



Electronics for the Future

设备端学习AI芯片 介绍资料

2022年9月27日
ROHM Co., Ltd.
Marketing Communication Department

人工智能 (Artificial Intelligence)

再现人类部分功能的
图像识别等技术

机器学习 (Machine Learning)

AI机械式学习的方法

深度学习：
从一系列连续的表示层中进行学习

神经网络 (Neural Network)

机器学习之一

AI的训练与推理

如果是猫和狗的图像识别AI，那么：

训练 . . . AI查看很多图像，掌握狗和猫的特征

推理 . . . AI看见图像时能判定它是猫还是狗

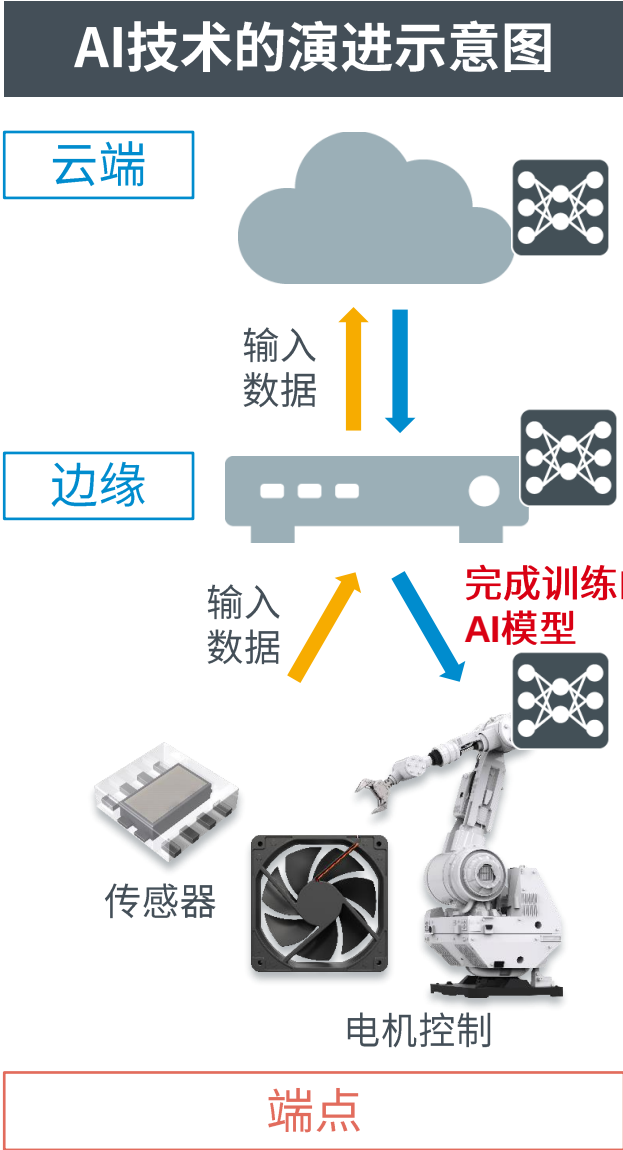
学习需要算力！



AI技术正在从云端发展到边缘再扩展到端点



	以往的云端AI	边缘AI	端点AI
AI 功能	训练和推理都在云端进行	训练在云端进行 推理在边缘进行	训练在云端进行 推理在端点进行
要求性能	• 出色的学习能力 • 高度安全性	• 网络负载降低 • 响应时间短 • 低功耗	• 网络负载为零 • 响应时间非常短 • 超低功耗
课题	• 通信成本和功耗增加 • 响应时间有波动 • 需要安全成本	• 在边缘需要高性能的FPGA和GPU • 问题较少 但响应时间有波动	• 受限于基于嵌入式MCU性能的AI模型

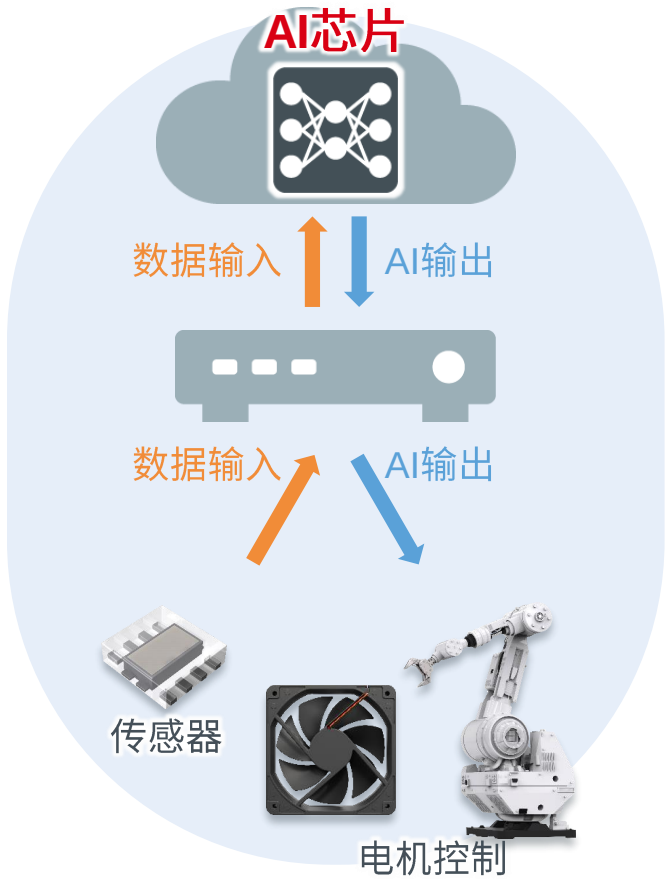


云端
计算设备

边缘
计算设备

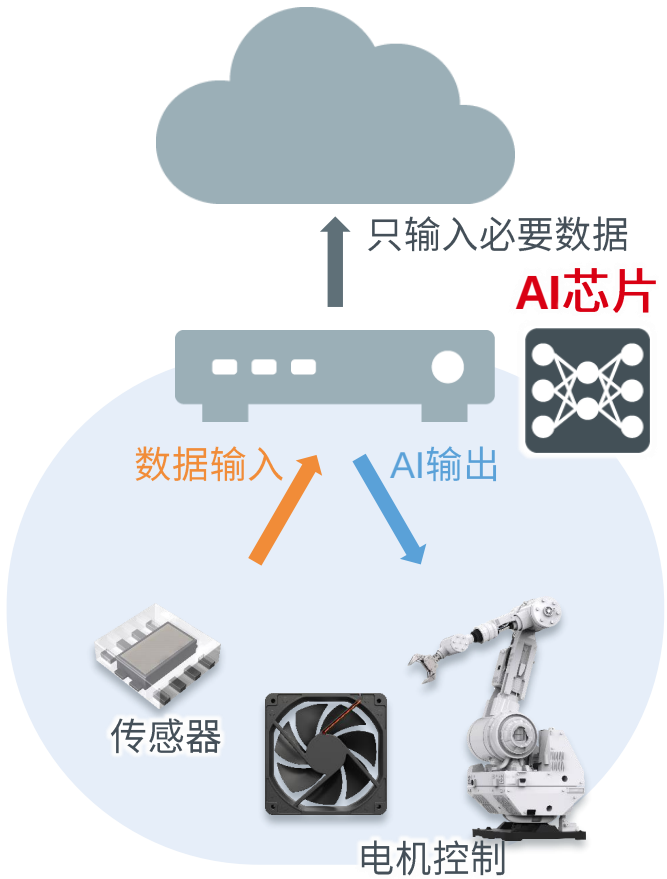
端点

云端AI系统



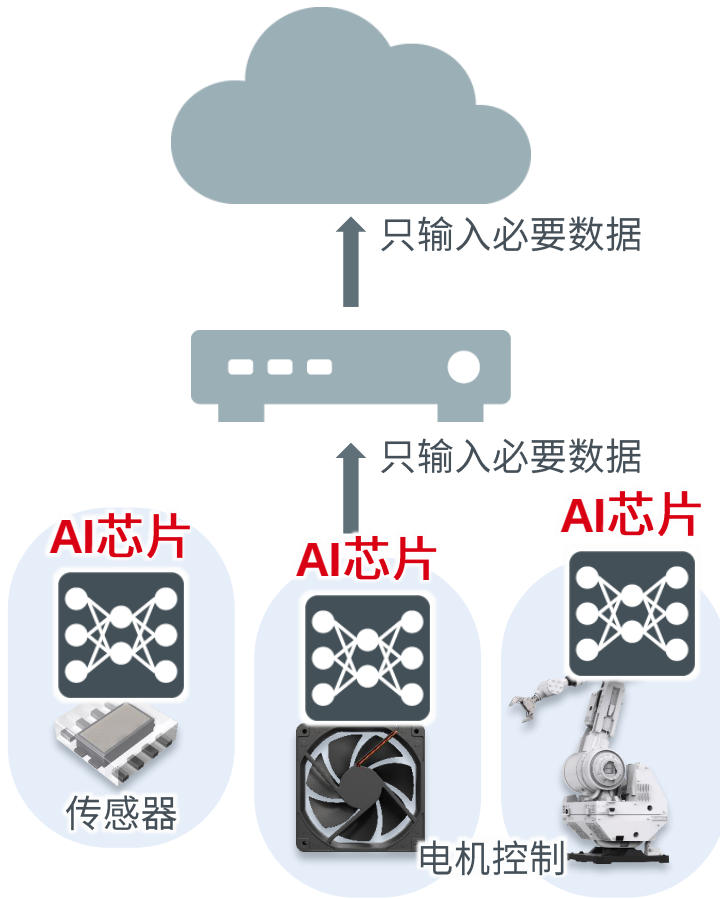
负载集中于
云计算设备的AI芯片

边缘AI系统



训练和推理的负载可分散给
边缘计算设备的AI芯片

端点AI系统



负载可分散于
端点的AI芯片

此次开发产品的主要目标应用

图像识别



要实现复杂的AI，
需要高性能的GPU和FPGA

各种设备的故障预测



相对较小的AI已足够，
比起高精度，更注重大小和成本

预测容易先出机械故障的电机故障

课题

- 需要逐个设备进行学习
- 当环境发生变化时需要重新学习
- 需要逐一设计IC产品

高效化

使用ROHM**自有技术(=设备端学习)**一举解决了这些课题

ROHM开发出省电、高速响应、超小型、
可实时学习的AI新解决方案

设备端学习算法



实际设备中的电路技术

设备端学习 在设备上快速进行AI训练的技术

- 可在芯片内进行训练，无需准备训练用的数据
- 无需在云端等进行预训练
- **可以在设备端学习**，因此抵抗波动和环境变化的能力强

+

AI加速器（AxlCORE-ODL）

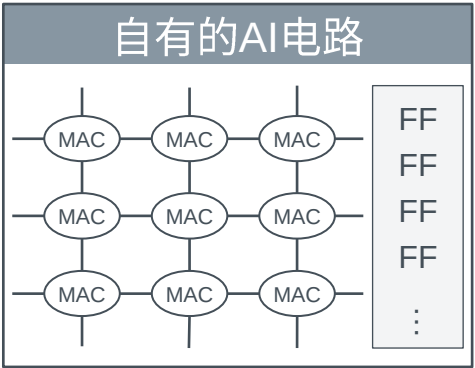
- 以**低成本的硬件电路**实现AI

小型CPU MatisseCORE

- AI配置可**通过软件灵活更改**

特点

- ① **低成本**
= 假设应用在边缘设备（分析权重计算）
- ② **跟随能力**
= 响应变化的模式（轻量遗忘机制）
- ③ **高精度**
= 即使有多个正常模式也能保持精度（集成学习）
- ④ **稳定运行**
= 嵌入后始终运行（抑制过度学习并实现稳定输出）



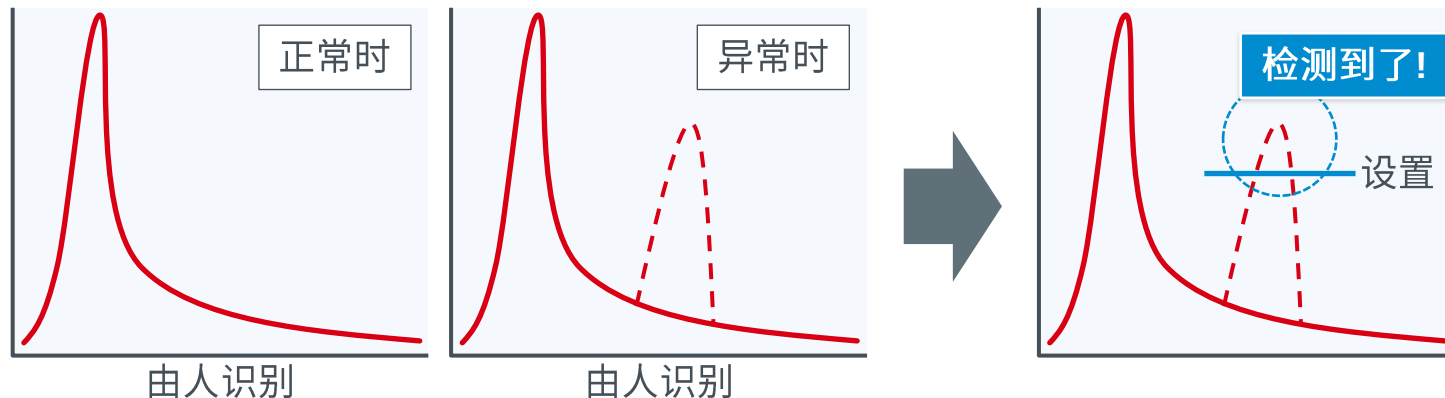
可在设备端学习不同安装场所的不同环境
(设备端学习)



无需对每个安装场所提前收集数据！
无需下载到每个设备！

AI可以对不同于正常工作时的变化进行量化,即使是未知的异常也可以检测出来

基于传统方法的故障预测



如果已经知道异常时的变化（例如出现特定峰值等），就可以检测出来

如果没有人为预先设置将输入什么数据或异常时会发生什么变化等算为异常，就无法检测出来

基于AI的故障预测



即使是意外的异常或异常时的变化是未知的，也可以检测出来

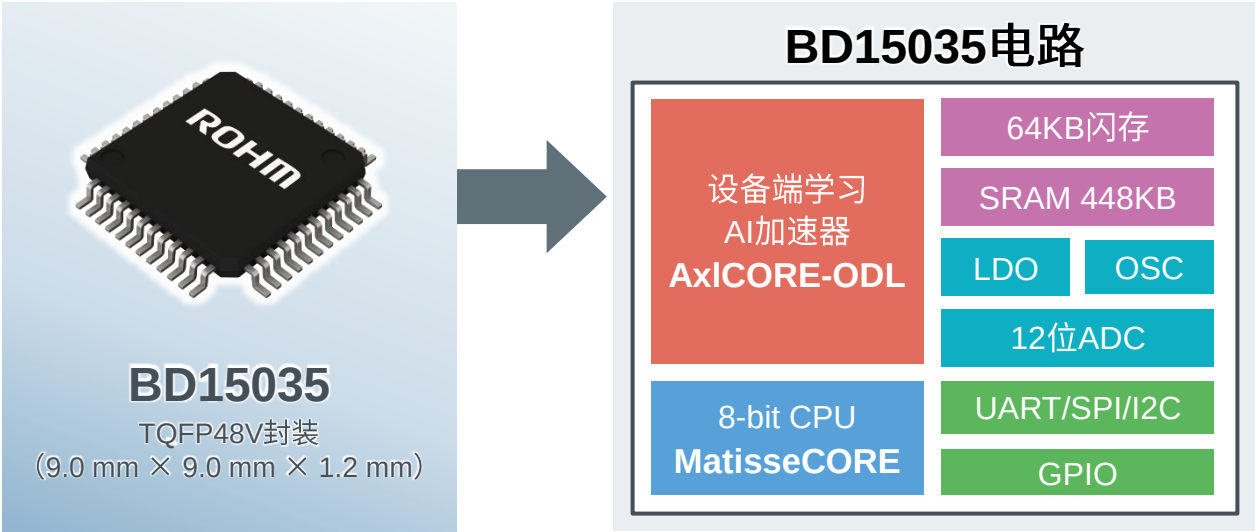
无论输入什么数据，AI可学会识别正常状态，并检测出异常（可推理）

设备端学习算法可在现场实时实现故障预测（无需云服务器）

已将设备端学习所需的AI加速器、CPU、输入I/F集成在1枚芯片中

主要的电路和功能

- 搭载AI加速器 “AxICORE-ODL”
 - 在AI技术基础上，采用设备端学习算法（三层神经网络）
 - 支持FFT和过滤处理
- 内置8位CPU “tinyMicon MatisseCORE™”
- 输入I/F配备UART、SPI、I2C以及12位ADC



特点

AI功能

- 支持设备端学习：无需预训练和通过云服务器进行分析（三层神经网络）

超低功耗

- 数十毫瓦的超低功耗：支持电池驱动和在端点工作

小型芯片

- 将AI功能重新构建为超小型AI加速器
- 小型且高效率的8位CPU

高速处理

- 通过AI加速器进行高速处理，给CPU带来的负载更少

实现在设备端实时故障预测（故障迹象检测）

各AI芯片与ROHM端点用AI芯片的性能比较



	云计算设备用AI芯片	边缘计算设备用AI芯片	端点用以往AI芯片	端点用ROHM AI芯片
要求性能	- 出色的学习能力 - 高度安全性	- 网络负载降低 - 响应时间短 - 低功耗	- 网络负载为零 - 响应时间非常短 - 超低功耗	- 网络负载为零 - 响应时间非常短 - 超低功耗
硬件组成	高性能GPU/ 机器学习专用处理器	嵌入式GPU和FPGA	MCU	AI加速器 + 内置Matisse的MCU
功耗	20W ~ 200W	2W ~ 10W	20mW ~ 1000mW	约30mW *特定应用工作时的实测值
响应时间	数秒 ~ 数十秒	数秒	毫秒	毫秒
训练	可	不可 * 使用完成训练的AI模型	不可 * 使用完成训练的AI模型	可
推理	可	可	可	可

使用仅数十毫瓦的低功耗AI芯片即可进行训练和推理，
实现在端点的实时故障预测

深度学习（具有几十个中间层）

※应用示例

- 代替人玩围棋或将棋
- 预测气象信息
- 识别监控摄像头和图像中的人等

三层神经网络

※应用示例

- 可以识别人体动作
- 例) 通过图像传感器检测一个人是躺卧还是站立

三层神经网络AI芯片 “BD15035”

（取决于存储器等的规格限制）

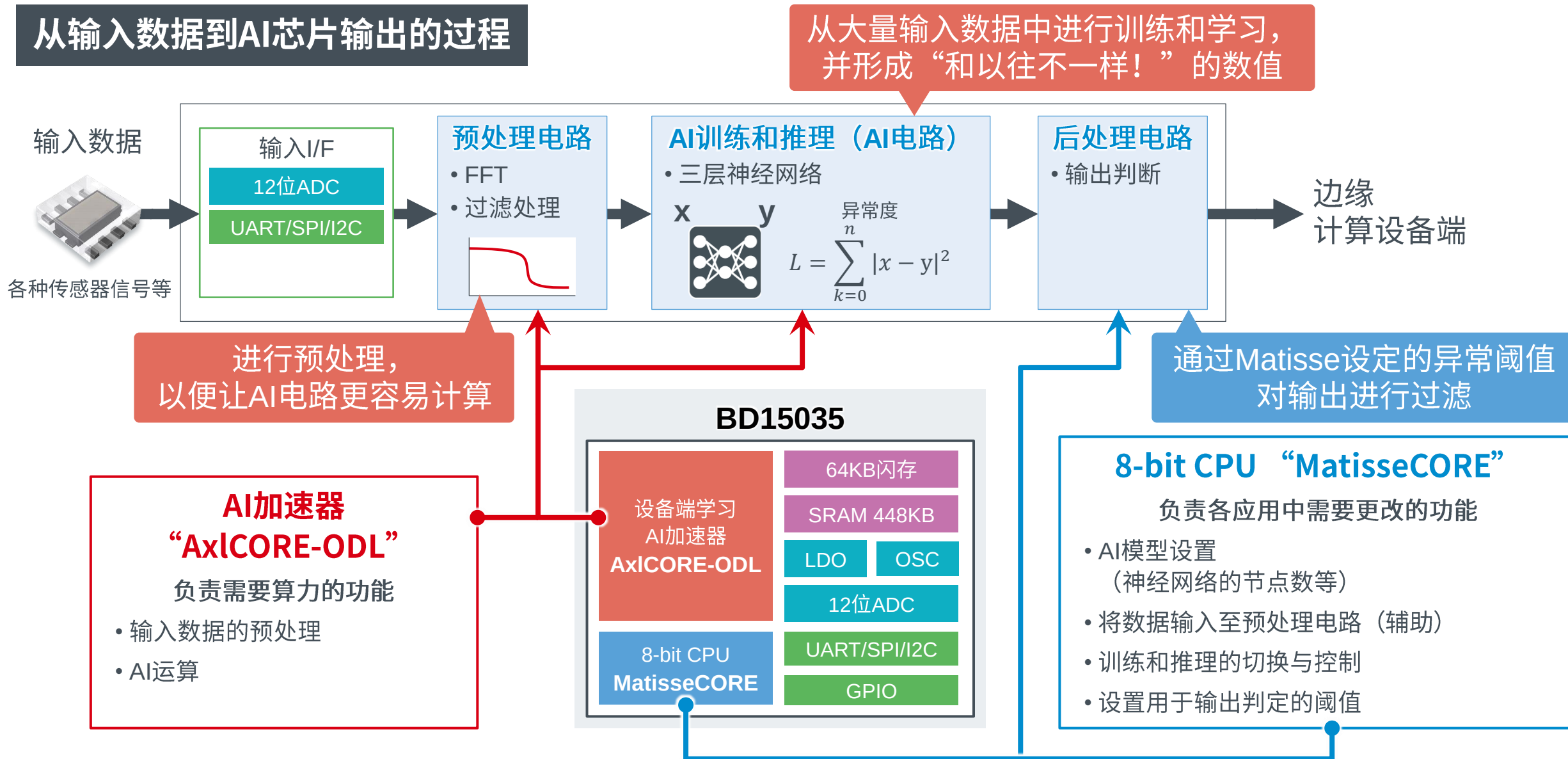
- 通过加速度、电流以及声音识别等进行故障预测

AI的“分层”示意图



AI芯片“BD15035”的处理和任务分担

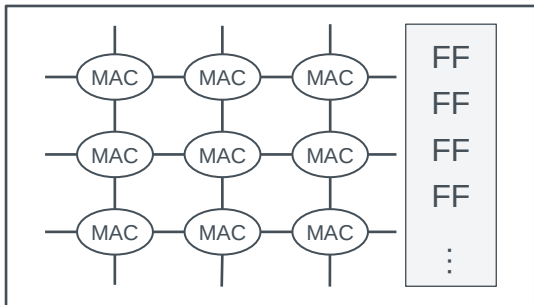
从输入数据到AI芯片输出的过程



将日本庆应义塾大学提供的设备端学习电路（AI电路）
利用AI加速器重新设计，门数更少

自有的AI电路

500万门



门数仅1/250
缩减至0.4%

ROHM的AI芯片（AI加速器+Matisse）

Matisse

AI运算
时序

时序器

运算器

SRAM

(通过训练得到的权重数据)

2项专利申请中
AxICORE-ODL

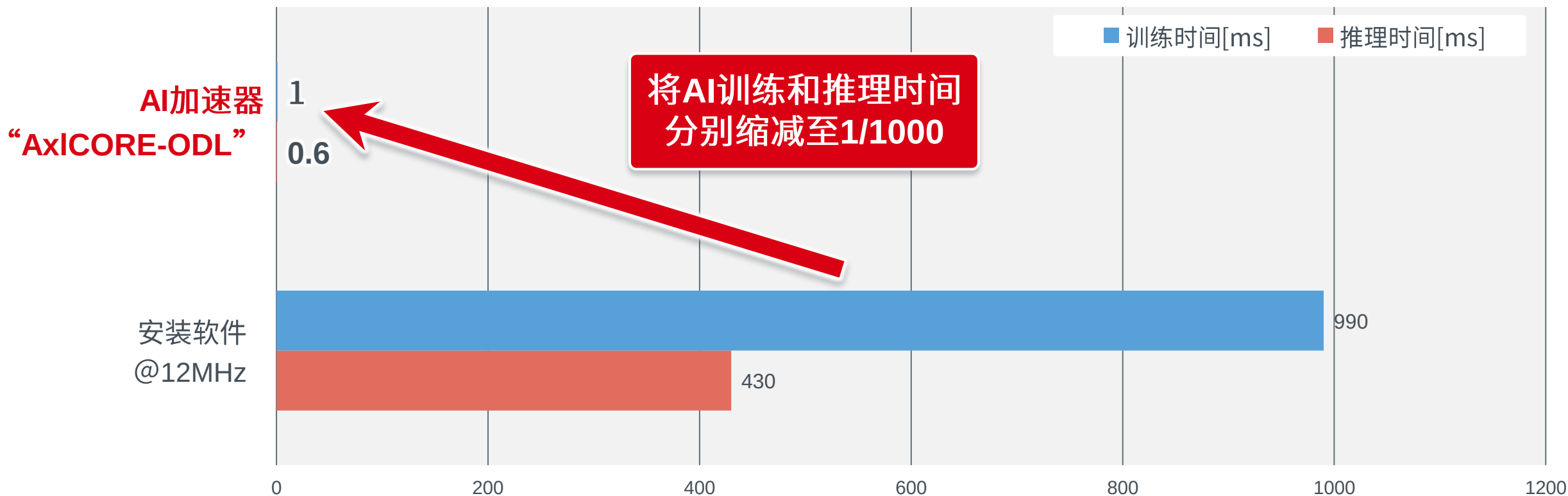
流水线处理

2万门

- 32bit定点数
- 由多个乘积累加运算器(MAC)和FF构成的电路非常大
- AI的结构是固定的

- ✓ 采用**bf16 16bit浮点数运算方式**，比二进制运算精度更高
(很多AI芯片为提高速度和节省内存而采用1-2bit的二进制运算方式，这会导致精度下降)
- ✓ 通过**Matisse设置AI运算时序**，将运算单元一体化
- ✓ 使**AI结构（输入数据数量、算法）可调**
- 可平衡处理时间和内存使用量，并改进算法
- ✓ 将获取来自SRAM的数据并进行计算和保存的过程进行流水线处理，使**处理速度提高到3倍**
- ✓ 利用设备端学习算法，可在芯片上**训练三层神经网络**
- ✓ 利用可无监督学习的自编码器，**无需预训练**即可进行异常检测

训练和推理的执行时间比较 (神经网络: 输入层设置为96个节点, 中间层设置为12个节点)



- ✓ **CPU负载低。**即使是低成本的8位CPU, 也能确保应用产品具备足够的处理能力
- ✓ **支持高速采样。**可检测出10kHz左右的高频范围出现的异常
- ✓ 时间序列数据预处理所必需的**FFT也可通过AI加速器 (内置在电路中) 来实现**

设备端AI芯片原型“BD15035”

- 内置AI加速器 “AxICORE-ODL”
- 内置8位高效CPU “tinyMicon MatisseCORE™”

BD15035

设备端学习
AI加速器
AxICORE-ODL

8-bit CPU
MatisseCORE

64KB闪存

SRAM 448KB

LDO

OSC

12位ADC

UART/SPI/I2C

GPIO

TQFP48V封装
(9.0 mm × 9.0 mm × 1.2 mm)

评估板

- 可连接Arduino用开发板
- 配备Wi-Fi/Bluetooth®模块
- 配备64kbit EEPROM



评估板使用示意图
(使用加速度传感器时)

例) 加速度传感器

故障预测
监控对象设备

获取数据

训练和推理

显示结果

发生异常 推理结果显示

正常振动

异常振动

加速度
信号

训练标志

异常值
(输出不同于以往的数值)

异常值
逐渐上升

CH0 @ input_x

CH1 @ input_y

CH2 @ input_z

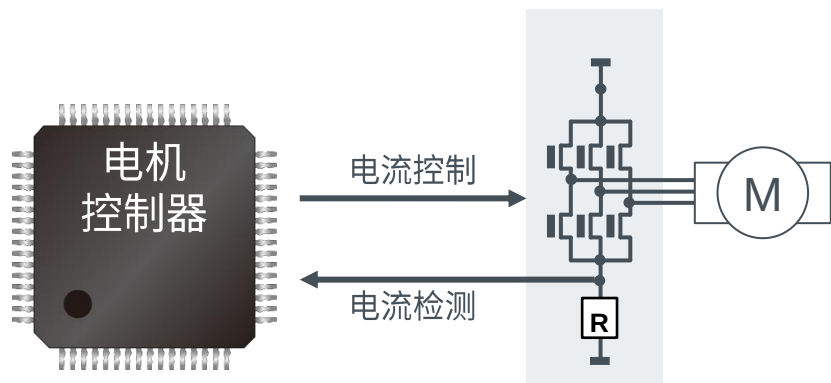
CH3 @ learning_loss

CH4 @ output_loss

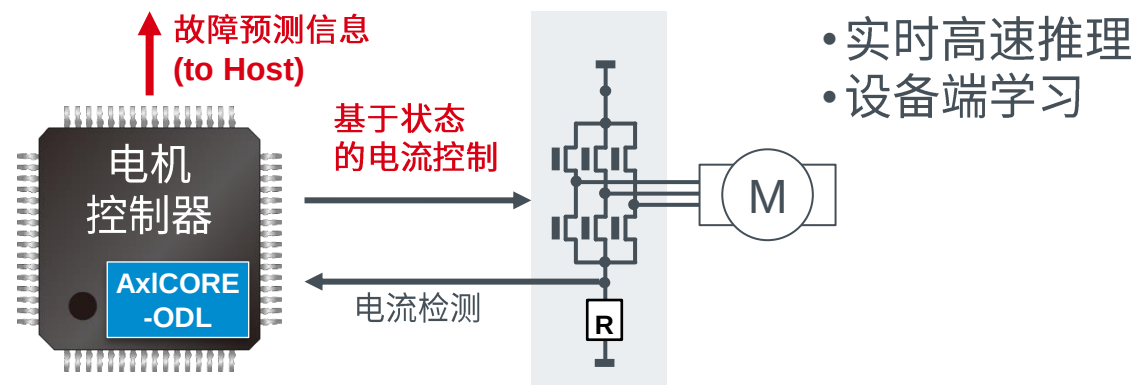
Timebase

可实现的应用案例1: 具有设备端学习AI功能的电机控制器

现有的电机控制环境

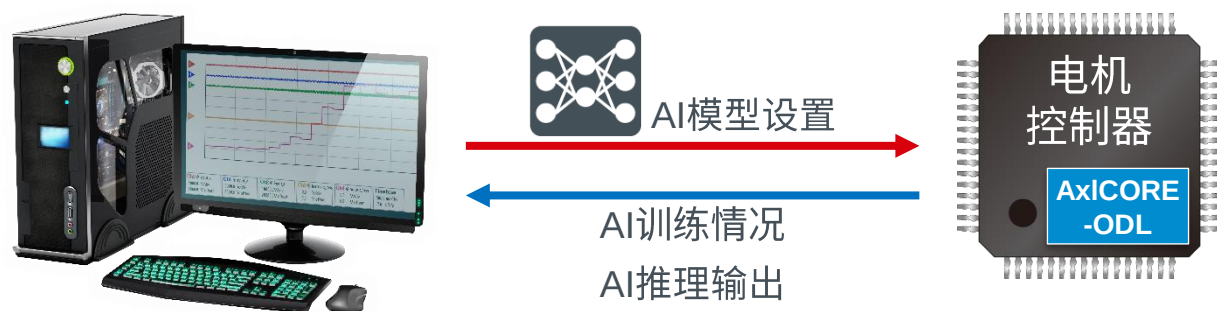


具有AI功能的电机控制环境



与传统的电机控制器IC相比，
成本更低而且很容易（无需增加部件）添加AI功能

ROHM正在开发一种可轻松完成从AI模型构建到评估的工具



无需设计复杂的模型和调整诸多参数

- 在监控AI芯片输出的同时轻松构建AI模型
- 用最少的参数调整AI模型
(输入数据数量、异常判定阈值)
- 只需按一下按钮，即可在设备端重新学习

可实现的应用案例2： 具有设备端学习AI功能的边缘通用微控制器

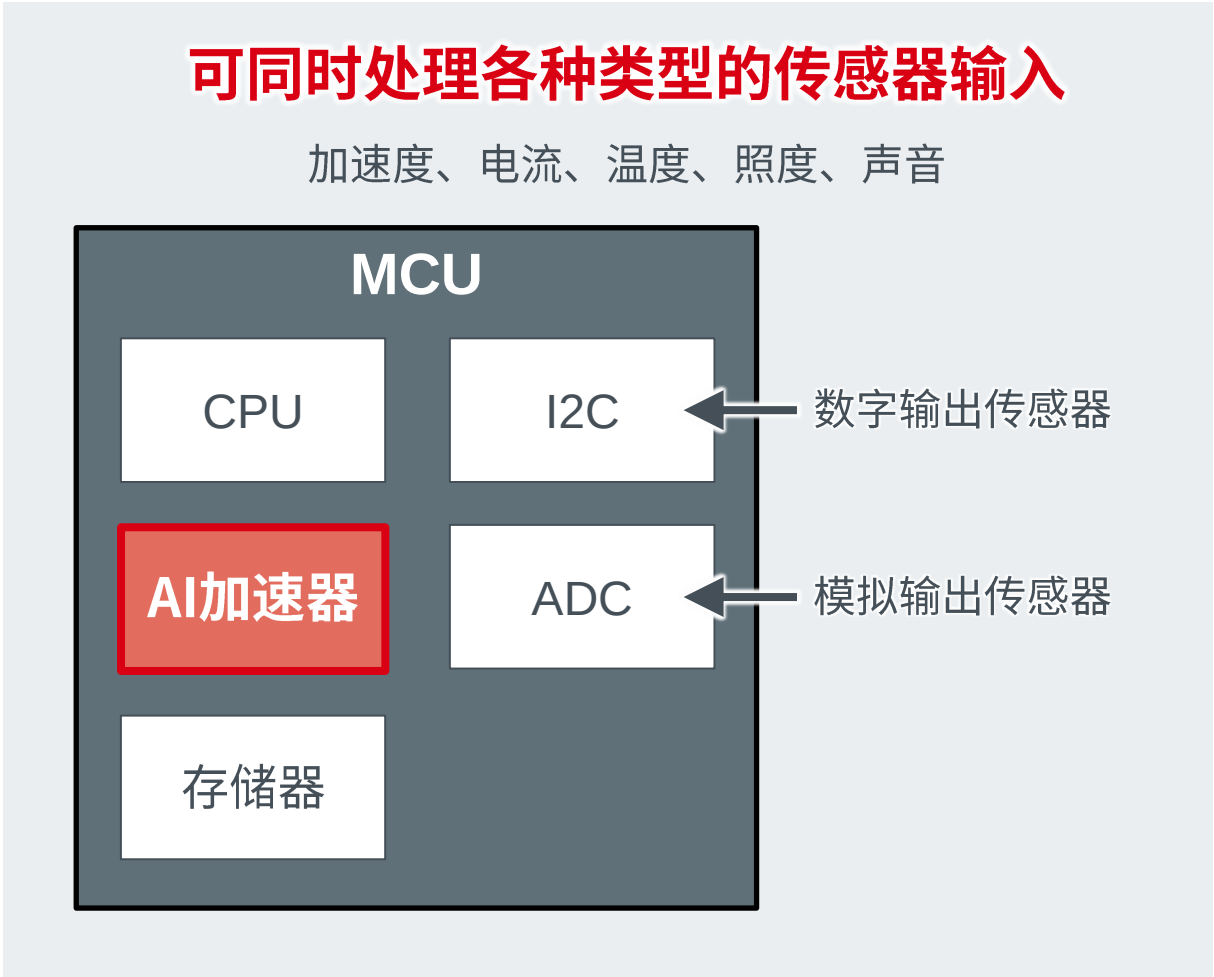


- 具有设备端学习AI加速器
并将其作为外围设备的通用微控制器
- 用小型硬件执行高速AI运算，
应用功能可通过软件灵活配置

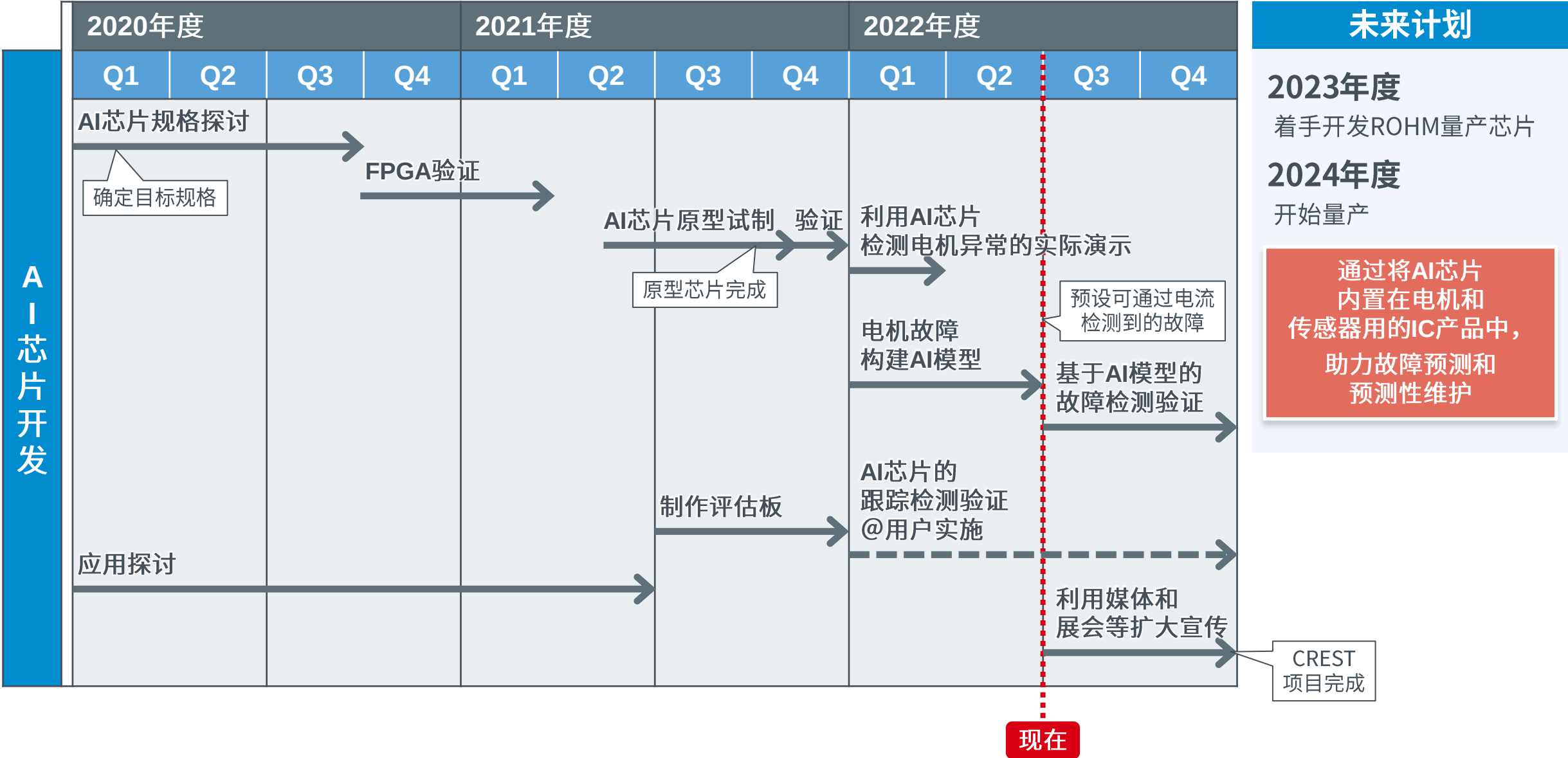
在工业设备、车载设备、家用电器等
边缘和端点MCU中
轻松添加AI功能（故障迹象检测等）

优点

- 所需的AI运算可以由硬件执行，
因此软件的负载更低，对应用产品的功能没有限制
- 通过替换应用产品现有的微控制器，即可新增AI功能
- 在设备端不仅可以进行推理，还可以进行训练，
因此可针对各安装场所轻松进行优化

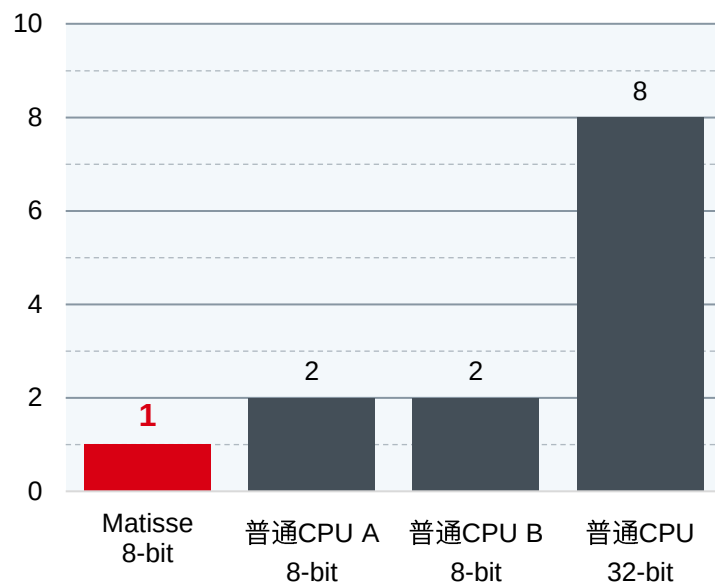


从芯片原型开发到未来向市场推出产品的时间表



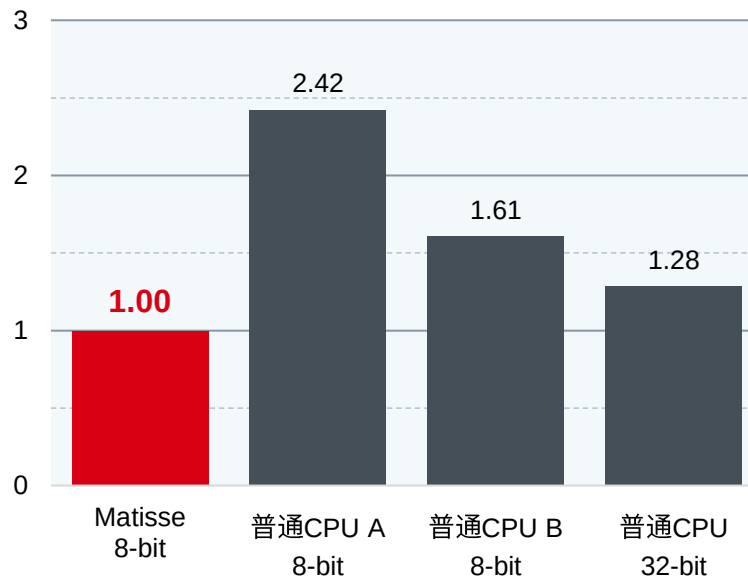
与普通小型CPU之间的性能比较 (设Matisse为1时)

门数比较



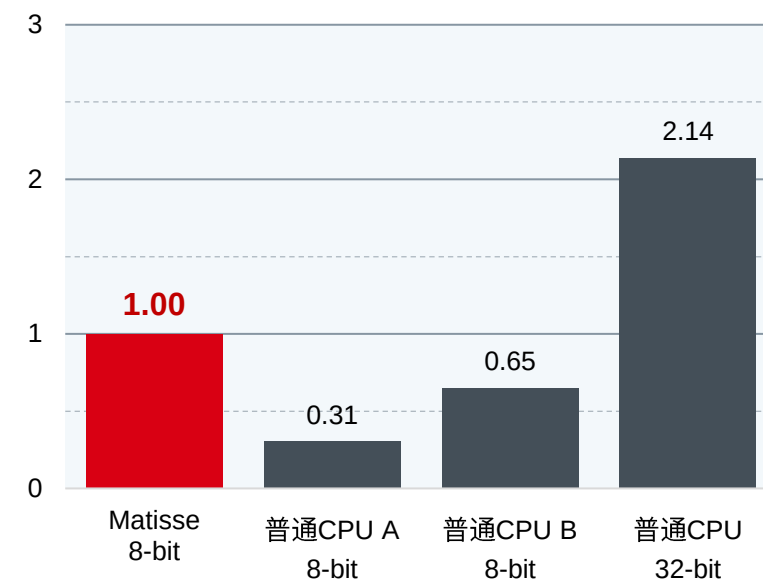
更节省面积

ROM大小比较 (8-bit运算程序时)



紧凑的
程序代码大小

处理性能比较 (Dhrystone)

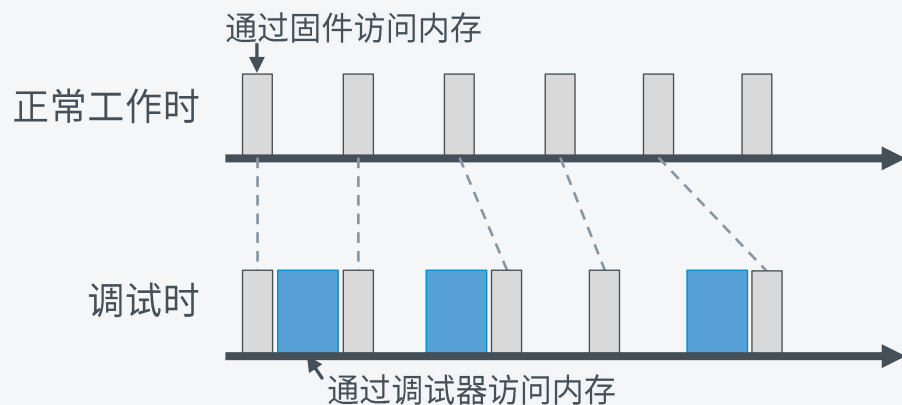


高速运算处理

**Matisse以高标准实现了更小的芯片面积和程序代码、
以及更高速的运算处理能力 (符合车载ASIL-D标准)**

～ 非常适合嵌入式应用的实时调试功能 ～

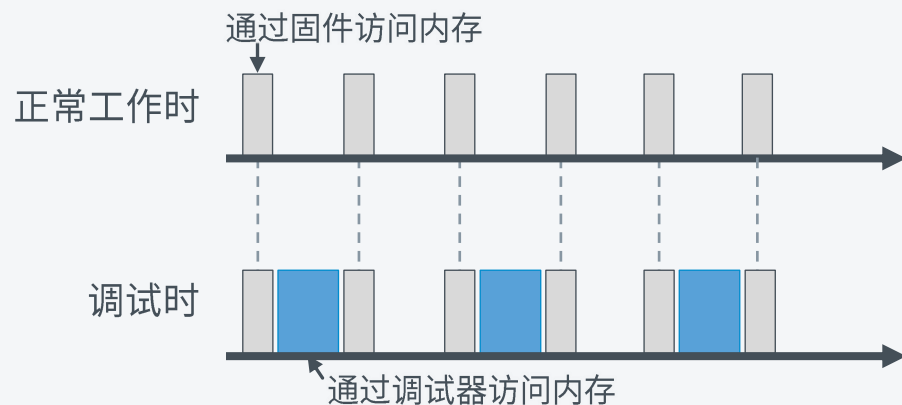
传统方法



调试器访问内存会妨碍固件工作

调试可能会导致
程序行为显著改变

Matisse的实时调试功能



通过调整调试器的内存访问，来避免妨碍固件的工作

调试时完全不需要
改变程序的行为

Matisse可以完全无损地提取和更改CPU内部信息

- 利用实时调试功能，**确保调试时的程序行为与正常工作时一样**
- 在电机控制等**无法停止运行来调试的应用中也可轻松调试**
- **可以实时提取**通常无法看到的**IC内部变量**并通过**波形显示**出来



波形显示用的软件
RapidScope™



Electronics for the Future

- 本资料中的内容旨在介绍ROHM的产品（以下称“ROHM产品”）。
- 在使用ROHM产品之前，请务必另行确认最新的规格书和技术规格书。
- 本资料中的信息不提供任何保证。客户或第三方万一因其中的信息错误或使用不当而造成损害，ROHM公司不承担任何责任。
- 本资料中列出的ROHM产品相关的典型工作和应用电路示例仅为示例，并非保证不侵犯与这些内容相关的第三方的知识产权及其他权利。
- 对于因使用上述技术信息而引起的任何纠纷，ROHM公司不承担任何责任。
- ROHM并未明示或暗示地授权实施或使用ROHM或其他公司的知识产权或其他任何权利。
- 本资料中的产品和技术中，当出口或向国外提供属于《外汇和对外贸易法》和其他出口法规管制的产品或技术时，应遵循这些法律法规并获得许可。
- 本资料中的内容为截至2022年9月的内容，如有更改，恕不另行通知。