

瞬变对AI加速卡供电的影响

Hamed Sanogo, 终端市场专家

摘要

图形处理单元(GPU)、张量处理单元(TPU)和其他类型的专用集成电路(ASIC)通过提供并行处理能力来实现高性能计算,以满足加速人工智能(AI)训练和推理工作负载的需求。

AI需要大量的算力,尤其是在学习和推理时。这种需求不断地将供电网络的边界推向前所未有的新水平。这些高密度工作负载变得愈加复杂,更高的瞬态需求推动配电网络的每个部分都必须高效运行。AI加速卡严格的功耗要求对系统性能也有影响。本文将讨论AI加速卡的配电网络要求,剖析瞬变的影响,并介绍ADI公司针对这些需求提出的多相供电解决方案。

简介

AI技术完全改变了计算架构,以复现模仿人脑的神经网络。AI看似已广泛存在,但实际上,驱动AI的技术仍在发展。专门用于AI计算的处理器加速器IC包括GPU、现场可编程门阵列(FPGA)、TPU和其他类型的ASIC。本文将它们统称为xPU。

随着AI技术部署快速推进,数据中心将继续批量购买AI加速卡。根据Gartner的报告,2021年AI芯片收入总计超过340亿美元,预计到2026年将增长至860亿美元。¹xPU采用大规模并行计算方案,与普通CPU相比,在AI性能方面实现了巨大飞跃。xPU拥有大量小内核,因此非常适合AI工作负载,有助于神经网络训练和AI推理。然而,xPU进行AI计算和移动数据通常会产生相对较大的功耗。简而言之,xPU是非常耗电的IC。其严格的功耗要求对AI加速卡提出了新的挑战,这也会影响系统性能。本文将分析AI加速卡的供电网络要求,并介绍ADI公司针对这些严格要求提出的多相供电解决方案。

AI带来的供电挑战

AI涉及许多方面,但能效不在其中。AI工作时,尤其是处理深度学习和推理等AI工作负载时,需要极高的计算功率。在系统层面,AI加速器对于提供近乎即时的结果(正是这些结果使其有价值)发挥着关键作用。所有xPU都有多个高端内核,这些内核由数十亿个晶体管构成,消耗数百安培电流。这些xPU的内核电压(V_{CORE})已降至低于1.0 V的水平。图1显示了AI加速卡的通用框图。本文将重点介绍为此类系统提出的多相控制器和相应的功率级IC。

AI加速卡所需的峰值电流密度对于任何主板来说都是非常沉重的负担,难以处理。工作负载的高度动态特性和极高的电流瞬变会导致非常高的 di/dt 和持续数微秒的尖峰电压瞬变,这些瞬变非常具有破坏性,可能对xPU造成损害。AI的平均工作负载会持续很长时间,解耦电容将无法始终提供满足即时需求的能量。本文的下一部分将介绍ADI公司提出的多相负载点(PoL)解决方案,它会消除典型AI加速器的瞬变,避免给整个配电网络产生压力。但首先,我们来讨论AI带来的电源设计挑战。

AI带来新的电源设计挑战

目前,AI功率需求远远超过传统供电网络的能力。xPU稳压器(VR)的要求与标准PoL稳压器有很大不同。业界发现,某些应用要求在小于1 V的电压下为xPU提供超过1000 A的电流。重要的是,电源必须非常稳定,产生的噪声非常小,同时消除所有电压瞬变可能性,以免导致xPU内部误触发。为了应对惊人的电流需求,高性能AI加速器VR PoL的设计必须满足某些关键要求。

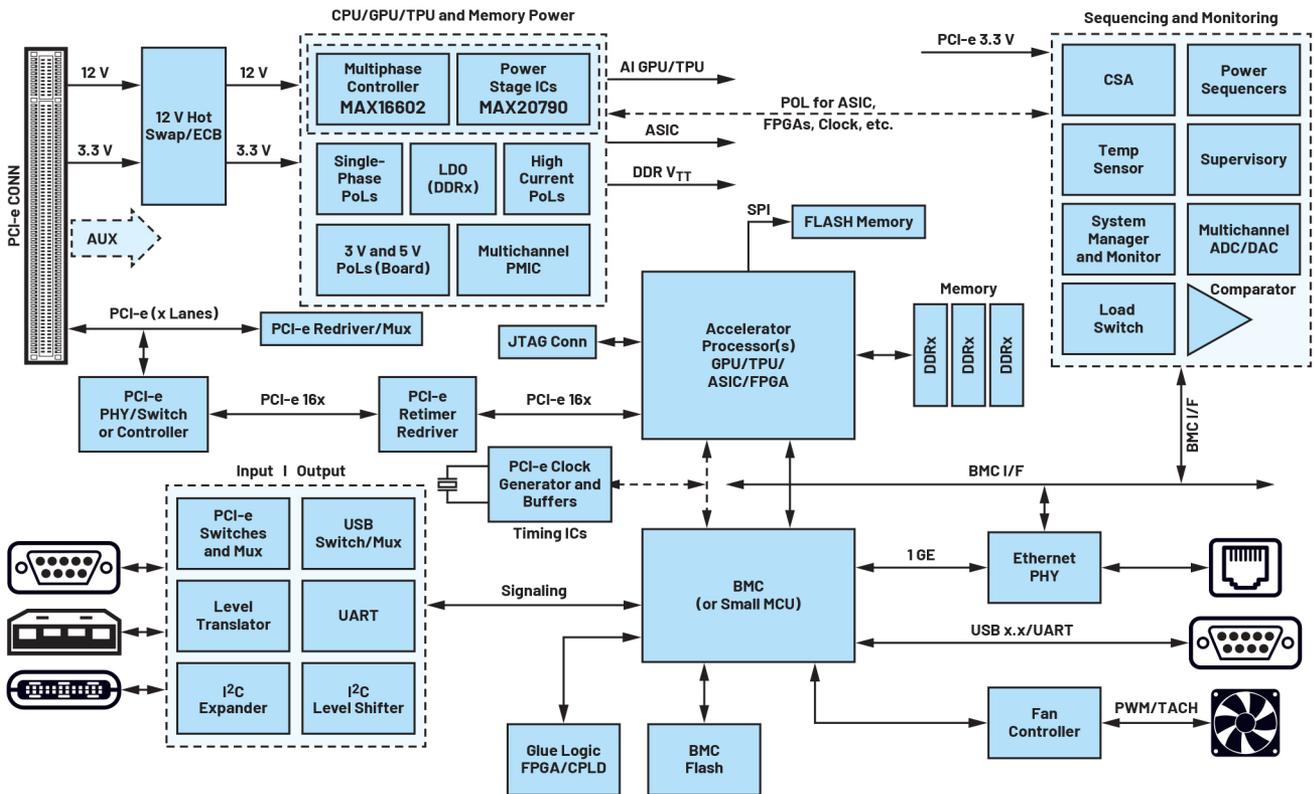


图1. 通用AI加速卡框图

电压尖峰和瞬变管理

AI加速卡的关键要求之一是VR的架构应能提供出色的瞬变电压管理。向任何系统提供千瓦级功率始终是首要挑战。输出电压（包括容差、纹波以及负载瞬态骤降和峰值）必须始终高于xPU最小电压以避免系统挂起，并且还必须始终低于xPU最大电压以免损坏xPU。加速卡的瞬态功率尖峰可能要求达到最大热功率目标的2倍甚至更高。

这里重要的是，PoL环路带宽须足够灵活，以处理所遇到的各类更快速瞬变。带宽越高，环路响应越快，电压偏差越小。实现快速瞬态电源轨较直接的一种方法是选择具有快速瞬态性能的稳压器。ADI AI V_{CORE}系列IC具有非常低的频率输出噪声、快速瞬态响应和高效率等特性。除此之外，ADI AI电源芯片组还支持负载线路，有助于电源设计人员有效管理AI工作负载引起的瞬变和尖峰。

长电源路径走线中的I²R损耗和热管理

随着AI xPU处理器电流不断提高，PoL供电解决方案的密度已成为关键要素。既要可靠地向xPU的每个部分供电，同时不用担心散发的热量会影响芯片的可靠性并导致热失控，现在变得极其困难。换言之，热管理是设计这种高功率电源所面临的重大挑战之一。传统的供电方法是将稳压器放置在xPU的一侧，以便将电力横向传输到处理器。这些走线的电阻哪怕再小，也可能引

起不可接受的电压(I²R)下降。PCB电源层电阻上的压降会随着xPU电流提高而成比例地增加。这意味着VR和BGA引脚之间几厘米的PCB电源走线会产生大量的损耗。PCB铜电源层中的此类损耗已成为计算稳压器设计效率和性能的主导因素。传统3芯片（分立式）供电解决方案需要大量高电流走线，与之相比，使用集成了电流和温度电路模块的单芯片功率级IC，可以大大减少PCB上的走线数量。

ADI价值主张：MAX16602 + MAX20790 + 耦合电感

AI稳压器的精度变得更加严格。效率和尺寸是重中之重。性能和功耗也受到严格审查。正如上一节所述，解决AI加速卡VR设计问题已成为一项艰巨的任务。设计人员非常清楚，若不能有效处理不必要的瞬态效应，就无法在所需电流中产生大的阶跃。解决这些瞬态效应还需要某种类型的高精度动态电压定位或负载线路方案。ADI公司大力投资AI市场，为48 V和12 V系统提供全套解决方案。本节介绍ADI AI多相电源芯片组，即MAX16602多相控制器和MAX20790功率级，以及我们获得专利的耦合电感(CL)技术，以帮助解决这些AI PoL设计挑战。图2显示了8相MAX16602CL8_EV设计的MAX16602、MAX20790和CL简化框图连接。这种相对简洁的设计实现了每相约88 A_{PK}的高电流传输能力。内部补偿和先进的控制算法，加上功率级中集成的电流检测电路以及耦合电感，使其成为拥有出色效率的小尺寸解决方案。

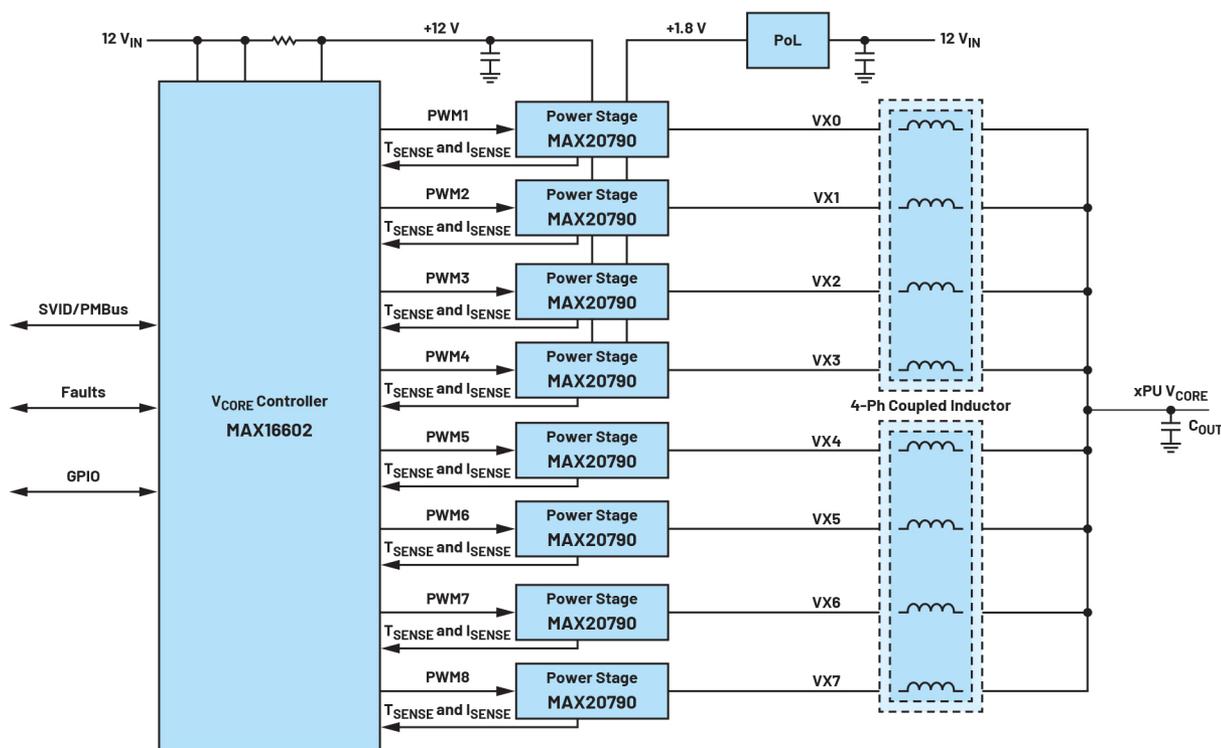


图2. 采用ADI高集成度电源芯片组的8相VR设计有助于实现高密度设计，同时减少外部连接

更高集成度的单芯片智能功率级IC

MAX20790是一款功能丰富的智能功率级IC，旨在与MAX16602（以及该产品系列中的其他几款ADI控制器）配合使用，以实现高密度多相稳压器。这是一种单芯片集成，几乎消除了分立式设计中常见的FET和驱动器之间的寄生电阻和电感，从而实现高开关速度，而且功率损耗明显低于传统方案。如果检测到开关节点(V_s)故障，功率级会立即关闭，并将故障ID传送给控制器。该智能功率级IC还有一个片内电流传感器。此电流检测电路模块显然优于使用电感直流电阻的方法。众所周知，DCR检测不准确，需要温度补偿才能使电流测量结果可信。

控制器IC

MAX16602是一款用于xPU V_{CORE} VR的多相控制器。该IC提供高密度、灵活且可扩展的解决方案，可为AI xPU供电。该器件支持脉冲宽度调制(PWM)并联，可控制多达16个相位。该IC的架构简化了设计，减少了组件数量，支持高级电源管理和遥测功能，并在整个负载范围内提高了节能效果。它实现了自主切相，在整个负载范围内保持高效率。完整芯片组是一个高效率多相降压转换器，具有广泛的状态和参数测量特性。保护和关断参数通过串行PMBus®接口进行设置和监测，甚至包括功率级IC中收集的故障。

以下是该ADI控制器支持的其他几个关键特性，这些特性对于任何AI供电方案都很重要。

高级调制方案

MAX16602利用高级调制方案(AMS)来提供更好的瞬态响应。该调制方案支持以极短的延迟开启和关闭相位。根据负载需求，当负载增加时，可以同时开启多个相位；当负载释放时，可以立即关闭多个相位。启用AMS后，系统闭环带宽可以扩展，而不会造成相位裕量损失。因此，PoL能够更好地响应AI VR的即时和动态电流需求。

负载线路控制

负载线路允许 V_{CORE} 根据输出电流在最小值和最大值之间变换。它实质上是为轻负载设置高 V_{CORE} 值，为重负载设置低值。主要是为了让控制环路可以处理更高的负载电流（这是让计算顺利进行所必需的）。ADI控制器在整个输出电流范围内提供准确的输出负载线路控制。输出电压定位利用来自功率级IC的无损电流检测信号进行，这些信号会反馈到控制器。负载线路是在控制器中通过对电压控制环路误差放大器的直流增益进行数字编程来设置。控制器的EC表和数据手册的表6中提供了各种直流负载线路特性，从0.105 mΩ到0.979 mΩ。图3显示了16相PoL设计在40 A至360 A负载阶跃和800 A/μs摆率下的瞬态曲线。结果表明过冲极小。

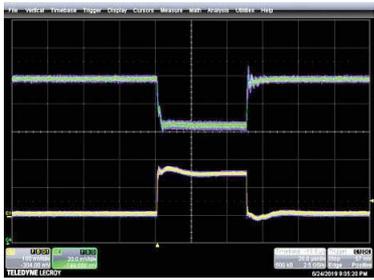


图3. 16相VR在40 A至360 A阶跃负载和800 A/ μ s摆率下的瞬态曲线

设计中添加有源电压定位可以降低对负载瞬态响应的要求，并更好地利用xPU总容差窗口。负载线路控制有助于降低给定阶跃负载的峰峰值输出电压偏差，同时可以减少输出轨上的bulk电容量。总电压波动将会减小，从而降低xPU崩溃或损坏的风险。请注意，MAX16602中的负载线路电路模块可以禁用。

耦合电感(CL)的优势

十多年来，ADI公司一直投资开发其专利CL技术。这项技术支持实现更高的密度、更大的带宽、更快的瞬变解决方案，与分立式实现方案相比，效率提高50%，磁性元件尺寸缩小1.82倍。CL能够有效地在稳态中用作大电感，在瞬态中用作小电感，除了减小电感尺寸外，还能节省 C_{OUT} 。²图4显示了ADI多相VR设计中常用的耦合电感系列。



图4. ADI多相VR设计常用的耦合电感系列

根据设计规格和优先级，耦合电感消除电流纹波的优势可用来换取更小的尺寸或更高的效率。¹较大的系统优势以及ADI产品的与众不同之处在于，AI PoL设计人员可以使用CL相对轻松地实现VR总尺寸较小的解决方案。几家知名磁性器件供应商拥有ADI的免费CL许可，可以为我们提供所需的元件。

顶部散热封装

顶部散热为表面贴装封装提供了另一种散热途径。MAX16602和MAX20790都是倒装芯片四方扁平无引线(FCQFN)封装，带裸露的顶部散热焊盘。FCQFN是一种先进的封装，可提供设计人员青睐的出色热性能。这种无引线封装不仅可以减少寄生电感，还能从器件的结直接向周围环境散热。MAX20790的结壳顶部(θ_{JC-TOP})热阻为0.25°C/W。AI电源设计利用顶端散热配置，可以提高系统的热性能和设计灵活性。

总而言之，ADI的多相功率转换和PoL产品提供高效率和高功率密度。图5显示了我们的16相MAX16602 + MAX20790 + CLH1110-4评估板的效率曲线以及偏置和电感损耗。ADI公司为各种AI加速器应用提供稳压器和其他电源转换解决方案。采用我们的多相控制器和集成功率级解决方案，有助于ADI客户满足严苛的动态xPU电源要求，应对当今AI应用带来的设计挑战。

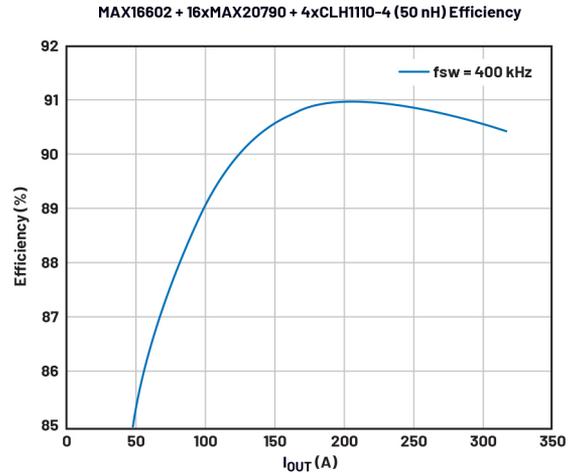


图5. 16相AI VR评估板设计的效率曲线

垂直供电

随着处理复杂AI功能的xPU问世，功耗随之急剧增加。具有高达650 A连续电流和超过1000 A峰值电流传输能力的VR开始普及。为AI处理器供电的挑战在于保持高效率。常规电源架构无法跟上这些非常耗电的AI xPU的步伐。VR芯片制造商和架构师正在从根本上研究不同的供电方法。业界正在讨论一种为AI xPU供电的新趋势，称为垂直供电，也称为背面供电。

VR必须尽可能靠近负载输入xPU电源引脚，以实现高电流输送。我们无法通过传统的横向供电方法实现这一目标。垂直供电将电源调节器移到处理器正下方，从而消除了PCB上可能产生的所有损耗。该结构将电源转换器、功率级、电容和磁性元件放置在PCB的背面，并通过过孔垂直地向xPU供电。换言之，电流传输是从xPU BGA阵列下方垂直进行。这是一条长度缩短的垂直路径，可显著降低阻抗并消除损耗。图6显示了安装在PCB另一侧、xPU下方的垂直供电模块架构。此示意图仅用于说明。

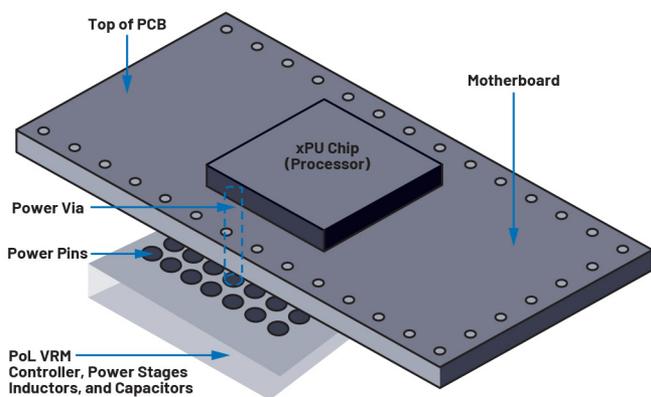


图6. 垂直供电模块架构 (仅用于说明目的)

构建垂直供电解决方案的难点包括解决模块的重量和安装问题。PCB另一侧的xPU下方比较适合放置高频解耦电容，用于储存能量以满足瞬时能量需求。垂直供电与ADI的CL技术相结合，可实现更高的电流密度、功率密度和更快的瞬态性能。垂直供电为ADI等PoL制造商提供了新的创新机会，并以自己的方式继续支持摩尔定律的发展。

结语

支持机器学习和深度学习的加速卡通过提供加速训练和推理工作负载所需的并行处理能力，将AI从理论变为现实。为高性能AI加速卡设计VR PoL是一项复杂的任务，尤其是在当前先进xPU的电源要求（表现在电流水平和电压精度方面）不断提高的情况下。

本文表明，xPU VR的要求与标准PoL调节器有很大不同。xPU供电轨具有极快的负载变化，需要动态电压定位或负载线路，并且必须很小。

ADI公司拥有广泛的AI xPU V_{CORE}解决方案系列，用于解决当今的这些问题。我们的电源解决方案能够以非常小的外形尺寸实现出色的效率。本文介绍的解决方案将多相控制器MAX16602和智能单芯片功率级MAX20790相结合，可提供非常高的电源转换效率、非常快的瞬态响应和非常准确的遥测报告。如需了解这些电源芯片组的更多信息或购买MAX16602CL8评估套件，请单击[此处](#)。

参考文献

¹ “[预测：2021-2027全球AI半导体](#)。” Gartner，2023年4月。

² “[利用耦合电感的优势](#)。” ADI公司。

作者简介

Hamed M. Sanogo是ADI公司全球应用部门的云和通信终端市场专家。Hamed拥有密歇根大学迪尔本分校的电子工程硕士学位，之后还获得了达拉斯大学的工商管理硕士学位。在加入ADI公司之前，毕业后的Hamed曾在通用汽车担任高级设计工程师，并在摩托罗拉系统担任过高级电气工程师以及Node B和RRH基带卡设计师。在过去的17年里，Hamed担任过不同的职务，包括FAE/FAE经理、产品线经理，目前是通信和云终端市场专家。

在线支持社区



访问ADI在线支持社区，中文技术论坛
与ADI技术专家互动。提出您的棘手设计问题、浏览常见问题解答，或参与讨论。

请访问ez.analog.com/cn

