# Cheshire：面向领域专用加速器即插即用的轻量级、支持 Linux 的 RISC-V 主机平台

Alessandro Ottaviano，*IEEE* 学生会员，Thomas Benz，*IEEE* 学生会员，Paul Scheffler，*IEEE* 学生会员，以及 Luca Benini，*IEEE* 会士

arXiv:2305.04760v2 [cs.AR] 6 Jul 2023

摘要——**物联网极端边缘和微型机器学习领域的功耗与成本限制，加之不断提高的性能需求，推动了异构架构的发展趋势。这类设计采用高能效的应用级主机处理器来协调计算专用的多核加速器，从而摊薄操作系统支持和外部通信所带来的架构成本。本文介绍*Cheshire*，这是一种轻量、模块化、支持 64 位 Linux 的主机平台，专为领域专用加速器的无缝即插即用而设计。其特色包括独特的低引脚数 DRAM 接口、可配置为暂存存储器的末级缓存，以及能够在加速器与 DRAM 之间高效搬运数据的 DMA 引擎。此外，它还提供多种可选 I/O 外设，包括通用异步收发器、串行外设接口、集成电路总线、视频图形阵列和通用输入输出端口。Cheshire 的可综合寄存器传输级描述包含其全部外设及全数字 DRAM 接口，并以免费开源形式提供。我们采用台积电 65 nm 互补金属氧化物半导体工艺实现并制造了 Cheshire，制成名为*Neo* 的硅验证芯片。在 1.2 V 电压下，Neo 的时钟频率最高可达 325 MHz，在数据密集型计算负载下总功耗不超过 300 mW。其 RPC DRAM 接口仅消耗 250 pJ/B，物理层面积仅为 3.5 kGE，并在 200 MHz 下实现了 750 MB/s 的峰值传输速率。**

索引词—**RISC-V、Linux、物理层、存储器控制器、领域专用架构、异构计算、超大规模集成电路**

## I. 引言

随着 Koomey 定律[1] [1] 逐渐放缓，计算机架构师转向硬件专用化，以满足机器学习（ML）和近传感器处理等数据密集型应用日益增长的能效需求。如今的异构架构将传统的应用级（AC）主机处理器与领域专用架构（DSA）或可编程众核加速器（PMCA） [2]–[5],相结合，在提高能效的同时，保持较强的可编程性并支持通用操作系统（GPOS）。因此，为了最大限度地提升异构系统的计算效率，架构师应尽可能增加分配给高效加速器的面积和能量资源，并尽量降低 AC 主机的成本与能耗。

异构架构已在移动和高性能应用中得到确立：经过硅验证的工业 [6]–[8]和学术界 [9], [10]片上系统（SoC）通常面向较大的功耗预算或多个 AC 核心。然而，异构性如今正渗透到物联网（IoT）的极端边缘和微型机器学习领域，这些领域对功耗和成本的限制极为严格。这些领域中的异构 SoC [11], [12] 通常使用轻量级 32 位微控制器作为管理处理器，但不具备 AC 能力，也不支持 GPOS。近期的 *HULK-V* [13] SoC 通过将一个支持 Linux 的 64 位 RISC-V 管理核心与一个用于高端微型机器学习应用的八核 PMCA 耦合，展示了功能更强大的主机。不过，它尚未经过硅验证，并且其 PMCA 的临时集成方式无法提供可复用的接口。此外，HULK-V 的外部 HyperRAM [14]内存接口带宽较低，限制了可扩展性，并可能导致主机与加速器之间出现内存瓶颈。

为解决这些不足，我们提出了*Cheshire*，这是一款节能、支持 Linux 的 64 位 RISC-V 主机平台，可无缝集成 [15]–[16] 等领域专用加速器（DSA）。Cheshire 采用完全模块化设计，提供可配置的互连、多种可选外设，以及一个直接存储器访问（DMA）引擎，以解耦主机与 DSA 之间的通信。它集成了首个可综合的开源减引脚数（RPC）DRAM 接口。RPC 是一种新近推出的低引脚数 DRAM 方案，相较于低功耗双倍数据速率（LPDDR）DRAM，其集成成本和工作量更低，同时在能效相当的情况下，带宽显著高于 HyperRAM。Cheshire 的可综合寄存器传输级（RTL）描述和 FPGA 实现均已开源[2]。我们的贡献如下：

- 一个精简、可定制且节能的 Linux 支持型 RV64GC 主机平台，可轻松与片上或芯粒 DSA 集成；我们提供可配置的高级可扩展接口 4（AXI4）互连和数字芯粒间（D2D）接口。
- 首个全数字、与工艺无关且兼容 RPC- DRAM 的存储器接口，仅需 22 个切换 IO，PHY 面积为 3.5kGE。该接口的存储器访问能耗仅为 250pJ/B，在 200MHz 下峰值传输速率可达 750MB/s，性能优于现有 HyperRAM 解决方案。据我们所知，这是首次在流片的开源 SoC 中对减引脚数 DRAM（RPC DRAM）进行表征。
- 一种灵活的存储器系统，可在仅 8 个周期内完成对外部 RPC DRAM 的访问，传输量为 32B；且仅在需要时才与 DSA 存储空间交互。
- 一款名为*Neo*的独立演示芯片，采用台积电 65 纳米工艺节点制造。在 1.2 V 电压下，Neo 的时钟频率最高可达 325 MHz，在数据密集型计算工作负载下的功耗低于 300 mW。

A. Ottaviano、T. Benz 和 P. Scheffler 对本研究的贡献相同。A. Ottaviano、T. Benz、P. Scheffler 和 L. Benini 均就职于瑞士苏黎世联邦理工学院集成系统实验室（IIS）。电子邮箱： {aottaviano,tbenz,paulsc,lbenini}@iis.ethz.ch。L. Benini 还任职于意大利博洛尼亚大学电气、电子与信息工程系（DEI）。[1]从 20 世纪 40 年代到 21 世纪初的观测结果表明，Koomey 定律指出，计算机的能源效率每 1.57 年翻一番。

[2]Cheshire 的网址为 https://github.com/pulp-platform/cheshire，其 RPC DRAM 接口的网址为 https://github. com/pulp-platform/rpc dram controller。

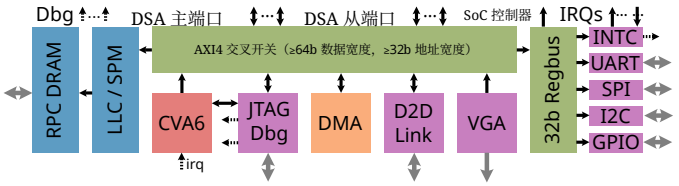


**图 1：** Cheshire 平台的架构。AXI4 交叉开关为连接至 DSA 提供数量可配置的**管理**端口和**从属**端口。

## II. 架构

在下一节中，我们将介绍 Cheshire 的整体架构，随后深入讨论全数字、技术无关的 RPC DRAM 接口。

### *A. Cheshire* 平台

Cheshire 基于高能效的 64 位 CVA6 [17]AC 处理器。它包含自主启动并运行 Linux 等通用操作系统所需的全部硬件，例如符合 RISC-V 标准的核心本地及平台中断控制器、用于访问外部存储和外设的多种可选标准 IO 接口，以及 DRAM 接口。片外 DRAM 至关重要，因为简单嵌入式 Linux 系统的 8–16 MB 内存占用 [18] 通常无法容纳于片上存储器中。

图 1 展示了 Cheshire 的架构。交叉开关 [19]采用 Arm 广泛使用的 AXI4 协议 [20] ，将 CVA6 处理器连接到 RPC DRAM、DSA 以及 Cheshire 内部的其他组件。交叉开关的地址宽度、数据宽度以及 AXI4 DSA 管理端口和从属端口的数量均可配置，以满足目标系统的带宽和寻址需求。不支持突发或乱序事务的简单从属设备通过轻量、可扩展的 Regbus [21] 解复用器连接，从而最大限度地降低交叉开关的面积和能耗。

Cheshire 的 RPC DRAM 通过一个可配置的末级缓存（LLC）进行连接。LLC 的每个路组均可在运行时单独配置为暂存存储器（SPM），以便在需要时为主机提供高速的内部静态随机存取存储器（SRAM）。一个符合 RISC-V 标准的调试模块由 JTAG 传输模块提供支持，可对 CVA6 以及任意数量的已配置外部 RISC-V 硬件线程（hart）进行实时外部调试，例如 PMCA DSA 中的硬件线程。同样，中断控制器也支持配置任意数量的外部中断源和目标。

Cheshire 提供了多种可选外设，包括用于高效数据传输的灵活 AXI4 DMA 引擎 [22] 、用于显示输出的 VGA 控制器，以及用于与片外系统通信的数字芯粒间链路。此外，它还提供了用于串行通信的通用异步收发器、GPIO 模块，以及用于访问外部外设的集成电路总线和串行外设接口主机。所有外设均通过 AXI4 或 Regbus 接口无缝集成，并
提供成熟完善的功能集，确保与现有 Linux 驱动程序完全兼容。另有一个*SoC* 控制端口连接 Cheshire 外部的片上设备，这些设备对于系统运行至关重要，例如时钟发生器、IO 多路复用器或时钟与电源域控制器。

Cheshire 内置了启动 ROM，可通过 JTAG、通用异步收发器或芯粒间链路进行被动预加载，也可从外部串行外设接口闪存、集成电路总线 EEPROM 或支持全局唯一标识符分区表（GPT）的 SD 卡自主启动。若经过配置，它会先调用外部 ROM，通过 SoC 控制端口设置周边芯片。使用 -Os 标志并进行全程序链接时

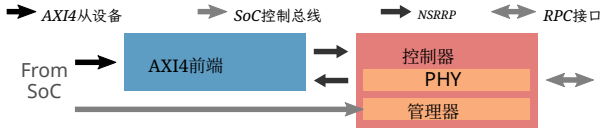


**图 2：** RPC DRAM 内存接口的架构。

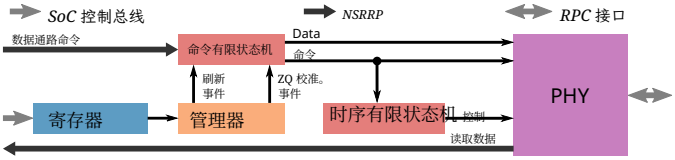


**图 3：** RPC DRAM 控制器的架构。

优化后，Cheshire 的启动 ROM 大小为 7.2 KiB。

虽然我们使用单个 CVA6 作为 Cheshire 的 AC 处理器，但也可以集成多个一致性 CVA6 核心或其他 AC 处理器；本文重点介绍精简型 AC 主机。

我们提供了 Cheshire 的开源 FPGA 实现，目前面向 Digilent 的 Genesys II 开发板，能够实现异构系统的快速原型开发与特性表征。

### *B. RPC DRAM* 接口

背景： RPC DRAM 被设计为仅使用最少数量的信号，即可提供 DDR3 级别的系统内带宽 [23]。其标志性特征是采用一种接口，将数据、地址和命令复用到一条公共数据总线（数据总线）上，从而最大限度地减少所需引脚数量。通过使用单个多功能引脚传输串行命令，命令/地址（CA）控制信号可以与数据同时发送，使其仅需为 16 位宽的数据总线使用 22 个切换信号，即可实现 DDR3 的高事务效率。选通引脚 *DQS*和*DQS#* 以与 DDR3 完全相同的前导码对数据总线进行采样。

目前已有研究提出了采用低引脚数接口的替代性 DRAM 解决方案。Cypress 的*HyperRAM* [14] 在 8 位共享总线上仅需 12 个可切换 IO 引脚。然而，其传输速率最高仅为 200MHz 或更低频率下的 400MB/s，且其自刷新机制限制了控制器端执行高级调度的能力。Antmicro 提出了一种基于 LiteDRAM 存储器控制器 [24]的 FPGA 行锤（Rowhammer）测试平台。该控制器支持多种双倍数据速率（DDR）存储器类型，并且近期新增了 RPC 物理接口电路（物理层）实现。尽管该物理层采用开源方式，但其仅面向 FPGA。

接口： 图 2 展示了我们的 RPC DRAM 接口。它由两部分组成：实现片外 RPC 协议 [23] 的控制器 ，以及实现符合 AXI4 规范的从属设备的 AXI4前端 。为便于适配 AXI4 以外的片上协议，控制器和前端通过一种通用接口连接，我们称之为非停顿请求-响应协议（*NSRRP*）；其数据宽度为 256b，即 RPC DRAM 标准中的一个字 。下文将详细讨论控制器和 AXI4 前端。

控制器：图3展示了控制器的内部架构。如图所示，控制器接收来自前端的数据通路命令。这些通用命令会传递给命令有限状态机，由其将命令分解为 RPC DRAM 专用命令。例如，一条通用的数据通路读取命令会被分解为：1）对相应存储体和行执行激活；2）读取连续的 $N$ 个 RPC DRAM 字；3）执行预充电以关闭存储体，并为下一次访问做好准备。除了前端发出的数据通路命令外，

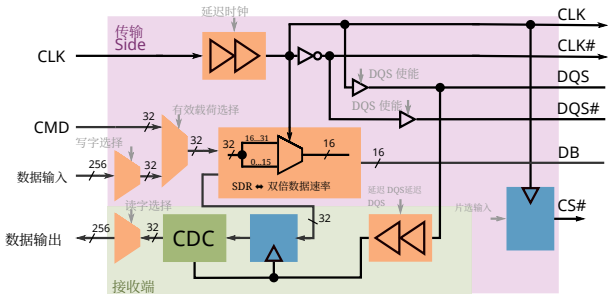


**图 4：** RPC物理层的架构。

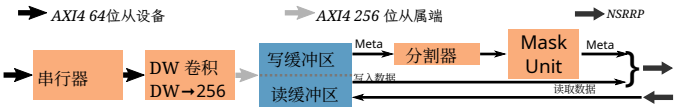


**图5：** RPC接口的AXI4前端架构。

命令有限状态机还会处理控制器内部管理命令模块发出的管理器命令。管理器有三项职责：1）启动时初始化RPC DRAM设备；2）定期刷新活动存储体；3）在必要时执行*ZQ*校准。为完成这些任务，管理器会使用可配置的时序参数，这些参数可通过存储器映射寄存器文件进行设置。

命令有限状态机将生成的 RPC DRAM命令传递给时序有限状态机，后者执行两项任务：1）为命令定时，确保命令遵守周期对齐和最小延迟等协议约束；2）为物理接口定时，包括控制片选信号、门控输出选通，以及将数据、掩码和命令多路复用到数据总线上。

物理接口电路（PHY）如图 4 所示，它实现了一个低功耗、纯数字、技术无关的 RPC DRAM物理层，且不包含内部时钟生成电路。如图所示，PHY 由两部分组成：发送端将数据发送到 RPC DRAM，接收端则接收来自 RPC DRAM 的数据。

发送端通过可配置的延迟线生成相位分别偏移 90 度和 270 度的时钟，以驱动选通信号 DQS 和 DQS#；这些信号由时序有限状态机选择性启用。任何要发送到 RPC DRAM 的 256b 数据字都会先序列化为 32b子字；随后，时序有限状态机通过多路复用器，在发送 32b 命令和 32b 数据子字之间进行仲裁。所选有效载荷通过图示的时钟驱动多路复用器从单倍数据速率（SDR）转换为双倍数据速率（DDR），并放置到数据总线上。

接收侧通过*DB*引脚接收来自 RPC DRAM 芯片的数据，同时接收与数据同相的选通信号*DQS* 。接收到的数据被转换为 SDR，随后由另一条可配置延迟线生成的延迟选通信号进行采样。经过时钟域跨越后，读出数据被打包成完整的 256 位字，然后通过 NSRRP 通道发送回 AXI4前端。

***AXI4* 前端**：如图 5 所示，该前端实现了符合 AXI4 规范的从属设备。由于 RPC DRAM控制器严格按顺序运行，传入的请求首先会被串行化。在当前设计中，来自不同 AXI4

ID 按先到先服务的顺序处理；我们计划在未来版本中使用 AXI4 的服务质量（QoS）信号实现传输优先级控制。序列化后，数据宽度转换器 会将 RPC DRAM 接口配置的数据宽度（Neo 中为 64 b）转换为 RPC 的 256 b 字长。

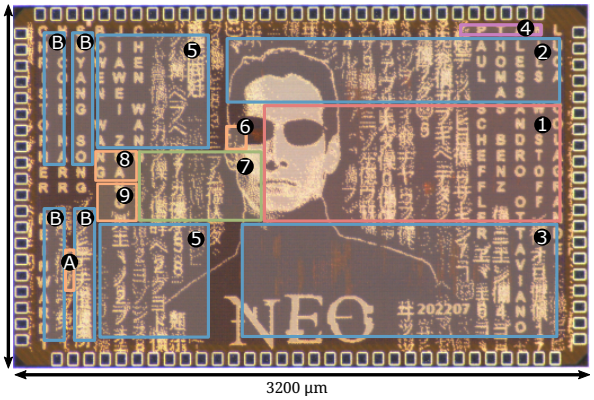


**图 6：** Neo的裸片照片： CVA6 和非核心区域， 数据缓存 指令缓存 芯粒间 路， 片上存 器， 频 锁定环， AXI4 互连和直接存储器访问， RP 管理器， RPC AXI 接口， RPC 控制器， RPC 缓冲区。

虽然 AXI4 允许在任意传输拍暂停传输，但 RPC 突发传输一旦启动便无法暂停。因此，读写操作都需要缓冲。写入数据会先缓冲，待缓冲区包含下一次写入所需的全部数据后再释放。读取数据会尽可能快速地转发到 AXI4 总线上，以尽量降低延迟；只有 AXI4 总线暂停传输时才会进行缓冲。

分割器在 2KiB 边界处分割 NSRRP 事务，以符合 RPC 协议。RPC DRAM 通过引入首个和末尾掩码来实现非对齐传输。这些掩码在写命令与相应数据之间传输。掩码单元根据 AXI4 请求中包含的选通信息生成这两个掩码。

## III. 评估

我们通过名为*Neo*的硅验证芯片来评估 Cheshire。该芯片采用台积电 65 纳米工艺节点和九层金属层设计、实现并制造。Neo 使用 Synopsys Design Compiler 2019.03 进行综合，并使用 Cadence Innovus 2019.10 进行实现，目标是在慢慢工艺角下以 125 °C 的温度和低-$V_T$单元实现 200MHz 系统时钟。我们将先介绍如何为 Neo 配置 Cheshire，然后从功能性能（第 III-B 节）和硅实现性能（第 III-C 节）两个方面评估 RPC DRAM 和 AXI4 DSA 接口。

### *A.* 硅验证芯片

图 6 展示了 Neo 的带注释裸片照片。其 6.4mm²裸片封装在 QFN64 封装中。我们使用 1.2 V 的核心电压和 1.5 V 的全局 IO 电压，以符合 RPC DRAM 接口。我们将 Cheshire 配置为不带 DSA 端口，因为该验证芯片中未集成任何加速器。Neo 配备了 128 KiB 的片上存储器、64 位数据和 48 位地址；其 CVA6 核心配置了 32KiB、8 路组相联的一级数据缓存和指令缓存。RPC DRAM 前端分别为读和写配置了 8 KiB 缓冲区。我们加入了一个片上频率锁定环（FLL），用于内部时钟生成与控制。

我们在图 7 所示的定制启动 PCB 上测试 Neo。该电路板既可连接工业级自动测试设备（ATE）（Advantest SoCV93000），也可独立运行。在独立运行模式下，电路板为 Neo 提供 32 kHz 参考时钟，用于锁定其内部频率锁定环（FLL），并提供消抖复位功能；它还配备了一组外设和连接器，以便启动和运行应用程序，其中包括一颗连接到 Neo 的 32MiB RPC DRAM 芯片（ EM6GA16LBXA-12H）

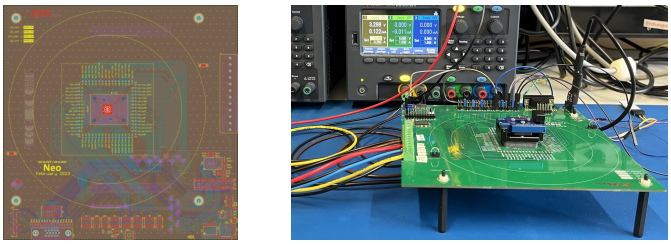

**图 7：** 用于表征*Neo* 独立运行状态的评估 PCB，并使用工业级 ATE 设备。（左）电路板布局（右）制成的 PCB。DRAM 走线长 25.23 cm，宽 0.15 mm，符合 JEDEC 标准的单端阻抗为 50 Ω，差分阻抗为 90 Ω。

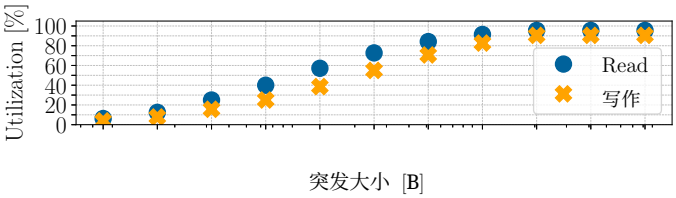


**图 8：** 读写操作中 RPC DRAM 总线的相对利用率。

并用于测试和测量。用户可通过通用异步收发器和视频图形阵列输出进行交互。

### *B.* 功能性能

功能评估通过周期精确的寄存器传输级仿真进行。对于所有给出的结果，我们都利用 DMA 引擎高效的数据移动能力，实现解耦且高吞吐量的主机与 DSA 之间的数据传输，并免去 CVA6 处理数据移动任务的负担。

RPC DRAM 规定的最高总线时钟频率为 933 MHz，优于其他竞争方案。Neo 配备 16 b 数据总线，并以 200 MHz 的工作频率采用 DDR 信号传输，因此可达到的峰值吞吐量为 $\Theta = \alpha \cdot 800\,\mathrm{MB/s}$，其中 $\alpha$ 表示总线的相对利用率。图 8 展示了读写数据通路所达到的 RPC 总线利用率。DMA 的配置为：从 8 B 开始，逐步增大突发传输大小，发起写入和读取传输。对于大小为 2 KiB 或更大的突发传输，AXI4前端的传输分割器会将其拆分为较小的传输，此时接口总线利用率趋于稳定，接近峰值利用率（$\alpha = 1$）。RPC 的前导码和后缀码会带来固定的协议规定开销，该开销与传输大小和控制器实现无关；而读取时的总线利用率平均比对应写入高 1.3×，这是因为读取数据会尽可能快地传送到 AXI4 总线，而写入则会延迟到缓冲足够的数据后才进行。

与最接近的竞争方案之一 HyperBus 相比，RPC DRAM 在可实现带宽和面积占用方面表现出明显优势；HyperBus 已在多项学术研究中实现 [12], [13]。HyperRAM 占用更多 PCB 面积，最高数据速率仅为 400MB/s，且最高频率仅为 200 MHz [14], ，因此限制了其在高带宽、高能效场景中的适用性。

### *C.* 芯片性能

面积分解： 图 6 展示了 Neo 的芯片裸片，并标出了主要组件的大致位置和尺寸。图 9 给出了 Cheshire 的精确面积分解（单位为 kGE），并逐步增加连接到其主 AXI4 交叉开关的 DSA 管理器—从属端口对的数量

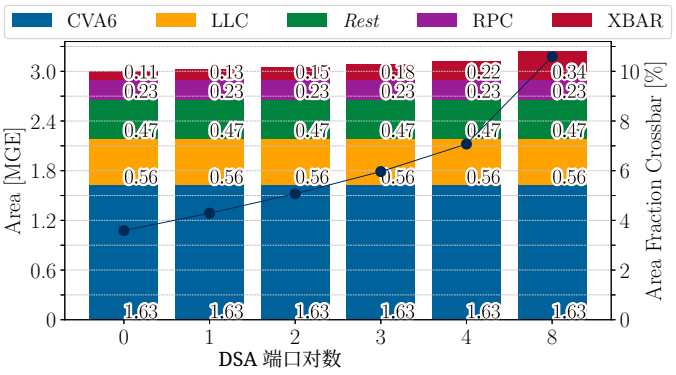


**图 9：** 采用台积电 65 nm 工艺实现的 Cheshire 面积细分，以及不同 DSA 端口对数量下交叉开关所占的比例。其余部分 表示直接存储器访问、外设和互连适配器的面积。

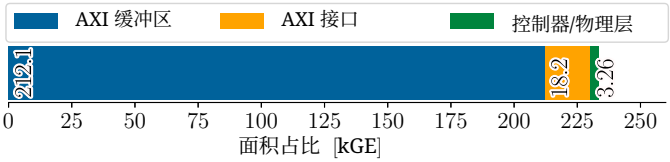


**图 10：** RPC DRAM 接口的面积细分。按 Neo 的配置，*AXI4* 缓冲区 和*AXI4* 接口占据了大部分面积。

。最左侧的柱形对应不含 DSA 端口的 Neo 配置。在所有考察的情况下，CVA6 都占据了 Cheshire 面积的主要部分，而 RPC DRAM控制器所占比例最高为 7.6%。随着 DSA 端口数量增加，全互连 AXI4 交叉开关的占比从 3.6 % 增至 10.6 %；与不含 DSA 端口的 Neo 配置相比，八对端口时面积最多增加 7.8 %。对于大量端口对而言，交叉开关的扩展性会成为限制因素；但包含许多管理器或从属端的 DSA 可以采用子互连、更宽的数据位宽或稀疏连接，以便于扩展。

图 10 展示了 Neo 的 RPC DRAM控制器的面积细分。管理器模块、命令/时序有限状态机和数字物理层仅占控制器面积的 1%，即 3.5kGE，这证实了 RPC 协议为其内存接口带来的面积开销极低。控制器的大部分面积开销来自用于存放 AXI4 数据拍的缓冲区。在 Neo 中，这些缓冲区的配置容量过大，以简化 AXI4前端的初始设计，但未来版本中可以进一步缩减其容量。尽管如此，我们的控制器仅占现有 65 nm 全引脚 DDR3 控制器面积的 6.3%，以及其海岸线长度的 33% [25]。

能效： 图 11 展示了在 bring-up 板上测得的 Neo 在不同场景和频率下的功耗；每种场景都考察了计算强度和内存强度各异的一个运行边界条件。本文重点评估 Cheshire 平台；有关 CVA6 处理器的更多基准测试，请参阅其论文 [17]。bring-up 板提供三个电源域及其供电：CORE、IO 和 RAM。CORE 为 Neo 的核心区域逻辑和 SRAM 供电，IO 为其焊盘供电，RA M 则为 RPC DRAM 内存芯片供电。

在*WFI* 场景中，CVA6 正在等待中断，不获取或解码指令，处于空闲状态；这提供了切换活动极少时的功耗基准。在*NOP* 场景中，CVA6 循环执行由空操作组成的代码体，为主动获取、分支和解码但几乎没有停顿的工作负载建立功耗下限。*2MM* 运行经过优化的双精度浮点矩阵

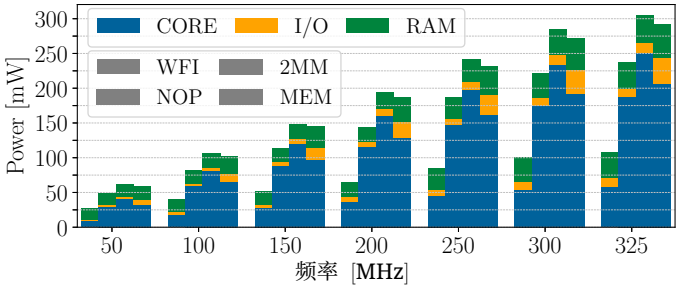


**图 11：** Neo 在四种工作负载下的功耗：*WFI*、*NOP*、*2MM*和*MEM*。功耗分为 Neo 的三个电源域。乘法运算的参数和结果均存储在 RPC DRAM 中，并将可复用的矩阵块保存在片上存储器中。*MEM* 使用 DMA 引擎向 RPC DRAM 写入高吞吐量数据突发。

当 CORE 电源电压为 1.2 V 时，Neo 的时钟频率最高可达 325 MHz；即使在*2MM*这类数据密集型计算场景中，功耗也保持在 300 mW 以内。所有功耗分量均按预期随频率线性变化；在几乎所有情况下，CORE 功耗都占主导地位；在 200 MHz 时，*MEM* 功耗的 69% 消耗在 CORE 中。由于本文提出的 RPC DRAM 接口版本未使用该技术的*DeepPower Down* 状态，因此所有基准测试均显示 RAM 空闲功耗。尽管如此，在高负载下，200 MHz 时 RPC DRAM 接口用于*MEM* 的 IO 功耗仍比现有的 65nm DDR3 接口低 45 % [25]。

我们使用*MEM* 计算 RPC DRAM 的传输效率。根据第 II 节的讨论，我们仅考虑写入方向，因为其代表缓冲方面的最坏情况。根据 III-B 节测得的最大带宽，RPC DRAM 接口每传输一个字节的能耗为 $\frac{P_{tot}}{\Theta} \simeq 250$ pJ B$^{-1}$。该结果与近期集成 HyperRAM 等较低带宽存储器的研究报告的每字节能耗相当 [12]。

## IV. 结论

本文介绍了 Cheshire，这是一个轻量、模块化的开源 64 位 Linux 主机平台，旨在无缝接入领域专用加速器（DSA）。Cheshire 具有独特的 RPC DRAM 接口、可配置为暂存存储器的末级缓存，以及 DMA 引擎，可实现加速器或 DRAM 与平台之间的高效数据传输。它还提供多种可选外设，包括通用异步收发器、串行外设接口、集成电路总线、GPIO 模块，以及用于显示输出的视频图形阵列。我们采用台积电 65nm 互补金属氧化物半导体工艺实现并制造了 Cheshire 硅验证芯片，命名为 *Neo*。在 1.2 V 电压下，即使在数据密集型计算场景中，Neo 的总功耗也不超过 300 mW，时钟频率最高可达 325MHz。其 DRAM 接口的能耗仅为 250 pJ/B，物理层面积仅为 3.5 kGE，并在 200MHz 下实现了 750 MB/s 的峰值传输速率。

## 致谢

我们感谢评审者提出的意见。感谢 C. Jinfan、V. Krishna、R. Zhou、N. Narr、C. Reinwardt、S. Song、B. Wang、C. Wang、J. Zhang、N. Wistoff 和 L. Colagrande 对本研究的贡献。我们还感谢 Etron Technology Inc.，特别是 R. Crisp，给予我们的支持。本研究部分得到了 FRACTAL（877056）项目的支持，该项目获得了 ECSEL 联合事业（JU）的资助；同时也得到了 TRISTAN（101095947）项目的支持，该项目获得了 HORIZON KDT-JU 计划的资助。

## 参考文献

[1] J. Koomey、S. Berard、M. Sanchez 和 H. Wong，“计算电气效率历史趋势的影响”，IEEE《计算机史年刊》，第 33 卷，第 3 期，第 46–54 页，2011 年。[2] Y. Chi、W. Qiao、A. Sohrabizadeh、J. Wang 和 J. Cong，“推动领域专用计算民主化”，《ACM 通讯》，第 66 卷，第 1 期，第 74–85 页，2022 年 12 月。[3] J. L. Hennessy 和 D. A. Patterson，“计算机体系结构的新黄金时代”，《ACM 通讯》，第 62 卷，第 2 期，第 48–60 页，2019 年 1 月。[4] Y. Zhao、R. Xie、G. Xin 和 J. Han，“面向 module-lwe 应用、带有 RISC-V 矩阵扩展的高性能领域专用处理器”，IEEE《电路与系统汇刊 I：正刊》，第 69 卷，第 7 期，第 2871–2884 页，2022 年。[5] R. Paludo 和 L. Sousa，“面向 Linux 的 RISC-V 全同态加密加速器的 NTT 架构”，IEEE《电路与系统汇刊 I：正刊》，第 69 卷，第 7 期，第 2669–2682 页，2022 年。[6] 树莓派基金会，“Rasperry Pi Zero”，https://www.raspberrypi.com/products/raspberry-pi-zero，2020 年。[7] M. Ditty，“Nvidia Orin 片上系统”，载于《2022 年 IEEE Hot Chips 第 34 届研讨会》，2022 年，第 1–17 页。[8] SiFive，“HiFive Unmatched”，https://www.sifive.com/boards/hifive-unmatched，2022 年。[9] J. Zuckerman、P. Mantovani、D. Giri 和 L. P. Carloni，“借助 RISC-V 和 ESP 推动异构多核 SoC 研究”，2022 年。[10] J. Balkind、K. Lim、M. Schaffner、F. Gao、G. Chirkov、A. Li、A. Lavrov、T. M. Nguyen、Y. Fu、F. Zaruba、K. Gulati、L. Benini 和 D. Wentzlaff，《BYOC：面向异构指令集架构研究的“自带核心”框架》。美国纽约州纽约市：ACM，2020 年，第 699–714 页。[11] A. Di Mauro、M. Scherer、D. Rossi 和 L. Benini，“Kraken：用于纳米无人机超高效视觉处理的直接事件/帧式多传感器融合 SoC”，载于《IEEE Hot Chips 第 34 届研讨会》，2022 年，第 1–19 页。[12] D. Rossi、F. Conti、M. Eggiman、A. D. Mauro、G. Tagliavini、S. Mach、M. Guermandi、A. Pullini、I. Loi、J. Chen、E. Flamand 和 L. Benini，“Vega：面向物联网终端节点的十核 SoC，具有 DNN 加速功能，并可从基于 MRAM 的状态保持休眠模式中认知唤醒”，IEEE《固态电路期刊》，第 57 卷，第 1 期，第 127–139 页，2022 年 1 月。[13] L. Valente、Y. Tortorella、M. Sinigaglia、G. Tagliavini、A. Capotondi、L. Benini 和 D. Rossi，“HULK-V：一种异构超低功耗、支持 Linux 的 RISC-V SoC”，载于《2023 年欧洲设计、自动化与测试会议暨展览会（DATE）》，2023 年。[14] Cypress，“HyperRAM 和八通道 xSPI RAM 存储器”，https://www.cypress.com/products/hyperram-octal-xspi-ram-memory。[15] A. Garofalo、M. Rusci、F. Conti、D. Rossi 和 L. Benini，“Pulp-nn：用于在基于 RISC-V 的并行超低功耗集群上进行边缘量化神经网络推理的计算库”，载于《第 26 届 IEEE 国际电子、电路与系统会议》，2019 年，第 33–36 页。[16] E. Manor 和 S. Greenberg，“用于 TensorFlow Lite for Microcontrollers 的定制硬件推理加速器”，《IEEE Access》，第 10 卷，第 73 484–73 493 页，2022 年。[17] F. Zaruba 和 L. Benini，“应用级处理的成本：采用 22 nm FDSOI 工艺、面向 Linux 的 1.7 GHz 64 位 RISC-V 核心的能耗与性能分析”，IEEE《超大规模集成电路系统汇刊》，第 27 卷，第 11 期，第 2629–2640 页，2019 年 11 月。[18] M. Opdenacker，“在不足 4 MB RAM 中运行 Linux”，
h
t
t
p
s://
b
o
o
t
li
n.com/pub/conferences/2017/jdll/opdenacker-embedded-linux-in-less-than-4mb-of-ram/，2017 年。[19] A. Kurth、W. Ronninger、T. Benz、M. Cavalcante、F. Schuiki、F. Zaruba 和 L. Benini，“ 用于高性能非一致性片上通信的开源平台 ”，IEEE 《计算机汇刊》，第 1–1 页，2021 年。[20] A. Limited，“AMBA AXI 和 ACE 协议规范，版本 H.c”，https://developer.arm.com/documentation/ihi0022/hc，2021 年。[21] 通用寄存器接口，PULP 平台贡献者。 [在线]。可访问：https://github.com/pulp-platform/register interface [22] T. Benz、M. Rogenmoser、P. Scheffler、S. Riedel、A. Ottaviano、A. Kurth、T. Hoefler 和 L. Benini，“高性能、节能的模块化 DMA 引擎架构”，2023 年。[23] Etron Technology，“256 Mb 高带宽 RPC DRAM”，
h
t
t
p
s://etrona
merica.com/wp-content/uploads/2019/05/EM6GA16LGDABMACAEA-RPC-DRAM Rev.-1.0.pdf，2019 年。[24] Antmicro，“用于缓解 Rowhammer 的开源 DDR 控制器框架”，https://antmicro.com/blog/2021/08/open-source-ddr-test-framework-for-rowhammer/，2021 年。[25] C. Sudarshan、J. Lappas、C. Weis、D. M. Mathew、M. Jung 和 N. Wehn，“面向精度可调计算的精简型低功耗、低延迟 DRAM 存储器控制器”，载于《嵌入式计算系统：架构、建模与仿真》，Springer 国际出版公司，2019 年，第 429–441 页。