附录1：（封面、封底用120克白色铜版纸打印，无须彩打。所有红字在打印之前必须删除）



**本科生毕业设计[论文]**

（华文中宋小初号加粗居中）

**（题目）非易失性内存文件系统的并行性优化研究**（黑体2号加粗居中）

院 系 计算机科学与技术学院

专业班级 CS1410

姓 名 彭 周 旋

学 号 U201414833

指导教师 冯 丹

2018年 4 月 8 日（华文中宋3号居中）

**学位论文原创性声明**

（黑体小2号加粗居中）

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

（宋体小4号）

作者签名： 年 月 日

**学位论文版权使用授权书**

（黑体小2号加粗居中）

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密囗，在 年解密后适用本授权书

2、不保密囗 。

（请在以上相应方框内打“√”）

（宋体小4号）

作者签名： 年 月 日

导师签名： 年 月 日

(注：此页内容装订在论文扉页)

**摘 要**

新兴的非易失性内存技术既具有与传统DRAM主存相近的低延迟和字节寻址特性，又具有类似硬盘等外存储设备的非易失存储特性，已逐渐对系统架构、系统软件、库和应用程序等产生深远的影响，并引起广泛的关注和研究。

非易失内存文件系统（Persistent Memeory File System，简称PMFS）是一个针对非易失内存（Persisitent Memory，以下简称PM）设计的轻量级的、符合POSIX文件系统标准的文件系统，该文件系统利用了PM的可字节寻址特性避免了面向块存储的开销，并使得应用程序通过内存映射I/O能够直接访问PM。PMFS的设计主要针对非易失内存进行优化，在作为文件系统的设计上许多地方采取了简化的策略（如分配器），存在许多性能上的问题。因此本文以PMFS为基础，重点对PMFS的分配器等进行优化研究，使得PMFS获得更好的并行性及更高的效率。

PMFS的分配器采用线性链表的方式组织文件系统的数据块，并在已分配的数据块上建立索引，在分配据块时通过索引来加速查找过程。由于PMFS中分配器的频繁调用，这种时间复杂度趋近于O(n)的组织方式存在严重的效率问题。本文中提出一种优化的方案可以将分配过程的时间复杂度降为O(1)，并且不提高回收过程的时间复杂度。在优化方案中，取消在已分配块上建立的索引，而在空闲块上建立索引，由于PMFS分配器分配块时有固定的三种大小：4K, 2M, 1G，因此可将连续的空闲块根据长度拆分成三条链表，在分配块时可直接在对应的链表取出表头即可满足要求，无需遍历查找，以达到提高效率的目的。实验结果表明，这种优化的方案以较小的空间代价，取得了性能上的明显提升。

（宋体小4号，行间距固定1.5倍行距，字符间距为标准）

**关键词：非易失内存**；**PMFS**；**分配器**

（黑体4号加粗） （宋体小4号）

**Abstract**（Time New Roman小2号加粗居中）

××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××.

××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××.

（Time New Roman小4号，行间距固定1.5倍行距，字符间距为标准）

**Key Words：**××××; ××××; ××××; ××××

（Time New Roman 4号加粗） （Time New Roman小4号）

**目 录**（黑体小2号加粗居中）

**摘要** Ⅰ

**Abstract** Ⅱ

**1□□□绪论** 1

1.1□□×××××× 1

1.2□□×××××× 3

1.2.1□□×××××× 7

1.3□□×××××× 10

**……**

**……**

**……**

3**□□□××××××** 20

3.1□□×××××× 20

3.2□□×××××× 23

3.2.1□□×××××× 25

3.3□□×××××× 30

**……**

**……**

**4□□□结论** 40

**致谢** 42

**参考文献** 44

**附录□××××××** 45

（章为宋体小4号加粗，其余宋体小4号，字母、阿拉伯数字为Time New Roman小4号）

**1 绪论**（黑体小2加粗居中）

**1.1 课题背景**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

随着非易失内存技术的兴起和普及，计算机存储体系结构正面临着全新的发展机遇。传统的存储体系按访存延时从低到高的顺序由缓存，主存和辅存等三个层次构成，然而，非易失内存的的出现将改变这一局面。非易失内存在存储体系中可以扮演多种角色：1）由于非易失内存可以提供与DRAM在同一个数量级上的性能，因此非易失内存可以作为当前主流的DRAM主存的替代品；2）作为存储体系结构中位于主存（DRAM）和外存（机械硬盘或固态硬盘）之间的一个全新的存储层级；3）作为一个独立的内存区域，与DRAM和硬盘等并行存在[6]。非易失内存已经对系统架构、系统软件、库和应用程序产生深远的影响[7,8,9,10]。由于非易失内存在存储体系中的灵活应用，计算机如何管理非易失内存的话题也逐渐得到重视。传统上，操作系统将易失性的内存（虚拟内存管理器VMM）和外存（文件系统，块设备驱动）的管理分开，而非易失内存兼顾内存的字节寻址和外存的非易失存储的特性，非易失内存的管理有多种选择：1）扩展VMM用于管理非易失内存；2）为非易失内存实现一个块设备从而使用现有的的文件系统来管理；3）实现一个针对非易失内存优化的、不经过块层的文件系统。Subramanya R Dulloor等人采取了第三种策略，实现了一个轻量级、符合POSIX文件系统标准的非易失内存文件系统，即PMFS[3]。该文件系统利用了非易失内存可字节寻址的特性避免了面向块存储的开销，并且使得应用程序通过内存映射I/O能够直接访问非易失内存。相较于另外两种方式，PMFS具有许多优势：1）支持传统应用程序。PMFS实现了许多存储密集型应用程序所依赖的完整的符合POSIX标准的文件系统接口；2）支持轻量级的文件系统。通过针对PM进行优化和避免块层，PMFS消除了复制开销，并为传统的应用程序提供了大量的好处；3）优化的内存映射I/O。PMFS通过直接将PM的页面映射到应用程序的地址空间中去以避免内存映射时的复制页面开销，同时PMFS实现了诸如透明的大页面支持等特征，以进一步优化内存映射I/O。本课题的研究正是基于PMFS展开的，尽管PMFS已经取得了可喜的成绩，但是其本身仍然存在许多可优化的地方，对于PMFS的研究也远远没有到尽头。本课题提出对PMFS的并行性（在Subramanya R Dullor论文中未提到的）进行研究，并尝试进行优化。对PMFS的进一步研究和优化，有利于非易失内存相关的技术以及PMFS本身更快的发展和普及。

（宋体小4号，行间距固定1.5倍行距，字符间距为标准）

**1.2 国内外研究现况及发展趋势**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

当前国际上针对非易失内存设计的文件系统主要有SCMFS，BPFS，Object-based File System以及PMFS。Wu Xiaojian等人根据SCM（Storage Class Memory）特性提出了专门设计应用于SCM设备的文件系统，简称SCMFS[12]，其中SCM设备直接连接到内存总线上，既用作内存设备，又可以作永久性存储设备。SCM使用传统文件系统接口，挂载在系统虚拟地址空间上，并通过内存管理模块MMU[13]完成虚拟地址到物理地址的映射，大大简化了SCMFS的开发。

基于可字节寻址的持久RAM（BPRAM）文件系统BPFS由Condit J等人提出[14]，BPFS根据BtrFS组织元数据的机制，使用两层树结构的空间框架，并使用短电路影子分页的技术提供可靠性保证。BPFS在文件管理方面比传统文件系统具有更好的性能，有较好的数据持久性和一致性，但其实现需要有相应的硬件支持，移植性不好，且文件系统实现使用的PCM比DRAM访问速度慢，使得文件系统影响了系统的整体性能。

Object-based File System由UCSC（University of California，Santa Cruz）的Yangwook Kang等人提出[15,16]，采用基于对象的方式设计文件系统，具有较好的可移植性。它将文件管理与存储管理分离，让存储管理根据应用与设备特性做优化，取代文件系统成为所有相关应用的底层，减少软件层次，简化应用程序设计。

Subramanya R Dullor等人提出的PMFS是一个轻量级的、符合POSIX文件系统标准的非易失内存文件系统。在PMFS的实现中，Subramanya R Dullor等人做了大量的研究和努力，使得PMFS相较于一些传统的文件系统在文件I/O、文件应用、内存映射I/O等方面的性能上取得了令人欣喜的成绩。PMFS的突出贡献包括利用处理器的分页和内存排序功能进行了优化，例如针对一致性而使用的精细日志记录，为更快的内存映射I/O而实现的透明大页面支持等；实现了一个硬件原语（PM写屏障）用于提供强一致性保证；使用处理器现有功能保护非易失内存免于杂散写入等。

而目前针对非易失内存的研究趋势，主要可以分为两类：（1）兼容传统主存控制器的非易失内存（Non-Volatile Memory, NVM）；（2）连接在I/O总线上的存储级内存（Storage-Class-Memory, SCM）。NVM的特点是可直接替换传统的DRAM内存或两者同时使用构成混合主存，其优势是NVM主存由硬件控制，只需针对NVM主存的特性对主存控制器进行适当的改进而无需引入系统级的改动。SCM是对介于传统DRAM主存和HDD外存之间的存储层次所使用的存储设备的统称。相比于DRAM，具备非易失、低价格、低功耗的优势；相比于HDD具备无运动部件、低延迟、高吞吐率的优势[4]。

**1.3 本课题主要工作**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

本论文主要研究非易失内存文件系统PMFS的分配器的组织方式和分配/回收策略，并探寻其中可进行并行性优化的要素，对二者提出优化的设计方案，并进行实现。

（1）优化分配器的组织方式。PMFS分配器将所有的数据块（包括已分配和空闲的）组织成一条线性链表，并在连续已分配块上建立的索引，当分配器进行块的分配或回收时就根据索引依次遍历该链表，直到找到满足需求的块。而在优化方案中，分配器在空闲块上建立索引，并根据空闲块的长度拆分为三条链表，消除了分配块时遍历空间带来的时间开销。

（2）优化的分配策略。PMFS分配器分配空闲块时每次都需要进行遍历，并且根据找到的空闲块位置采取左对齐、右对齐或居中等不同的策略，而居中策略会导致产生更多的碎片。在优化方案中，由于新的组织方式的良好特性，在分配时每次都直接从表头取空闲块而无需遍历，并且每次在需求块小于空闲块长度时都可以采取左对齐策略，减少碎片化。

**1.4 本文组织结构**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

全文一共分为六章，组织内容安排如下：

第一章阐述非易失内存及非易失内存文件系统的研究背景和意义，并介绍了国内外关于非易失内存文件系统的研究现况，以及非易失内存技术的发展趋势，说明了本文的主要工作。

第二章简要介绍了非易失内存文件系统PMFS，着重讲述PMFS的应用背景、总体布局，研究PMFS的分配器组织方式和分配/回收策略，发现PMFS分配器性能方面的不足，针对组织方式和分配策略提出优化方案。

第三章提出对PMFS分配器进行优化的各种不同策略，经过详细的分析和论证后确定最终优化方案，并详细阐述了PMFS组织方式和分配策略的优化设计方案。

第四章根据上述提出的优化方案，详细讲述了PMFS分配器组织方式和分配策略优化的具体实现。

第五章介绍了PMFS文件系统的测试方法和测试平台，对优化前后的性能进行了测试和对比，并详细分析了文件系统测试结果。

第六章对全文进行总结并展望未来工作。

.......

**2 PMFS文件系统概述**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

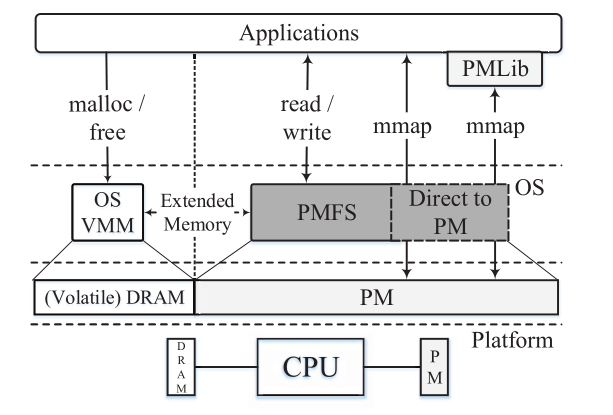
**2.1 PMFS的应用背景**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

由于新兴的非易失内存技术兼具大容量、字节编址、低延迟及存储非易失等特性，及其在传统存储体系结构层次中的参与方案的多样性和灵活性，计算机对于非易失内存的管理也呈现出多种选择。传统上，操作系统将内存和外存的管理分开，用虚拟内存管理器（Virtual Memory Manager，VMM）管理内存，用文件系统和块设备驱动等来管理外存储设备。然而非易失内存兼具内、外存的特性，在管理上，既可以扩展虚拟内存管理器来管理非易失内存，也可以为现有的文件系统（如Ext4）使用非易失内存实现块设备驱动，还可以在不经过块层的情况下实现一个针对非易失内存的文件系统。

正是在上述背景下，SR Dulloor等人设计了一个针对非易失内存进行优化的、符合POSIX标准的文件系统，即PMFS。并且提出PMFS作为管理非易失内存的系统软件层相较于其他方式具有许多优势。一方面，PMFS实现了完整的符合POSIX标准的文件系统接口，对传统的应用程序有更好的支持；另一方面，通过针对非易失内存进行优化和避免块层，PMFS消除了复制开销，支持轻量级的文件系统。此外，在传统的文件系统实现中，内存映射I/O将会先复制访问的页面到DRAM中，而PMFS可以通过直接将非易失内存的页面映射到应用程序的地址空间中以避免这种开销，PMFS还实现了其他的一些特征，以进一步优化内存映射I/O。

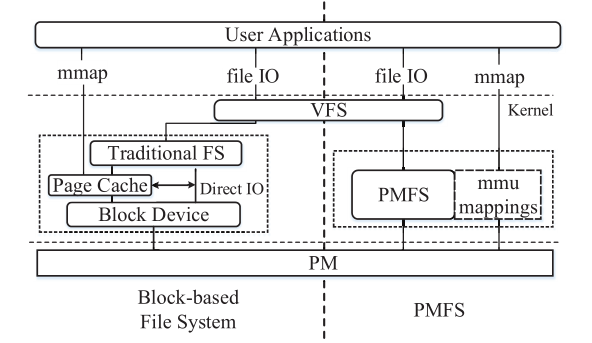
**2.2 PMFS在内核中的架构及空间布局**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

PMFS及其所管理的非易失内存在系统体系结构中的架构如图2-1所示。在该架构中，PMFS假设处理器能够同时支持易失性的DRAM和非易失内存PM，并且非易失内存作为主存扩展的一部分。操作系统的VMM继续管理DRAM，而PMFS负责管理PM。PMFS通过mmap实现应用程序对PM的直接访问，并且将为简化使用PM而设计的编程模型和库集成为PMLib。



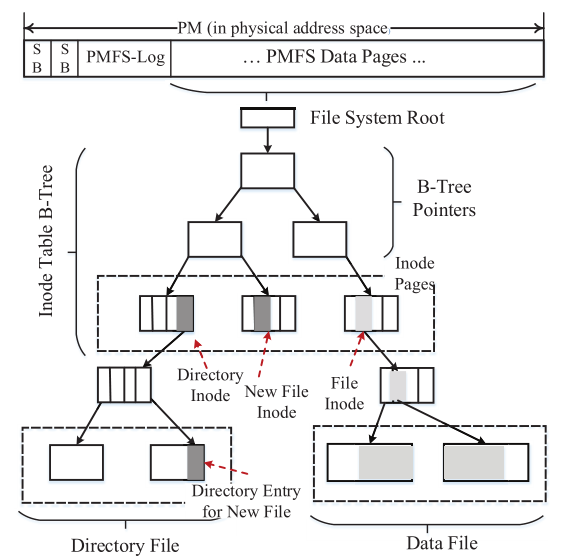
**图2-1 非易失内存系统体系架构**

PMFS设计之初有三个目标：1）优化字节寻址的存储。PMFS利用PM的字节寻址能力来避免传统文件系统基于块设计的开销；2）实现应用程序对PM的高效访问。PMFS通过避免不必要的复制和软件开销优化了文件读、写以及mmap；3）保护PM免于杂散写入。PMFS为了性能将整个PM映射到内核虚拟地址空间中，但这样可能导致PM遭受杂散写入而永久性损坏。为避免如此，PMFS使用处理器中的写保护控制功能实现了一个低开销写保护方案原型。与传统的基于块的文件系统对比，PMFS的设计避免了块设备层，并因此消除了页面复制的开销，最关键的是，PMFS提供给应用程序对PM的直接访问，图2-2展示了一个使用PMFS的系统的高级软件架构，包括与传统文件系统的对比。



**图2-2 PMFS与传统文件系统的对比**

PMFS的空间布局如图2-3所示。超级块及其冗余备份之后是日志（PMFS-Log）和动态分配的页面。和许多其他文件系统一样，PMFS中的元数据使用B树进行组织，B树是索引大量稀疏数据的最佳选择之一。B树用于表示inode表和inode中的数据。



**图2-3 PMFS的数据布局**

**2.3 PMFS分配器的设计与不足**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

**2.3.1 分配器的调用过程**

**2.3.2 分配器的组织方式**

**2.3.3 分配/回收策略**

**2.3.4 分配器效率分析**

**2.4 优化PMFS分配器的设计思路**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**2.5 本章小结**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**3 PMFS块组织方式及分配器优化方案**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**3.1□××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

××××××××××××××××××××××××××××××××××××××××××××××××，其×××××可表示如下：

 (3-1）

 (3-2）

×××××××××××××××××××××××××××× (如表3-1所示)

表3-1□××××××××××

|  |  |  |  |
| --- | --- | --- | --- |
| ××××× | ××× | ××× | ××× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |

（表标题：位于表格上方，黑体小4号，字母、阿拉伯数字为Time New Roman小4号，表内容：宋体5号，字母、阿拉伯数字为Time New Roman 5号）

××××××××××××××××××××××××××× (如图3-1所示)



**图3-1□××**××××××××

（图标题：位于图下方，黑体小4号，字母、阿拉伯数字为Time New Roman小4号）

--------章与章之间插入分页符----------

**4 PMFS块组织方式及分配器优化实现**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**4.1 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**4.2 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**4.3 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**4.4 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**5 性能测试与分析**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**5.1 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**5.2 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**5.3 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**5.4 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**6 总结和展望**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**2.1 ××××××**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

×××××××××××××××××××××××××××××××××

**致 谢**

**参考文献** (黑体小2号加粗居中)

[1]□Persistent Memory. <https://en.wikipedia.org/wiki/Persistent_memory>.

[2]□PMFS source code. <https://github.com/linux-pmfs/pmfs>.

[3]□Subramanya R Dulloor, Sanjay Kumar, Anil Keshavamurthy, Philip Lantz, Dheeraj Reddy, Rajesh Sankaran, Jeff Jackson. System Software for Persistent Memory. In EuroSys’14 Proceedings of the Ninth European Conference on Computer Systems,Article No.15, Amsterdam, The Netherlands – April 14-16,2014.

[4]□孙广宇,舒继武,王鹏.面向非易失内存的结构和系统级设计与优化综述.《华东师范大学学报（自然科学版）》,2014,5:72-81.

[5]□BAI Tongxin, DOND Zhenjiang, Cai Manyi, FAN Xiaopeng, XU Chengzhong, LIU Lixia.A Survey of System Software Techniques for Emerging NVMs.《ZTE Communications》,2017,1:35-42.

[6]□Sorin, Daniel J. Persistent Memory. Computer(00189162), 2017,50(3):12.

[7]□J. Coburn, A. M. Caulfield, A. Akel, L. M. Grupp, R. K.Gupta, R. Jhala, and S. Swanson. NV-Heaps: Making Persistent Objects Fast and Safe with Next-generation, Non-volatile Memories. In Proceedings of the Sixteenth International Conferenceon ArchitecturalSupportfor Programming Languages and Operating Systems, ASPLOS XVI, pages 105–118, 2011.

[8]□J. Condit, E. B. Nightingale, C. Frost, E. Ipek, B. Lee, D. Burger, and D. Coetzee. Better I/O Through Byte-addressable, Persistent Memory. In Proceedings of the ACM

SIGOPS 22nd Symposium on Operating Systems Principles,SOSP ’09, pages 133–146, 2009.

[9]□M. K. Qureshi, V. Srinivasan, and J. A. Rivers. Scalable High Performance Main Memory System Using Phase-change Memory Technology. In Proceedings of the 36th

Annual International Symposium on Computer Architecture,ISCA ’09, pages 24–33, 2009.

[10]□H. Volos, A. J. Tack, and M. M. Swift. Mnemosyne:Lightweight Persistent Memory. In Proceedings of the Six-teenth International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS XVI, pages 91–104, 2011.

[11]□FREITAS R F, WILCKE W W. Storage-class memory: The next storage system technology[J]. IBM Journal of Research and Development, 2008, 52(4.5): 439-447.

[12]□Wu X, Reddy A L. SCMFS: a file system for storage class memory. in: Proceedings of 2011 International Conference for High Performance Computing, Networking,Storage andAnalysis. Seatle,WA:IEEE, 2011.39~50

[13]□Dalinger I. Formal Verification of a Processor with Memory Management Units.

Vdm Verlag Dr Müller, 2008,1(1):33~35

[14]□Condit J, Nightingale E B, Frost C, et al. Better I/O through byte-addressable,

persistent memory. Proceedings of Sosp, 2009,2(3):133~146

[15]□Kang Y, Yang J, Miller E L. Object-based SCM: An efficient interface for Storage Class Memories. in:Proceedings of the 27th Symposium on Mass Storage Systems and Technologies. Denver, CO:IEEE, 2011.1~12

[16]□D. Nagle M. E. Factor, S. Iren, et al. The ANSI T10 object-based storage standard

and current implementations. IBM Journal of Research and Development, 2008,52(4):401~411

( 宋体小4号)

……

……

……

**附录** (黑体小2号加粗居中)

×××××××××××××××××××××××××××

( 宋体小4号)

……

……

……