



借助自适应计算提速 您的 AI 赋能边缘解决方案

自适应系统模块 (SOM)
正式发布

概要

AI 赋能的应用正越来越多地部署于边缘应用。自适应 SOM 提供了一款综合全面的可量产化平台，用以在边缘构建 AI 赋能的加速应用，包括城市、工厂和医院中的智能视觉应用。这些应用要求以低时延、低功耗和小封装尺寸执行大量数据处理。要想实现这一综合要求，则必须为整体应用（包括 AI 流程和非 AI 流程）进行提速。

随着 AI 模型快速演进，加速平台必须具备灵活应变的能力，这样才能取得当今乃至未来 AI 技术的最佳实现。在本电子书中，我们将介绍边缘端的 AI 并阐释 SOM 缘何成为一款出色的量产解决方案。此外，我们还将介绍自适应计算，并详细讲解自适应 SOM 提供的优势。最后，我们将详细介绍选择 AI 赋能的边缘解决方案时应考虑的各种因素。

目录

第 1 章： AI 赋能边缘计算的崛起	1
第 2 章： 系统模块 (SOM) 介绍	4
第 3 章： 边缘端数据处理的挑战	9
第 4 章： 自适应计算介绍	13
第 5 章： 自适应系统模块 (SOM)	16
第 6 章： 自适应 SOM 对硬件开发者的益处	20
第 7 章： 自适应 SOM 对软件开发者的益处	23
第 8 章： 选择边缘解决方案时的考量因素	26
第 9 章： 总结	30

第 1 章

AI 赋能边缘计算的崛起

SOM 用例：农作物
健康分析

需要具备超低功耗的高性能 AI 推断, 从而在不影响电池寿命的情况下最大限度发挥功能。



AI 赋能的应用正越来越普遍地部署于边缘和终端。众多行业与地区均在开发数量庞大的，采用最新 AI 技术的应用。

随着自动化视觉应用可管理安全并在需要帮助时告警应急服务，城市将愈发智能。而伴随工厂智能化与自动化水平的提升，工业物联网应用对高性能 AI 推断处理的需求也日益高涨。甚至零售业也开始向智能零售转型，其所带来的自动化购物体验是 10 年前无法实现的。所有

这些应用都必须以极高可靠性在严苛的条件下运行 10 年乃至更长时间。这些应用不仅需要高性能，还要求必须以高效、可靠及紧凑的外形尺寸交付。

AI 边缘芯片组收入，全球市场：2019-2025

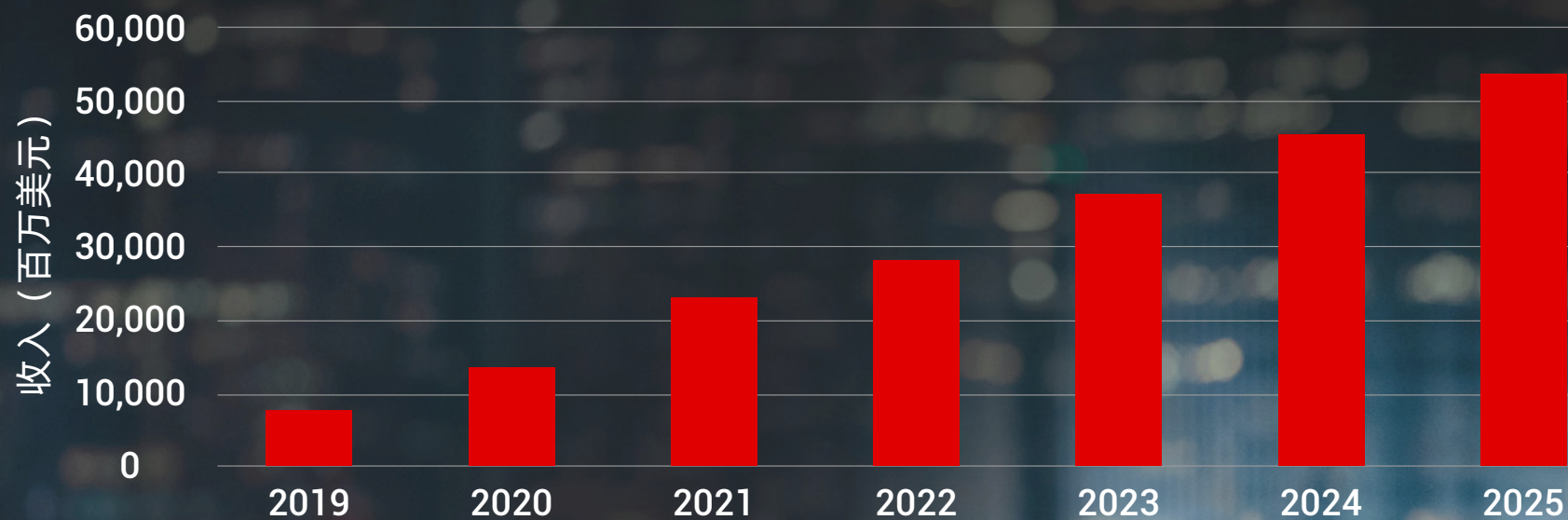


图 1：到 2025 年，AI 边缘芯片组全球收入预计将超过 500 亿美元。（来源：Omdia）

AI 赋能的边缘计算日渐增长趋势也为市场研究公司 Omdia 所印证。据 Omdia 估算，到 2025 年，AI 边缘芯片组的全球收入将超过 500 亿美元。

第 2 章

系统模块 (SOM) 介绍

SOM 用例：3D 打印手臂控制

3D 打印领域正快速演进。通过对可扩展的运动轴数施加精准的、明确性的控制，设计师正寻求提升 3D 打印工序的速度与品质

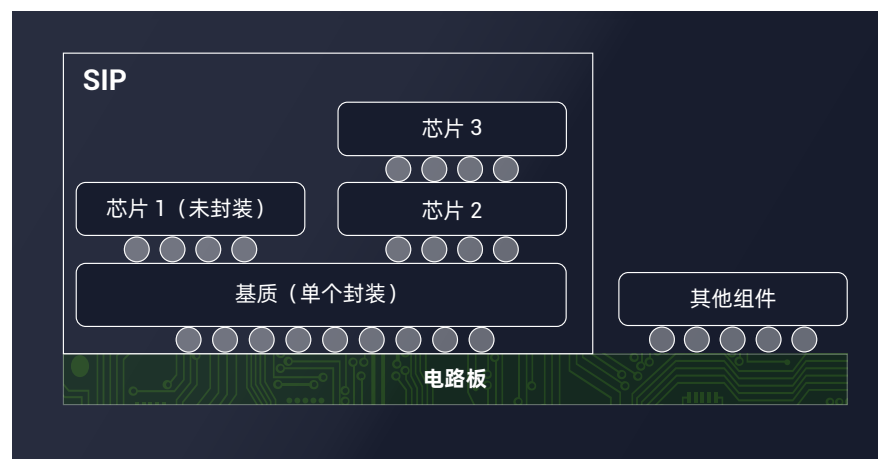
构建一款 AI 赋能边缘应用的过程中会遇到大量可用的方案。在众多行业中，通用方法是硬件设计团队开展“芯片先定”的策略，即先选择特定芯片，然后为应用开发完全可定制的电路板。尽管这种方法能切实提供高度优化的实现方案，但需要耗费大量开发与成本才能达到量产就绪状态。

为节省“芯片先定”策略的研发支出和时间，设计团队可以考虑采用集成度更高的解决方案，例如多芯片模块 (MCM)、系统级封装 (SIP)、单板计算机 (SBC) 或系统模块 (SOM) 等。



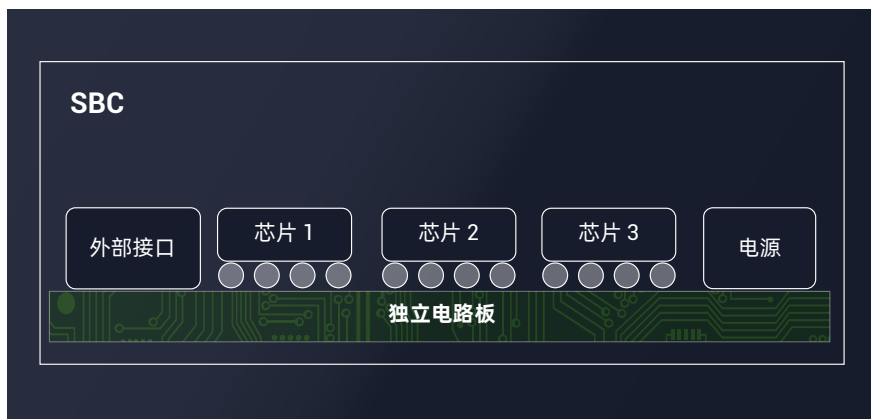
多芯片模块 (MCM)

MCM 将多个（通常已封装的）芯片器件融入在单个较大型的封装内。这种方法减少了电路板上单个器件的数量，因而具备简化电路板设计的优势。尽管 MCM 通常拥有成本和性能优势，但其只适用于一些特定应用，而且仍需开发定制的复杂电路板。



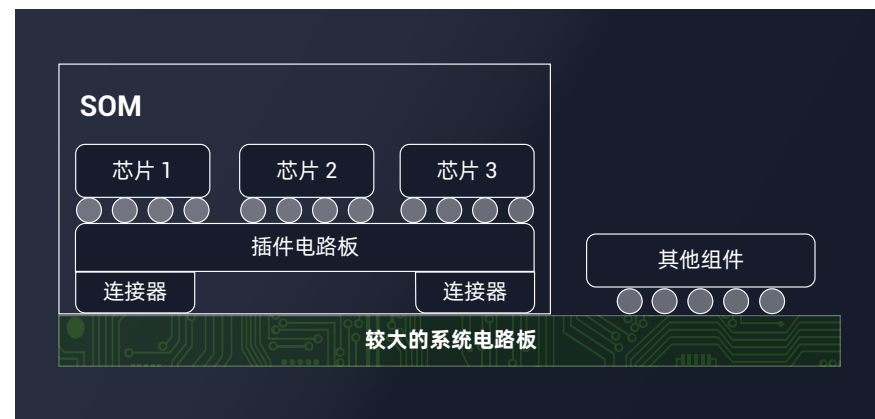
系统级封装 (SiP)

SiP 与 MCM 相似，不同的是，单个硅裸片往往不进行封装。此外，一些 SiP 还将裸片彼此堆叠成 3D 阵列。SiP 还可以包含电阻器和电容器等分立组件，以减少在电路板上装配这些组件的需要。不过 SiP 与 MCM 相似，仍然需要设计完全定制化的电路板，即要求完整的“芯片先定”设计流程。



单板计算机 (SBC)

SBC 是在电路板上构建并作为独立产品交付的完整计算系统。通常情况下，这些系统包括一个小型微处理器、存储器和输入/输出 (I/O)。但是 SBC 往往不以量产部署为开发或验证目的。



系统模块 (SOM)

SOM 是一种与 SBC 相似的计算系统，但其设计初衷是连接更大型的解决方案，而非独立运行。在构建到高性能边缘应用中时，SOM 相较于前文所述其他解决方案类型具有众多优势。SOM 提供了一个完整的可量产化计算平台，与“芯片先定”开发相比，可节省大量开发时间和成本。SOM 能插入更大型的边缘应用中，既能为定制实现方案带来灵活性与易用性，又缩短了现成解决方案的上市时间。



“利用 SOM 取代‘芯片先定’开发策略可以节省大量开发成本。开发用时更短，并且开发过程中遇到的错误也更少，二者均有利于加快产品的上市时间。此外，SOM 经过预先验证，可以降低风险，以及帮助工程师们更加轻松地实现其定制设计。工程师们不必着眼于 SOM 上的基础设施，而只需聚焦于其所关注的关键差异因素即可。”

— Bryan Fletcher, 安富利技术营销总监

因此，SOM 是可量产化的、业经验证的设计，其包含具体应用领域所需的所有芯片组件和接口连接器。

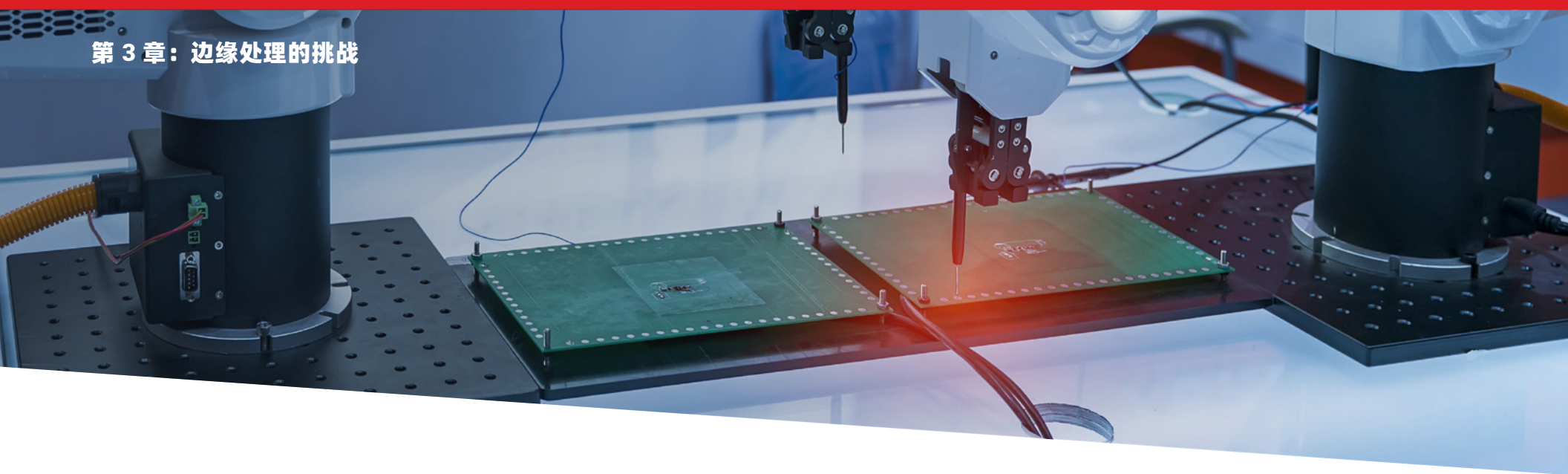


第 3 章

边缘处理的挑战

SOM 用例： 数字化病理学

放射学、病理学、皮肤医学和眼科医学的典型流程是处理较大的图像尺寸，其间需要复杂的图像处理，尤其是借助计算密集与存储器密集型 AI 工作流程进行处理。



边缘计算通常受制于功耗、占板面积及成本。随着处理需求的增加，在边缘计算的限制内提供所需性能水平的难度也呈指数级增长。

CPU 已在边缘端实现了大量改进，但近年来改进速度有所放缓。未经加速的 CPU 只能勉强为下一代 AI 赋能的边缘应用提供所需性能，尤其是在考虑到严格的时延要求情况下更是捉襟见肘。

为了在边缘实现先进的 AI 应用，则需要特定领域架构 (DSA)。与未加速的 CPU 相比，DSA 可提供高度优化的应用实现方案，以及确定性与低时延。

在边缘端实现高效的 AI 赋能应用需要整体应用加速。

AI 推断需要非 AI 的预处理和后处理，而所有这些都要求更高的性能。为了高效处理所需的数据，可以专门设计一个合适的 DSA，既处理应用的 AI 推断部分，同时也处理应用的非 AI 部分（即整体应用流程）。因此，在边缘（和其他位置）实现高效的 AI 赋能应用需要整体应用加速。

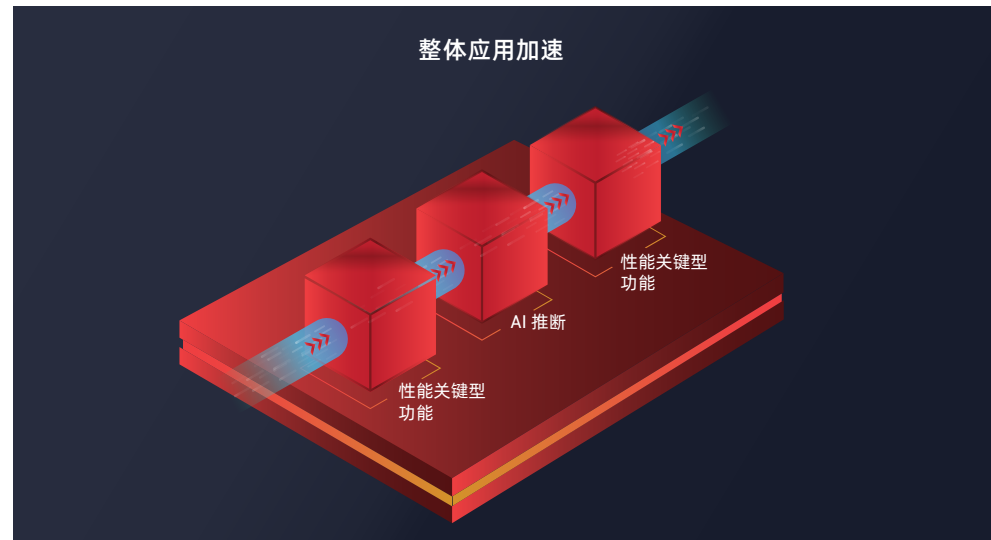


图 2：AI 推断通常是大型处理系统的组成部分

尽管一些针对 AI 边缘应用的 ASSP 已被开发出来，然而与所有固定芯片解决方案类似，它们依然存在局限性。主要在于 AI 创新的节奏极快，致使 AI 模型的淘汰速度远快于非 AI 技术。用来实现 AI 的固定芯片器件会因更新颖、更高效的 AI 模型的出现而迅速过时，而推出一款固定芯片器件的流片可能需要花费数年时间，届时，AI 模型的技术水平已经发生重大发展。

对于边缘应用而言，安全性和功能安全也正变得愈发重要，因其时常导致潜在的现场更新，成本高昂。无法适应这方面变化的固定硬件系统将迅速沦为遗留系统。边缘端的 AI 赋能系统必须具备足够的灵活性，以延长其工作寿命并实现投资回报最大化。

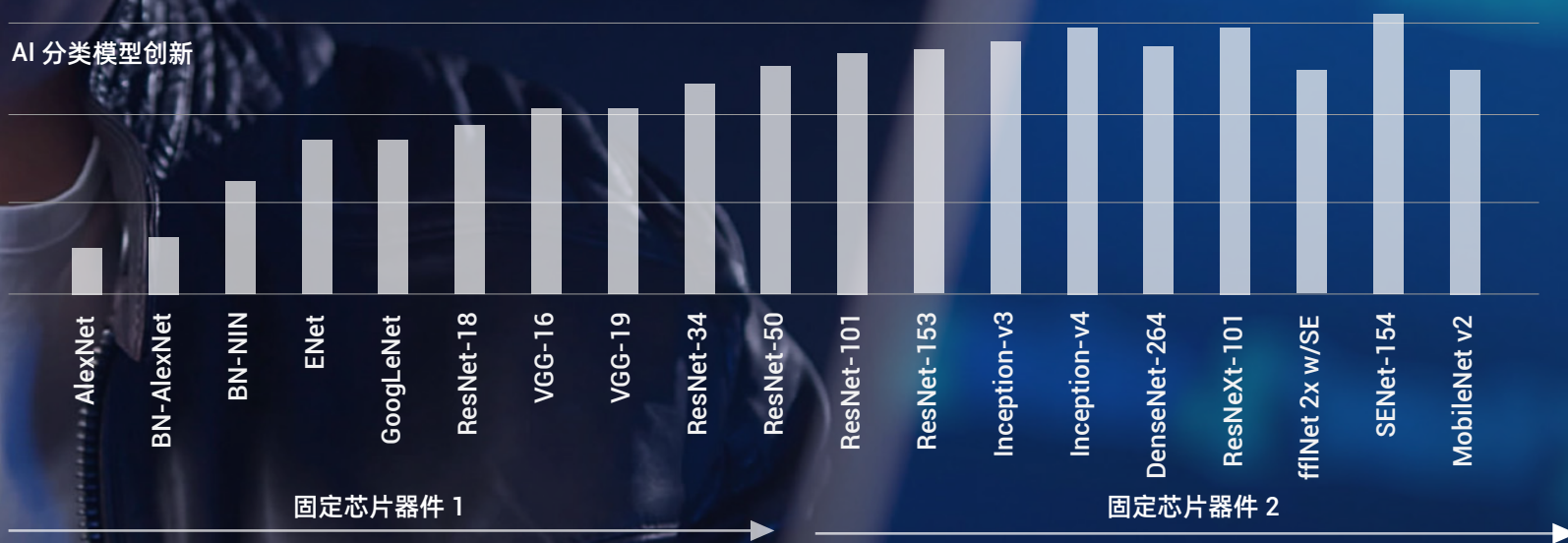


图 3：AI 模型快速演进 – 速度远快于芯片开发周期

A background image showing two industrial robotic arms in a dark factory setting, performing a welding task. Bright orange sparks are being emitted from the welding points, creating a dynamic and high-contrast scene. The robot arms are metallic and complex in structure, with various joints and cables visible.

第 4 章

自适应计算介绍

SOM 用例：
工厂中的机器人控制

一款真正的工业解决方案需能支持更宽泛的工作温度范围、严苛的工作环境以及更长的生命周期，这样才能放心用于不容停机的应用中。



对于 AI 赋能的边缘应用而言，自适应计算是最具前景的支持技术之一。自适应计算采用的硬件，如现场可编程门阵列 (FPGA)，能针对具体应用进行高度优化。除了 FPGA 以外，近期还有一些新型自适应硬件推出，包括含有 FPGA 架构的自适应系统级芯片 (SoC)，它结合了一个或多个嵌入式 CPU 子系统。

然而，自适应计算绝不只是硬件。它还涵盖了一套综合全面的设计与运行时 (Run-time) 软件集。这些软件结合起来，造就了一款独特的自适应平台，在此之上可以开发高度灵活且高效的系统。

**自适应计算允许为应用专门开发硬件，
而且如果工作负载或标准发生演进，
仍能根据需要灵活适应。**

借助自适应计算，无需 ASIC 等定制芯片器件涉及的设计时间和前期成本就能实现 DSA。这便于为任何给定领域迅速开发灵活的、经优化的解决方案，包括 AI 赋能的边缘应用。

自适应 SoC 是此类特定领域处理的理想选择，因其结合了全面的嵌入式 CPU 子系统所具备的灵活性，以及自适应硬件的最佳数据处理能力。



自适应计算允许为应用专门开发硬件，而且如果工作负载或标准发生演进，仍能根据需要灵活适应。

第 5 章

自适应系统模块 (SOM)



SOM 用例： 访问控制

AI 赋能的边缘应用（如访问控制）需要高性能、低时延的实现方案，但必须维持在限定的功耗和占板面积要求之内。

正如我们所看到的，SOM 是一种面向边缘应用的理想平台。然而，要想达到现代 AI 赋能的应用所需性能，则需要进行加速。而自适应计算恰能为边缘端 AI 应用提供所需的加速。

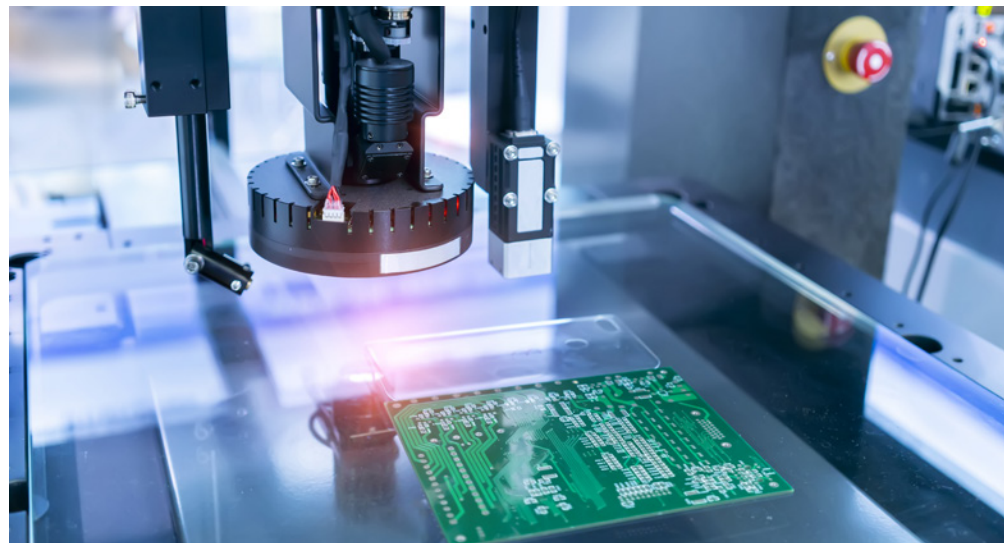
典型的应用开发团队由各类硬件工程师和软件工程师组成。电路板设计师负责为所需的应用设计定制电路板。如果需要使用自适应计算（包括 FPGA），通常由 RTL 设计师负责利用 Verilog 和 VHDL 等硬件描述语言 (HDL) 配置自适应器件。软件开发者则使用 C++ 等语言和常用于嵌入式视觉应用的 OpenCV 等框架编写运行于嵌入式 CPU 上的代码。

第 5 章：自适应系统模块 (SOM)

传统上，三个专业领域都需使用定制电路板来开发应用，让定制硬件位于自适应 SoC 内部、软件运行于嵌入式 CPU 上。

如前文所述，设计、制造含有应用所需的自适应 SoC 和其他物理组件的定制电路板的做法称为“芯片先定”开发方法。几十年来，电路板设计师始终采用“芯片先定”开发方法来发挥自适应计算的优势。实际上，自适应 SoC 已通过这种方法部署在众多 AI 赋能的边缘应用上，其中包括数千万辆汽车上。

部分应用仍需定制硬件组件才能与自适应 SoC 进行接口连接，因而需要“芯片先定”设计。不过，越来越多 AI 赋能的边缘应用需要类似的硬件组件和接口，即便对于



大相径庭的最终应用而言，也是如此。随着行业向标准化接口和通信协议发展，同样一套组件能适用于各种类型的应用，尽管它们的处理需求相差甚远。

第 5 章：自适应系统模块 (SOM)

面向 AI 赋能的边缘应用的自适应 SOM，融合了自适应 SoC 以及行业标准的接口和组件，使得开发者即便不具备或仅略微掌握硬件知识，也受益于自适应计算技术。自适应 SoC 既能执行 AI 处理也能执行非 AI 处理，从而实现整体应用。

自适应 SOM 上的自适应 SoC 无需定制 PCB 即可实现高度定制化。其设计旨在集成到更大的系统中，且使用预定义的外形尺寸。

借助自适应 SOM，无需进行“芯片先定”设计，就能充分发挥自适应计算的优势。尽管如此，自适应 SOM 也只是解决方案的一部分，因为软件也是一项关键考量因素。我们将在第 7 章介绍软件考量因素。

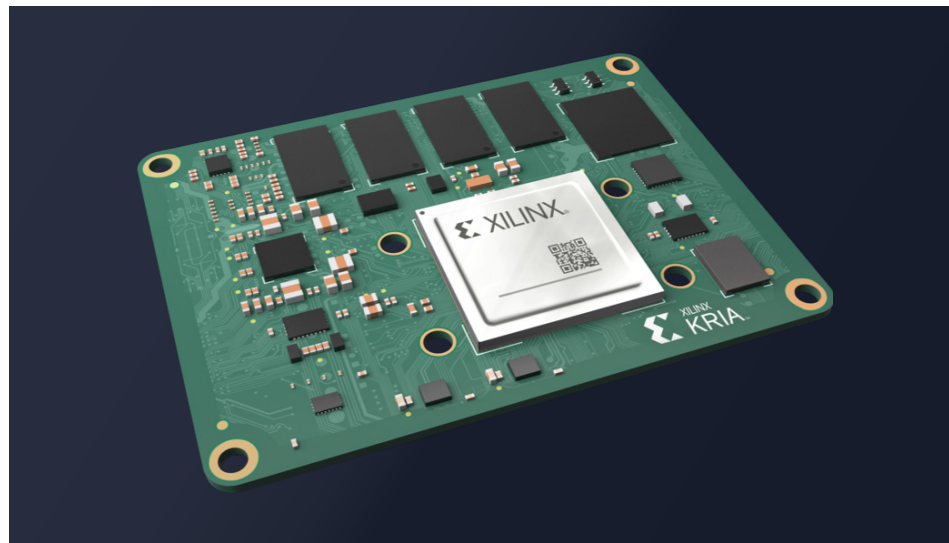


图 4：自适应 SOM 示例

第 6 章

自适应 SOM 对硬件 开发者的益处

SOM 用例： 自动导引车

为了提高生产力并将停机时间降至最短，物流公司需要最先进的 AI 模型，且必须以紧凑、低功耗且坚固耐用的解决方案进行交付。



通常情况下，硬件开发者在构建 AI 赋能的边缘硬件时会选定主要芯片器件，然后开发边缘外形尺寸电路板，以容纳所选器件。正如前面讨论过的，这种“芯片先定”开发方式成本高昂、流程复杂且开发周期漫长。

首先，开发者需要评估和选定器件、构建原型，然后验证该架构能与所需的软件和计划部署的 AI 模型充分兼容。仅评估过程就要耗费数月时间。一旦评估完成，则需开发、集成和制造成品电路板，然后才能进行部署。



图 5：SOM 可助力加快市场投放速度，并降低开发成本

自适应 SOM 支持开发者使用现成的、可量产化的解决方案，从而省去大量的开发成本和时间。上图显示，与典型“芯片先定”方式相比，基于 SOM 的设计可助力嵌入式视觉应用提前 9 个月完成。

自适应 SOM 对于硬件开发者大有裨益，因为它支持在设计流程后期修改设计，而基于固定芯片技术的 SOM 则没有这样的优势。

第 7 章

自适应 SOM 对软件 开发者的益处

SOM 用例：外科
手术机器人

工业控制、通信、机器视觉、机器学习、
人机界面、网络安全与安全性
均为关键技术考量因素。



```
for (type in types) {  
  // types is a map of types/handlers  
  if (types[type] === "object") {  
    // (types-Object, selector, data)  
    if (typeof selector !== "string") {  
      // (types-Object, data)  
      data = data || selector;  
      selector = undefined;  
    }  
    for (type in types) {  
      on( elem, type, selector, data );  
    }  
    return elem;  
  }  
}
```

自适应 SOM 的优势不仅限于硬件开发者。通过采用针对底层自适应 SoC 的预构建配置，软件开发者也能加快其设计周期。

有鉴于软件工具、库和框架的近期发展，一些设计团队已能在无需硬件工程师介入的情况下利用自适应计算。现有综合软件平台使得软件工程师无需进行专门的硬件定制，也能使用整个自适应 SoC 的功能。

第 7 章：自适应 SOM 对软件开发者的益处

加速 API 和预构建加速软件平台等解决方案可在自适应 SoC 的供应商网站上免费寻获，也可从第三方软件厂商处获取。两种方式都能获得可执行的通用 AI 推断功能一系列 AI 模型。

对于软件开发者而言，自适应 SOM 可以在其熟悉的 Python、C++、TensorFlow 和 PyTorch 等环境下，为其提供易于使用、开箱即用的体验。

除了预构建平台和 API，综合全面的软件工具还能实现自适应硬件的完全定制，从而进一步提升灵活性和优化水平。

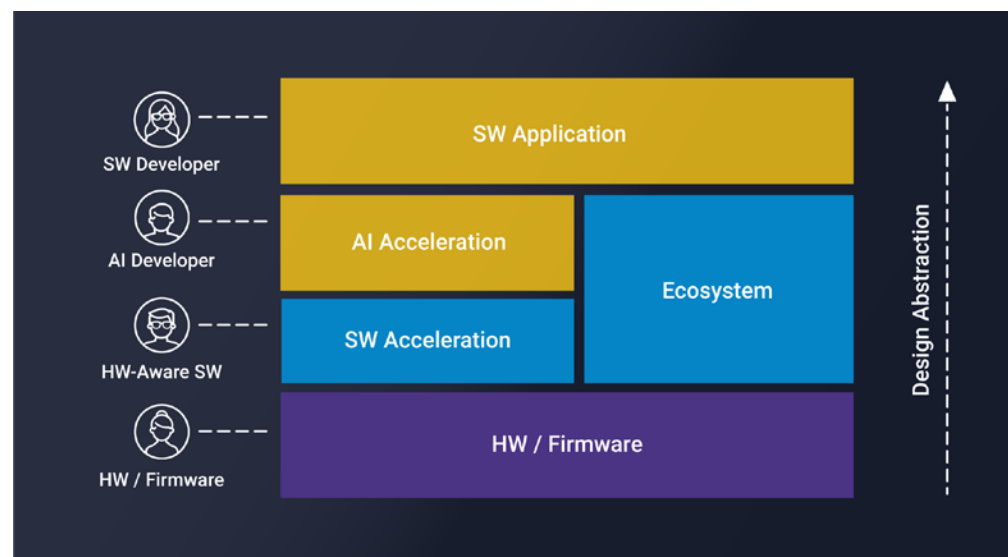


图6：自适应 SOM 可以在多个抽象层级进行编程

第 8 章

选择边缘解决方案时的 考量因素

SOM 用例： 电动机控制风力涡轮机

无论地处世界何处、陆地抑或海上，
在几十年的工作周期内，风力涡轮机
都需要远程、安全地进行升级



为边缘应用评估解决方案时，应考虑多个因素：

开发与成本

成功的设计团队会寻求一款可以缩短开发时间、快速制作原型的交钥匙解决方案。而该解决方案不仅在于硬件，还应当包含操作系统、板卡级支持包、参考设计、开放的系统导向文档，以及易于理解的说明和指导。借助预

构建的入门套件，开发者可以在不改变硬件和软件的前提下，即刻投身到开发能直接投产的应用中去，从而进一步缩短设计周期。而降低开发成本则尤其有利于中小型企业从自适应计算中获益。

性能功耗比

该解决方案应提供最佳性能功耗比。对于通常运行在不超过 10 瓦的预定功耗下的边缘应用而言，这一点尤为重要。这款解决方案应提供配有库且能权衡性能与功耗的特定领域架构，以满足特定应用的需求。低时延下的性能未必始终与高时延下的性能相当。例如，为了达到更高的吞吐量，GPU 通常会对数据进行批处理，但会以牺牲时延为代价。低时延下的吞吐量往往达不到峰值吞吐量。

安全性

该解决方案应能支持 IEC 62443 级网络安全，即工业物联网的安全标准。这意味着系统至少应建立在硬件征信根系统 (Root-of-Trust) 上，且通过软件应用的远程认证支持安全、合规的开机。此外，安全解决方案还应当能适应边缘设备的工业生命周期（通常为 10 到 20 年），以应对同一时期内不断演进的威胁。

面向未来

除了安全性，该平台还应具备未来兼容的能力，这样才能支持软件与硬件更新，即便在系统部署到实地后也能进行空中 (OTA) 更新。该解决方案必须满足工业级要求，生命周期至少达到 10 年。鉴于与安全性和工业生命周期类似的原因，该解决方案还应该能适应不断变化的 AI 模型，因为近年来 AI 方法层出不穷，并且还将在未来多年继续快速演进。

生态系统

该解决方案应拥有规模可观的生态系统，从而提供各种类型的服务和解决方案，包括预先与选定 AI 平台或替代性 AI 平台预集成的补充硬件、软件以及工具。同时，该解决方案应涵盖加速应用库，使得软件开发也能迅速驾驭设计，就如同参考设计对硬件开发者所起的作用一样。

自适应 SOM 既能提供基于 SOM 的解决方案的上市时间优势，又具备自适应计算才能实现的应用优化能力。这使之成为实现 AI 赋能的边缘应用的理想方式。



“我们近来已看到机器学习和 AI 应用对 SOM 的大量关注。许多客户不想置身于具体的复杂设计中。SOM 能帮助他们快速进入市场并节省开发成本。”

– Immanuel Rathinam,
副总监兼研发与业务运营主管;
iWave Systems Technologies Pvt Ltd.



第 9 章

总结

SOM 用例：
自动化多层泊车系统

可靠性在许多应用中都至关重要，
停运将对业务运行造成重大冲击



复杂的 AI 赋能的工作负载正日渐向边缘迁移。这些应用需要以低时延、低功耗和小尺寸执行大量的处理。为了满足这一要求，必须对整体应用进行加速（同时包括 AI 功能和非 AI 功能）。

随着 AI 模型快速演进，加速平台也必须具备灵活应变的能力，这样才能取得当今乃至未来 AI 技术的最佳实现。

SOM 提供了一款理想的边缘处理平台。与自适应 SoC 结合使用时，所产生的自适应 SOM 可为 AI 赋能的边缘应用提供一个全面的、可量产化的平台。

企业采用自适应 SOM 后，将从性能、灵活性及快速开发的独特组合中受益。企业无需构建自己的电路板，就能获得自适应计算所带来的效益。在赛灵思近期推出 Kria™ 自适应 SOM 产品组合以前，这是无法在边缘端实现的。

第 9 章：总结

Kria 自适应 SOM 围绕赛灵思 Zynq® UltraScale+™ MPSoC 架构进行构建，为开发者提供了一个交钥匙自适应计算平台。通过标准化系统的核心组成部分，开发者可以有更多时间聚焦于功能特性，使其技术区别于竞争对手。

欲进一步了解自适应 SOM 如何支持您的应用，请访问赛灵思自适应 SOM 页面：china.xilinx.com/kria

了解更多详情：

[借助 Kria SOM 实现嵌入式设计简化](#)

[Kria K26 SOM：边缘视觉 AI 的理想平台](#)

关于赛灵思：

赛灵思提供自适应平台。我们的自适应 SoC、FPGA、加速器卡和系统级模块赋予领先企业创新和快速部署的自由。我们与我们的客户紧密合作，从云端到边缘打造可扩展、差异化且智能化的解决方案。在这个创新节奏不断加快的世界中，越来越多的创新者相信赛灵思能帮助他们以最佳的效率和性能，更快将产品投放市场。如需了解更多信息，敬请访问赛灵思中文网站：china.xilinx.com。



图 7：赛灵思 Kria KV260 视觉 AI 入门套件

进一步了解： 赛灵思 Kria 自适应 SOM

欢迎与我们联系！
sales@xilinx.com

公司总部
Xilinx, Inc.
2100 Logic Drive
San Jose, CA 95124
USA
Tel: 408-559-7778
www.xilinx.com

赛灵思欧洲
Bianconi Avenue
Citywest Business Campus
Saggart, County Dublin
Ireland
Tel: +353-1-464-0311
www.xilinx.com

日本
Xilinx K.K.
Art Village Osaki Central Tower 4F
1-2-2 Osaki, Shinagawa-ku
Tokyo 141-0032 Japan
Tel: +81-3-6744-7777
japan.xilinx.com

Asia Pacific Pte. Ltd.
赛灵思亚太地区
5 Changi Business Park
Singapore 486040
Tel: +65-6407-3000
www.xilinx.com

印度
Xilinx India Technology Services Pvt. Ltd.
Block A, B, C, 8th & 13th floors,
Meenakshi Tech Park, Survey No. 39
Gachibowli(V), Seri Lingampally (M),
Hyderabad -500 084
Tel: +91-40-6721-4747
www.xilinx.com

© Copyright 2021 年赛灵思公司版权所有。Xilinx、赛灵思标识、Artix、ISE、Kintex、Kria、Spartan、Versal、Virtex、Vitis、Vivado、Zynq 及本文提到的其它指定品牌均为赛灵思在美国及其它国家的商标。AMBA、AMBA Designer、ARM、ARM1176 JZ-S、CoreSight、Cortex、PrimeCell 均属于 ARM 在欧盟和其它国家或地区的商标。“PCIe”和“PCI Express”均为 PCI-SIG 拥有的商标，且经授权使用。所有其它商标均是各自所有者的财产。

在美国印刷 WW4/9/21

