Elevating AI Server Performance: Precision Timing Solutions for Enhanced Efficiency and User Experience

提升AI伺服器效能：精準定時解決方案，增進效率與用戶體驗

SMART | CONNECTED | SECURE

Peter Chiou

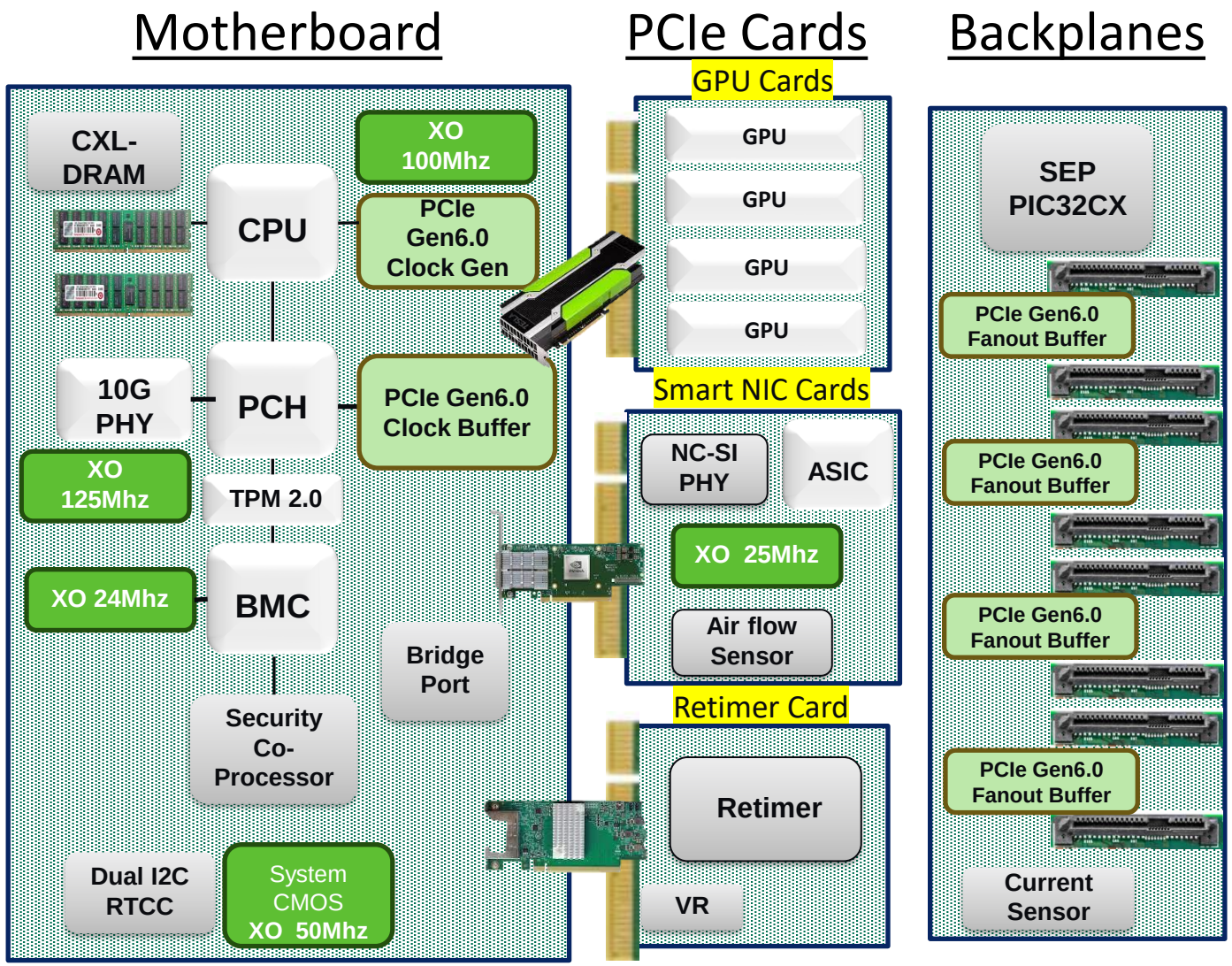
May 2024

AI vs. General Servers and Timing Challenge

Features	AI Servers	General Servers
Primary Use	Running machine learning and deep learning workloads	Data storage, running applications, providing network services
Design Aim	High computational capabilities, fast data processing speeds	Efficiency, high stability, reliable data storage
Computational Units	High number (for parallel processing, rapid matrix multiplications)	Fewer (operate at slower speeds)
Data Transmission	Efficient, low latency	Strong capabilities, but slower speeds
Applications	AR/VR, autonomous driving (where low latency is crucial)	General purpose

- **High-speed interfaces:**
  - > 25 Gbps I/O for communication between processors and memory.
  - Require extremely precise timing references with minimal jitter
- **Power efficiency:**
  - Timing circuits need to deliver precision while minimizing power
- **Scalability:**
  - Timing solutions need to be scalable for future growth in data rates and core count
- **Synchronization:**
  - Timing solution to support multiple processors or accelerators working in parallel
- **Integration with emerging technologies:**
  - To support multiple FPGAs for AI workloads.

Timing Component Highlight for AI server



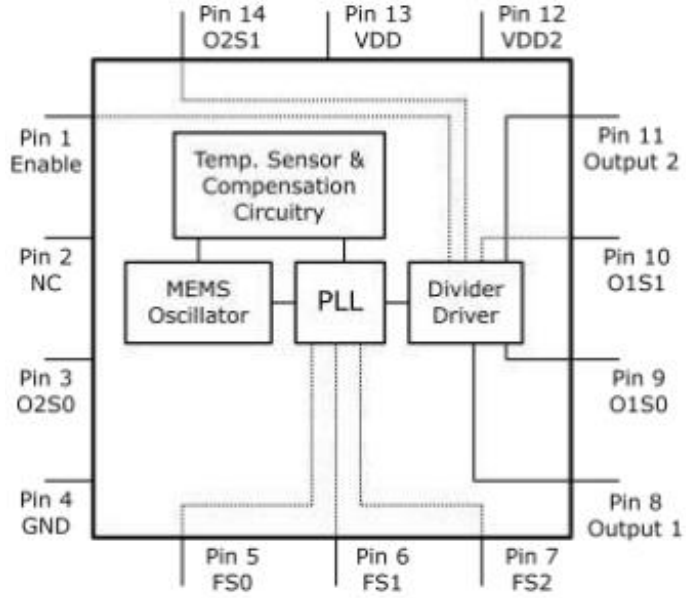
- **Timing Solutions for AI Server rely on**
  - High Precision & Efficiency
  - High Reliability & Availability
  - Very low Jitter
  - Low Skew

MICROCHIP

MEMS Timing In AI Revolution

- **AI & Precision Timing Synergy:**
  - Precision timing is foundational for AI-driven innovations, enhancing everything from cloud data centers to autonomous vehicles.
- **MEMS Over Quartz :**
  - Silicon MEMS technology, offering compactness and precision, is replacing quartz in timekeeping, crucial for AI's rapid computations.
- **Critical Applications :**
  - MEMS-based timing is vital in harsh environments and is used in sectors like automotive, aerospace, and telecommunications.
- **Future of AI Integration:**
  - Precision timing will be key in AI advancements, impacting various industries and enabling efficient, reliable electronic functions

MEMS

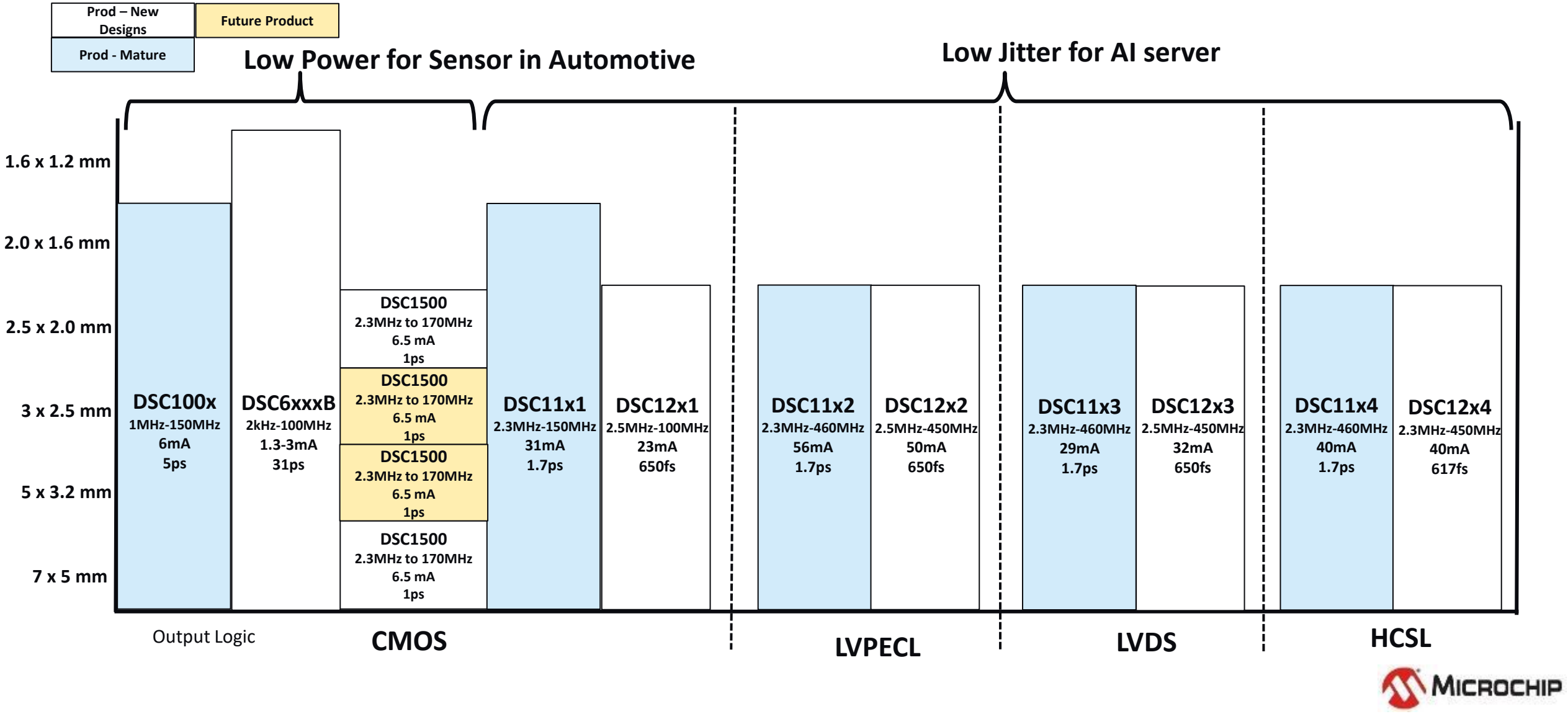


- Leverages Silicon processes
- Miniaturization
- Configurable
- More robust

WWW→IOT, IPv4→IPv6, AI→?

MICROCHIP

MCHP Single Output MEMS Oscillators



Microchip TSS: Total System Solution

Product Validation:

- Designs the Functional Validation Board (FVB)

System Applications:

- Reuses FVB as Customer Evaluation Board

Reference Design:

- Sanitized version of the FVB schematics

MICROCHIP

Data Center

Microchip and 3rd Party Solutions

MICROCHIP

Investment in Data Center - PCIe 6.0

Microchip Device P/N	Microchip Device Type	Description
ZL40294	LPHCSL Buffer (Intel DB2000)	20 LPHCSL outputs
ZL30291	PCIe Clock Generator (Intel CK440)	19 LPHCSL CLK Gen
SY756xx/SYA756xx	Buffer, I2C, individual OE 85Ω or100Ω Termination SYA - AEC-Q100	2,4,8,12 output LPHCSL
ZL30282/ZL3026x	PCIe Clock Generator With SSC	Up to 10 output CLK Gen HCSL, LVDS, PECL, CMOS
ZL30281	PCIe Clock Generator	3 output PCIe clock gen
DSC1200/DSA1200	Oscillator DSA - AEC-Q100	HCSL, LVDS, PECL, CMOS
DSC50x/DSA50x In Development	MEMS PCIe CLK GEN With SSC DSA - AEC-Q100	6 differential 12 CMOS LPHCSL, LVDS, PECL
New Projects in Review	DB1206 12-output PCIe Buffer	DB1206 listed
	Low Voltage buffer family LPHCSL 1.2 – 1.8V	Pin-to-Pin with TI

AI Server / Accelerator

Microchip and 3rd Party Solutions

Quick guidelines for selecting Microchip MEMS

Low Power (1.3mA)	DSC6xxx
Smallest (1.6x1.2mm)	DSC61xx or DSC60xx
Smallest multi-output (3 output 1.6x1.2mm)	DSC613
Spread Spectrum	DSC63xx
Best Stability (+/-10ppm)	DSC10xx or DSC11xx
Best Jitter (600fs)	DSC12xx
PCIE (Gen 1/2/3/4)	DSC11xx or DSC557xx (multi-output)
PCIE (Gen 5/6)	DSC12xx or DSC500xx (multi-output)
Best Jitter / Power combination	DSC15xx

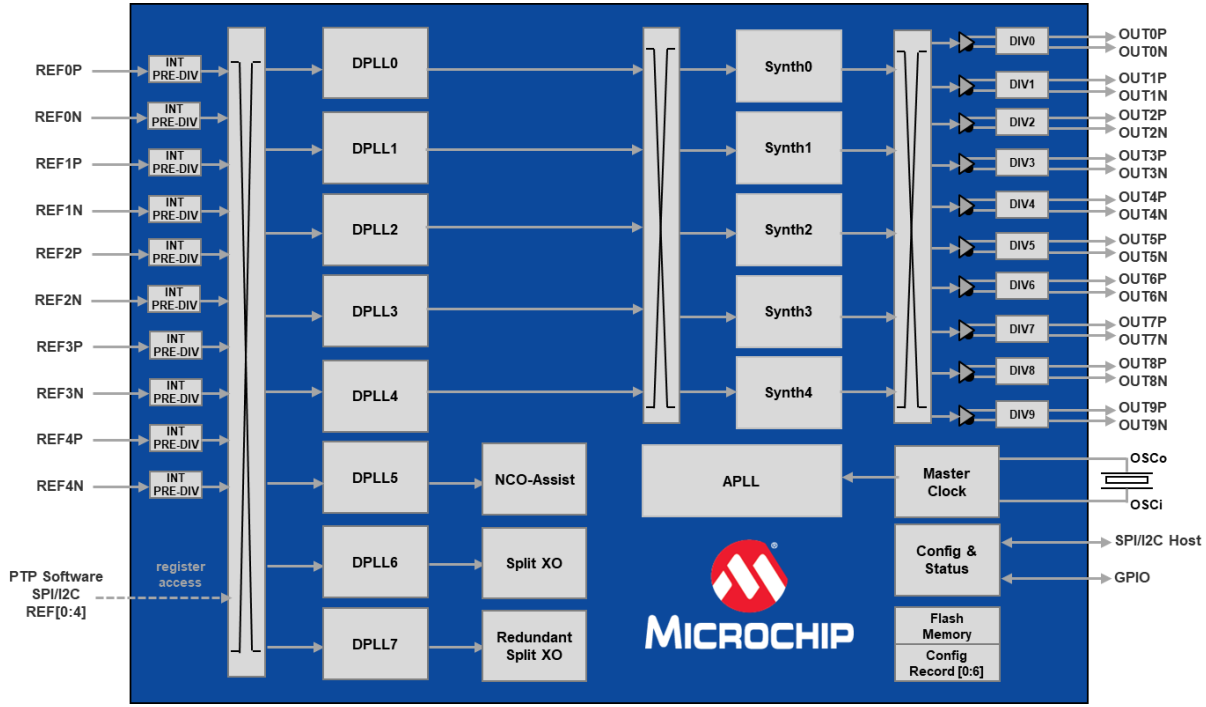
5G

Microchip and 3rd Party Solutions

MICROCHIP

Azurite Family

- Multi-channel frac_N synthesizers
- 100fs$_{RMS}$ typical jitter performance
- 100ps O-O alignment
- 100ps I-O delay variation
- Fast Lock to PPS
- Enhanced chip-to-chip interface (A-Series)
  - miToDBasic™
  - miToDSync™

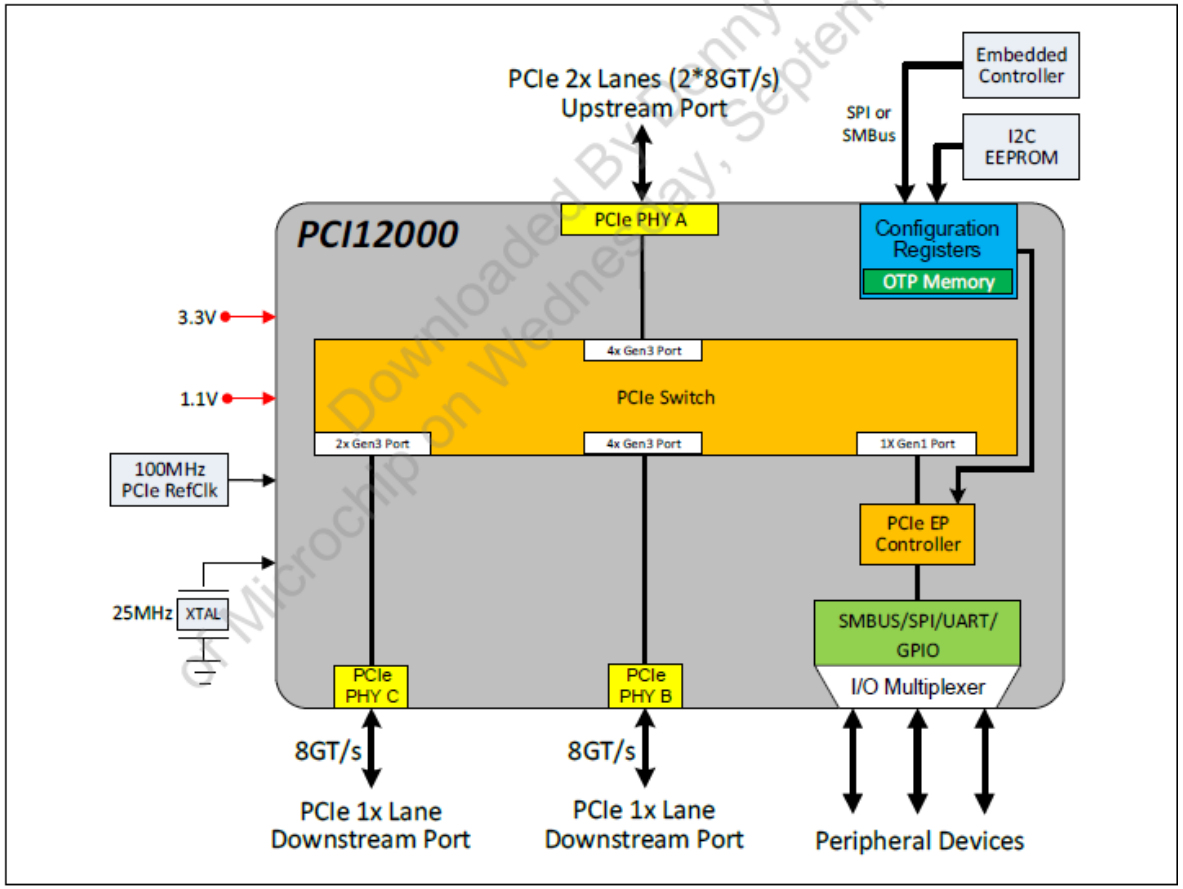


Part #	Application	# of DPLLs with NCO	DPLL BW, Hz	# Phy Inputs	Max # of Outputs	# Low-Jitter APLLs	Max Jitter, fs RMS	Input Freq. Range	Output Freq. Range	Pkg. Size, mm
ZL3064xy	Line Card/JA	x=1,2,3,4,5	14-470	5D/10S	10D/20S	5	150	1-1.25G	0.5-850M	9x9, 7x7
ZL3063xy	SyncE	x=1,2,3,4,5	0.1m-470	5D/10S	10D/20S	5	150	1-1.25G	0.5-850M	9x9, 7x7
ZL80032	SyncE	2	0.1m-470	5D/10S	10D/20S	5	150	1-1.25G	0.5-850M	9x9
ZL3073xy	IEEE1588	x=1,2,3,4,5	0.1m-470	5D/10S	10D/20S	5	150	1-1.25G	0.5-850M	9x9, 7x7
ZL80732	IEE1588	2	0.1m-470	5D/10S	10D/20S	5	150	1-1.25G	0.5-850M	9x9

ADAS / Autonomous Driving

Qualcomm Snapdragon Ride ADAS/AD

QDRIVE 3.0 is Qualcomm's scalable autonomous driving platform based on the Snapdragon Ride 8195 processor. Microchip components are validated with Snapdragon Ride Platform making it very easy for OEM and Tier-I customers to adopt in their design. Microchip's Gen 4 PCIe Switch and its development tools/APIs have been tested and validated with the platform, which helps in rapid prototyping while adopting the Snapdragon Ride solution



- **Target Market Segments**
  - Automotive
- **Target Application**
  - ADAS, AD, Infotainment with Snapdragon Ride 8195
  - Targeting OEMs, Tier-1s and Design Houses
- **Value Proposition**
  - Scalable, high performance ADAS/AD platform
  - Lower power than traditional platforms.
  - MCHP Gen 4 PCIe switch pre-optimized with QCOM's
  - Snapdragon Ride Platform

TCG Solution	Automotive PCIe Gen4/5
2 or 4 output Buffer	ZL40262/264
OX	VXM7-25Mhz
Clock Gen	DSA12xx

Most open, scalable, and customizable ADAS/AD solution

Our DSA 12xx – Smallest Differential Oscillator

Applications:

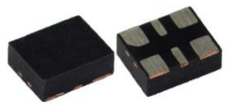
- Vital for generative AI applications like ChatGPT
- Key in 5G/6G communications, aerospace, defense, and AR/VR.
- Synchronizes deep learning processes for accurate outputs

Challenges to address:

- Integration into dense modules
- Functionality under harsh conditions
- Solving unique density problems

- Meets PCIe Gen1/2/3/4/5/6 clock jitter spec
- Any frequency between 2.3MHz to 450MHz
- LVCMOS/LVPECL/LVDS/HCSL output formats
- 2.5x2.0mm 6-L package available , 2.5 to 3.3V VDD
- ±20ppm/ ±25ppm / ±50 ppm stability
- Up to -40°C to +125 °C temperature range for LVCMOS and LVDS output
- Dual frequencies through Frequency Select input
- AEC-Q100 Automotive grade product available

Most Compelling Features	Benefits
~0.65ps integrated RMS phase jitter (12k-20MHz)	Excellent jitter margin for high performance networking applications
Smallest deferential oscillator: 2.5x2.0mm	Save board space, ideal SERDES clock for SiP module design
Two pre-programmed frequencies selected by FS input	Improve design flexibility, BOM consolidation.



MICROCHIP

One More Thing : AIHO

• Why Smart Oscillators

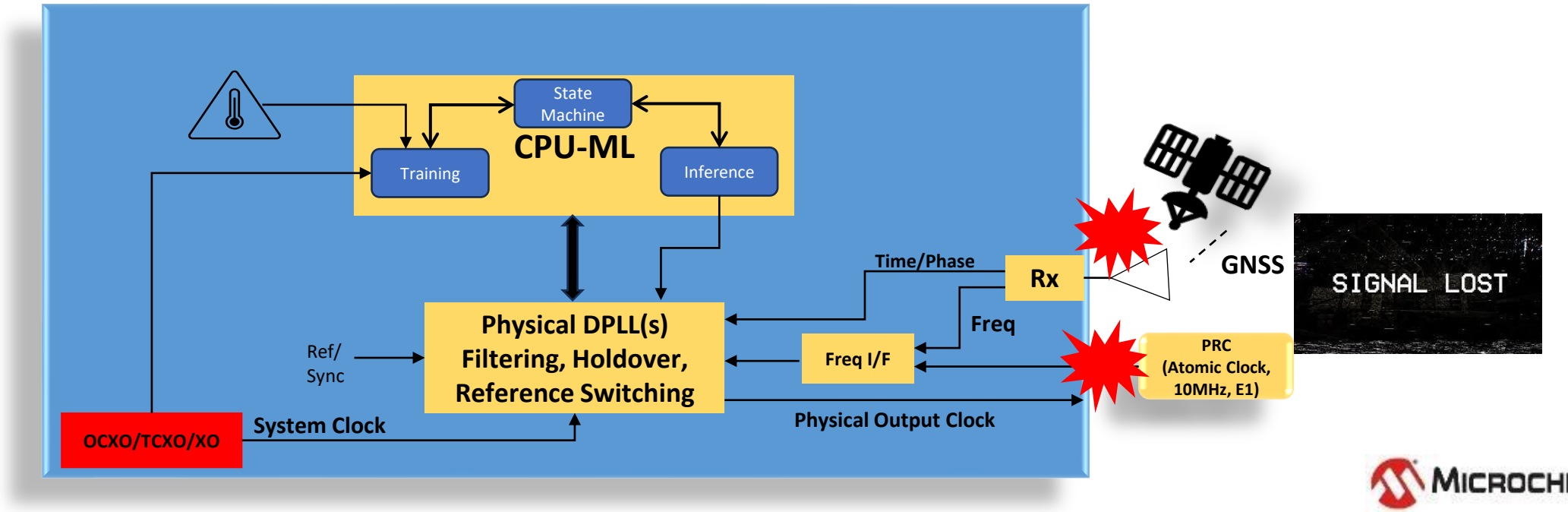
Frequency drift of an oscillator output clock is impacted by:

- **Temperature**
- **Pressure**
- **Magnetism**
- **Humidity**

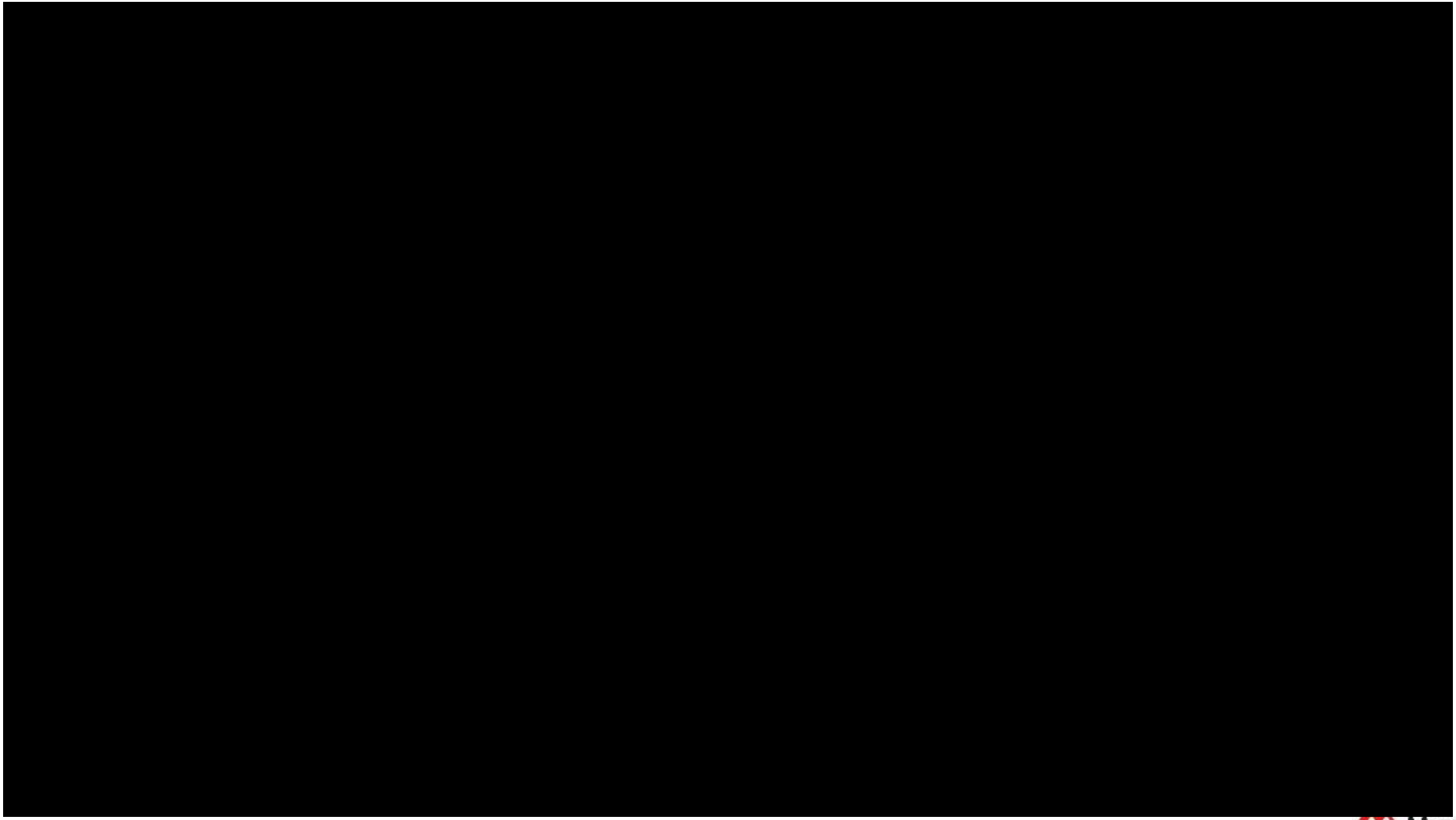
These environmental variables are used by the ML solution to learn the oscillator behavior then compensate and remove frequency wander during reference loss (holdover) periods.

• Applications

- **Service providers require reliable holdover accuracy during loss of GM or GNSS failures. These requirements can reach a 12ppt holdover accuracy, which corresponds to a maximum phase drift of 1 microsecond per 24 hours.**
- **This type of requirements are associated with very expensive oscillators. Relatively cheaper oscillators can be used with a ML-engine to provide the same level of accuracy.**



MICROCHIP



ML Reliability in our Smart OX

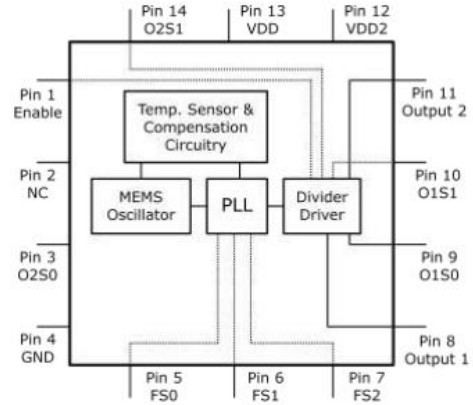
- **The Specifications:**

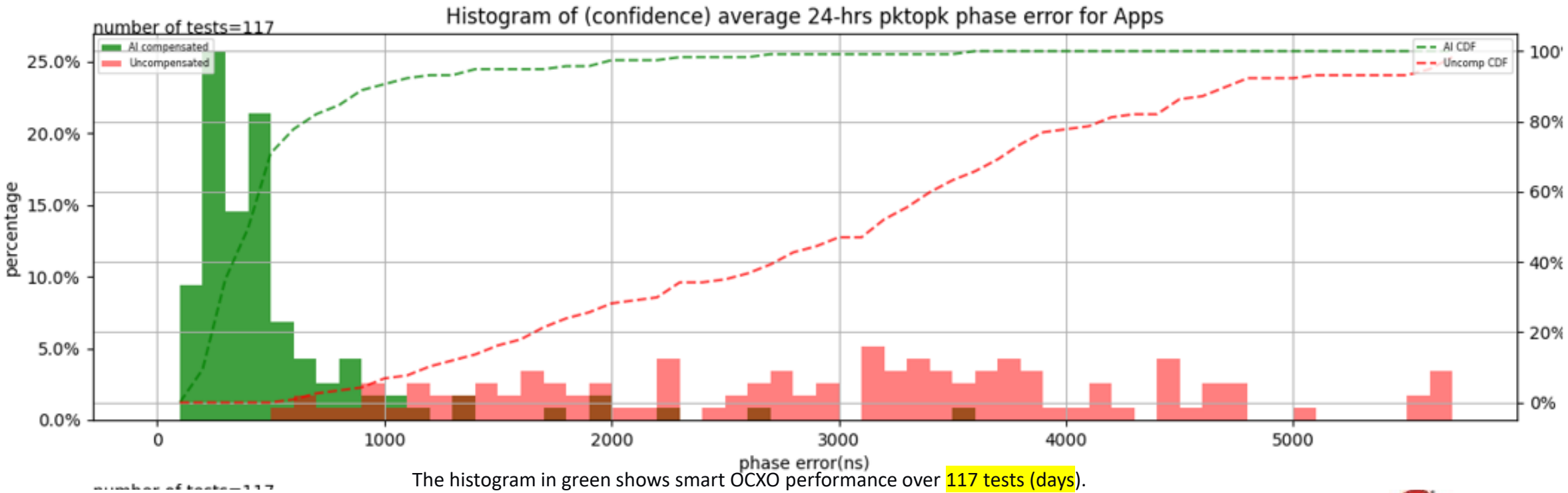
The ML-solution is programmable such that Training is performed in the background with controlled captured feature length (hours) and training time.

Parallel multi-algorithm engine that choses best performance solution for current environment conditions.

- **Reliability**:

Long term analysis of the solution over thousands of hours and a bank of OCXOs showed that the developed smart-OCXOs drift less than 1 microseconds per 24-hours 90% of the time.





The histogram in green shows smart OCXO performance over 117 tests (days).

MICROCHIP

Q&A