

# FUJITSU-MONAKA: Next-Generation Processor for Sovereign, Sustainable AI & HPC



\* This presentation is based on results obtained from a project subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

# Market Context and FUJITSU-MONAKA Value Proposition

- Growing AI & HPC demand is reshaping requirements for computing infrastructure.

## Key Trends

### Workload diversification Evolving compute demands

- Single large-scale  
→ Diversified & high- throughput workloads
- AI infrastructure and HPC-AI convergence

### Power constraints

- Rapid growth in compute & power consumption
- Need to maximize performance under power limits

### Europe-specific requirement

- Need for control over data, infrastructure, and operations
- Driven by geopolitics, export controls, supply chain risks, and vendor lock-in concerns

## FUJITSU-MONAKA Value



### Powering Diverse AI/HPC Workloads

- Up to 2× higher performance in AI workloads
- High-throughput CPU architecture
- Open AI/HPC ecosystem



### Over 50% TCO Reduction at the Same Performance

- High performance per watt



### Sovereign & Trusted computing

- Control over data and infrastructure
- Secure execution and data protection in use

# CPUs as the Foundation for Evolving AI/HPC Systems

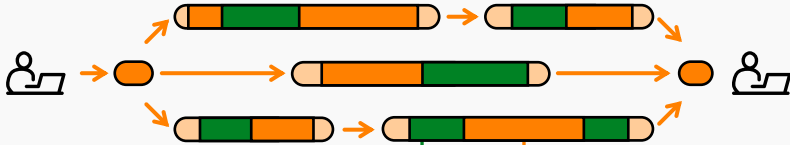
CPU continues to expand its role in AI/HPC systems alongside growing GPU adoption

- AI: Agentic AI drives greater CPU usage for orchestration and peripheral processing
- HPC: CPUs remain essential for high-throughput and high-precision workloads

## AI Workloads



### Multi-Agent collaborative execution

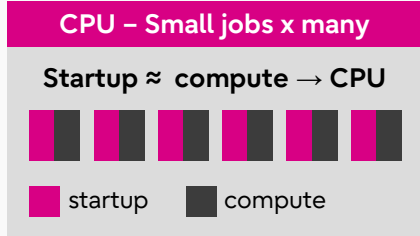


**AI workload** CPU GPU  
LLM, Speech, Video, Vision, etc.

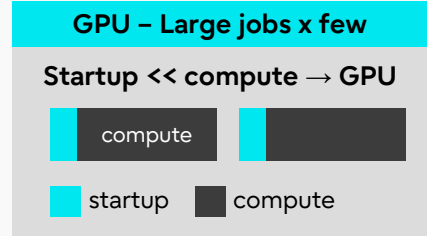
**General-purpose workloads CPU**  
Orchestration, RAG, DB, Code execution, etc.

## HPC Workloads

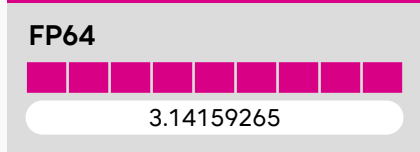
### Precision Small-Job Processing



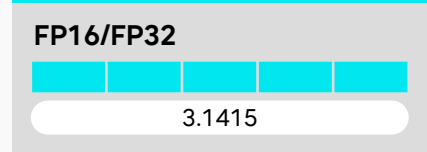
### Large-Scale Parallel Processing



### CPU - High precision

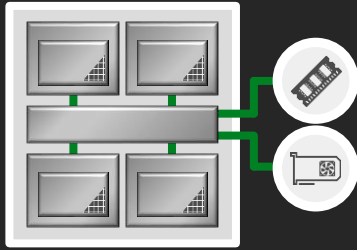


### GPU - Speed over precision



# FUJITSU-MONAKA Processor Overview

## FUJITSU-MONAKA



### Armv9-A Architecture



### 3D chiplet

- Core die 2nm
- SRAM die/IO die 5nm



### Ultra low voltage for energy-efficiency



### DDR5 12 channels



### Liquid / Air-cooling

subject to change without notice



### Arm SVE2-256bit for AI and HPC



### 144 cores x 2 sockets (288 cores per node)



### Confidential Compute Architecture for security



### PCI Express 6.0 (CXL3.0)

## FUJITSU-MONAKA

### High-Performance and Energy-Efficient CPU for a Carbon-Neutral Digital Society

#### High-Performance

Achieving high-speed computing centered on AI workloads (2x competitors CPUs).

#### Power-Efficiency

Reducing CO<sub>2</sub> emissions and electricity costs (2x competitor CPUs).

#### Safety & Security

Leveraging mainframe RAS technologies.

#### Ease of Use

Leveraging Armv9 Software ecosystem.



Selected PoC starts from 2026 Summer.  
Product shipments begin in 2027.

# FUJITSU-MONAKA Technology Features

Fujitsu's Made-in-Japan technology powers performance and efficiency

## High-Performance per Cost: Optimized Multi-Die Architecture

### Core Die: 2nm process

- Die integrating multiple compute cores

### SRAM+IO Die: 5nm process

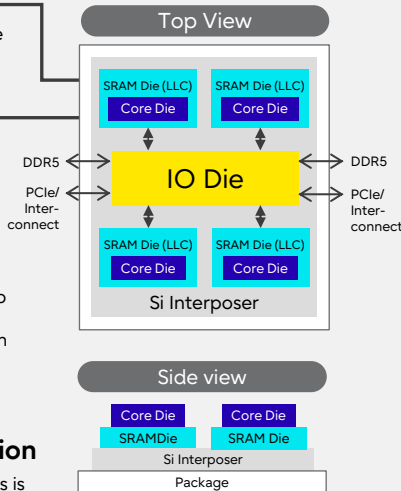
- Die integrating last-level cache (LLC)

### 3D stacking

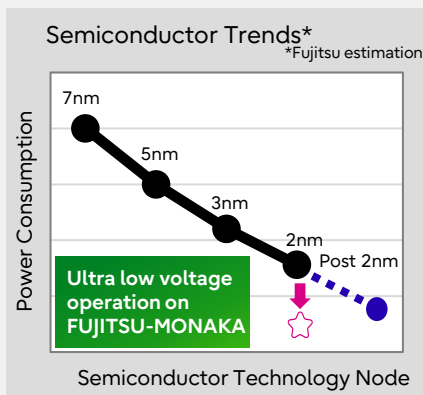
- Core die stacked on top of SRAM die
- Tightly coupled through TSV

### Heterogeneous process integration

- The costly 2 nm process is applied only to the core die
- Limiting the cutting-edge process area to less than 30%

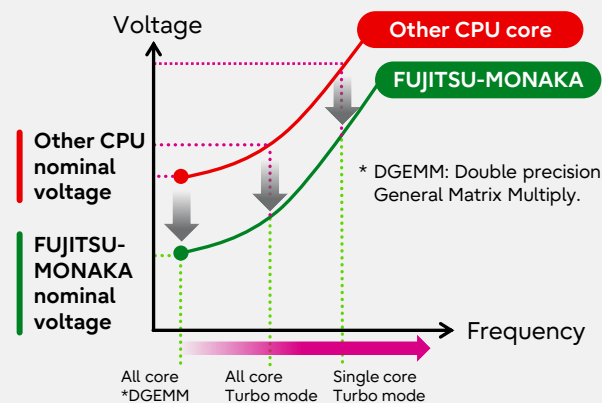


## For High-Performance per Watt: Ultra Low Voltage Technology



### Unique Ultra Low Voltage libraries for Core Die

- FUJITSU SRAM cells
  - FUJITSU custom logic cells
- Achieves lower voltage operation **reduce core operating voltage.**



### Advantage of Ultra Low Voltage design

- FUJITSU-MONAKA micro-architecture enables lower operating voltage for core at any operating frequency. (i.e. any workload)
- Since  $\text{power} \propto \text{voltage}^2$ , FUJITSU-MONAKA core can drastically reduce power consumption, realizing **superior power efficiency.**

# FUJITSU-MONAKA Servers

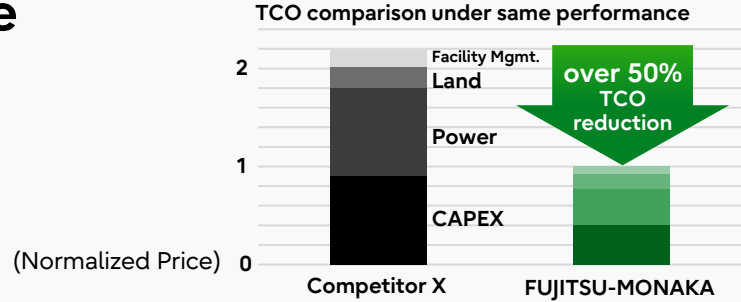
## System Architecture / Business Value



- **Over 50% TCO reduction** in data centers

Reduces Total Cost of Ownership (TCO) in data centers while providing necessary computing power.

※ Based on projected 2027 performance from SPECrate2017\_int



- **Liquid-cooling and air-cooling servers** for Diverse Data Center Environments

### High-performance, high-density liquid-cooling server

High Performance Computing

Core Data Center

#### Sustains peak performance

through DLC, removing thermal constraints

- 4x Dual CPU nodes / 2U chassis (19-inch)
- 500W, up to 90% Liquid-cooling coverage



### Flexible air-cooling server

Regional Data Center

Edge Computing

#### Fits within power/air-cooling limits

with ample PCIe slots and drive bays

- 1x Dual CPU nodes / 2U chassis (19-inch)
- 350W, Air-cooling



### Common Node Capabilities of FUJITSU-MONAKA Servers

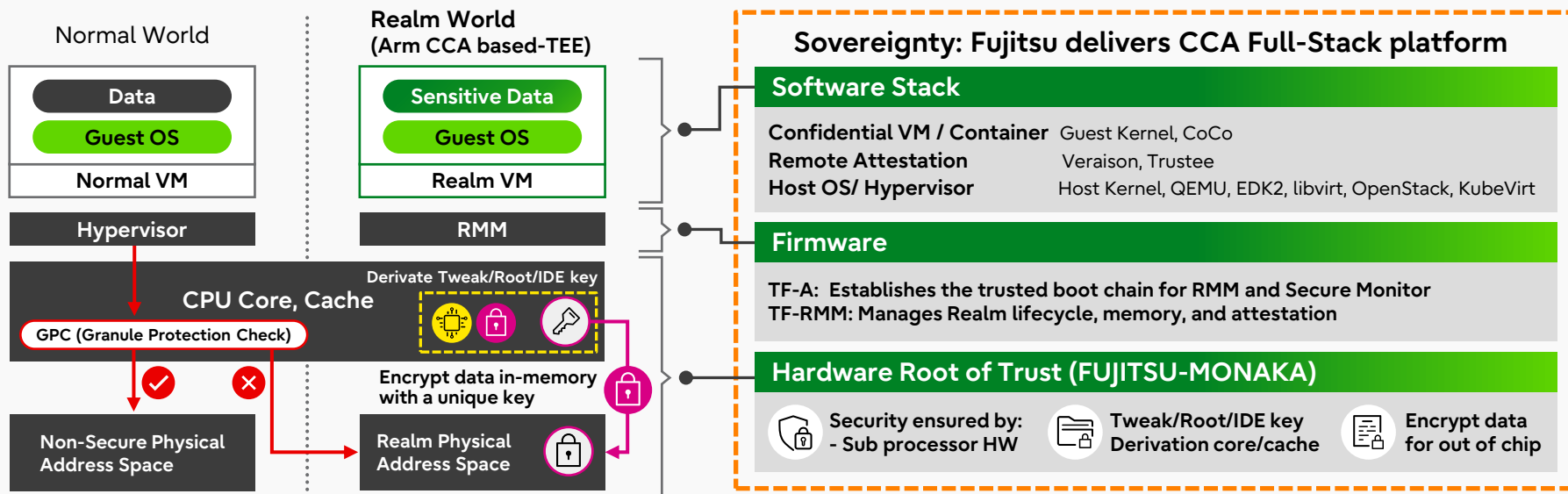
- High clock frequency (over 3.3 GHz)
- 24x DDR5 RDIMM slots, up to 6 TiB (21 GB per core) at up to 8800 MT/s

# Arm CCA with FUJITSU-MONAKA

\*CCA = Confidential Compute Architecture

## Fujitsu delivers a secure and sovereign CCA platform to meet Europe's requirements

- **Security:** Protects data-in-use within VMs, using Arm Confidential Compute Architecture (CCA)
- **Integrity:** Prevents environment spoofing through hardware-based remote attestation
- **Compatibility:** VM-based CCA technology enables seamless migration of existing environment



# Building the OSS Ecosystem for FUJITSU-MONAKA

## Advancing AI/HPC Software Readiness Through Collaboration with OSS Communities

- Delivering High-Performance Execution Environment for Diverse AI/HPC Workloads
- Ensuring Performance and Continuous Validation of Major OSS Applications and ISVs

Submitted 200+ PRs to 30+ OSS communities, doubling contribution in a year for FUJITSU-MONAKA

### AI Frameworks/Libraries

ML DL

OpenBLAS PyTorch oneDNN

Improved performance by **1.2-3x** through introducing SVE, Int8, thread throttling and others

LLM

vLLM llama.cpp OpenVINO

Improved inference performance by **2-3x** through supporting quantization and enhancing framework interoperability

DP

Milvus PostgreSQL

Improved similarity search by **1.3x** and encoding/decoding by **17x** through using SVE and OpenMP

### Application Software

OSS

Achieved **1.9-2.1x** performance over projected 2027 CPUs with OSS applications: OpenFOAM, FrontISTR, LAMMPS

ISVs

Expanding validation activities with major ISVs

### System Software

Linux PAPI/libpfm Collaboration with Arm, Red Hat, SUSE

**Review&Test: 90+**  
Working on community development for FUJITSU-MONAKA and Armv9.

GCC LLVM

**Submitted bug report: 350+**  
Working on performance improvements and Fortran quality enhancements.

# FUJITSU-MONAKA Accelerates Industrial AI

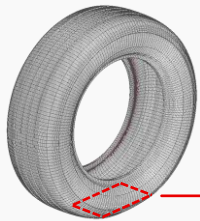
## Achieve cost-effective and high-accuracy CAE design in the real world

- Rapid, low-cost design evaluation with surrogate models
- Improve design accuracy in the early stages of product development

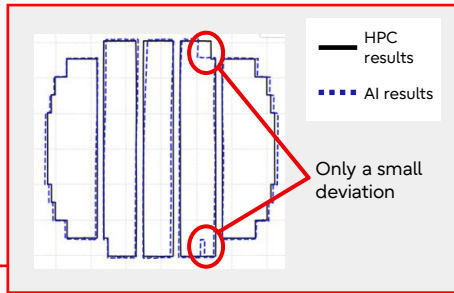
### AI model for structural analysis

- Achieved **9X** speed-up over HPC simulation
- Evaluation Accuracy: **87.7%**
- PoC in collaboration with DUNLOP

Reference



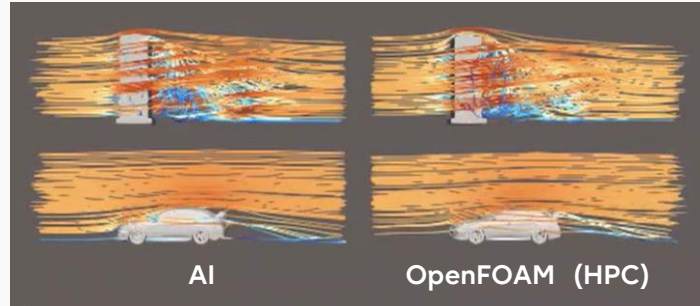
Tire Deformation



Deformation at the Tire Contact Patch

### AI model for fluid dynamics

- Achieved **5X** speed-up over HPC simulation
- Evaluation Accuracy: **91.1%**



Building

Car

AI

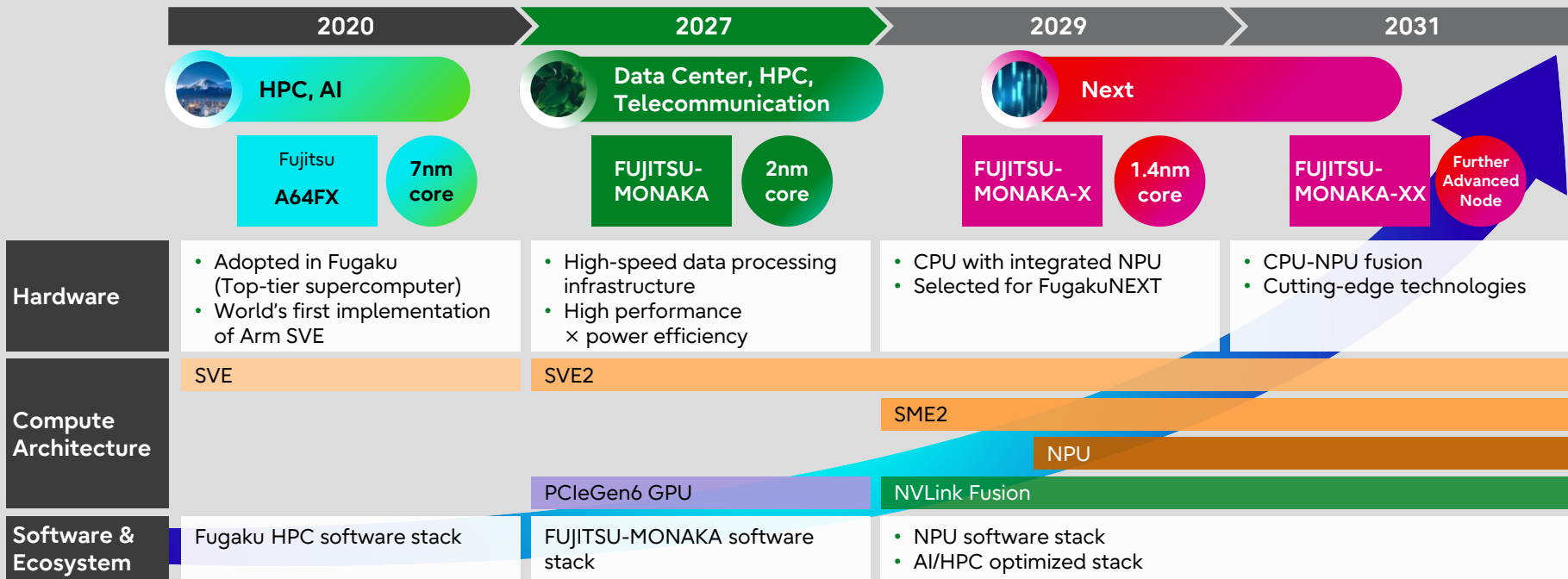
OpenFOAM (HPC)

# AI Evolution and FUJITSU-MONAKA Technology Roadmap



AI workloads evolve from HPC-centric to AI-native computing, and FUJITSU-MONAKA addresses this through multi-generation innovation

## Development Roadmap



**Thank you**

