

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023 年 5 月 11 日 (11.05.2023)



(10) 国际公布号
WO 2023/077751 A1

(51) 国际专利分类号:
G06F 9/4401 (2018.01)

江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(21) 国际申请号: PCT/CN2022/090021

(22) 国际申请日: 2022 年 4 月 28 日 (28.04.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202111301938.3 2021年11月4日 (04.11.2021) CN

(71) 申请人: 苏州浪潮智能科技有限公司
(**INSPUR SUZHOU INTELLIGENT TECHNOLOGY CO., LTD.**) [CN/CN]; 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(72) 发明人: 王鑫(**WANG, Xin**); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 刘哲(**LIU, Zhe**); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 赵金(**ZHAO, Jin**); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 张乾坤(**ZHANG, Qiankun**); 中国

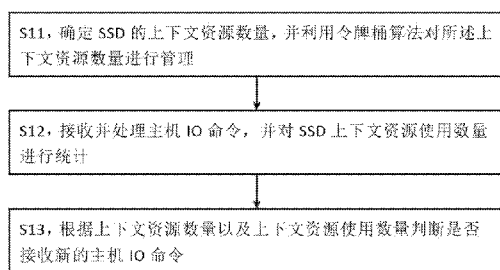
(74) 代理人: 北京市万慧达律师事务所(**BEIJING WANHUIDA LAW FIRM**); 中国北京市海淀区中关村南大街1号友谊宾馆颐园写字楼2楼, Beijing 100873 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,

(54) **Title:** METHOD AND APPARATUS FOR IMPROVING QUALITY OF SERVICE OF SSD, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 提升SSD服务质量方法、装置、计算机设备和存储介质



S11 Determine the number of context resources of an SSD, and manage the number of context resources by using a token bucket algorithm
S12 Receive and process a host IO command, and count the number of used context resources of the SSD
S13 Determine, according to the number of context resources and the number of used context resources, whether to receive a new host IO command

图 1

(57) **Abstract:** The present application relates to a method and apparatus for improving the quality of service of an SSD, and a computer device and a storage medium. The method comprises: determining the number of context resources of an SSD, and managing the number of context resources by using a token bucket algorithm; receiving and processing a host IO command, and counting the number of used context resources of the SSD; and determining, according to the number of context resources and the number of used context resources, whether to receive a new host IO command.

(57) **摘要:** 本申请涉及一种提升 SSD 服务质量的方法、装置、计算机设备和存储介质。所述方法包括: 确定 SSD 的上下文资源数量, 并利用令牌桶算法对所述上下文资源数量进行管理; 接收并处理主机 IO 命令, 并对 SSD 上下文资源使用数量进行统计; 根据所述上下文资源数量以及所述上下文资源使用数量判断是否接收新的主机 IO 命令。

RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

- 包括国际检索报告(条约第21条(3))。

提升 SSD 服务质量方法、装置、计算机设备和存储介质

相关申请的交叉引用

本申请要求于 2021 年 11 月 04 日提交中国专利局，申请号为 202111301938.3，申请名称为“提升 SSD 服务质量方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术领域，特别是涉及一种提升 SSD 服务质量的方法、装置、计算机设备和存储介质。

背景技术

大数据时代背景下，大数据保存和获取对存储设备提出了更高的性能要求，SSD 作为主流存储设备应运而生，从上层应用到底层存储，每一个环节的性能都直接或间接的影响着终端用户体验，因此，如何提高数据存储和读取的性能显得至关重要。

在计算机领域，相对于进程而言，上下文即为进程执行时的环境，也就是各种变量和数据，如：所有的寄存器变量、进程打开的文件、内存信息等。也可以将上下文理解为环境的一个快照，是一个用来保存状态的对象。通常，在一个程序中大多数函数都不是独立运行的，而是需要同其他函数的一部分进行交互，且需要其他外部环境变量的支持，而上下文即用于给外部环境的变量赋值，使函数能够正常运行。

发明人意识到，SSD 对主机 IO 的处理需要上下文来管理，处理 IO 的过程因为要有元数据的更新和获取等操作，所以此过程相对比较复杂，而为了提高 SSD 的性能这些操作需要实现并发操作和状态机的管理，这就需要上下文来记录和管理每个状态机，受限于硬件资源和能力的限制，这些上下文资源是有限的，当处理不同主机 IO 的工作负载时，可能会有拿不到上下文资源的情况，这就会使状态机进入挂起状态，处理这些处于挂起状态的状态机，排队累积效应恶化了对应 IO 命令的相应速度，从而影响了 SSD 的 QOS。

发明内容

本申请提供了一种提升 SSD 服务质量的方法，该方法包括：

确定 SSD 的上下文资源数量，并利用令牌桶算法对上下文资源数量进行管理；

接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统计；和

根据上下文资源数量以及上下文资源使用数量判断是否接收新的主机 IO 命令。

在其中一个实施例中，上述方法还包括对 SSD 的状态机挂起数量进行统计，其中根据上下文资源数量以及上下文资源使用数量判断是否接收新的主机 IO 命令，包括：

根据上下文资源数量、上下文资源使用数量以及状态机挂起数量判断是否接收新的主机 IO 命令。

在其中一个实施例中，SSD 的上下文资源包括第一上下文资源和第二上下文资源，第一上下文资源用于管理可接收的主机 IO 命令数量，第二上下文资源用于管理 IO 命令拆分出的数据结构。

在其中一个实施例中，确定 SSD 的上下文资源数量，包括：

初始化第一上下文资源数量和第二上下文资源数量；

根据第一上下文资源初始数量以及第二上下文资源初始数量进行读写 IOPS 测试；

保持第一上下文资源数量不变以及不断增加第二上下文资源数量，并继续进行 IOPS 测试直至吞吐量不再增加；

将吞吐量最大所对应的最小第二上下文资源数量作为第二上下文资源数量；和

根据第二上下文资源数量确定第一上下文资源数量。

在其中一个实施例中，第二上下文资源数量与第一上下文资源数量的比例为 1:32。

在其中一个实施例中，对 SSD 的状态机挂起数量进行统计，包括：

统计第二上下文资源使用数量，当第二上下文资源使用数量达到第二上下文资源数量时，如果 SSD 接收到新的主机 IO 命令，则 SSD 的状态机挂起数量加 1，如果 SSD 处理完一个主机 IO 命令，则 SSD 的状态机挂起数量减 1。

在其中一个实施例中，根据上下文资源数量、上下文资源使用数量以及状态机挂起数量判断是否接收新的主机 IO 命令，包括：

在第一上下文资源使用数量达到第一上下文资源数量或状态机挂起数量达到预设阈值时，SSD 停止接收新的主机 IO 命令。

本申请还提供了一种提升 SSD 服务质量的装置，装置包括资源确定模块、资源统计模块、判断模块，其中：

资源确定模块，用于确定 SSD 的上下文资源数量，并利用令牌桶算法对上下文资源数量进行管理；

资源统计模块，用于接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统

计；

判断模块，用于根据上下文资源数量以及上下文资源使用数量判断是否接收新的主机 IO 命令。

本申请还提供了一种计算机设备，包括存储器及一个或多个处理器，存储器中储存有计算机可读指令，上述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行上述任意一项实施例提供的提升 SSD 服务质量的方法的步骤。

本申请最后还提供了一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行上述任意一项实施例提供的提升 SSD 服务质量的方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对现有技术和实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请根据一个或多个实施例提供的提升 SSD 服务质量的方法的流程示意图；

图 2 为本申请根据一个或多个实施例提供的提升 SSD 服务质量的方法的流程示意图；

图 3 为本申请根据一个或多个实施例提供的提升 SSD 服务质量的装置的结构框图；

图 4 为本申请根据一个或多个实施例提供的计算机设备的内部结构图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请的主要构思在于：

在处理 HOST 主机 IO 的 SQ 时，利用令牌桶算法管理 SSD 内部的上下文资源，当桶内令牌不足或超额使用时，不再接收处理 SQ 内新的 IO 命令。

在一个实施例中，如图 1 所示，提供了一种提升 SSD 服务质量的方法，以该方法应用于计算机设备为例进行说明，包括以下步骤：

步骤 S11，确定 SSD 的上下文资源数量，并利用令牌桶算法对上下文资源数量进行管理；

具体的，上下文资源的数量即令牌桶容量根据后端 nand flash 处理能力和并发度确定。

在一个实施例中，SSD 的上下文资源包括第一上下文资源和第二上下文资源，第一上下文资源用于管理可接收的主机 IO 命令数量，第二上下文资源用于管理 IO 命令拆分出的数据结构。

具体的，X 和 Y 表示的是两种上下文资源，X 用于管理 IO cmd 命令，Y 用于管理 IO 命令拆分出来的数据结构资源，举例说明：主机下来的区块大小为 128K，下来的这 128K 用 X 上下文来管理，SSD 接下 128K 后，会把 128K 拆分成 32 个 4K（SSD 内部的数据管理单元是 4K），32 个 4K 用 Y 上下文资源来管理。

在一个实施例中，步骤 S11 中确定 SSD 的上下文资源数量，包括：

步骤 S111，初始化第一上下文资源数量和第二上下文资源数量；

步骤 S112，根据第一上下文资源初始数量以及第二上下文资源初始数量进行读写 IOPS 测试；

步骤 S113，保持第一上下文资源数量不变以及不断增加第二上下文资源数量，并继续进行 IOPS 测试直至吞吐量不再增加；

步骤 S114，将吞吐量最大所对应的最小第二上下文资源数量作为第二上下文资源数量；

步骤 S115，根据第二上下文资源数量确定第一上下文资源数量。

具体的，X 资源根据吞吐量测试时主机的最大压力（以顺序读写的吞吐量来计算），最大压力下 $4\text{jobs} \times 256\text{IOD} = 1024$ 个，SSD 内部初始 X 资源为 1024 个，这样保证至少处理一次主机的 IOD（IOD 为 HOST 的 IO 队列深度）。

Y 资源根据 FTL 和后端处理能力，假设后端处理能力的并发度为 n ，为了让 IO 命令 ping-pang 起来，将后端带宽压满，前端上下文资源的数量为 $2n \sim 3n$ ，实际可根据测试情况，在随机读写的 IOPS 测试时，不断的加大上下文资源数量，直到增大到吞吐量不再增加。

在一个实施例中，第二上下文资源数量与第一上下文资源数量的比例为 1:32。

具体的，如果主机下来的区块大小为 4K，那个 X 资源与 Y 资源是一一对应的关系。

步骤 S12，接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统计；

具体的，在 SSD 固件中增加上下文资源数量的统计，刚上电时初始化为 1 中确定的资源数 X，SSD 接收并处理 IO 时，会申请上下文资源，申请一次上下文资源，X 减 1，当 IO 完成释放资源时，X 加 1。

步骤 S13，根据上下文资源数量以及上下文资源使用数量判断是否接收新的主机 IO 命令。

在一个实施例中，响应于第一上下文资源使用数量达到第一上下文资源数量，SSD 停止接收新的主机 IO 命令。

在一个实施例中，若第一上下文资源使用数量达到第一上下文资源数量，SSD 停止接收新的主机 IO 命令。

具体的，当上下文资源 X 耗尽时，不再接收处理新的 IO 命令，把 IO 命令堵在 SQ 内，避免处理大量挂起引发的额外开销。

在一个实施例中，步骤 S3 之前还包括：对 SSD 的状态机挂起数量进行统计。

在本实施例中，如图 2 所示，提供了另一种提升 SSD 服务质量的方法，包括：

步骤 S21，确定 SSD 的上下文资源数量，并利用令牌桶算法对上下文资源数量进行管理；

具体的，涉及步骤 S21 的具体内容同步骤 S11 的内容。

步骤 S22，接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统计；

具体的，涉及步骤 S22 的具体内容同步骤 S12 的内容。

步骤 S23，对 SSD 的状态机挂起数量进行统计；

在一个实施例中，步骤 S23 包括：

统计第二上下文资源使用数量，当第二上下文资源使用数量达到第二上下文资源数量时，如果 SSD 接收到新的主机 IO 命令，则 SSD 的状态机挂起数量加 1，如果 SSD 处理完一个主机 IO 命令，则 SSD 的状态机挂起数量减 1。

步骤 S24，根据上下文资源数量、上下文资源使用数量以及状态机挂起数量判断是否接收新的主机 IO 命令。

在一个实施例中，若第一上下文资源使用数量达到第一上下文资源数量或状态机挂起数量达到预设阈值时，SSD 停止接收新的主机 IO 命令。

具体的，当上下文资源 X 耗尽或状态机挂起数量 Z 达到一定阈值时，不再接收处理

新的 IO 命令，把 IO 命令堵在 SQ 内，避免处理大量挂起引发的额外开销，其中挂起数量阈值根据实际测试的经验值设定，例如为 10。

此外，对挂起状态的处理需要采用 FIFO 的管理模式，保证 IO 的命令的处理顺序，保证先到的 IO 命令能尽快得到处理；FIFO 管理是先挂起的状态机先得到处理，保证顺序。

上述提升 SSD 服务质量的方法中，借助令牌桶算法提升了 SSD 的 QOS，保证了用户的良好体验，提升了产品竞争力。

应该理解的是，虽然图 1-2 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 1-2 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在一个实施例中，如图 3 所示，提供了一种提升 SSD 服务质量的装置，包括：装置包括资源确定模块、资源统计模块、判断模块：

资源确定模块，用于确定 SSD 的上下文资源数量，并利用令牌桶算法对上下文资源数量进行管理；

资源统计模块，用于接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统计；

在其中一个实施例中，SSD 的上下文资源包括第一上下文资源和第二上下文资源，第一上下文资源用于管理可接收的主机 IO 命令数量，第二上下文资源用于管理 IO 命令拆分出的数据结构。

判断模块，用于根据上下文资源数量以及上下文资源使用数量判断是否接收新的主机 IO 命令。

在其中一个实施例中，资源统计模块还用于对 SSD 的状态机挂起数量进行统计。

在其中一个实施例中，判断模块还用于根据上下文资源数量、上下文资源使用数量以及状态机挂起数量判断是否接收新的主机 IO 命令。

在其中一个实施例中，资源确定模块用于：

初始化第一上下文资源数量和第二上下文资源数量；

根据第一上下文资源初始数量以及第二上下文资源初始数量进行读写 IOPS 测试；

保持第一上下文资源数量不变以及不断增加第二上下文资源数量，并进行 IOPS 测试直至吞吐量不再增加；

将吞吐量最大所对应的最小第二上下文资源数量作为第二上下文资源数量；

根据第二上下文资源数量确定第一上下文资源数量。

在其中一个实施例中，资源确定模块中第二上下文资源数量与第一上下文资源数量的比例为 1:32。

在其中一个实施例中，资源确定模块用于：

统计第二上下文资源使用数量，当第二上下文资源使用数量达到第二上下文资源数量时，如果 SSD 接收到新的主机 IO 命令，则 SSD 的状态机挂起数量加 1，如果 SSD 处理完一个主机 IO 命令，则 SSD 的状态机挂起数量减 1。

在其中一个实施例中，判断模块用于：

若第一上下文资源使用数量达到第一上下文资源数量或状态机挂起数量达到预设阈值时，SSD 停止接收新的主机 IO 命令。

关于提升 SSD 服务质量的装置的具体限定可以参见上文中对于提升 SSD 服务质量的方法的限定，在此不再赘述。上述提升 SSD 服务质量的装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是终端，其内部结构图可以如图 4 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口、显示屏和输入装置。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机可读指令。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种提升 SSD 服务质量的方法。该计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏，该计算机设备的输入装置可以是显示屏上覆盖的触摸层，也可以是计算机设备外壳上设置的按键、轨迹球或触控板，还可以是外接的键盘、触控板或鼠标等。

本领域技术人员可以理解，图 4 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布

置。

本申请实施例还提供了一种非易失性计算机可读存储介质，该非易失性计算机可读存储介质存储有计算机可读指令，该计算机可读指令计算机程序被一个或多个处理器执行时可实现上述任意一个实施例提供的 SSD 服务质量的方法的步骤。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权 利 要 求 书

1. 一种提升 SSD 服务质量的方法，其特征在于，包括：
确定 SSD 的上下文资源数量，并利用令牌桶算法对所述上下文资源数量进行管理；
接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统计；和
根据所述上下文资源数量以及所述上下文资源使用数量判断是否接收新的主机 IO 命令。
2. 根据权利要求 1 所述的提升 SSD 服务质量的方法，其特征在于，所述方法还包括对 SSD 的状态机挂起数量进行统计，其中所述根据所述上下文资源数量以及所述上下文资源使用数量判断是否接收新的主机 IO 命令，包括：
根据所述上下文资源数量、所述上下文资源使用数量以及所述状态机挂起数量判断是否接收新的主机 IO 命令。
3. 根据权利要求 2 所述的提升 SSD 服务质量的方法，其特征在于，所述 SSD 的上下文资源包括第一上下文资源和第二上下文资源，所述第一上下文资源用于管理可接收的主机 IO 命令数量，所述第二上下文资源用于管理 IO 命令拆分出的数据结构。
4. 根据权利要求 3 所述的提升 SSD 服务质量的方法，其特征在于，所述确定 SSD 的上下文资源数量，包括：
初始化第一上下文资源数量和第二上下文资源数量；
根据第一上下文资源初始数量以及第二上下文资源初始数量进行读写 IOPS 测试；
保持第一上下文资源数量不变以及不断增加第二上下文资源数量，并继续进行 IOPS 测试直至吞吐量不再增加；
将吞吐量最大所对应的最小第二上下文资源数量作为第二上下文资源数量；和
根据所述第二上下文资源数量确定第一上下文资源数量。
5. 根据权利要求 4 所述的提升 SSD 服务质量的方法，其特征在于，第二上下文资源数量与第一上下文资源数量的比例为 1:32。
6. 根据权利要求 4 所述的提升 SSD 服务质量的方法，其特征在于，对 SSD 的状态机挂起数量进行统计，包括：
统计第二上下文资源使用数量，当第二上下文资源使用数量达到第二上下文资源数量时，如果 SSD 接收到新的主机 IO 命令，则 SSD 的状态机挂起数量加 1，如果 SSD 处理完一个主机 IO 命令，则 SSD 的状态机挂起数量减 1。
7. 根据权利要求 4 所述的提升 SSD 服务质量的方法，其特征在于，所述根据所述上

下文资源数量、所述上下文资源使用数量以及所述状态机挂起数量判断是否接收新的主机 IO 命令，包括：

在第一上下文资源使用数量达到第一上下文资源数量或状态机挂起数量达到预设阈值时，SSD 停止接收新的主机 IO 命令。

8. 一种提升 SSD 服务质量的装置，其特征在于，所述装置包括资源确定模块、资源统计模块、判断模块，其中：

所述资源确定模块，用于确定 SSD 的上下文资源数量，并利用令牌桶算法对所述上下文资源数量进行管理；

所述资源统计模块，用于接收并处理主机 IO 命令，并对 SSD 上下文资源使用数量进行统计；和

所述判断模块，用于根据所述上下文资源数量以及所述上下文资源使用数量判断是否接收新的主机 IO 命令。

9. 一种计算机设备，其特征在于，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1-7 任意一项所述的方法的步骤。10. 一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1-7 任意一项所述的方法的步骤。

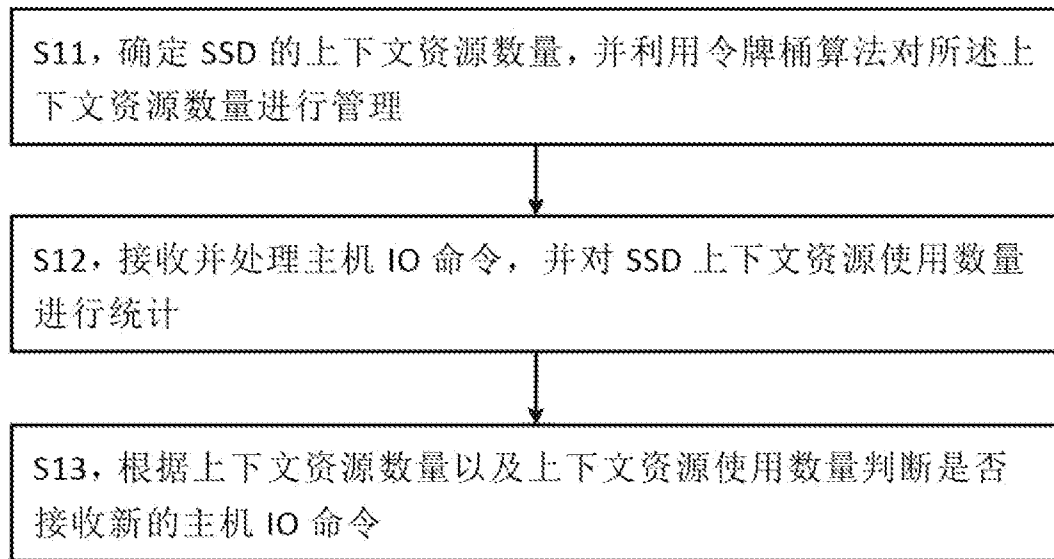


图 1

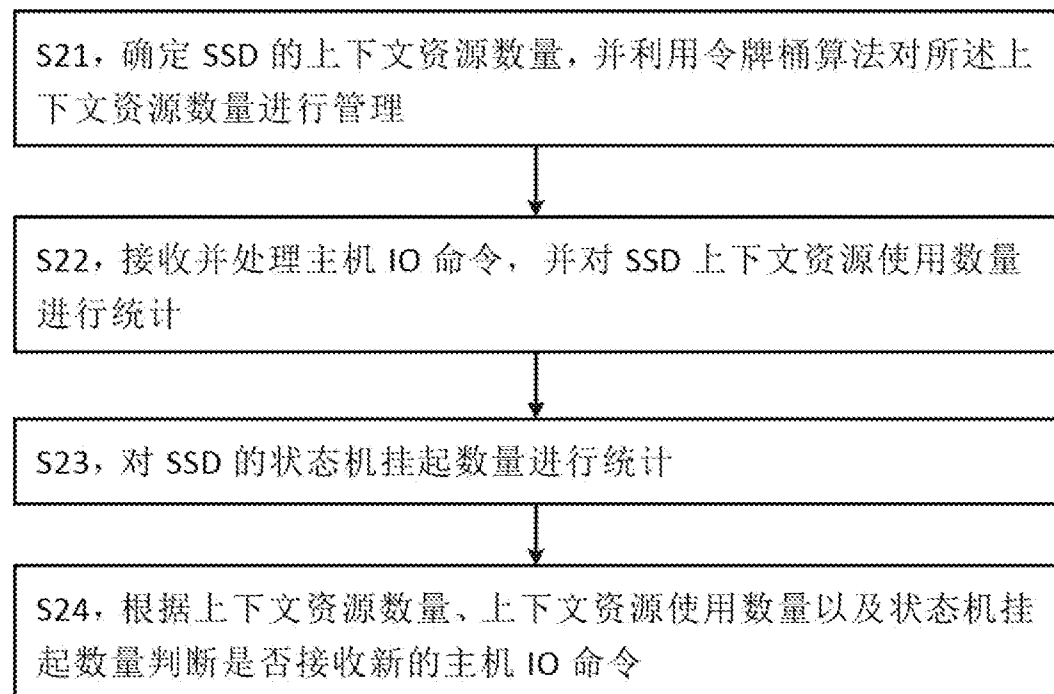


图 2

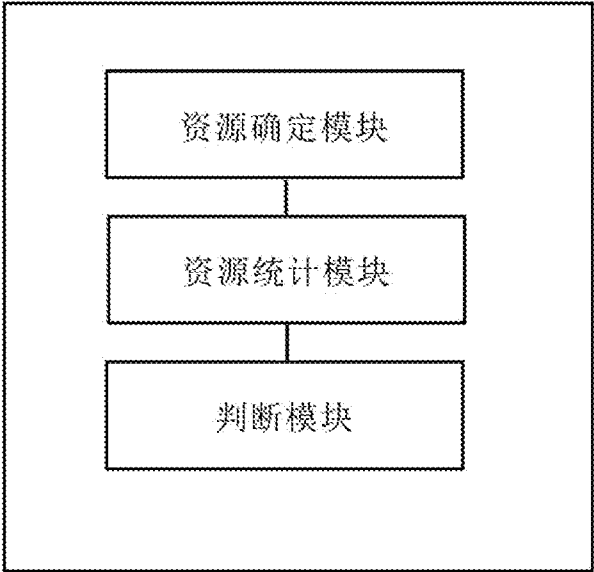


图 3

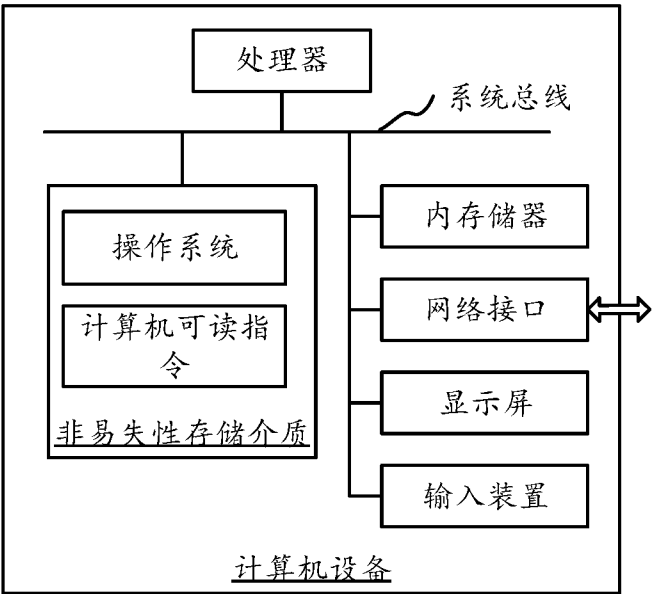


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/090021

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/4401(2018.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT; CNABS; ENTXT; VEN; CNKI; IEEE: 固态硬盘, 上下文, 令牌桶, 输入输出, 新, 状态机, 挂起, 拆分, 每秒进行读写操作的次数, 吞吐量, SSD, context, token bucket, input output, IO, new, state machine, suspend, split, divide, IOPS, throughput

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113741991 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 03 December 2021 (2021-12-03) claims 1-10	1-10
X	CN 112306415 A (CHENGDU BIWIN STORAGE TECHNOLOGY CO., LTD.) 02 February 2021 (2021-02-02) description, paragraphs 34-104, and figure 7	1, 2, 8-10
X	US 10268419 B1 (EMC IP HOLDING COMPANY L.L.C.) 23 April 2019 (2019-04-23) description, columns 1-3 and 5-8	1, 2, 8-10
Y	CN 112306415 A (CHENGDU BIWIN STORAGE TECHNOLOGY CO., LTD.) 02 February 2021 (2021-02-02) description, paragraphs 34-104, and figure 7	3, 9-10
Y	US 10268419 B1 (EMC IP HOLDING COMPANY L.L.C.) 23 April 2019 (2019-04-23) description, columns 1-3 and 5-8	3, 9-10
Y	CN 105900061 A (HUAWEI TECHNOLOGIES CO., LTD.) 24 August 2016 (2016-08-24) description, paragraphs 76-171	3, 9-10
A	US 2010223624 A1 (IBM) 02 September 2010 (2010-09-02) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

11 July 2022

Date of mailing of the international search report

22 July 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/090021

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113741991	A	03 December 2021	CN	113741991	B	18 February 2022
CN	112306415	A	02 February 2021	None			
US	10268419	B1	23 April 2019	None			
CN	105900061	A	24 August 2016	US	2018018293	A1	18 January 2018
				WO	2016061766	A1	28 April 2016
				US	2017220500	A1	03 August 2017
				EP	3128420	A1	08 February 2017
				JP	2017514210	A	01 June 2017
				IN	201647027878	A	07 October 2016
				CN	105900061	B	16 January 2018
US	2010223624	A1	02 September 2010	US	8424014	B2	16 April 2013

国际检索报告

国际申请号

PCT/CN2022/090021

A. 主题的分类

G06F 9/4401 (2018.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNTXT;CNABS;ENTXT;VEN;CNKI;IEEE:固态硬盘, 上下文, 令牌桶, 输入输出, 新, 状态机, 挂起, 拆分, 每秒进行读写操作的次数, 吞吐量, SSD, context, token bucket, input output, IO, new, state machine, suspend, split, divide, IOPS, throughput

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 113741991 A (苏州浪潮智能科技有限公司) 2021年12月3日 (2021 - 12 - 03) 权利要求1-10	1-10
X	CN 112306415 A (成都佰维存储科技有限公司) 2021年2月2日 (2021 - 02 - 02) 说明书第34-104段, 附图7	1-2, 8-10
X	US 10268419 B1 (EMC IP HOLDING CO LLC) 2019年4月23日 (2019 - 04 - 23) 说明书第1-3、5-8栏	1-2, 8-10
Y	CN 112306415 A (成都佰维存储科技有限公司) 2021年2月2日 (2021 - 02 - 02) 说明书第34-104段, 附图7	3, 9-10
Y	US 10268419 B1 (EMC IP HOLDING CO LLC) 2019年4月23日 (2019 - 04 - 23) 说明书第1-3、5-8栏	3, 9-10
Y	CN 105900061 A (华为技术有限公司) 2016年8月24日 (2016 - 08 - 24) 说明书第76-171段	3, 9-10
A	US 2010223624 A1 (IBM) 2010年9月2日 (2010 - 09 - 02) 全文	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2022年7月11日

国际检索报告邮寄日期

2022年7月22日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

杨牛

电话号码 86-(20)-28950388

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/090021

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	113741991	A	2021年12月3日	CN 113741991 B	2022年2月18日
CN	112306415	A	2021年2月2日	无	
US	10268419	B1	2019年4月23日	无	
CN	105900061	A	2016年8月24日	US 2018018293 A1	2018年1月18日
				WO 2016061766 A1	2016年4月28日
				US 2017220500 A1	2017年8月3日
				EP 3128420 A1	2017年2月8日
				JP 2017514210 A	2017年6月1日
				IN 201647027878 A	2016年10月7日
				CN 105900061 B	2018年1月16日
US	2010223624	A1	2010年9月2日	US 8424014 B2	2013年4月16日