



(12)发明专利申请

(10)申请公布号 CN 110196624 A

(43)申请公布日 2019.09.03

(21)申请号 201910132389.8

(22)申请日 2019.02.22

(30)优先权数据

62/634,836 2018.02.24 US

15/967,476 2018.04.30 US

(71)申请人 三星电子株式会社

地址 韩国京畿道

(72)发明人 平展

(74)专利代理机构 北京市柳沈律师事务所

11105

代理人 钱大勇 陈芳

(51)Int.Cl.

G06F 1/20(2006.01)

G06N 3/06(2006.01)

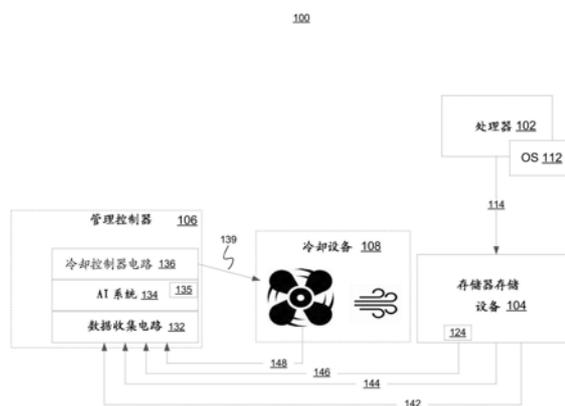
权利要求书3页 说明书12页 附图4页

(54)发明名称

用于服务器和固态设备的人工智能冷却方法

(57)摘要

根据一个一般方面,一种装置可包括存储器存储设备。存储器存储设备可以包括多个存储器单元,该多个存储器单元被配置为存储数据。存储器存储设备可以包括第一温度传感器,该第一温度传感器被配置为检测存储器单元的温度。存储器存储设备可以包括人工智能系统,该人工智能系统被配置为至少部分地基于第一温度传感器和存储器存储设备的工作负荷来动态地确定要由存储器存储设备采用的热管理设置。存储器存储设备可以响应于由人工智能系统确定的热管理设置来动态地改变操作参数的集合。



1. 一种存储装置,包括:
存储器存储设备,包括:
多个存储器单元,被配置为存储数据;
第一温度传感器,被配置为检测所述存储器单元的温度;和
人工智能系统,被配置为至少部分地基于所述第一温度传感器和所述存储器存储设备的工作负荷,来动态地确定要由所述存储器存储设备采用的热管理设置;
其中,所述存储器存储设备响应于由所述人工智能系统确定的所述热管理设置,来动态地改变操作参数的集合。
2. 根据权利要求1所述的装置,其中,所述人工智能系统包括神经网络;以及,
其中,所述神经网络被配置为预测在随后时段期间所述存储器存储设备的温度,以及至少部分地基于在所述随后时段期间的所预测的温度而选择所述热管理设置。
3. 根据权利要求2所述的装置,其中,至少部分地经由在离线模式下加载的训练数据来训练所述神经网络;以及
其中,还基于当所述存储器存储设备操作于在线模式下时所述存储器单元的温度和工作负荷来训练所述神经网络。
4. 根据权利要求2所述的装置,其中,所述人工智能系统被配置为,响应于在所述随后时段期间的温度将会超过第一阈值的预测,在所述随后时段之前启动冷却序列。
5. 根据权利要求1所述的装置,其中,所述装置还包括第二温度传感器,第二温度传感器被配置为检测所述存储器存储设备的温度;以及
其中,所述人工智能系统被配置为至少部分地基于所述第二温度传感器来确定所述热管理设置。
6. 根据权利要求1所述的装置,其中,所述操作参数包括从包含以下的组中所选择的参数:读取/写入单元的大小、擦除单元的大小、读取缓冲器的大小、写入缓冲器的大小、扼制阈值、存储器单元利用量、高速缓存大小以及垃圾收集的频率。
7. 根据权利要求1所述的装置,其中,所述人工智能系统被配置为预冷却所述存储器存储设备。
8. 根据权利要求1所述的装置,其中,所述人工智能系统被配置为基于期望的功率效率来确定要由所述存储器存储设备采用的热管理设置。
9. 一种计算设备,包括:
存储器存储设备,被配置为存储数据;
控制器,被配置为管理所述计算设备的物理状态;以及
冷却设备,被配置为至少部分地冷却所述计算设备;
其中,所述控制器包括人工智能系统,所述人工智能系统被配置为至少部分地基于所述存储器存储设备的温度和所述存储器存储设备的工作负荷,来动态地确定要由所述冷却设备采用的热管理设置;以及
其中,所述冷却设备被配置为基于由所述人工智能系统确定的所述热管理设置,来动态地尝试升高或降低所述计算设备的温度。
10. 根据权利要求9所述的计算设备,其中,所述控制器还包括数据收集电路,所述数据收集电路被配置为接收以下作为输入:

所述冷却设备的工作负荷；
所述存储器存储设备的温度，和
所述存储器存储设备的工作负荷；以及

其中，所述人工智能系统被配置为确定至少部分地基于所述冷却设备的所述工作负荷来选择所述热管理设置。

11. 根据权利要求9所述的计算设备，其中，所述人工智能系统包括神经网络；并且其中，所述神经网络被配置为预测在随后时段期间所述存储器存储设备的温度，以及至少部分地基于在所述随后时段期间的所预测的温度而选择所述热管理设置。

12. 根据权利要求9所述的计算设备，其中，所述存储器存储设备的工作负荷包括来自操作系统的数据访问请求的工作负荷以及来自所述存储器存储设备的数据访问履行的工作负荷。

13. 根据权利要求12所述的计算设备，其中，所述人工智能系统被配置为至少部分基于来自操作系统的数据访问请求的工作负荷来预冷却所述存储器存储设备。

14. 根据权利要求9所述的计算设备，其中，所述冷却设备包括风扇；

其中，所述控制器包括风扇速度控制器，所述风扇速度控制器被配置为设置所述风扇的速度；以及

其中，所述热管理设置包括期望的风扇速度。

15. 根据权利要求9所述的计算设备，其中，所述人工智能系统被配置为预测对于所述存储器存储设备要维持小于或等于期望的温度所需要的、所述计算设备内的空气流动的速率。

16. 根据权利要求9所述的计算设备，还包括：

发热设备，被配置为输出温度信号；以及

其中，所述人工智能系统被配置为至少部分地基于所述发热设备的温度信号来确定所述热管理设置，并且其中，所述热管理设置被配置为将所述存储器存储设备和所述发热设备两者维持在各自的期望温度。

17. 一种信息处理系统，包括：

处理器，被配置为至少部分地指令存储器存储设备的工作负荷；

存储器存储设备，被配置为存储数据，并且

包括第一人工智能系统，所述第一人工智能系统被配置为至少部分地基于所述存储器存储设备的至少一部分的检测的温度以及所述存储器存储设备的工作负荷，来动态地确定要由所述存储器存储设备采用的局部热管理设置；

冷却设备，被配置为至少部分地冷却所述系统；以及

管理控制器，被配置为管理所述系统的物理状态，并且包括：

第二人工智能系统，被配置为至少部分地基于所述存储器存储设备的温度和所述存储器存储设备的工作负荷，来动态地确定要由所述冷却设备采用的系统热管理设置。

18. 根据权利要求17所述的系统，其中，所述第一人工智能系统和所述第二人工智能系统中的至少一个被配置为至少部分地基于估计的未来的工作负荷来预冷却所述存储器存储设备。

19. 根据权利要求17所述的系统，其中，所述存储器存储设备响应于由所述第一人工智

能系统确定的所述局部热管理设置,来动态地改变操作参数的集合;

其中,至少部分地经由在离线模式下加载的训练数据来训练所述第一人工智能系统;
以及

其中,还基于在线模式下的存储器存储设备操作来训练所述第一人工智能系统。

20. 根据权利要求17所述的系统,其中,所述第二人工智能系统被配置为至少部分地基于所述处理器的温度来动态地确定所述系统热管理设置。

用于服务器和固态设备的人工智能冷却方法

[0001] 对相关申请的交叉引用

[0002] 本申请要求于2018年2月24日提交的、题为“用于服务器和固态设备的人工智能冷却方法”的、序列号为62/634,836的临时专利申请和于2018年4月30日提交的、题为“用于服务器和固态设备的人工智能冷却方法”的、序列号为15/967,476的临时专利申请的优先权。该在先提交的申请的主题通过引用并入于此。

技术领域

[0003] 本说明书涉及计算技术,并且更具体地,涉及一种用于服务器和固态设备(SSD)的人工智能(AI)冷却方法。

背景技术

[0004] 例如,从移动设备、物联网(IoT)传感器、电动汽车、商业交易等产生了越来越多的数据。该数据需要数据中心能够具有高密度和高性能存储处理越来越高等级的数据。该数据的提高的存储和处理要求系统等级的、针对存储设备的有效的冷却和高效的热解决方案。

[0005] 传统上,通过定义如何冷却计算设备的静态设置来冷却计算设备。传统上,这些设置是基于建模或“最佳猜测(best guess)”分析而预先确定的(例如,在工厂处或跨设备的集合)。因此,热设置未被优化并且不能适应变化的条件。

[0006] 人工神经网络(ANN)或连接性(connectionist)系统通常是受到构成动物的大脑的生物神经网络模糊地启发的计算系统。这样的系统通过考虑示例来“学习”任务(即,逐步改进关于任务的性能),而通常无需任务特定的编程。例如,在图像识别中,这样的系统可以通过下述来学习识别包含猫的图像:分析已经手动标记为“猫”或“没有猫”的示例图像,并且使用结果来识别其他图像中的猫。这样的系统在没有关于猫的任何先验知识(例如,猫有毛皮、尾巴、胡须和猫状的脸部)的情况下学习。相反,这样的系统从其处理的学习材料中发展出其自己的相关特征集。

[0007] 神经网络通常基于被称为人工神经元的连接的单元或节点的集合。人工神经元之间的每个连接(简化版本的突触)可以将信号从一个人工神经元传输到另一个人工神经元。接收信号的人工神经元可以对其进行处理,然后信号通知连接到其的人工神经元。在通常的神经网络实施方式中,人工神经元之间的连接处的信号是实数,并且每个人工神经元的输出通过其输入之和的非线性函数来计算。人工神经元和连接通常具有随着学习进行而调节的权重。权重增加或减少连接处信号的强度。人工神经元可以具有阈值,使得仅当聚集的信号超过该阈值时才发送信号。通常,人工神经元按层组织。不同的层可以对其输入执行不同类型的变换。信号可能在遍历各层多次之后从第一(输入)层行进到最后(输出)层。

发明内容

[0008] 根据一个一般方面,一种装置可包括存储器存储设备。存储器存储设备可以包括

多个存储器单元,该多个存储器单元被配置为存储数据。存储器存储设备可以包括第一温度传感器,该第一温度传感器被配置为检测存储器单元的温度。存储器存储设备可以包括人工智能系统,该人工智能系统被配置为至少部分地基于第一温度传感器和存储器存储设备的工作负荷来动态地确定要由存储器存储设备采用的热管理设置。存储器存储设备可以响应于由人工智能系统确定的热管理设置来动态地改变操作参数集合。

[0009] 根据另一个一般方面,一种计算设备可以包括存储器存储设备,该存储器存储设备被配置为存储数据。计算设备可以包括控制器,该控制器被配置为管理计算设备的物理状态。计算设备可以包括冷却设备,该冷却设备被配置为至少部分地冷却计算设备。控制器可以包括人工智能系统,该人工智能系统被配置为至少部分地基于存储器存储设备的温度和存储器存储设备的工作负荷来动态地确定要由冷却设备采用的热管理设置。冷却设备可以被配置为基于由人工智能系统确定的热管理设置来动态地尝试升高或降低计算设备的温度。

[0010] 根据另一个一般方面,一种系统可以包括处理器,该处理器被配置为至少部分地指令存储器存储设备的工作负荷。该系统可以包括存储器存储设备,该存储器存储设备被配置为存储数据。存储器存储设备可以包括第一人工智能系统,该第一人工智能系统被配置为至少部分地基于存储器存储设备的至少一部分的检测到的温度和存储器存储设备的工作负荷来动态地确定要由存储器存储设备采用的局部热管理设置。该系统可以包括冷却设备,该冷却设备被配置为至少部分地冷却系统。该系统可以包括管理控制器,该管理控制器被配置为管理系统的物理状态。该管理控制器可以包括第二人工智能系统,该第二人工智能系统被配置为至少部分地基于存储器存储设备的温度和存储器存储设备的工作负荷来动态地确定要由冷却设备采用的系统热管理设置。

[0011] 在附图和以下说明中阐述了一个或多个实施方式的细节。根据说明书和附图并且根据权利要求,其他特征将是明显的。

[0012] 基本上如至少一个附图中所示和/或如结合至少一个附图所描述,一种用于计算技术的系统和/或方法,并且更具体地,一种用于服务器和固态设备(SSD)的人工智能(AI)冷却方法如在权利要求中更完整地阐述。

附图说明

[0013] 图1是根据所公开的主题的系统的示例实施例的框图。

[0014] 图2是根据所公开的主题的系统的示例实施例的框图。

[0015] 图3是根据所公开的主题的系统的示例实施例的框图。

[0016] 图4是可以包括根据所公开的主题的原理所形成的设备的信息处理系统的示意性框图。

[0017] 在各个图中的相同的附图标记指示相同的元件。

具体实施方式

[0018] 以下将参考在其中示出了一些示例实施例的附图来更加全面地描述各种示例实施例。然而,本公开的主题可以体现为许多不同的形式,并且不应被解释为限于本文中所阐述的示例实施例。相反,提供这些示例实施例,以使得本公开将是彻底的和完整的并且将本

公开的主题的范围完全地传达给本领域技术人员。在附图中,为了清楚,可以夸大层和区域的大小和相对大小。

[0019] 将理解,当元件或层被称为“在另一元件或层上”、“连接到另一元件或层”或者“耦合到另一元件或层”时,其可以直接地在另一元件或层上、连接到另一元件或层或者耦合到另一元件或层,或者可以存在居于中间的元件或层。相反,当元件被称为“直接地在另一元件或层上”、“直接地连接到另一元件或层”或者“直接地耦合到另一元件或层”时,不存在居于中间的元件或层。相同的附图标记始终指代相同的元件。当在本文中使用时,术语“和/或”包括相关列举项目中的一个或多个的任何和所有组合。

[0020] 将理解,尽管在本文中可以使用术语第一、第二、第三等来描述各种元件、组件、区域、层和/或部分,但是这些元件、组件、区域、层和/或部分不应受这些术语的限制。这些术语仅用于将一个元件、组件、区域、层或部分与另一个元件、组件、区域、层或部分相区分。因此,在不背离本公开的主题的教导的情况下,下面讨论的第一元件、组件、区域、层或部分可以被称为第二元件、组件、区域、层或部分。

[0021] 为了描述的方便,在本文中可以使用诸如“在……下方”、“在……以下”、“下方”、“在……以上”、“上方”等的空间相对术语来描述一个元件或特征与另外的(一个或多个)元件或特征的关系,如附图所示。将理解,除了附图中所描绘的方位之外,空间相对术语旨在包含在使用或操作中的设备的不同方位。例如,如果附图中的设备被翻转,则被描述为“在其他元件或特征下方”或“在其他元件或特征以下”的元件将被定向为“在其他元件或特征之上”。因此,示例性术语“在……以下”可以包含在……以下和在……以上的方位两者。设备可以以其他方式定向(旋转90度或在其他方位),并且相应地解释本文中所使用的空间相对描述符。

[0022] 同样地,为了描述的方便,在本文中可以使用诸如“高”、“低”、“上拉”、“下拉”、“1”、“0”等的电气术语来描述相对于其他电压电平或电流的电压电平或电流,如附图所示。将理解,除了附图中所描绘的电压或电流之外,电气相对术语旨在包含在使用中或操作中的设备的不同参考电压。例如,如果附图中的设备或信号被反转,或者附图中的设备或信号使用其他参考电压、电流或电荷,则被描述为“高”或“上拉”的元件将是相较于与新的参考电压或电流的“低”或下拉。因此,示例性术语“高”可以包含相对低的电压或电流或者相对高的电压或电流两者。该设备可以另外地基于不同的电气参考框架,并且相应地解释本文中所使用的电气相对描述符。

[0023] 在本文中所使用的术语仅用于描述特定示例实施例的目的,并且不旨在限制本公开的主题。如在本文中使用的,单数形式“一”、“一个”和“该”也旨在包括复数形式,除非上下文另外明确地指示。将进一步理解,当在本说明书中使用时,术语“包括”和/或“包含”指定所陈述的特征、整数、步骤、操作、元件和/或组件的存在,但是不排除一个或多个其他特征、整数、步骤、操作、元件、组件和/或分组的存在或添加。

[0024] 在本文中参考作为理想化示例实施例(和中间结构)的示意图示的截面图示来描述示例实施例。因此,将预期作为例如制造技术和/或公差的结果的与图示形状的偏离。因此,示例实施例不应被解释为限于本文中所示出的区域的特定形状,而是包括例如由制造导致的形状偏差。例如,被示出为矩形的注入区域通常在其边缘处将具有圆形的或弯曲的特征和/或注入浓度的梯度,而不是从注入区域到非注入区域的二元改变。同样地,通过注

入形成的掩埋区域可以导致在掩埋区域与通过其进行注入的表面之间的区域中的一些注入。因此,附图中示出的区域本质上是示意性的,以及其形状不旨在示出元件的区域的实际形状并且不旨在限制本公开的主题的范围。

[0025] 除非另外地定义,否则在本文中所使用的所有术语(包括技术术语和科学术语)具有与本公开的主题所属领域的普通技术人员通常理解的含义相同的含义。将进一步理解的是,诸如在常用词典中定义的那些术语应该被解释为具有与其在相关领域的上下文中的含义一致的含义,并且将不以理想化或过于形式的含义被解释,除非在本文中明确地如此定义。

[0026] 在下文中,将参考附图详细地解释示例实施例。

[0027] 图1是根据所公开的主题的系统100的示例实施例的框图。在各种实施例中,系统100可以包括计算设备。应理解,以上仅是一个说明性示例,所公开的主题不限于此。

[0028] 在各种实施例中,系统100可以包括计算设备,诸如,例如,膝上型计算机、桌面型计算机、工作站、个人数字助理、智能电话、平板型计算机和其他适合的计算机或者其虚拟机或虚拟计算设备。在一些实施例中,系统100可以包括服务器或机架安装设备。在这样的实施例中,系统100可以在数据中心或其他高密度计算环境中找到。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0029] 在各种实施例中,系统100可以包括处理器102,处理器102被配置为执行一个或多个机器可执行指令或软件、固件或者其组合。在一些实施例中,系统100可以包括被配置为临时地、永久地、半永久地或其组合地存储一个或多个数据的存储器、存储介质或存储器存储设备104。

[0030] 在各种实施例中,处理器102可以执行或运行操作系统(OS) 112或其他形式的软件。处理器102/OS112可以对存储器存储设备104进行大量数据访问(读取和/或写入命令)。在这样的实施例中,来自处理器102的初始数据访问(例如,请求数据、存储数据)可以是数据访问请求,并且存储器存储设备104完成数据访问(例如,返回数据、报告成功或失败)可以被称为数据访问完成。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0031] 通常,这些数据访问可以被称为存储器存储设备104的工作负荷114。在各种实施例中,如果数据访问的数量高,则工作负荷114可能高,反之亦然。此外,在各种实施例中,工作负荷114可以包括大量数据访问,其导致由存储器产生更多或更少的热(例如,与读取相比更多的写入、与对单个高速缓存线的数据访问相对地更多的对多个高速缓存线的数据访问)。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0032] 在各种实施例中,存储器存储设备104可以包括固态驱动器或盘(SSD)或者其他形式的存储器技术。通常,SSD是固态存储设备,其使用集成电路组件(assembly)作为存储器来持久地存储数据。

[0033] 在另一个实施例中,所示出的存储器存储设备104可以是计算系统或设备100内产生热的任何设备。这种发热设备的示例可以包括图形处理器、芯片组、处理器102本身等。然而,为了图示的方便,发热设备将被称为存储器存储设备104。将理解,以上仅是一个示例性示例,所公开的主题不限于此。

[0034] 在所示的实施例中,系统100可以包括冷却设备108。在各种实施例中,冷却设备108可以包括风扇,其将空气吹过存储器存储设备104,从而将由存储器存储设备104产生的

热移动到其他地方(并且冷却存储器存储设备104)。在另一个实施例中,冷却设备108可以包括液体冷却设备,该液体冷却设备通过水、油或其他液体的循环和/或浸没(以其他方式部分地浸没)在水、油或其他液体中来冷却。在又一个实施例中,冷却设备108可以包括压电泵。应理解,以上仅是几个说明性示例,所公开的主题不限于此。再次地,为了图示的方便,冷却设备108将被最经常地称为风扇,但是应理解,以上仅是一个说明性示例,所公开的主题不限于此。

[0035] 在所示的实施例中,冷却设备108可以被配置为提供可变等级的冷却。例如,冷却设备108的风扇速度可以增加或减少;冷却设备108可以包括多个风扇,其中可以打开或关闭任何数量的风扇;冷却设备108可以旋转或提供来自多个源的冷空气(例如,环境空气、压缩空气、空调空气等)。应理解,以上仅是几个说明性示例,所公开的主题不限于此。在所示的实施例中,冷却设备108可以被配置为动态地改变分别地提供给系统100或存储器存储设备104的冷却等级。

[0036] 在各种实施例中,存储器存储设备104可以在期望的环境温度为大约或大致40°C或更低的环境中进行操作。在各种实施例中,存储器存储设备104的温度可以随着工作负荷增加或字符(character)改变(到更密集的程度)而升高,并且可以随着工作负荷减轻(例如,在强度或字符方面)而降低。在一些实施例中,存储器存储设备104的温度可以增加超过100°C。随着温度升高,这可能会降低可用性能或增加功率使用。这可能是不期望的。

[0037] 在所示的实施例中,系统100可以包括管理控制器或处理器106。在各种实施例中,管理控制器106可以包括基板管理控制器(BMC),该基板管理控制器(BMC)可以是使用传感器并且通过独立的连接与系统管理员通信来监视计算机、网络服务器或其他硬件设备(例如,系统100)的物理状态的专用服务处理器。在一些实施例中,管理控制器106可以是处理器102(例如,片上系统(SoC))的一部分。

[0038] 在所示的实施例中,管理控制器106可以包括人工智能(AI)系统或电路134。通常, AI系统134可以被配置为动态地调节提供给存储器存储设备104或系统100的冷却量,或者动态地调节存储器存储设备104或系统100的温度。

[0039] 在各种实施例中, AI系统134可以包括神经网络(例如,深度前馈神经网络、深度循环神经网络)。应理解,以上仅是一个说明性示例,所公开的主题不限于此。在另一个实施例中, AI系统134可以包括符号AI、认知模拟、启发式或其他形式的AI。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0040] 这与传统的系统不同,在传统的系统中,系统的冷却设置或参数是静态的且固定的,并且不随系统改变而改变。在各种实施例中,静态设置可以是基于规则的(例如,如果温度超过阈值,则提高风扇速度)。传统上,温度设置是反应式的,并且仅在温度已经升高之后改变冷却设置(例如,风扇速度)。

[0041] 在所示的实施例中,人工智能(AI)系统134可以被配置为不只是仅基于在给定时刻存储器存储设备104或系统100的温度来调节冷却或热管理设置135。在所示的实施例中,人工智能(AI)系统134可以被配置为不仅动态地或自适应地(与静态规则或设置相对)而且还预测性地调节冷却设置或热管理设置135,以使得可以预测温度的升高(或降低)并且可以在实际温度升高/降低之前改变冷却或热管理设置135。在各种实施例中, AI系统134可以被配置为提供对热波动的快速响应。应理解,以上仅是几个说明性示例,所公开的主题不限

于此。

[0042] 在所示的实施例中,管理控制器106可以包括数据收集电路132。在这样的实施例中,数据收集电路132可以被配置为从多个源收集或接收数据或信号并且将其(以或原始的、或过滤的、或处理过的形式)输入到AI系统134。

[0043] 在所示的实施例中,数据收集电路132可以被配置为接收关于工作负荷142的信息或数据。在各种实施例中,这可以由处理器102请求的工作负荷114或数据访问、由存储器存储设备104经历的整体的工作负荷、和/或由存储器存储设备104实行的数据访问或由存储器存储设备104提供的响应的工作负荷。

[0044] 在所示的实施例中,数据收集电路132可以被配置为接收关于所消耗的或提供给存储器存储设备104的电力量(amount of power) 144的信息或数据。在所示的实施例中,数据收集电路132可以被配置为接收关于存储器存储设备104的温度146的信息或数据。在各种实施例中,这可以包括关于存储器存储设备104的特定部分(例如,存储器单元或内部控制器)的温度和/或存储器存储设备104作为整体的温度的细节。在所示的实施例中,数据收集电路132可以被配置为接收关于冷却设备108的状态、工作负荷(例如,风扇速度) 148或设置的信息或数据。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0045] 在所示的实施例中,AI系统134可以获取这些数据142、144、146和148,并且确定对于给定时刻哪些冷却或热管理设置135是期望的。

[0046] 在所示的实施例中,管理控制器106可以包括冷却控制器电路136。在各种实施例中,冷却控制器电路136可以被配置为与冷却设备108接口(interface)。在这样的实施例中,冷却控制器电路136可以将所确定的冷却或热管理设置135转换(如果需要)为冷却设备108可以理解的命令或信号139。例如,如果热管理设置135指示需要更多冷却,则冷却控制器电路136可以(经由信号139)施加更多电压并且增加冷却设备108的风扇速度。应理解,以上仅是一个说明性示例,所公开的主题不限于此。

[0047] 如上所述,在各种实施例中,AI系统134可以维护这些设置的历史(或作为数据,或作为神经网络中的权重)。在各种实施例中,AI系统134可以使用该历史来预测系统100或存储器存储设备104的未来温度可以是什么。在这样的实施例中,AI系统134可以主动地和预测性地尝试预冷却或降低系统100或存储器存储设备104的温度,使得当温度确实上升时,从较低的位置开始并且因此在较低的位置处达到峰值。

[0048] 例如,如果工作流程142类似于在先前的较高温度之前的工作流程,则AI系统134可以识别该工作流程并且预测可能发生未来的较高温度。在一个这样的实施例中,工作流程142可以指示正在访问的存储器地址,该存储器地址在上次繁重工作流程开始之前访问过,或者正在发生的存储器请求模式,该存储器请求模式具有与之前的高工作负荷事件相同的模式,等等。在各种实施例中,AI系统134可以在存储器存储设备104处理工作流程142(例如,读取/写入请求)并且产生热量之前看到它们,或者AI系统134可以采用模式识别或其他预测技术。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0049] 在这样的实施例中,AI系统134可以动态地将热管理设置135改变为更积极的姿态。然后,冷却设备108可以相应地改变其行为(例如,增加风扇速度、泵送更多的水)。然后,存储器存储设备104可以降低其温度,或者可以降低温度升高的速率。例如,与在不具有AI系统134的预测能力的情况下将发生的冷却相比,冷却可以更快地发生。在这样的实施例

中,可以减少用于冷却系统100或存储器存储设备104的能量的量。

[0050] 在各种实施例中,当系统100可操作时,可以实时地或以在线模式来训练AI系统134。在这样的实施例中,当系统100进行操作并且产生数据(例如,数据142、144、146和148)时,关于对于给定的情形要采取的适当措施或者要选择热管理设置135是什么,AI系统134的权重或其他决策制定机制可以进行学习或被训练。

[0051] 在各种实施例中,当系统100关机、停止操作或者以另外期望的方式操作时,可以(例如,在存储器存储设备104中、在非易失性存储器中、或外部地)存储这些训练值或权重。在这样的实施例中,当系统100再次变为可操作时,可以将这些训练值或权重重新加载或重新存储到AI系统134中。

[0052] 在另一个实施例中,AI系统134可以在离线模式下(即,不基于由系统100实时提供的数据)训练,或者使用不是由系统100创建的训练数据来训练。在一个这样的实施例中,典型的系统100可以用于训练AI系统134,然后可以将那些训练值或权重复制到或重复到多个系统(例如,整个数据中心或服务器群(server farm))。在一个这样的实施例中,这些AI系统134可以具有存储在管理控制器106或系统100的固件(未示出)中的训练值或权重的默认(default)集合。

[0053] 在各种实施例中,AI系统134可以通过以上的混合来训练。在一个这样的实施例中,可以采用默认训练值或权重来离线加载AI系统134。可以通过所存储的来自可操作的系统100的先前实例的训练值来增强这些默认训练值。然后,当系统100运行或操作时,可以实时地或以在线模式训练AI系统134。在这样的实施例中,如果在每次系统100启动时AI系统134以未训练开始,则AI系统134可以尽可能地训练或者可以具有更程度的训练。

[0054] 图2是根据所公开主题的系统200的示例实施例的框图。在各种实施例中,系统200可以包括处理器202和存储器存储设备204。

[0055] 如上所述,处理器202可以被配置为执行操作系统203或其他软件。处理器202可以至少部分地控制要由存储器存储设备204运行的或执行的数据访问或其他操作的工作流程242。

[0056] 如上所述,存储器存储设备204可以被配置为存储数据。存储器存储设备204可以包括具有闪速(FLASH)存储器的固态设备(SSD),但不限于此。此外,所公开的主题不限于存储介质或存储器存储设备204,而是可以应用于任何计算设备组件(例如,图形处理器、芯片组)。然而,存储器存储设备204将用作说明性示例。应理解,以上仅是一个说明性示例,所公开的主题不限于此。

[0057] 在所示的实施例中,存储器存储设备204可以包括被配置为实际地存储数据的多个存储器单元212。存储器存储设备204还可以包括控制器处理器214,该控制器处理器214被配置为执行工作负荷242、或者由处理器请求的或对于常规事务(housekeeping)(例如,垃圾清理)所需要的操作。在各种实施例中,控制器处理器214可以被配置为基于操作参数238的集合来执行这些操作。

[0058] 在这样的实施例中,这些操作参数238可以包括诸如以下内容:读取/写入单元的大小、擦除单元的大小、读取缓冲器的大小、写入缓冲器的大小、扼制(throttling)阈值、存储器单元利用率、高速缓存大小以及垃圾收集频率。在各种实施例中,这些操作参数238可以至少部分地影响存储器存储设备204的性能和/或功耗。结果,这些操作参数238可以影响

由存储器存储设备204产生的热量。

[0059] 在各种实施例中,存储器存储设备204可以包括多个热传感器或温度传感器,其被配置为检测或指示存储器存储设备204的一个或多个部分的温度。在所示的实施例中,存储器存储设备204可以包括设备温度传感器226,设备温度传感器226被配置为指示存储器存储设备204整体(或者至少放置传感器226的位置)的当前温度。在所示的实施例中,存储器存储设备204可以包括控制器温度传感器224,控制器温度传感器224被配置为指示控制器处理器214的当前温度。在所示的实施例中,存储器存储设备204还可以包括存储器单元温度传感器222,存储器单元温度传感器222被配置为指示存储器单元212的当前温度。

[0060] 在所示的实施例中,存储器存储设备204可以包括AI系统、模型或电路216。在各种实施例中,AI系统216可以包括神经网络或其他形式的AI,如上所述。在这样的实施例中,AI系统216可以被配置为选择或确定热设置232的集合来管理或尝试管理存储器存储设备204的温度。

[0061] 在所示的实施例中,AI系统216可以与图1的AI系统非常类似地工作。然而,图1的AI系统可以尝试整体地管理图1的计算设备或服务器的温度,或者在各种实施例中可以尝试管理图1的存储组件的温度,而图2的AI系统216仅针对存储器存储设备204的温度。在各种实施例中,系统等级AI(例如,图1的AI系统)可以与设备等级AI(例如,AI系统216)一起工作。在这样的实施例中,两个AI系统(例如,图1的AI系统和图2的AI系统)可以彼此通信(例如,设备等级AI可以向系统级AI提供早前的温度数据或预测)。在另一个实施例中,它们可以独立地或隔离地工作。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0062] 在所示的实施例中,AI系统216可以接收来自工作负荷242以及传感器222、224和226中的至少一个的输入。在各种实施例中,AI系统216可以从存储器单元温度传感器222接收输入。在另一个实施例中,AI系统216还可以从设备温度传感器226接收输入。在又一个实施例中,AI系统216可以从所有三个传感器222、224和226接收输入。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0063] 在这样的实施例中,基于这些输入242、222、224和226,AI系统216可以确定期望的冷却或热设置232的集合。如上所述,在各种实施例中,这可以至少部分地基于工作负荷242的历史或模式。作为确定期望的冷却或热设置232的集合的一部分,AI系统216可以产生或预测未来的温度234和期望的温度236。在这样的实施例中,如果不采取冷却动作或者不改变热设置232集合,则未来温度234将成为实际温度。在各种实施例中,该产生和确定可以与以上描述类似地发生。

[0064] 在所示的实施例中,AI系统216可以被配置为动态地改变正在操作的或操作参数238中的一个或多个。在这样的实施例中,作为新的操作参数238的结果,控制器处理器214可以减少(或增加)由存储器存储设备204产生的热量的方式动作。例如,控制器处理器214可以使得数据访问执行得更慢、或者垃圾收集被推迟、或者数据被预先取回。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0065] 与以上描述类似,在各种实施例中,可以在离线模式、在线模式或其组合中训练AI系统216。在各种实施例中,与以上描述类似,AI系统216可以被配置为通过提供早期的温度调节来预冷却或预测性地调节存储器存储设备204的热设置。此外,在各种实施例中,以上描述的其他特征可以应用于设备等级AI系统216,并且这里所描述的特征可以应用于系统

等级AI。

[0066] 在各种实施例中, AI系统216可以被配置为基于除最低预测温度之外的结果来选择热设置232。例如, 可以选择期望的功率效率或性能水平。在一个这样的实施例中, 如果预测的未来温度234低于或者在给定的阈值, 则AI系统216可以指令存储器存储设备204提高性能, 即使产生更多的热量。在各种实施例中, 即使超过阈值热量(如果存在阈值热量), 处理器202也可以指示(例如, 通过命令或工作负荷242模式)性能是关键。在这样的实施例中, 可能不只是降低温度, 而是优化或均衡温度。应理解, 以上仅是几个说明性示例, 所公开的主题不限于此。

[0067] 图3是根据所公开主题的系统300的示例实施例的框图。在各种实施例中, 系统300可以包括计算设备。应理解, 以上仅是一个说明性示例, 所公开的主题不限于此。

[0068] 在各种实施例中, 系统300可以包括计算设备, 诸如, 例如, 膝上型计算机、桌面型计算机、工作站、个人数字助理、智能电话、平板型计算机和其他适当的计算机或者其虚拟机或虚拟计算设备。在一些实施例中, 系统300可以包括服务器或机架安装设备。在这样的实施例中, 系统300可以在数据中心或其他高密度计算环境中找到。应理解, 以上仅是几个说明性示例, 所公开的主题不限于此。

[0069] 如上所述, 系统300可以包括(运行OS112的)处理器102、存储器存储设备104、冷却设备108和管理控制器306。在所示的实施例中, 管理控制器306可以包括数据收集电路332、AI系统334和冷却控制器336, 类似于以上描述的那些。

[0070] 与图1的系统类似, 管理控制器306可以被配置为从存储器存储设备104和冷却设备108接收热相关数据148和346(例如, 工作负荷、热传感器数据)。管理控制器306可以被配置为采用AI系统334处理该数据148和346, 以及使用确定的热设置335的集合来动态地并且可能地预测地调节冷却设备108的操作。

[0071] 另外, 在所示的实施例中, 系统300还可以包括第二发热设备304(其中, 存储器存储设备104是第一发热设备)。在各种实施例中, 发热设备304可以包括另一个处理器(例如, 图形处理器、协处理器(co-processor))、芯片组、第二存储设备。在这样的实施例中, 发热设备304可以提供功工作、消耗电力以及产生热。

[0072] 在这样的实施例中, 管理控制器306可以被配置为从发热设备304接收热相关数据356(例如, 工作负荷、热传感器数据)。管理控制器306可以被配置为采用AI系统334来处理该数据, 以及使用确定的热设置335的集合来动态地并且可能地预测地来调节冷却设备108的操作。在这样的实施例中, 管理控制器306(和AI系统334)可以管理多个发热设备104和304的温度。在各种实施例中, 发热设备(例如, 设备104和304)的数量可以大于两个。

[0073] 在所示的实施例的变型中, 系统300可以包括第二冷却设备308。在这样的实施例中, 第一冷却设备108可以与存储器存储设备104相关联, 而第二冷却设备308可以与发热设备304相关联。在这样的实施例中, 第二冷却设备308可以向管理控制器306提供输入数据358, 类似于第一冷却设备108所做的。

[0074] 在这样的实施例中, 管理控制器306或冷却控制器336可以向第二冷却设备308提供命令或信号339以便控制它(例如, 调节其风扇速度)。如上所述, 这些命令或信号339可以基于由AI系统334确定的热设置335。在这样的实施例中, 可以分别地控制第一冷却设备108和第二冷却设备308。在这样的实施例中, 提供给装置104和304的冷却量或装置104和304的

期望的温度可以不同。

[0075] 在各种实施例中,冷却设备的数量可以大于两个。在又一个实施例中,每个发热设备(例如,设备104和304)可以不直接与单个冷却设备相关联,并且它们的数量可以不同。在这样的实施例中,一个冷却设备可以与多个发热设备相关联。在这样的实施例中,冷却设备可以与系统300的物理区域相关联,并且仅与发热设备相关联,这是因为该发热设备在给定的物理区域内。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0076] 图4是信息处理系统400的示意性框图,信息处理系统400可以包括根据所公开主题的原理形成的半导体器件。

[0077] 参考图4,信息处理系统400可以包括根据所公开的主题的原理构造的设备中的一个或多个。在另一个实施例中,信息处理系统400可以采用或运行根据所公开的主题的原理的一个或多个技术。

[0078] 在各种实施例中,信息处理系统400可以包括计算设备,诸如,例如,膝上型计算机、桌面型计算机、工作站、服务器、刀片式服务器、个人数字助理、智能电话、平板型计算机和其他适当的计算机或者其虚拟机或虚拟计算设备。在各种实施例中,信息处理系统400可以由用户(未示出)使用。

[0079] 根据所公开的主题的信息处理系统400还可以包括中央处理单元(CPU)、逻辑或处理器410。在一些实施例中,处理器410可以包括一个或多个功能单元块(FUB)或组合逻辑块(CLB)415。在这样的实施例中,组合逻辑块可以包括各种布尔(Boolean)逻辑运算(例如,NAND、NOR、NOT、XOR)、稳定(stabilizing)逻辑器件(例如,触发器、锁存器)、其他逻辑设备或其组合。可以以简单或复杂的方式配置这些组合逻辑运算来处理输入信号以实现期望的结果。应理解,虽然描述了同步组合逻辑运算的一些说明性示例,但是所公开的主题不限于此并且可以包括异步运算、或者其混合。在一个实施例中,组合逻辑运算可以包括多个互补金属氧化物半导体(CMOS)晶体管。在各种实施例中,这些CMOS晶体管可以被布置为执行逻辑运算的门;但是应理解可以使用其他技术,并且这些技术也在所公开的主题的范围内。

[0080] 根据所公开的主题的信息处理系统400还可以包括易失性存储器420(例如,随机存取存储器(RAM))。根据所公开的主题的信息处理系统400还可以包括非易失性存储器430(例如,硬驱、光学存储器、NAND或闪速存储器)。在一些实施例中,易失性存储器420、非易失性存储器430或者其组合或部分可以被称为“存储介质”。在各种实施例中,易失性存储器420和/或非易失性存储器430可以被配置为以半永久或基本上永久的形式存储数据。

[0081] 在各种实施例中,信息处理系统400可以包括一个或多个网络接口440,一个或多个网络接口440被配置为允许信息处理系统400成为通信网络的一部分并且经由通信网络进行通信。Wi-Fi协议的示例可以包括但不限于:电气和电子工程师协会(IEEE)802.11g、IEEE 802.11n。蜂窝协议的示例可以包括但不限于:IEEE 802.16m(又称先进无线-MAN(城域网))、先进长期演进(LTE)、增强型数据速率GSM(全球移动通信系统)演进(EDGE)、演进的高速分组接入(HSPA+)。有线协议的示例可以包括但不限于:IEEE 802.3(又称以太网)、光纤信道、电力线通信(例如,HomePlug、IEEE 1901)。应理解,以上仅是几个说明性示例,所公开的主题不限于此。

[0082] 根据所公开的主题的信息处理系统400还可以包括用户接口单元450(例如,显示器适配器、触觉接口、人类接口设备)。在各种实施例中,该用户接口单元450可以被配置为

接收来自用户的输入和/或向用户提供输出。其他类型的设备也可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的感官反馈，例如视觉反馈、听觉反馈或触感反馈；并且可以以任何形式接收来自用户的输入，包括声学、语音或触感输入。

[0083] 在各种实施例中，信息处理系统400可以包括一个或多个其他设备或硬件组件460（例如，显示器或监视器、键盘、鼠标、相机、指纹读取器、视频处理器）。应理解，以上仅是几个说明性示例，所公开的主题不限于此。

[0084] 根据所公开的主题的信息处理系统400还可以包括一个或多个系统总线405。在这样的实施例中，系统总线405可以被配置为通信地耦合处理器410、易失性存储器420、非易失性存储器430、网络接口440、用户接口单元450以及一个或多个其他设备或硬件组件460。由处理器410处理的数据或从非易失性存储器430的外部输入的数据可以存储在非易失性存储器430或易失性存储器420中。

[0085] 在各种实施例中，信息处理系统400可以包括或执行一个或多个软件组件470。在一些实施例中，软件组件470可以包括操作系统 (OS) 和/或应用。在一些实施例中，OS可以被配置为向应用提供一个或多个服务，并且在应用与信息处理系统400的各种硬件组件（例如，处理器410、网络接口440）之间进行管理或充当中介。在这样的实施例中，信息处理系统400可以包括一个或多个本机 (native) 应用，其可以本地安装（例如，安装在非易失性存储器430内）并且被配置为由处理器410直接地执行并直接地与OS交互。在这样的实施例中，本机应用可以包括预编译的机器可执行代码。在一些实施例中，本机应用程序可以包括脚本解释器（例如，C shell (csh)、AppleScript、AutoHotkey）或虚拟执行机 (VM)（例如，Java虚拟机、Microsoft公共语言运行时），该脚本解释器被配置为将源代码或目标代码转换成可执行代码，然后由处理器410执行。

[0086] 可以使用各种封装技术来封装 (encapsulate) 上述半导体器件。例如，根据所公开主题的原理构造的半导体器件可以使用层叠封装 (POP) 技术、球栅阵列 (BGA) 技术、芯片规模封装 (CSP) 技术、塑料式引线芯片承载封装 (PLCC) 技术、塑料双列直插封装 (PDIP) 技术、华夫管芯 (die in Waffle) 封装技术、晶片形式的管芯封装技术、板上芯片 (COB) 技术、陶瓷双列直插封装 (CERDIP) 技术、塑料公制四方扁平封装 (PMQFP) 技术、塑料四方扁平封装 (PQFP) 技术、小外型封装 (SOIC) 技术、缩小外型封装 (SSOP) 技术、薄型小外型封装 (TSOP) 技术、薄型四方扁平封装 (TQFP) 技术、系统级封装 (SIP) 技术、多芯片封装 (MCP) 技术、晶片级制造封装 (WFP) 技术、晶片级处理堆栈封装 (WSP) 技术，或者本领域技术人员将会知道的其他技术中的任何一个进行封装。

[0087] 方法步骤可以由执行计算机程序的一个或多个可编程处理器来执行，以通过对输入数据进行操作并且产生输出来执行功能。方法步骤也可以由专用逻辑电路执行，并且装置可以实现为专用逻辑电路，专用逻辑电路例如FPGA（现场可编程门阵列）或ASIC（专用集成电路）。

[0088] 在各种实施例中，计算机可读介质可以包括指令，该指令当被执行时，使得设备执行方法步骤的至少一部分。在一些实施例中，计算机可读介质可以包括以下中：磁介质、光学介质、其他介质或其组合（例如，CD-ROM、硬驱、只读存储器、闪速驱动）。在这样的实施例中，计算机可读介质可以是有形且非暂时性地体现的制品。

[0089] 虽然已经参考示例的实施例描述了所公开的主题的原理，但是对于本领域技术人

员将明显的是,在不背离这些公开的概念的精神和范围的情况下,可以对其进行各种改变和修改。因此,应理解,以上实施例不是限制性的,而仅是说明性的。因此,所公开的概念的范围将由所附权利要求及其等同物的最宽泛的可允许解释来确定,并且不应受前述描述的约束或限制。因此,应理解,所附权利要求旨在覆盖落入实施例范围内的所有这些修改和改变。

100

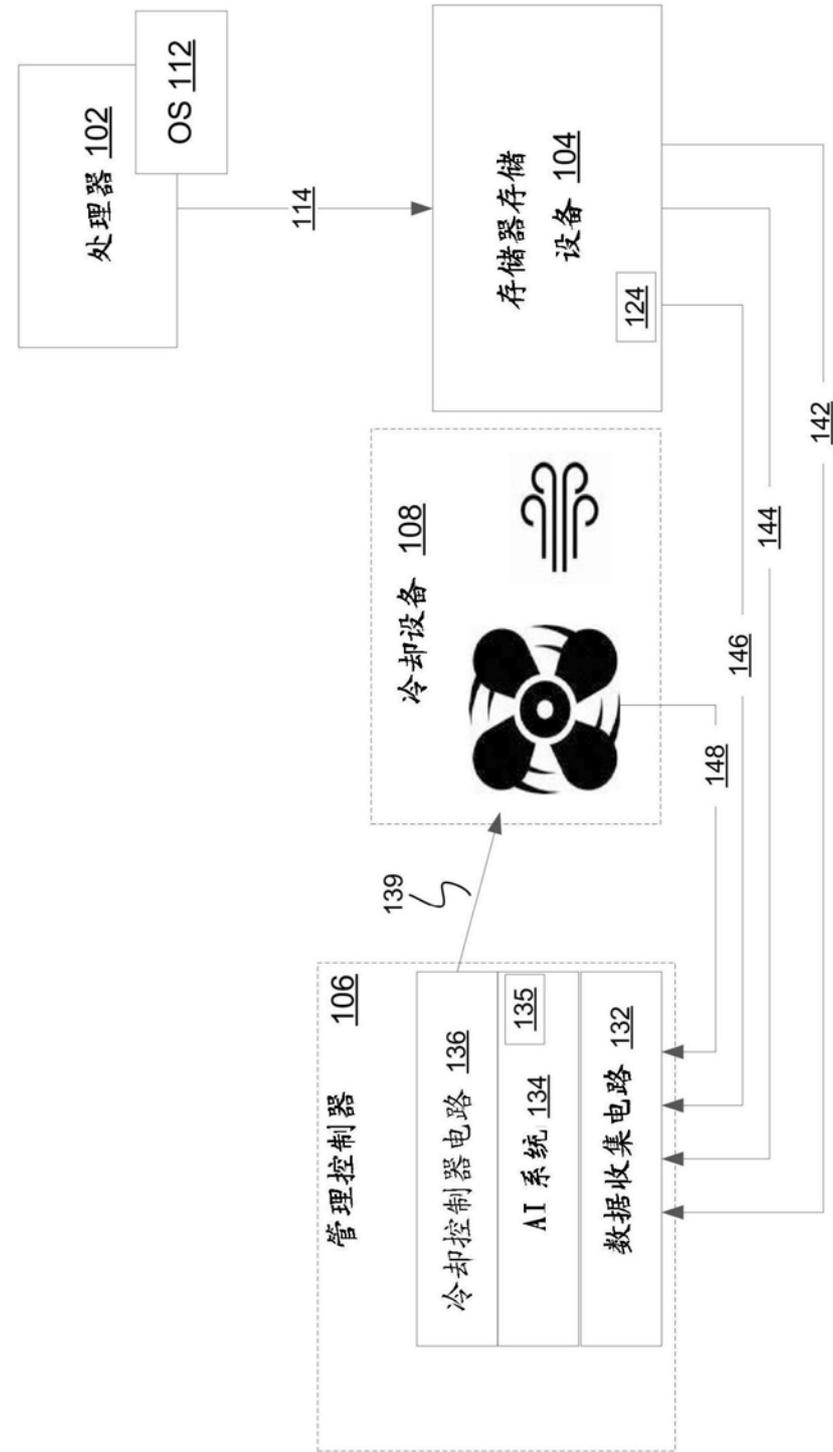


图1

200

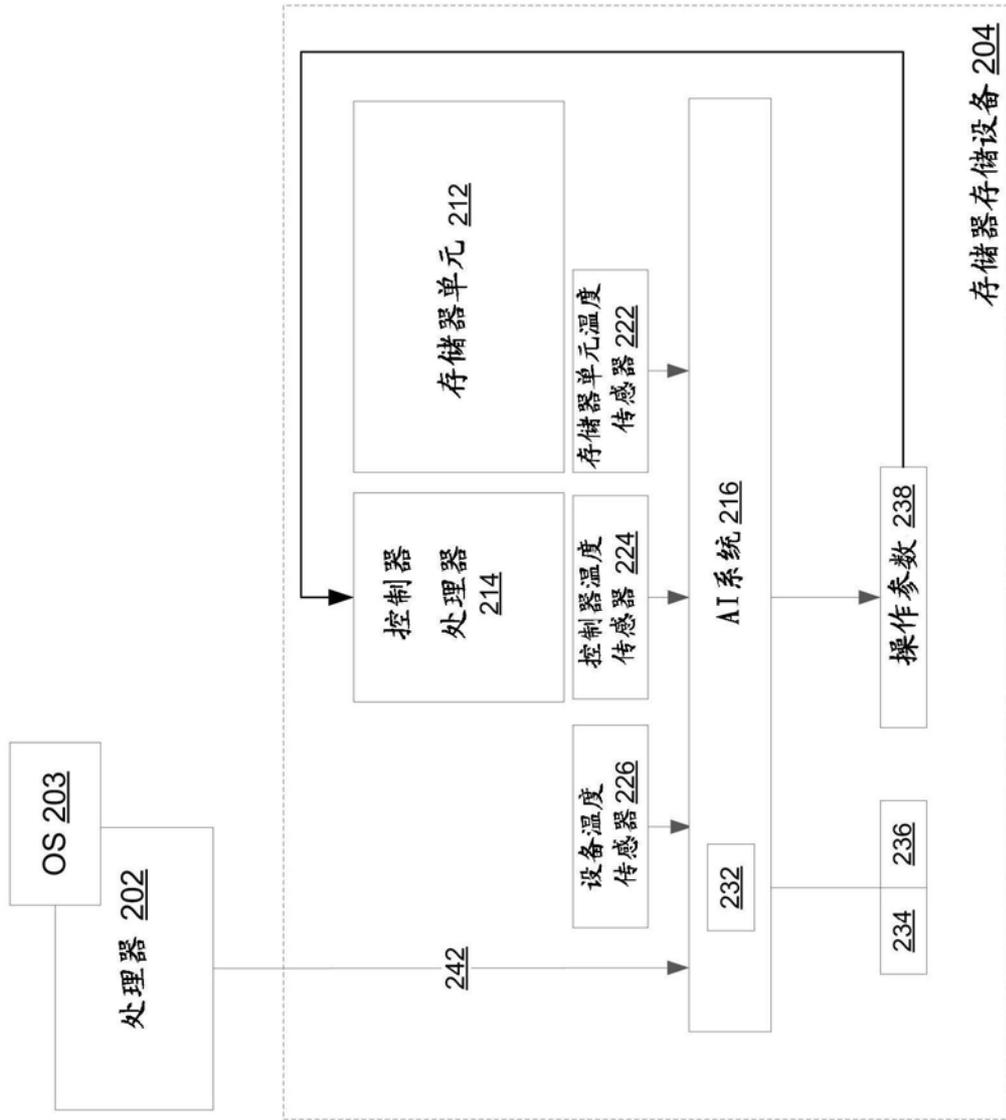


图2

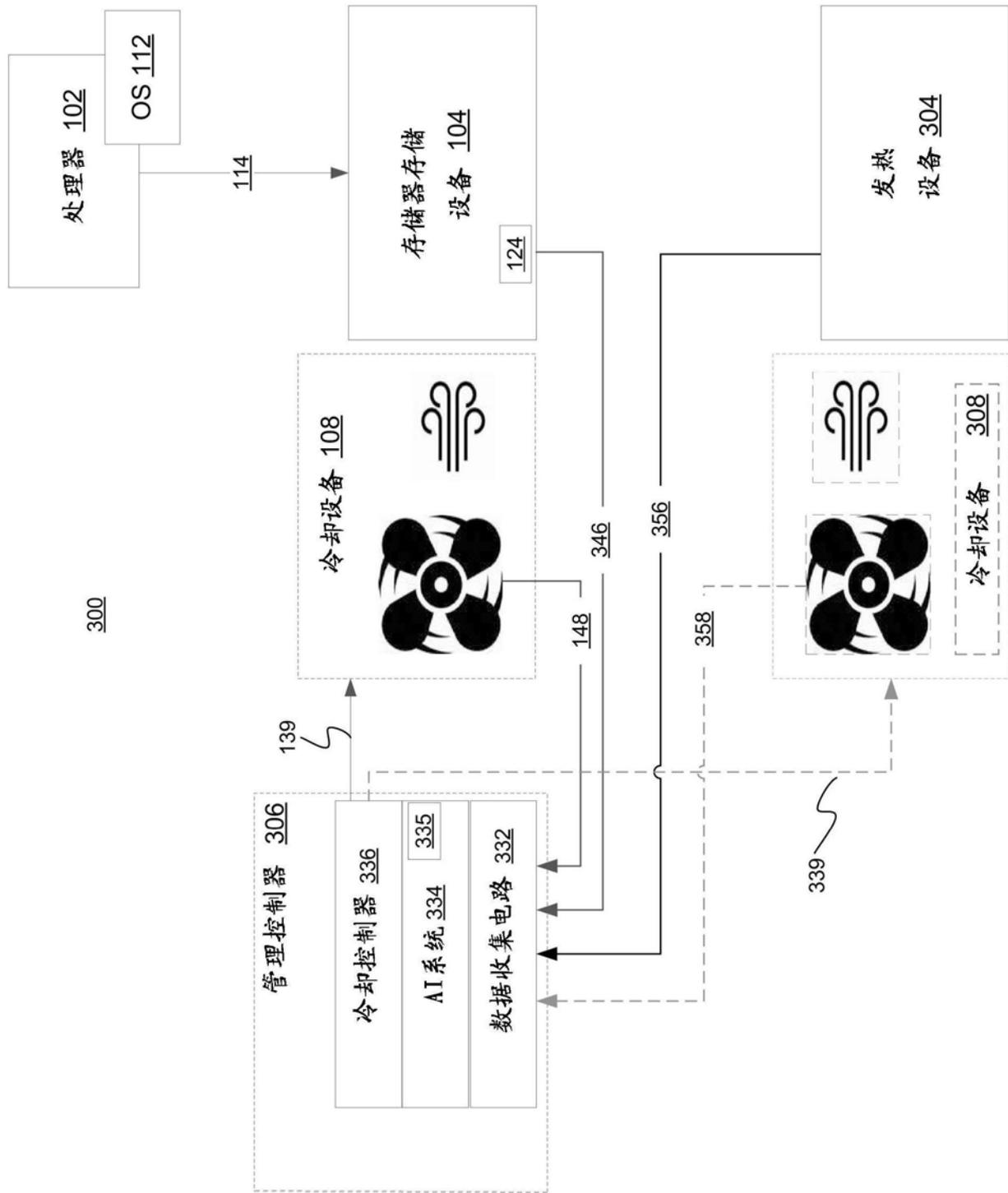


图3

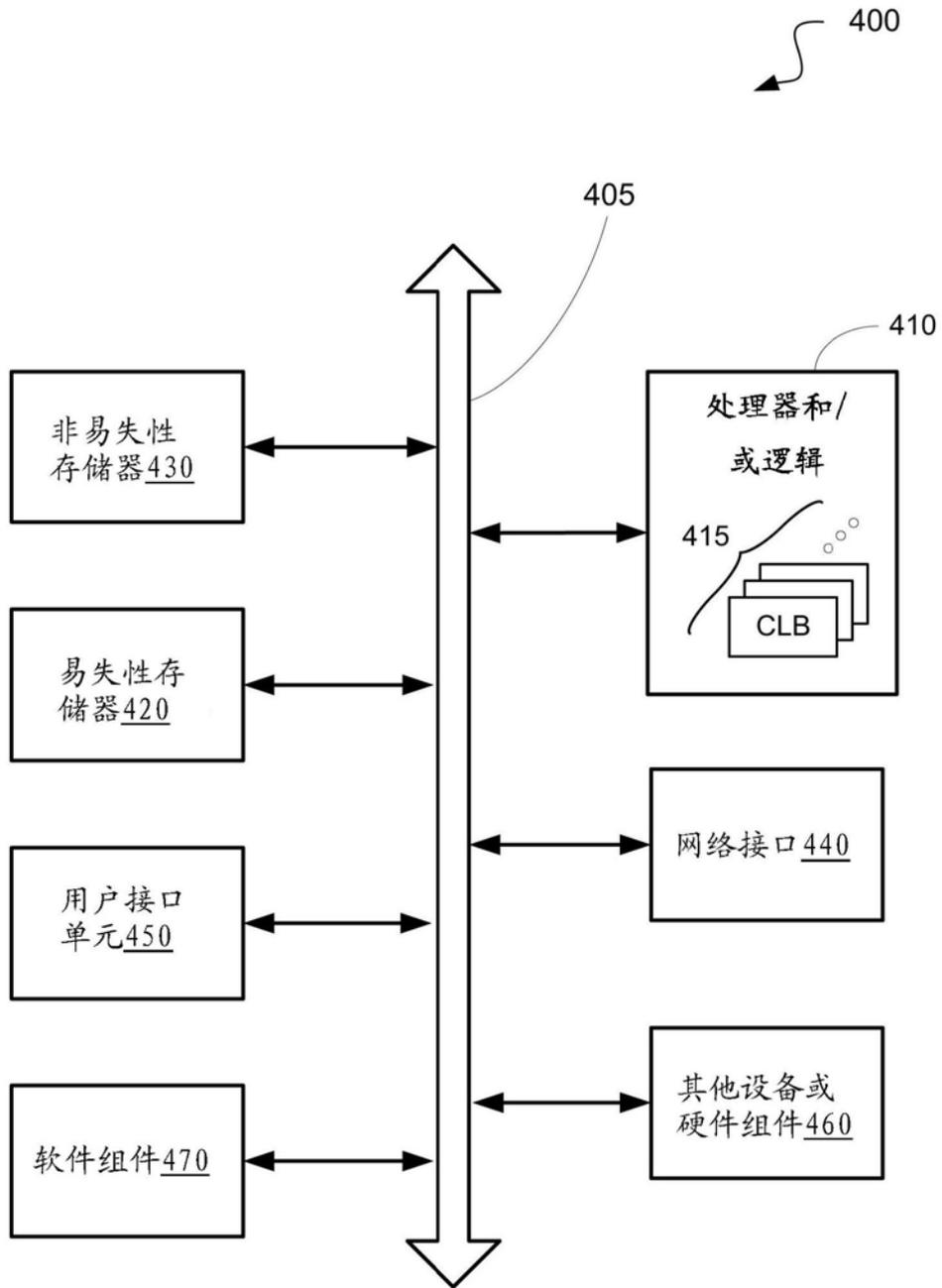


图4