architect

NETSPEED SYSTEMS

NetSpeed Technology: Choice of Next-Gen Application Leaders



NetSpeed Systems

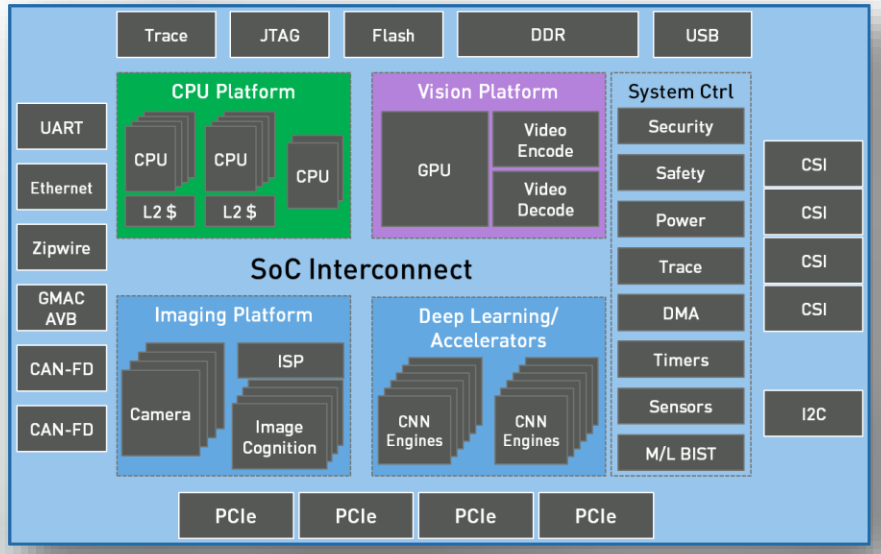
The NetSpeed Difference

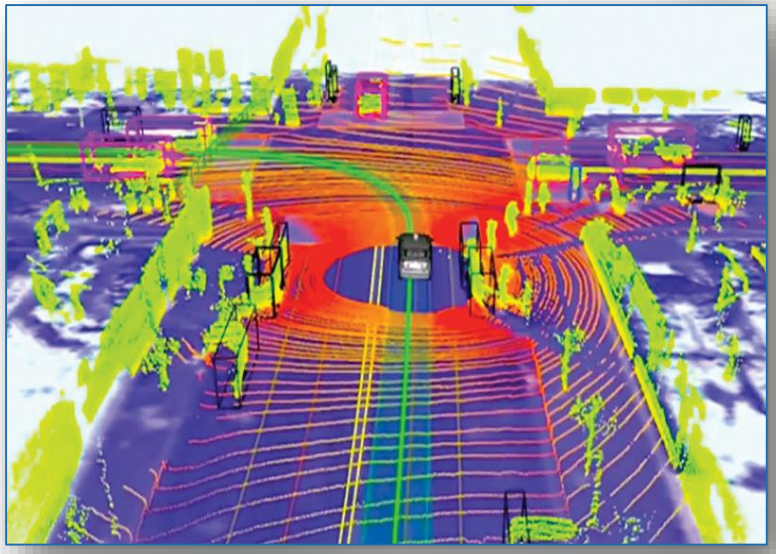
- Delivering Intelligent Interconnect technologies to silicon
- Advanced Machine Learning algorithms to optimize SoC designs
- Innovative team with a proven track record of silicon execution

Adopted By Industry Leaders

NETSPEED SYSTEMS

Summary





HETEROGENEOUS ARCHITECTURES

- Cache coherency critical for delivering high performance

NETSPEED TECHNOLOGY

- Scalable coherency solution with uniform view of system memory

REALTIME SENSOR PROCESSING

- Complex IP interactions needing sophisticated QoS schemes

NETSPEED TECHNOLOGY

- Advanced QoS schemes with traffic isolation & BW allocation

SAFETY & RELIABILITY

- Functional Safety needs to be architected, not added-on

NETSPEED TECHNOLOGY

- ASIL-D Ready: Robust safety mechanisms architected in IP

NETSPEED SYSTEMS


NETSPEED
SYSTEMS