FD-SOI Substrates for Edge Computing

Michael REIHA
FD-SOI Business Unit Manager

Exane BNP Paribas - Edge Computing Conference
December 15th, 2020

Disclaimer

Soitec megatrends

Semiconductor megatrends

- 5G
- AI
- EE*

*Energy Efficiency

Differentiated engineered substrates to serve our strategic end markets

Key figures in H1'21

- 254 M€ sales
- 30.4% EBITDA margin
- 102 M€ operating cash flow

Outline

1 What is Edge computing?

2 How is Edge computing utilized?

3 Why is FD-SOI seamless for Edge computing?

Outline

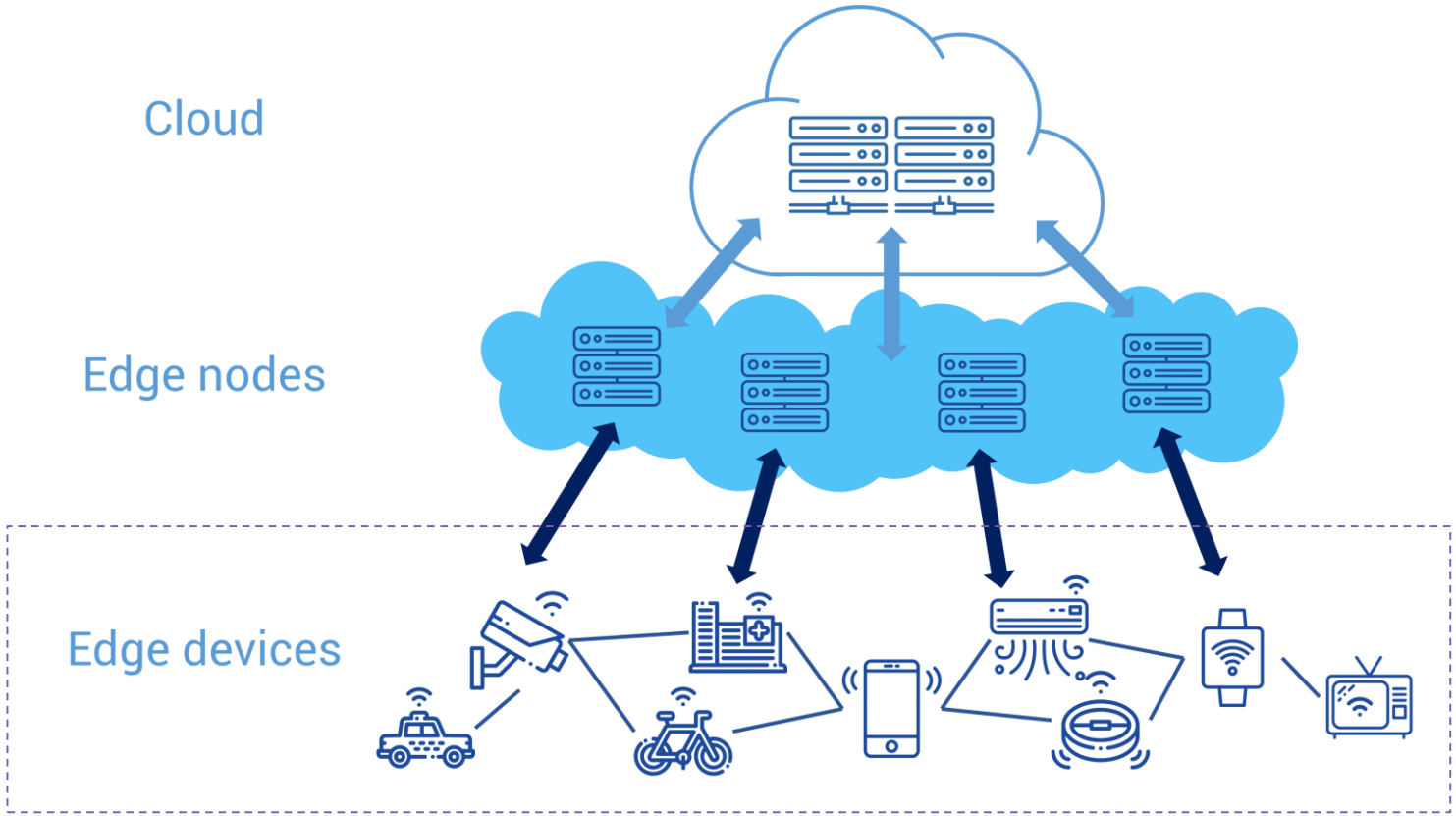
1 **What is Edge computing?**

2 How is Edge computing utilized?

3 Why is FD-SOI seamless for Edge computing?

Edge computing – Decentralized data processing and analysis

Edge computing integrates intelligence to edge devices
Data is processed and analyzed in real time near the sensor node



Source: www.alibabacloud.com

Edge computing – Intelligent analysis autonomous from the cloud

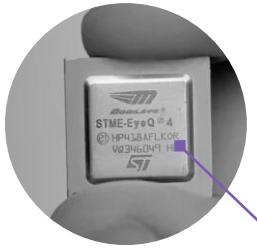
Real time analytics, increasing privacy/safety, providing
new value & experience

Sense

Things (devices) with various sensors

Think

Intelligence, analysis, management



- Data Processing
- Device Control
- Data Analysis

Connect

Wired & wireless connectivity

Act

Thing/user/cloud interaction

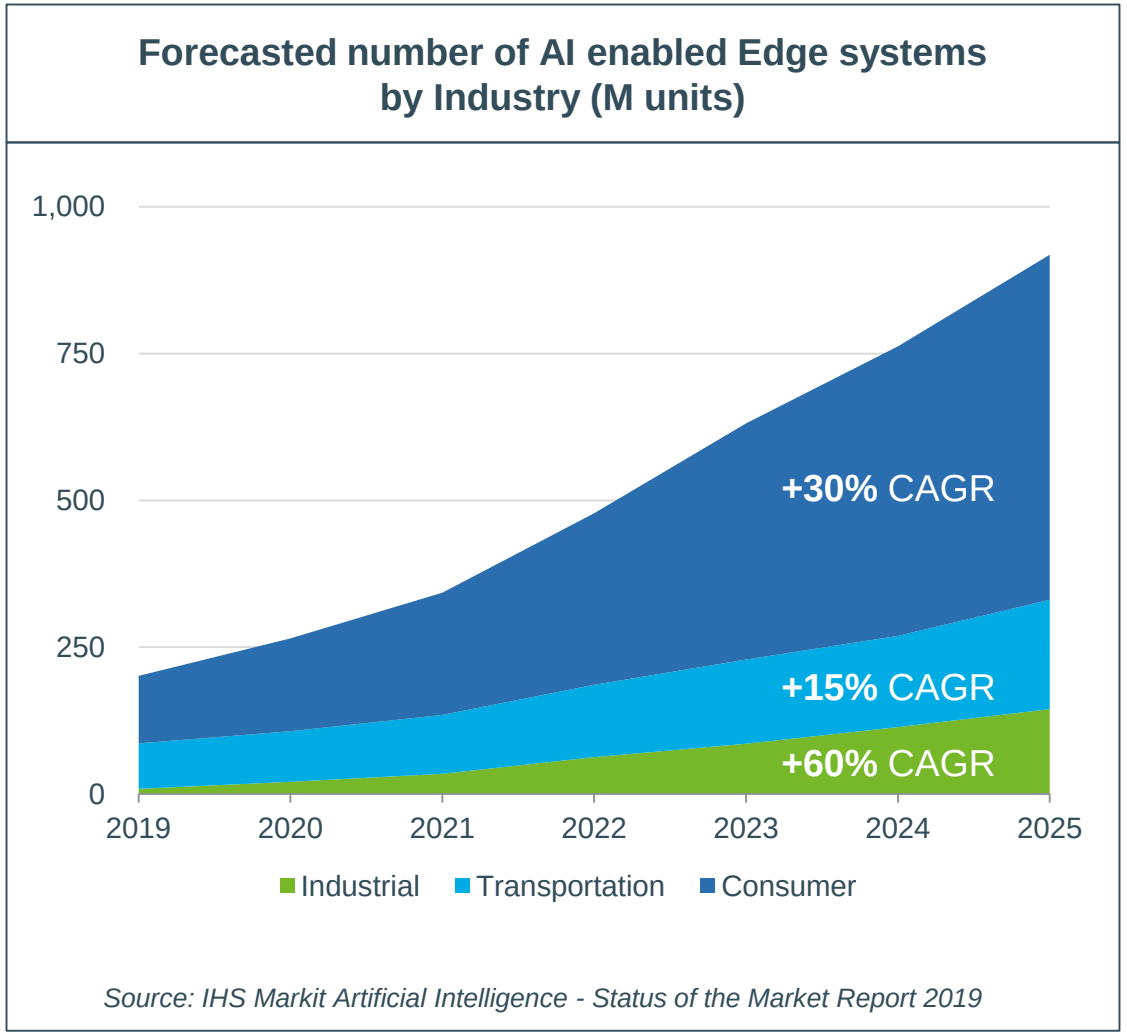
Edge computing – Evolution from cloud to on-device Edge computing

Before	*Now*	*Future*
Cloud	**Edge**	**On-device Edge**
AI training in the Cloud + Inference in the Cloud	AI training in the Cloud + Inference at the Edge	AI training at the Edge + Inference at the Edge
		

Edge computing – Chip features and applications

	Low Complexity	Medium Complexity	High Complexity
Chip Features	Dynamic Always-ON / Ubiquitous	Energy Efficient Computing	Highly Reliable / Predictive
Examples of Applications	Smart City / Smart Home (Sensors)	Smart Devices / Wearables	Smart Vehicles / Smart Machines

Edge computing – Towards a trillion of connected ‘things’



Source: IHS Markit Artificial Intelligence - Status of the Market Report 2019

Consumer



Transportation


Industrial



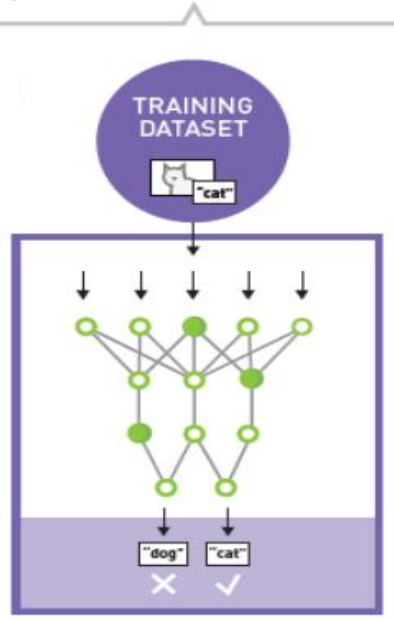
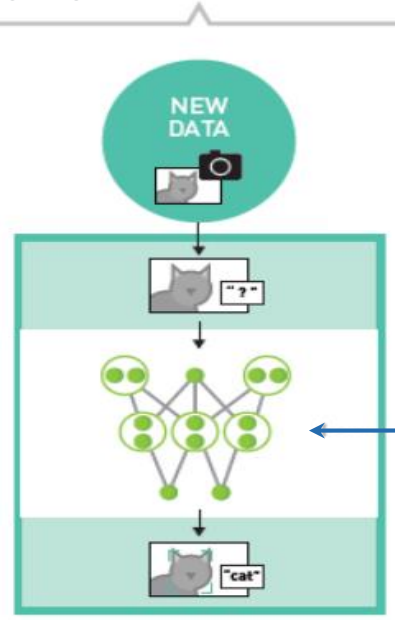
Outline

1 What is Edge computing?

2 How is Edge computing utilized?

3 Why is FD-SOI seamless for Edge computing?

Edge computing – Training vs. Inference

	Training **Creating AI** model from existing data	**Inference** **Applying AI** model to new data
		 *New data into trained model*
Evolution	Cloud → Cloud and Edge	Cloud and Edge → On-device Edge
Limitations	Power ceiling due to thermal inefficiency	Energy limited due to battery capacity
Requirements	High-Performance (TOPS)	Energy efficiency (mJ / frame), Competitive cost
Main Architectures	CPU, GPU, TPU, FPGA	Low Power FPGA, NPU, MCU
Technologies	FinFET	FD-SOI, 2.5D-3D packaging

Source: Nvidia, Soitec

Edge computing – Requires new paradigm for efficient design

Throughput vs Latency
Inferences per second, frame rate,
sub-ms delay

Hardware cost of ownership
On-chip storage
Number of processing elements
Chip area
Process technology

Robustness
Soft-error rate
(20x better than bulk technologies)
Environmental factors (up to 125°C)

Energy and Power
Battery lifetime (+10 years),
Energy/operation, Memory bandwidth

Accuracy vs Efficiency
Range of deep neural networks models
Interconnect design
Architecture selection

Connectivity and Integration
Seamless integration of Radio and DSP
Embedded compute IP

Source: Soitec, industry data

Outline

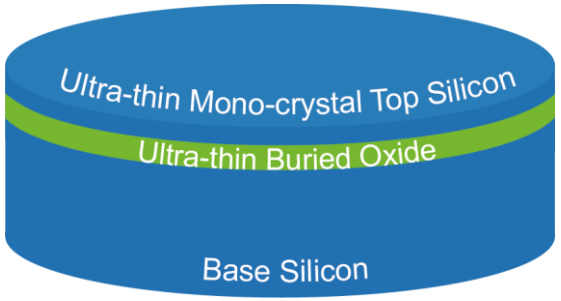
1 What is Edge computing?

2 How is Edge computing utilized?

3 Why is FD-SOI seamless for Edge computing?

FD-SOI substrate structure and manufacturing sites

FD-SOI substrate structure



Ultra-thin top silicon & box enabling fully-depleted transistor operation

300 mm high volume manufacturing in France and in Singapore

Bernin 2, France

Awarded "Factory of the year 2020" in France by L'Usine Nouvelle, thanks to Industry 4.0 initiatives

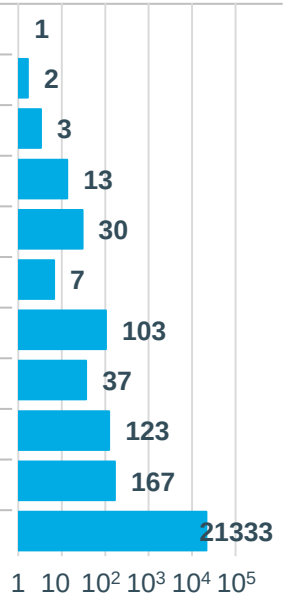
Pasir Ris, Singapore

Edge computing – Why FD-SOI?

FD-SOI is a power-efficient & flexible mixed-signal platform which can enable analog/RF integration for edge computing applications

Cost of Edge acceleration

Operation	Energy (pJ)	Relative Energy Cost
8b Add	0.03	1
16b Add	0.05	2
32b Add	0.1	3
16b FP Add	0.4	13
32b FP Add	0.9	30
8b Multiply	0.2	7
32b Multiply	3.1	103
16b FP Multiply	1.1	37
32b FP Multiply	3.7	123
32b SRAM Read (8kB)	5	167
32b DRAM Read	640	21333



Source: Horowitz, ISSCC 2014

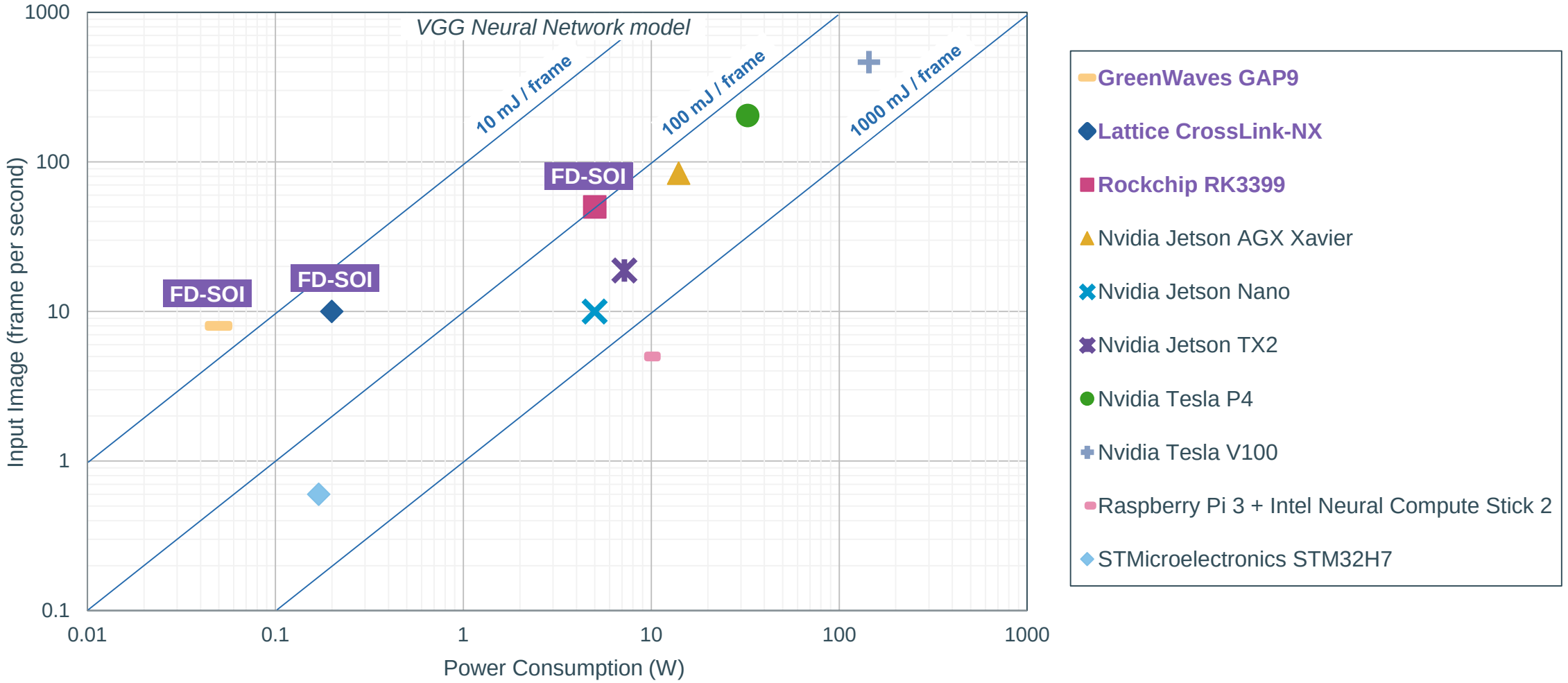
Edge computing requirements

- Limit data size (8 bit)
- Reduce read/write and Energy/MAC
- On-chip memory
- Energy-efficient architecture
- Reduce number of convolutions

FD-SOI value proposition

- High speed devices for analog compute
- Low power devices for efficient data conversion
- Low energy eNVM for on/near memory compute
- Lowest power connectivity (BLE, NB-IoT, WiFi)
- Design-to-cost technology

Edge computing – FD is the ideal platform for edge inference

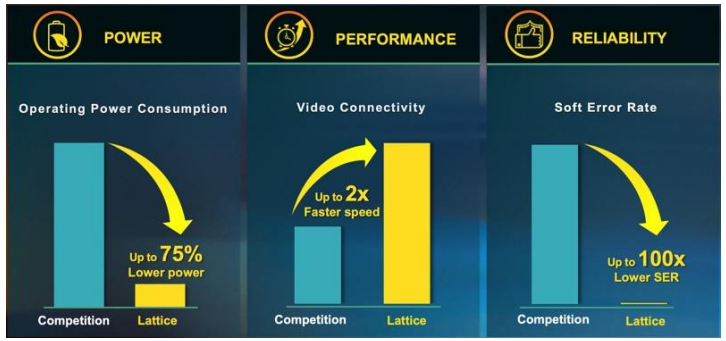


Source: Soitec, industry data

Edge computing – Edge-based Integrated Circuits using FD-SOI

Low Power FPGA

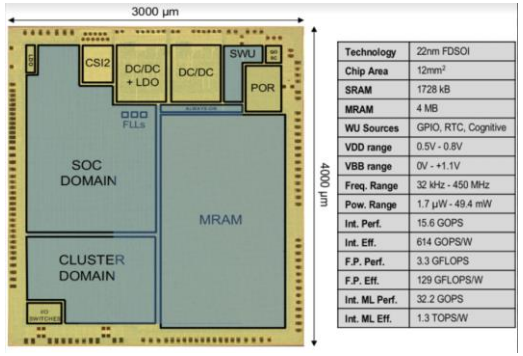
› **CrossLink-NX™** built on the **28FDS** Lattice Nexus platform for Vision Processing Applications



Source: Lattice Semiconductor

IoT Application Processor

› **GAP9 IoT,** *state-of-the-art* Application Processor in **22FDX** for the Next Wave of Intelligence at the Very Edge

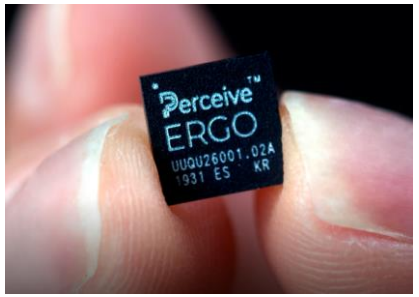


Technology	22nm FDSOI
Chip Area	12mm²
SRAM	1728 kB
MRAM	4 MB
WU Sources	GPIO, RTC, Cognitive
VDD range	0.5V - 0.8V
VBB range	0V - +1.1V
Freq. Range	32 kHz - 450 MHz
Pow. Range	1.7 µW - 49.4 mW
Int. Perf.	15.6 GOPS
Int. Eff.	614 GOPS/W
F.P. Perf.	3.3 GFLOPS
F.P. Eff.	129 GFLOPS/W
Int. ML Perf.	32.2 GOPS
Int. ML Eff.	1.3 TOPS/W

Source: GreenWaves Technologies

Edge Inference Processor

› **Ergo** delivers **4+ TOPS** sustained and **55 TOPS/W,** capable of processing large neural networks in **20mW,** in **22FDX**

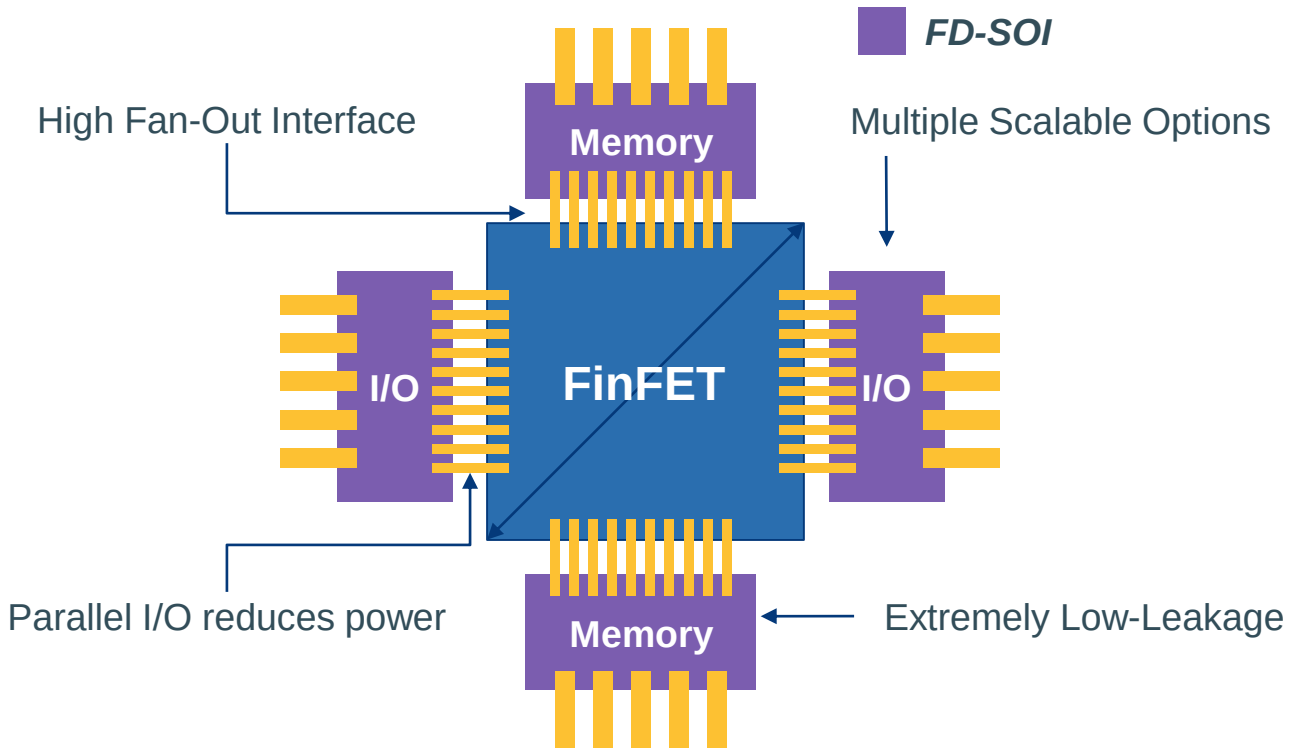


Source: Perceive

Future FD-SOI opportunities for Edge computing applications

FinFET delivers performance

FD-SOI delivers thermal & power efficiency



Heterogeneous packaging offers

- Improved thermal efficiency
- Scalable architectures via a chiplet-based design
- Enhanced cost utilization with FinFET
- Lower power budget potentials
- Architecture renewal for bandwidth vs energy tradeoff
- More competitive cost structure
- Better production yield

Source: Soitec

A growing number of FD-SOI applications for Edge computing

Consumer

Facial recognition

Wearables

Voice recognition processor

Transportation

Vision processing for autonomous drones

Vision processing for ADAS

MCUs for automotive

Industrial

Smart sensors for agriculture

Smart meters

Industrial robots

Summary – FD-SOI for Edge computing

1 **Efficiency** Edge computing reduces network complexity, planning but requires lowered energy per frame

2 **FD-SOI** Natural platform for Edge: *Energy & Cost Efficiencies, Robustness vs. Environmental factors*

3 **Scaling** Moore's law improves gate-density, peak performance but not end-to-end Edge architectures

4 **Challenges** Consistent Edge *experience* requires predictable performance – *and reliable technology*

5 **Strategy** Soitec advancing FD-SOI to lower power potentials and planar nodes, *extending the Edge*

Edge computing – Glossary

ADAS	Advanced Driving Assistance Systems
ASIC	Application Specific Integrated Circuit
Body-bias	Body bias is a technique used to dynamically adjust the threshold voltage of a CMOS transistor
CPU	Central Processing Unit
eMRAM	embedded Magnetic Random Access Memory
eNVM	embedded Non-Volatile Memory
FD-SOI	Fully Depleted Silicon on Insulator
FPGA	Field-Programmable Gate Array
GPU	Graphics Processing Unit
Inference	Applying a deep learning model to make predictions on a new data
MAC	Multiplier-accumulator
MCU	Microcontroller Unit
mmW	millimeter Wave
MRAM	Magnetic Random Access Memory
PCM	Phase Change Memory
PVT	Process-Voltage-Temperature
TOPS	Trillions or Tera Operations per Second
Training	Creating a deep learning model from a dataset
TPU	Tensor Processing Units
VGG	VGG is a convolutional neural network model proposed by K. Simonyan and A. Zisserman

Thank you

Follow us on:

Soitec
@Soitec_FR / @Soitec_EN
Soitec

www.soitec.com