

# 多核嵌入式架构 使能毫瓦级嵌入式AI应用

张焱

首席软件工程师 兼客户项目经理

[www.greenwaves-technologies.com](http://www.greenwaves-technologies.com)



# GreenWaves Technologies

- 成立于2014年11月
- 由行业内资深团队共同建立
- 现总部位于法国格勒诺布尔
- 员工遍布法国，中国，意大利，瑞士，美国，韩国
- 与苏黎世联邦理工（ETH-Z）及意大利博洛尼亚大学共同实验室（PULP）深度合作
- 于2018年二月发布了公司首款芯片 GAP8
- 于2018年五月开始发售开发板
- GAP9已在RISC-V Summit发布



# 嵌入式技术与AI的结合

## 感知

人类的感知...



超人的感知...

高频震动信号

雷达

红外

超声

等等

## 分析与解释

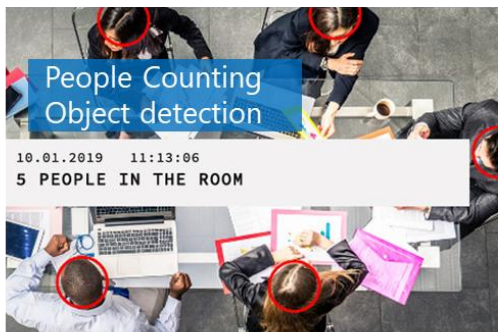


## 执行与连接



# 嵌入式AI应用

智能办公与  
能源管理



大型机械  
监护



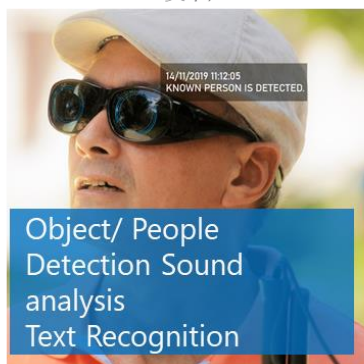
安全与监护



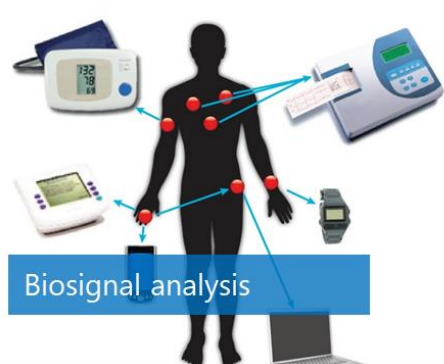
监控与  
出入控制



可穿戴  
设备



医疗传感器



小型机器人  
和玩具机器人



智能家电



# 嵌入式端侧产品限制



## 嵌入式端侧 AI

- 体积小
- 远离电源
- 供电条件苛刻
- 设备数量多
- 安装与维护成本



# GAP系列AIoT处理器，从边缘端开启AI

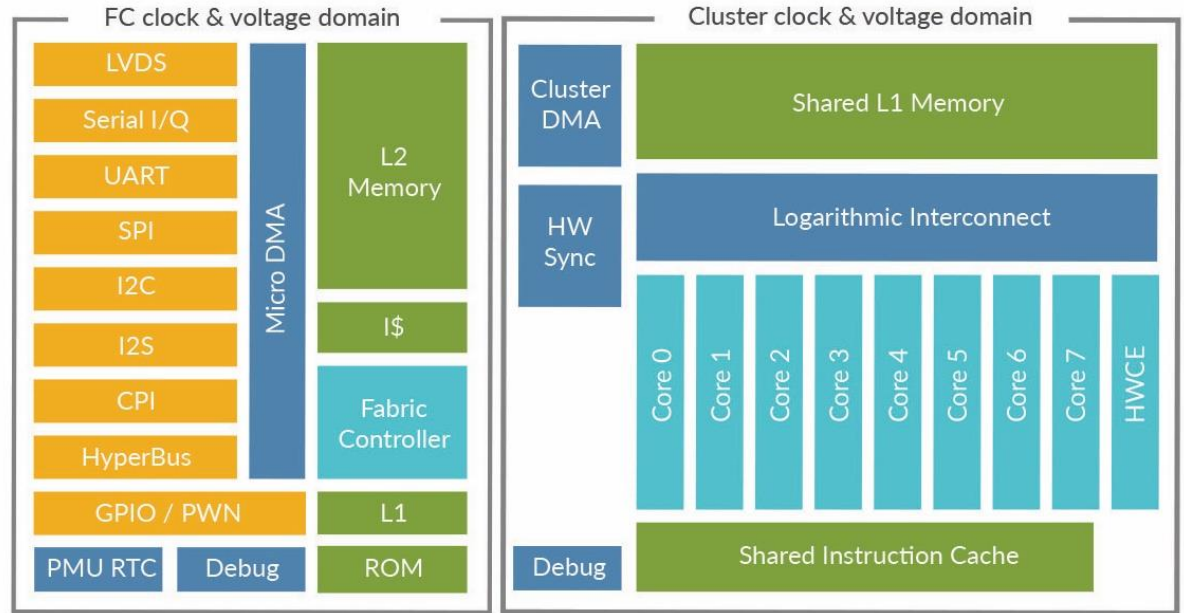
- **MCU级功耗**
  - 高效的多核并行架构
  - 灵活而全面的架构（包含扩展指令集架构）
  - 高速的多级内存
- **灵活性**
  - 灵活调整芯片主频及电压
  - 极快地状态切换
- **可编程**
  - 适用于真实世界各种问题 – 不仅卷积神经网络（CNN）
  - 可适用各种最新，最好的算法 – 二进制网络
  - 单一算法模型可应用于不同架构之上

智能传感设备的单芯片解决方案



# GAP8 – 智能物联网处理器

- GAP系列收款处理器
- 基于PULP（超低功耗并行架构）研究项目5年的研究成果
- 采用台积电55nm超低功耗制程。
- 加入了9种扩展指令集
- 8核并行架构
- 1微控制内核
- 2018年Q3开始发售不同的HDKs
- 已量产



# 基于GAP8的毫瓦级AI应用

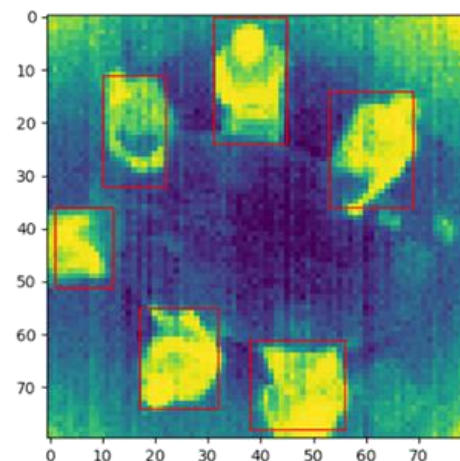
- QVGA 人脸识别

- 人脸检测部分:  $\sim 25\text{ms}$   $\sim 1\text{mW}$  / 帧 / 秒
- 人脸识别部分:  $400\text{ms}$   $22\text{mW}$  / 帧 / 秒
- LFW( Labelled Faces in the Wild ) 数据集下可达百分之93的准确率



- IR摄像头人体检测

- 80 x 80 IR 图像 - LynRED ThermEye
- 图像处理 + 人体检测
- $62\text{ms}$   $\sim 4.4\text{mW}$  / 帧 / 秒
- 可达百分之99的准确率 (使用内部收集数据集)





# GAP9 – GAP系列新一代智能物联网处理器

- GAP8
  - 结合市场领先的SoC架构...
  - ...与市场主流的台积电55nm超低功耗制程
- GAP9
  - 利用从GAP8获得的经验来调整架构
  - 采用市场领先的半导体工艺Global Foundries 22nm FDX
- GAP9 为行业市场带来了前所未有的能效
  - 比起GAP8，计算复杂度提升10倍
  - 功耗是GAP8的五分之一
  - 更灵活，可编程程度更高



# GAP9 – 架构的进化

- 能力的提升

- 可以解决更复杂的问题:

- 内部RAM容量提升至1.5MB
    - Cluster L1 最高带宽提升至24GB/s
    - L2带宽最高提升至7.2 GB/s
    - 硬件支持NN权重参数的压缩, 打包与解包

- 更多的运行状态

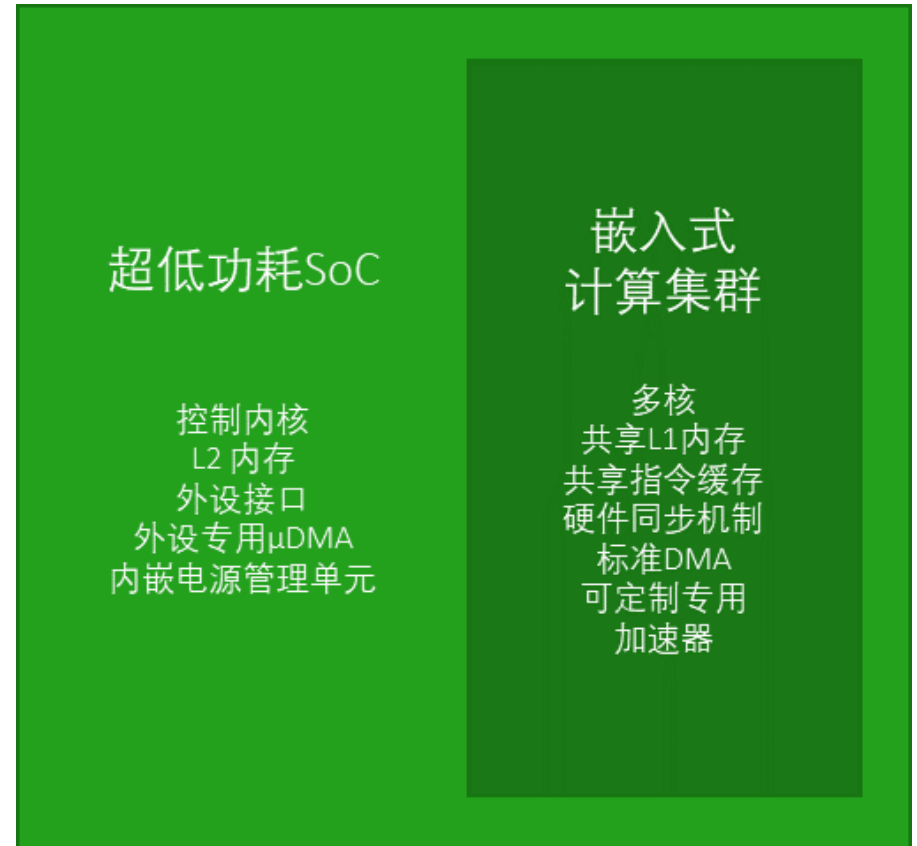
- 芯片主频最高400MHz
    - 更灵活多变的能源管理方式

- 功能更丰富

- 支持高精度IEEE 32/16bit浮点运算, 备用16/8bit浮点运算
    - 双向多通道音频接口
    - 增加CSI2摄像头接口

- 安全性提升

- 支持硬件256/128 bit AES编解码
  - 硬件可编程反克隆功能 (PUF)



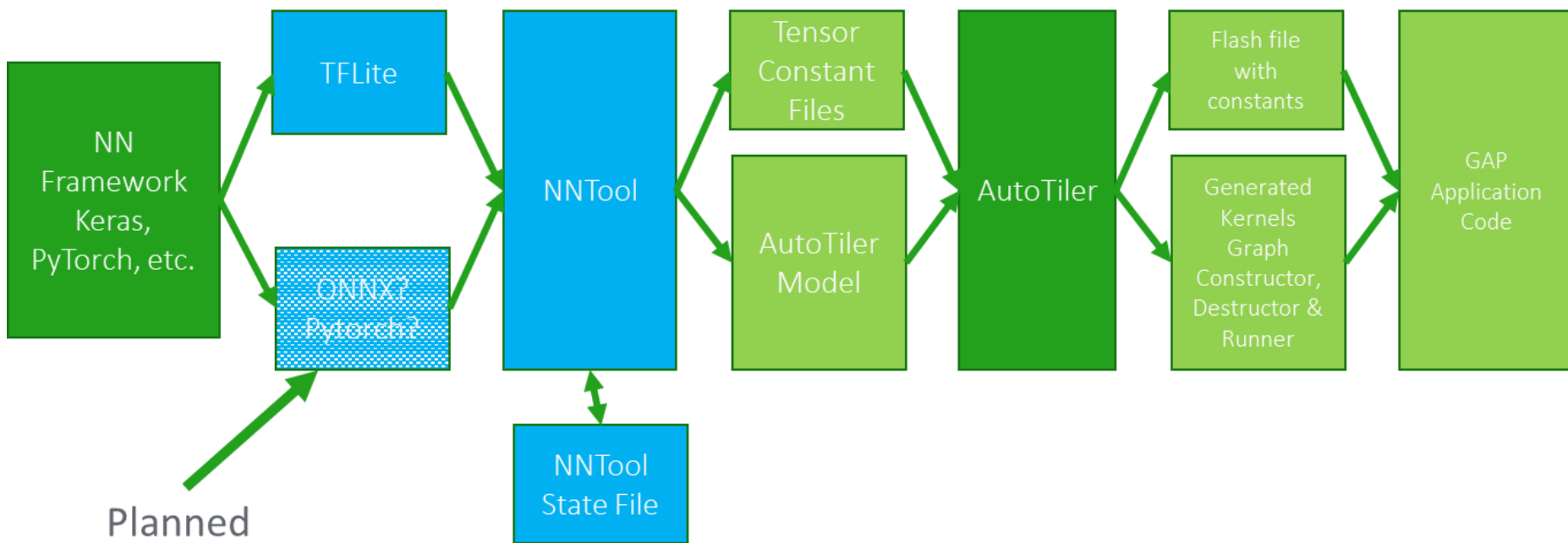
# 先进的架构只是成功的一半，另一半靠工具

客户的需求是：**二次开发成本最小化。**

客户的疑虑：

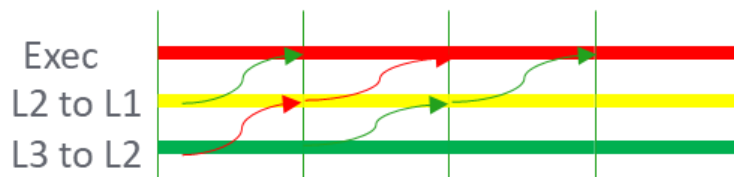
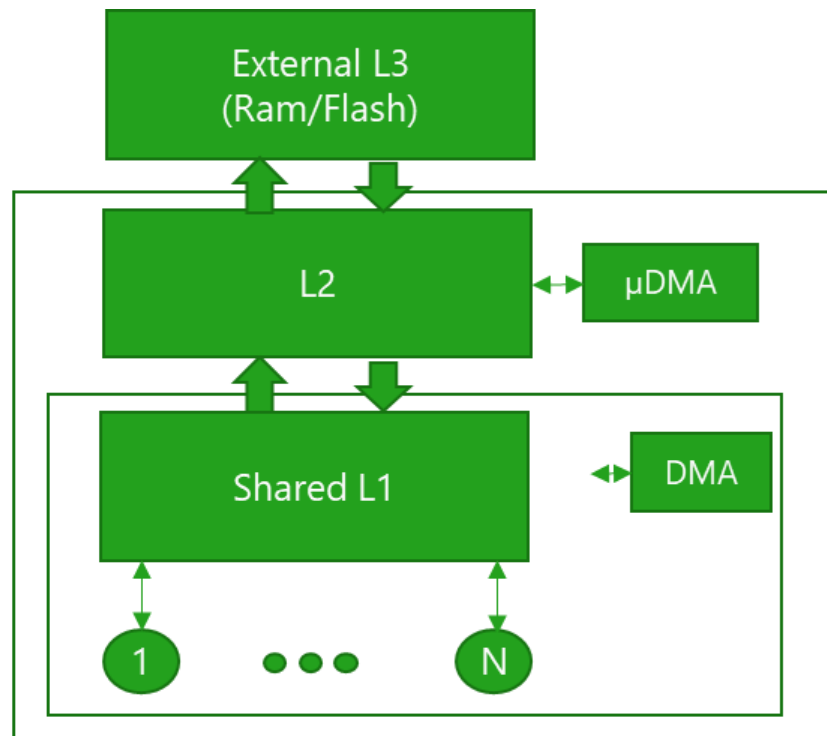
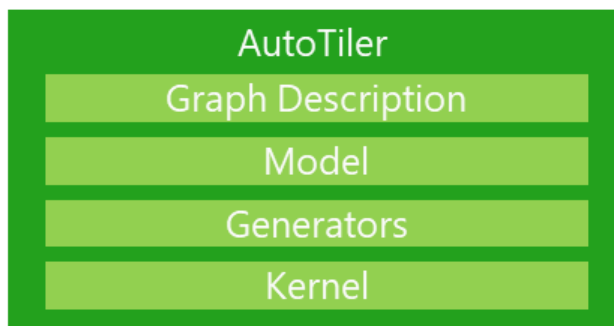
- 算法如何并行化？我单核的算法都写不好，8个核怎么玩？
- 数据流处理太复杂了，异步程序开发门槛太高。
- 我的算法移植后性能如何？表现如何？是否能满足我的需求？
- 需要多少内存？
- 需要多少人？什么样的人才？多少时间才能完成呢？
- 怎么用好GAP的架构？发挥出其优势？
- 如何量化神经网络？（什么？量化是什么？）
- 等等。。。

# GAPFlow 工具套件



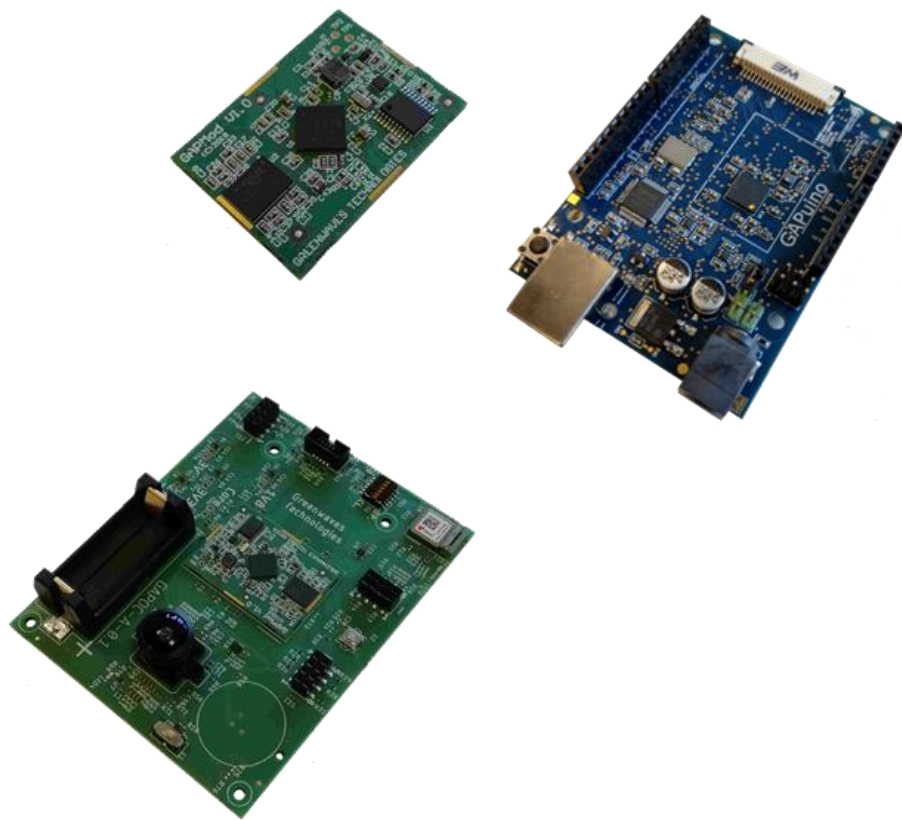
# GAP AutoTiler 代码生成器 – 数据流管理

- 数据缓存不适合流数据 – 缓存效率低
- 机器学习/信号处理数据流的大小在编译时即可知。
- 自动将数据进行切割并生成完整数据流控制，内存控制，以及对多核算法调用的代码。



# GAP 系列让低功耗AI应用落地成为可能

- GAP8已量产
- GAP8开发板，核心板，以及计算机视觉/IR视觉参考设计板已开始发售。
- GAP9开发板将于2020年年中向固定用户投放。
- GAP9虚拟平台已提供给部分用户测试，开发
- GAP系列SDK与工具均托管于GitHub上
- 欢迎大家来RISC-V展台，1E58D展位进一步交流



Real ... Now ...

Thank you!

We are looking forward to see you at 1E58C !

Questions?