[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 9/455 (2006.01)
G06F 12/10 (2006.01)

[12] 发明专利说明书

专利号 ZL 200680014922.1

[45] 授权公告日 2009 年 12 月 16 日

[11] 授权公告号 CN 100570563C

[22] 申请日 2006.5.4
[21] 申请号 200680014922.1
[30] 优先权
[32] 2005. 5. 5 [33] US [31] 11/122,801
[86] 国际申请 PCT/EP2006/062046 2006.5.4
[87] 国际公布 WO2006/117394 英 2006.11.9
[85] 进入国家阶段日期 2007.11.1
[73] 专利权人 国际商业机器公司
地址 美国纽约
[72] 发明人 W·J·阿姆斯特朗
R·L·阿恩特 M·J·克里甘
D·R·恩格布雷特森
T·R·马齐尼 N·纳亚尔
[56] 参考文献
US2002/0082824A1 2002.6.27
EP0423453A2 1991.4.24
审查员 王 荣
[74] 专利代理机构 北京市中咨律师事务所
代理人 于 静 李 峥

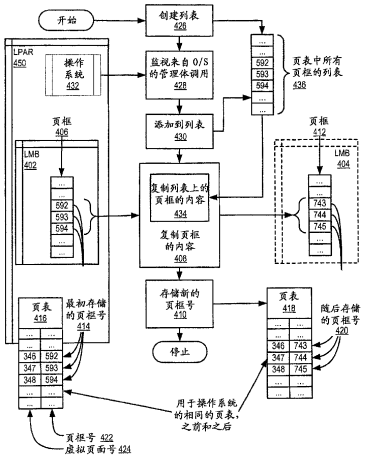
权利要求书 12 页 说明书 17 页 附图 8 页

[54] 发明名称

在具有动态逻辑分区的计算环境中管理计算机存储器

[57] 摘要

在具有动态逻辑分区的计算机中管理计算机存储器，其相对于逻辑分区中的操作系统透明地操作。 描述了用于在具有动态逻辑分区的计算机中管理计算机存储器的示例性方法、系统和产品，其包括：通过管理体，从逻辑分区(“LPAR”)的一个逻辑存储块(“LMB”)中的页框，将具有用于所述 LPAR 中的操作系统的页表中的页框号的页框的内容复制到所述 LMB 外部的页框。 实施例通常包括：在所述页表中存储新的页框号，包括通过所述管理体，为复制了其内容的每个页框存储标识出向其复制了内容的页框的新的页框号。 在典型的实施例中，复制页框的内容和存储新的页框号是相对于所述操作系统透明实现的。



1. 一种用于在具有动态逻辑分区的计算机中管理计算机存储器的方法，所述方法包括:

通过管理体，从逻辑分区的一个逻辑存储块中的页框，将具有用于所述逻辑分区中的操作系统的页表中的页框号的页框的内容复制到所述逻辑存储块外部的页框；以及

在所述页表中存储新的页框号，包括通过所述管理体存储标识出复制了内容的页框的新的页框号;

其中复制页框的内容和存储新的页框号是相对于所述操作系统透明实现的。

2. 根据权利要求 1 的方法，其进一步包括:

通过所述管理体创建所述页表中所有页框的列表;

通过所述管理体监视从所述操作系统到所述管理体的、将页框添加到所述页表的调用，此时所述管理体正在复制页框的内容和存储新的页框号;以及

将添加到所述页表的页框添加到所述列表;

其中复制页框的内容进一步包括复制所述列表上的页框的内容。

3. 根据权利要求 1 或 2 的方法，其中具有超过一个尺寸的存储页面被映射到所述逻辑存储块的页框，所述方法进一步包括:

将存储器管理中断从所述操作系统导引到所述管理体；以及

将用于所述操作系统的存储器管理操作从用于所述操作系统的页表切换到临时可选页表;

其中复制页框的内容进一步包括：复制与被映射到所述逻辑存储块的页框的页面中的最小页面具有相同尺寸的分段中的页框的内容。

4. 根据权利要求 3 的方法，其中复制页框的内容进一步包括:

从所述临时可选页表中删除同样处在用于所述操作系统的页表中的页框；以及

在用于所述操作系统的页表中存储这样的删除页框的状态比特。

5. 根据权利要求 1、2、4 中任何一项的方法，其中所述逻辑存储块的页框中的至少一个被映射用于直接存储器存取，并且复制页框的内容进一步包括:

在复制被映射用于直接存储器存取的页框的内容时，通过所述管理体来阻闭直接存储器存取操作；以及

在直接存储器存取映射表中为所述逻辑存储块的被映射用于直接存储器存取的每个页框存储标识复制了内容的页框的新的页框号。

6. 根据权利要求 3 的方法，其中所述逻辑存储块的页框中的至少一个被映射用于直接存储器存取，并且复制页框的内容进一步包括:

在复制被映射用于直接存储器存取的页框的内容时，通过所述管理体来阻闭直接存储器存取操作；以及

在直接存储器存取映射表中为所述逻辑存储块的被映射用于直接存储器存取的每个页框存储标识复制了内容的页框的新的页框号。

7. 根据权利要求 1、2、4 中任何一项的方法，其进一步包括：创建一段空闲的相连存储器，其既大于逻辑存储块并且又大得足以容纳页表。

8. 根据权利要求 3 的方法，其进一步包括：创建一段空闲的相连存储器，其既大于逻辑存储块并且又大得足以容纳页表。

9. 根据权利要求 5 的方法，其进一步包括：创建一段空闲的相连存储器，其既大于逻辑存储块并且又大得足以容纳页表。

10. 根据权利要求 6 的方法，其进一步包括：创建一段空闲的相连存储器，其既大于逻辑存储块并且又大得足以容纳页表。

11. 根据权利要求 7 的方法，其中创建一段空闲的相连存储器进一步包括: 通过所述管理体为两个或更多的相连逻辑存储块重复实现以下步骤:

通过所述管理体，将处在用于所述逻辑分区中的操作系统的页表中的逻辑存储块的页框的内容从所述逻辑存储块中的页框复制到所述逻辑存储块外部的页框;

在所述页表中存储新的页框号，包括通过所述管理体存储标识出复制

了内容的页框的新的页框号；以及

将所述逻辑存储块添加到空闲存储器的列表。

12. 根据权利要求 1、2、4 中任何一项的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框；以及

存储新的页框号进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号。

13. 根据权利要求 3 的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框；以及

存储新的页框号进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识复制了内容的页框的新的页框号。

14. 根据权利要求 5 的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临

时页框;

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框;以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框;以及

存储新的页框号进一步包括:对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者,存储标识复制了内容的页框的新的页框号。

15. 根据权利要求 6 的方法,其进一步包括:改善逻辑存储块对处理器的亲和性,其中:

复制所述逻辑存储块的页框的内容进一步包括:

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框;

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框;以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框;以及

存储新的页框号进一步包括:对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者,存储标识复制了内容的页框的新的页框号。

16. 根据权利要求 7 的方法,其进一步包括:改善逻辑存储块对处理器的亲和性,其中:

复制所述逻辑存储块的页框的内容进一步包括:

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框;

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框;以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框;以及

存储新的页框号进一步包括:对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者,存储标识复制了内容的页框的新的页框

号。

17. 根据权利要求 8 的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框；以及

存储新的页框号进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识复制了内容的页框的新的页框号。

18. 根据权利要求 9 的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框；以及

存储新的页框号进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识复制了内容的页框的新的页框号。

19. 根据权利要求 10 的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框；以及

存储新的页框号进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识复制了内容的页框的新的页框号。

20. 根据权利要求 11 的方法，其进一步包括：改善逻辑存储块对处理器的亲和性，其中：

复制所述逻辑存储块的页框的内容进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框；以及

存储新的页框号进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识复制了内容的页框的新的页框号。

21. 一种用于在具有动态逻辑分区的计算机中管理计算机存储器的装置，所述装置包括：

通过管理体从逻辑分区的一个逻辑存储块中的页框，将具有用于所述逻辑分区中的操作系统的页表中的页框号的页框的内容复制到所述逻辑存储块外部的页框的装置；以及

在所述页表中存储新的页框号的装置，包括通过所述管理体存储标识出复制了内容的页框的新的页框号；

其中复制页框的内容和存储新的页框号是相对于所述操作系统透明地的。

22. 根据权利要求 21 的装置，其进一步包括：

通过所述管理体创建所述页表中所有页框的列表的装置；

通过所述管理体监视从所述操作系统到所述管理体的、将页框添加到所述页表的调用的装置，此时所述管理体正在复制页框的内容和存储新的页框号；以及

将添加到所述页表的页框添加到所述列表的装置；

其中复制页框的内容的装置进一步包括复制所述列表上的页框的内容的装置。

23. 根据权利要求 21 或 22 的装置，其中具有超过一个尺寸的存储页面被映射到所述逻辑存储块的页框，所述装置进一步包括：

将存储器管理中断从所述操作系统导引到所述管理体的装置；以及

将用于所述操作系统的存储器管理操作从用于所述操作系统的页表切换到临时可选页表的装置；

其中复制页框的内容的装置进一步包括：复制与被映射到所述逻辑存储块的页框的页面中的最小页面具有相同尺寸的分段中的页框的内容的装置。

24. 根据权利要求 23 的装置，其中复制页框的内容的装置进一步包括：

从所述临时可选页表中删除同样处在用于所述操作系统的页表中的页框的装置；以及

在用于所述操作系统的页表中存储这样的删除页框的状态比特的装置。

25. 根据权利要求 21、22、24 中任何一项的装置，其中所述逻辑存储块的页框中的至少一个被映射用于直接存储器存取，并且复制页框的内容的装置进一步包括：

在复制被映射用于直接存储器存取的页框的内容时，通过所述管理体来阻闭直接存储器存取操作的装置；以及

在直接存储器存取映射表中为所述逻辑存储块的被映射用于直接存储器存取的每个页框存储标识出复制了内容的页框的新的页框号的装置。

26. 根据权利要求 23 的装置，其中所述逻辑存储块的页框中的至少一

个被映射用于直接存储器存取，并且复制页框的内容的装置进一步包括:

在复制被映射用于直接存储器存取的页框的内容时，通过所述管理体来阻闭直接存储器存取操作的装置；以及

在直接存储器存取映射表中为所述逻辑存储块的被映射用于直接存储器存取的每个页框存储标识出复制了内容的页框的新的页框号的装置。

27. 根据权利要求 21、22、24 中任何一项的装置，其进一步包括创建一段空闲的相连存储器的装置，所述相连存储器既大于逻辑存储块并且又大得足以容纳页表。

28. 根据权利要求 23 的装置，其进一步包括创建一段空闲的相连存储器的装置，所述相连存储器既大于逻辑存储块并且又大得足以容纳页表。

29. 根据权利要求 25 的装置，其进一步包括创建一段空闲的相连存储器的装置，所述相连存储器既大于逻辑存储块并且又大得足以容纳页表。

30. 根据权利要求 26 的装置，其进一步包括创建一段空闲的相连存储器的装置，所述相连存储器既大于逻辑存储块并且又大得足以容纳页表。

31. 根据权利要求 27 的装置，其中创建一段空闲的相连存储器的装置进一步包括: 通过所述管理体为两个或更多的相连逻辑存储块重复实现以下步骤的装置:

通过所述管理体，将处在用于所述逻辑分区中的操作系统的页表中的逻辑存储块的页框的内容从所述逻辑存储块中的页框复制到所述逻辑存储块外部的页框的装置;

在所述页表中存储新的页框号的装置，包括通过所述管理体，为复制了内容的每个页框存储标识复制了内容的页框的新的页框号；以及

将所述逻辑存储块添加到空闲存储器的列表的装置。

32. 根据权利要求 21、22、24 中任何一项的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中:

复制所述逻辑存储块的页框的内容的装置进一步包括:

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置;

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

33. 根据权利要求 23 的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

34. 根据权利要求 25 的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

35. 根据权利要求 26 的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识复制了内容的页框的新的页框号的装置。

36. 根据权利要求 27 的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

37. 根据权利要求 28 的装置，其进一步包括能够改善逻辑存储块对处

理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

38. 根据权利要求29的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

39. 根据权利要求30的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

40. 根据权利要求 31 的装置，其进一步包括能够改善逻辑存储块对处理器的亲和性的装置，其中：

复制所述逻辑存储块的页框的内容的装置进一步包括：

将所述逻辑存储块的页框的内容复制到所述逻辑存储块外部的临时页框的装置；

将第二逻辑存储块的页框的内容复制到所述逻辑存储块的页框的装置；以及

将所述临时页框的内容复制到所述第二逻辑存储块的页框的装置；以及

存储新的页框号的装置进一步包括：对于所述逻辑存储块的内容以及对于所述第二逻辑存储块的内容这二者，存储标识出复制了内容的页框的新的页框号的装置。

在具有动态逻辑分区的计算环境中管理计算机存储器

技术领域

本发明涉及数据处理，并且更具体地，涉及用于在具有动态逻辑分区的计算机中管理计算机存储器的方法、系统和产品。

背景技术

常常将 1948 年 EDVAC 计算机系统的开发援引为计算机时代的开始。从那一时期开始，计算机系统就已经演进成非常复杂的设备。如今的计算机比诸如 EDVAC 的早期系统更精密复杂得多。计算机系统通常包括硬件和软件组件、应用程序、操作系统、处理器、总线、存储器、输入/输出设备等的组合。随着半导体加工和计算机体系机构的进步推动计算机的性能变得越来越高，更为精密复杂的计算机软件已经演进成利用硬件的更高性能，致使如今的计算机系统比几年前的系统更强有力得多。

现今存在这样的趋势，即发展在处理器数、输入/输出（“I/O”）槽数，以及存储量方面日趋大型的系统。尽管计算机硬件设计上的进步继续在这些物理资源的大小方面提供了快速增长，然而一些主要的应用和子系统却在可扩缩性方面落后。因此存在这样的趋势，即利用分区为系统提供物理分区或逻辑分区，以便基本计算机系统本身提供功能的粒度(granularity)。物理分区提供了通常来说相对粗糙的分区粒度，因为分区出现在诸如多芯片模块（“MCM”）、底板、子板、母板，或者其它系统板这样的物理边界（physical boundaries）处。在逻辑分区系统中，分区的粒度通常要细粒得多，例如单 CPU 或者甚至是 CPU 的一小部分、一小块存储器，或者 I/O 槽而不是整个 I/O 总线。利用逻辑分区，可以将给定的一组计算机资源细化成多个比物理分区多的逻辑分区。

逻辑分区 LPAR（“LPAR”）是计算机资源的子集，其可以托管（host）

操作系统（“O/S”）的实例。通过专门的硬件寄存器和称作管理体（hypervisor）的可信固件组件来实现 LPAR。这些组件一起在每个逻辑分区周围构建出紧密体系结构的“盒子（box）”，将分区操作限制在分派给该分区的一组专门的处理器、存储器和 I/O 资源。如今，随着计算机系统变得越来越大，在给定的硬件系统上运行操作系统的若干实例的能力（以便每个 O/S 实例加上其子系统很好地扩缩或实现）支持对硬件的最优使用并且转换到成本节约。尽管静态分区有助于调谐整个系统性能，然而如今的逻辑分区系统还可以提供“动态重构”能力－使得硬件资源、处理器、存储器、I/O 槽等能够移动到 LPAR 或从 LPAR 移动，或者从一个 LPAR 移动到另一 LPAR，而无需重新引导。动态重构通过提供以适时的（timely）方式将硬件资源动态移动到贫穷的（needy）O/S 以匹配工作负荷需求的能力而启用了一种改进的解决方案。

然而，如今典型的动态重构工具依赖于 LPAR 中管理体与操作系统之间的合作或协同（具有一些缺陷的计算机操作的模式）。例如，在存储器的动态重构中，O/S 可以保持 O/S 不会释放的栓定（bolted）或插定（pinned）页框（page frame）。很多不同的操作系统可以在相同时刻在同一系统上分离的 LPAR 中运行。例如，IBM 的 POWER$_{TM}$ 管理体支持三种不同的操作系统。所支持的操作系统中的一种或多种可能完全不支持与管理体的这种合作所需要的功能。另外，在协同方案中，对存储器的管理在作为 O/S 的错误或恶意实例的合作方案中变得更加复杂，不仅可能根本不合作，而且实际上可能在某种意义上对有效的计算机资源管理产生危害。

发明内容

提供了用于在具有动态逻辑分区的计算机中管理计算机存储器的方法、系统和产品，其相对于逻辑分区中的操作系统透明地操作。描述了用于在具有动态逻辑分区的计算机中管理计算机存储器的示例性方法、系统和产品，其包括：通过管理体，从逻辑分区（“LPAR”）的一个逻辑存储块（“LMB”）中的页框，将具有用于所述 LPAR 中的操作系统的页表

中的页框号的页框的内容复制到所述 LMB 外部的页框。

本发明的实施例通常包括：在所述页表中存储新的页框号，包括通过所述管理体，为复制了其内容的每个页框存储标识出向其复制了内容的页框的新的页框号。在典型的实施例中，复制页框的内容和存储新的页框号是相对于所述操作系统透明实现的。

典型的实施例还包括：通过所述管理体创建所述页表中所有页框的列表；通过所述管理体监视从所述操作系统到所述管理体的、将页框添加到所述页表的调用，此时所述管理体正在复制页框的内容和存储新的页框号；将添加到所述页表的页框添加到所述列表；并且其中复制页框的内容是通过复制所述列表上的页框的内容来实现的。

在一些实施例中，具有超过一个尺寸的存储页面被映射到 LMB 的页框。这样的实施例通常包括：将存储器管理中断从所述操作系统导引（vector）到所述管理体，以及将用于所述操作系统的存储器管理操作从用于所述操作系统的页表切换到临时可选页表。在这样的实施例中，复制页框的内容通常是通过复制与被映射到所述 LMB 的页框的页面中的最小页面具有相同尺寸的分段中的页框的内容来实现的。在这样的实施例中，复制页框的内容可以通过从所述临时可选页表中删除同样处在用于所述操作系统的页表中的页框以及在用于所述操作系统的页表中存储这样的删除页框的状态比特来实现。

在一些实施例中， LMB 的页框可以被映射用于直接存储器存取（“DMA”）。在这样的实施例中，复制页框的内容可以包括：在复制被映射用于 DMA 的页框的内容时，通过所述管理体来阻闭（blocking）DMA 操作，以及在 DMA 映射表中为所述 LMB 的被映射用于 DMA 的每个页框存储标识出向其复制了内容的页框的新的页框号。

实施例可以包括：创建一段空闲的相连存储器（free contigous memory），其既大于 LMB 并且又大得足以容纳页表。创建一段空闲的相连存储器可以通过由所述管理体为两个或更多的相连 LMB 重复实现以下步骤来完成：通过所述管理体，将处在用于所述 LPAR 中的操作系统的页

表中的LMB的页框的内容从所述LMB中的页框复制到所述LMB外部的页框；在所述页表中存储新的页框号，这包括通过所述管理体，为复制了其内容的每个页框存储标识出向其复制了内容的页框的新的页框号；以及将所述LMB添加到用于所述系统的空闲存储器的列表。

实施例还可以包括改善LMB对处理器的亲和性（affinity）。在这样的实施例中，复制所述LMB的页框的内容可以包括：将所述LMB的页框的内容复制到所述LMB外部的临时页框，将第二LMB的页框的内容复制到所述LMB的页框，以及将所述临时页框的内容复制到所述第二LMB的页框。在这样的实施例中，存储新的页框号可以包括：对于所述LMB的内容和对于所述第二LMB的内容这二者，存储标识出向其复制了内容的页框的新的页框号。

根据以下对如附图（其中相同的参考标号一般表示本发明的示例性实施例的相同部分）中所说明的本发明的示例性实施例的较为详细的描述，本发明的前述以及其它的特征和优点将显而易见。

附图说明

图1阐明了包括示例性计算机的自动计算机器的框图，该示例性计算机用于根据本发明的实施例管理具有动态逻辑分区的计算机存储器；

图2阐明了用于根据本发明的实施例管理具有动态逻辑分区的计算机存储器的另外的示例性计算机的框图；

图3阐明了根据本发明的实施例对计算机存储器进行管理的、具有动态逻辑分区的另外的示例性计算机系统的框图；

图4阐明了对示例性方法进行说明的流程图，该示例性方法用于根据本发明的实施例在具有动态逻辑分区的计算机中管理计算机存储器；

图5阐明了对另外的示例性方法进行说明的流程图，该示例性方法用于在具有动态逻辑分区的计算机中管理计算机存储器；

图6阐明了对另外的示例性方法进行说明的流程图，该示例性方法用于在具有动态逻辑分区的计算机中管理计算机存储器；

图 7 阐明了对创建一段空闲的相连存储器的示例性方法进行说明的流程图；以及

图 8 阐明了对改善 LMB 对处理器的亲和性的示例性方法进行说明的流程图。

具体实施方式

从图 1 开始，参照附图描述了根据本发明的实施例，用于在具有动态逻辑分区的计算机中管理计算机存储器的示例性方法、系统和产品。依照本发明在具有动态逻辑分区的计算机中管理计算机存储器一般是利用自动计算机器，即，利用计算机来实现的。因此，为了进一步解释，图 1 阐明了自动计算机器的框图，其包括用于根据本发明的实施例管理具有动态逻辑分区的计算机存储器的示例性计算机（152）。图 1 的计算机（152）包括至少一个计算机处理器（156）或“CPU”以及通过系统总线（160）连接至处理器(156)和计算机的其它组件的随机访问存储器(168)(“RAM”)。在实际情况下，根据本发明的实施例用于在具有动态逻辑分区的计算机中管理计算机存储器的系统通常包括超过一个的计算机处理器。图 1 的例子中的 RAM（168）是以被称为逻辑存储块或“LMB”的分段（101-110）来管理的。

RAM（168）中存储有应用程序（158），即用于实现执行的线程的用户级(user-level)数据处理的计算机程序指令。根据本发明的实施例，RAM（168）中还存储有管理体（102），即用于管理为了在具有动态逻辑分区的计算机中管理计算机存储器而改进的 LPAR 中的资源的一组计算机程序指令。RAM（168）中还存储有操作系统（154）。根据本发明的实施例，计算机中有用的操作系统包括 $UNIX_{TM}$、$Linux_{TM}$、Microsoft NT_{TM}、AIX_{TM}、IBM 的 $i5/OS_{TM}$，以及本领域的技术人员可以想到的其它操作系统。操作系统（154）和应用程序（158）置于 LPAR（450）中。在图 1 例子中，在 RAM（168）中示出了操作系统（154）、应用程序（158），以及管理体（102），但是读者应当理解，这样的软件的组件还可以存储在非

易失性存储器（166）中。

图 1 的系统支持动态逻辑分区并且一般可以操作以便通过以下来管理计算机存储器，即通过管理体（102），从逻辑分区（“LPAR”）的一个逻辑存储块（“LMB”）中的页框，将具有用于该 LPAR 中的操作系统的页表中的页框号的页框的内容复制到该 LMB 外部的页框，以及在该页表中存储新的页框号，这包括通过管理体，为复制了其内容的每个页框存储标识出向其复制了内容的页框的新的页框号。在图 1 的系统中，复制页框的内容和存储新的页框号可以是相对于操作系统（154）透明实现的。

图 1 的计算机（152）包括通过系统总线（160）耦合于处理器（156）和计算机（152）的其它组件的非易失性计算机存储器（166）。可以将非易失性计算机存储器（166）实现为硬盘驱动器（170）、光盘驱动器（172）、电可擦可编程只读存储空间（所谓的“EEPROM”或“闪速”存储器）（174）、RAM 驱动器（未示出），或者本领域的技术人员可以想到的任何其它种类的计算机存储器。

图 1 的示例计算机包括一个或多个 I/O 接口适配器（178）。计算机中的输入/输出接口适配器实现面向用户的输入/输出，例如，通过用于控制通往诸如计算机显示屏的显示设备（180）的输出以及来自诸如键盘和鼠标的用户输入设备（181）的用户输入的计算机硬件以及软件驱动器。在本说明书中，一般将实现连接 I/O 适配器的 I/O 的 I/O 硬件资源称为“I/O 槽”。

图 1 的示例性计算机（152）包括用于实现数据通信的通信适配器（167）。这样的数据通信可以通过串行通过 RS-232 连接、通过诸如 USB 的外部总线、通过诸如 IP 网络的数据通信网络，以及以本领域的技术人员可以想到的其它方式来实现。通信适配器实现硬件级的数据通信，由此，一个计算机直接地或通过网络将数据通信发送至另一计算机。根据本发明的实施例，对确定目的地的可用性有用的通信适配器的例子包括：用于有线拨号通信的调制解调器、用于有线网络通信的以太网（IEEE 802.3）适配器，以及用于无线网络通信的 802.11b 适配器。

为了进一步解释，图 2 阐明了用于根据本发明的实施例管理具有动态

逻辑分区的计算机存储器的另外的示例性计算机（152）的框图。构造图 2 以便进一步解释对这样的系统中的物理存储器的管理，即该系统用于根据本发明的实施例在具有动态逻辑分区的计算机中管理计算机存储器。图 2 的系统中的物理存储器连同处理器芯片一起置于多芯片模块（“MCM”）（202）中的存储芯片（204）中。而 MCM 在底板（206，208）上实现，底板（206，208）又通过系统总线（160）耦合用于数据通信。底板上的 MCM 通过底板总线（212）耦合用于数据通信，并且 MCM 上的处理器芯片和存储芯片通过如 MCM（222）上的参考标记（210）所说明的 MCM 总线耦合用于数据通信，MCM（222）扩展了对 MCM（221）的图示。

多芯片模块或“MCM”是衬底（substrate）上装配有两个或更多的裸集成电路（裸片（bare dies））或“芯片大小的组件”的电子系统或子系统。在图 2 的例子中，MCM 中的芯片是计算机处理器和计算机存储器。衬底可以是，例如，印刷电路板或者具有互连图案（interconnection pattern）的厚的或薄的薄膜陶瓷或硅。衬底可以是 MCM 组件的整体部分或者可以被安装在 MCM 组件内。MCM 在计算机硬件体系结构中是有用的，因为其代表了专用集成电路（“ASIC”）与印刷电路板之间的封装等级（packaging level）。

图 2 的 MCM 说明了硬件存储器分离或“亲和性”的级别。MCM（222）上的处理器（214）可以访问位于以下的物理存储器：

- 在相同的 MCM 上的存储芯片（216）中，其中该 MCM 具有访问该存储芯片的处理器（214），
- 在相同的底板（208）上的另一 MCM 上的存储芯片（218）中，或者
- 在另一底板（206）上的另一 MCM 中的存储芯片（220）中。

访问分离于 MCM 的存储器比访问具有处理器的同一 MCM 上的存储器要花费更长的时间，因为用于访问这样的存储器的计算机指令以及从这样的存储器返回数据必须遍历更多的计算机硬件、存储器管理单元、总线驱动器，更不用说本身在现今的计算速度上就有所考虑的总线地带（bus

land）和线路的长度。出于相同的原因，访问分离于相同底板的存储器花费甚至更长的时间。因此，认为在相同的 MCM（其具有访问存储器的处理器）上的存储器比分离于该 MCM 的存储器具有更紧密的亲和性，并且认为在相同底板（其具有访问处理器）上的存储器比在另一底板上的存储器具有更紧密的亲和性。如此描述计算机体系机构是为了进行解释，而不是对计算机存储器的限制。可以将若干 MCM 安装在印刷电路板上，例如，在将印刷电路板插入底板的情况下，从而创建图 2 中未说明的亲和性的附加级别（additional level of affinity）。本领域的技术人员可以想到的计算机体系机构的其它方面可能影响处理器－存储器亲和性，而所有这样的方面都属于根据本发明的实施例在动态逻辑分区情况下的存储器管理的范围之内。

为了进一步解释，图 3 阐明了具有动态逻辑分区的另外的示例性计算机系统的框图，该示例性计算机系统根据本发明的实施例对计算机存储器进行管理。如以上所提及的，逻辑分区是一种计算机设计特征，其通过使得有可能在单个计算机上并发地运行多个独立的操作系统映像来提供灵活性。

图 3 的系统包括管理体（102）以及可以运行用于 LPAR（450，452，454）中的应用软件的执行的多个线程（302）的三个操作系统（154）和三个处理器（156）。使用三个例子是为了进行解释，而非用于限制。事实上，本领域的技术人员可以认识到，诸如所说明的系统的系统可以操作任何数目的 LPAR、操作系统、处理器，以及仅受到系统中物理资源的实际数量的限制的线程。线程（302）在组织于虚拟地址空间中的虚拟存储器地址上操作。处理器（156）访问组织于真实地址空间中的物理存储器。

每个操作系统映像（154）均需要可以以真实寻址模式来访问的一系列存储器。在该模式中，不进行虚拟地址转换，并且地址从地址 0 开始。操作系统通常将该地址范围用于启动内核码（startup kernel code）、固定内核结构，以及中断向量。由于不能够允许多个分区共享位于物理地址 0 处的相同的存储范围，因此每个 LPAR 必须具有其自己的真实模式寻址范围。

管理体为每个LPAR分派唯一的真实模式地址偏移和范围值，并且然后将这些偏移和范围值设置到分区中每个处理器中的寄存器中。这些值映射到已经被专门分派给那一分区的物理存储器地址范围。当分区程序以真实寻址模式访问指令和数据时，硬件在访问物理存储器之前自动地将真实模式偏移值添加到每个地址。以这样的方式，每个逻辑分区编程模型看起来似乎都访问物理地址0，即使地址被透明地重定向（redirected）到另一地址范围。硬件逻辑通过在分区中运行的操作系统代码来阻止对这些寄存器的修改。对访问所分派的范围以外的真实地址的任何尝试均导致寻址异常中断，其通过分区中的操作系统异常处理体来处理。

操作系统使用另一种类型的寻址，虚拟寻址，以便为用户应用线程提供超过安装在系统中的物理存储器的数量的有效地址空间。操作系统通过将很少使用的程序和数据从存储器向外编页（paging）到磁盘，并且根据需要将它们带回到物理存储器来实现这一功能。

当应用以虚拟寻址模式访问指令和数据时，它们并不知道其地址正在被使用页面转换表（416）的虚拟存储器管理转换。这些表格（在本说明书中一般将其称为“页表”）驻留于系统存储器中，并且每个分区均具有代表其本身而由管理体管理的自身专有的页表。处理器使用这些表格（通过对管理体的调用）来将程序的虚拟地址（424）透明地转换成其中页面已经被映射到物理存储器的物理地址（422）。如果在线程访问存储器的页面时页框已经被从物理存储器向外移出到磁盘上，则操作系统接收到页面故障。

在非LPAR操作中，操作系统直接创建和维护页表条目，使用真实模式寻址来访问表格。在逻辑分区操作中，页面转换表位于仅管理体可访问的保留物理存储区域中。换句话说，分区的页表位于分区的真实模式地址范围之外。仅可以通过管理体来修改为处理器提供其页表的物理地址的寄存器。

将虚拟地址实现为虚拟页面号（424）和虚拟页面内的偏移的组合。将真实地址实现为标识出真实存储器的页面的页框号（422）和该页面内的偏移的组合。虚拟地址的偏移也是虚拟地址被映射到的真实地址的偏移。页

表将虚拟地址映射到真实地址，但由于偏移是相等的，因此页表仅映射虚拟页面号和对应的页框号。偏移并不包括在页表中。

当操作系统（154）需要创建页面转换映射时，其在处理器（156）上执行对管理体（102）的调用，处理器（156）将执行转移到管理体。管理体创建代表分区的页表条目并且将其存储在页表中。线程还可以进行管理体调用来修改或删除现有的页表条目。页表条目仅映射到特定的物理存储区域，称为逻辑存储块或“LMB”，其被以粒状分段（granular segment）分派给每个 LMB。这些 LMB 提供了对 LPAR 的虚拟页面地址空间进行备份的物理存储器。因此，LPAR 的存储器一般由从物理存储器中的任何地方以任何顺序分派的 LMB 组成。

I/O 硬件使用直接存储器存取（“DMA”）操作来在 I/O 槽（407）中的 I/O 适配器与系统存储器中的页框（406）之间移动数据。DMA 操作使用类似于页表的地址浮动（address relocation）机制。I/O 硬件将 I/O 槽中的 I/O 设备所生成的地址（425）转换成物理存储器地址。I/O 硬件利用存储在物理存储器中的 DMA 映射(650)(有时被称为转换控制条目(“TCE”)表）来进行该转换。如同页表的情况，DMA 映射驻留于不可由分区访问而仅可由管理体访问的系统存储器的物理地址区域中。通过调用管理体服务，分区程序可以创建、修改或删除用于分派给该分区的 I/O 槽的 DMA 映射条目。当 I/O 硬件将 I/O 适配器 DMA 地址转换成物理存储器时，所得到的地址落入分派给那一分区的物理存储空间内。

为了进一步解释，图 4 阐明了对示例性方法进行说明的流程图，该示例性方法用于根据本发明的实施例在具有动态逻辑分区的计算机中管理计算机存储器，其包括：通过管理体创建(426)页表中所有页框的列表(436)。有利地，相对快地进行对根据本发明的实施例的存储器管理功能的实现，以便降低导致从在用户应用中执行的线程的观点来看过多的存储故障（memory fault）和延迟的风险。扫描通过大型数据结构的页表、寻找所映射的页面是耗时的。在实施实际的存储器管理操作时，期望在可快速访问的结构中存储有受影响的页框的简明列表。举例来说，这样的列表可以

通过在后台独立运行的管理体过程来构建，直到汇集了该列表。因此，图4的方法有利地包括：通过管理体来监视（428）从操作系统到管理体的调用，该调用将页框添加到页表（416），而此时管理体正在复制页框的内容和存储新的页框号。图4的方法还包括：将添加到页表的页框添加（430）到列表（436）。

图4的方法包括：通过管理体，从LPAR的一个LMB（402）中的页框（406），将具有用于该LPAR（450）中的操作系统（432）的页表（416）中的页框号（422）的页框的内容复制（408）到该LMB（402）外部的页框（412）。用点线轮廓示出LMB（404）以示强调，尽管所有受影响的页框均组织于LMB中，然而在作为存储器管理操作的主体的LMB（402）外部的页框（412）的位置无关紧要，只要它们不在主体LMB（402）中。在图4的方法中，如以上所提及的，复制（408）页框的内容是通过复制（434）列表（436）上页框的内容来实现的。图4的方法还包括在页表（418）中存储（410）新的页框号，这包括通过管理体，为复制了其内容的每个页框存储标识出向其复制了内容的页框的新的页框号。

利用页表（416，418）说明了这些存储器管理操作的效果。页表（416，418）是在图4的方法中的存储器管理操作之前（416）和之后（418）所说明的同一页表。在存储器管理操作之前，页表将虚拟页面号346、347和348映射到置于LMB（402）中的页框592、593和594。在图4的例子中的存储器管理操作之后，页表将虚拟页面号346、347和348映射到置于LMB（402）外部的页框592、593和594。由于将页框592、593和594的内容复制（而不是移动）到页框743、744和745，因此页框592、593和594的内容不受影响。然而，先前映射到它们的虚拟页面现在在别处被映射到其它的页框。这有效地释放出LMB（402）的页框用于其它用途。可以将其列为空闲的，用于为新的LPAR安装大的页表、用于改善处理器－存储器亲和性，或者用于本领域的技术人员可以想到的其它方面。

在图4的方法中，复制页框的内容和存储新的页框号是相对于操作系统透明实现的。下一次操作系统在访问一个被重新映射的虚拟页面中经历

存储故障时，位于 LMB（404）中新页框处的物理存储器的内容会与其在图 4 的方法中存储器管理操作被应用之前相同。在实现图 4 的方法时，管理体并不对请求释放资源的操作系统（432）进行调用，并且操作系统从未发觉页表条目已经受到影响。

为了进一步解释，图 5 阐明了对另外的示例性方法进行说明的流程图，该示例性方法用于根据本发明的实施例在具有动态逻辑分区的计算机中管理计算机存储器，其中超过一个尺寸的存储页面被映射到 LMB（402）的页框（406）。如以上所提及的，LPAR 可以支持超过一种的操作系统，每种类型的操作系统均可以支持不同的页面尺寸，并且每个操作系统均可以支持超过一个的页面尺寸。有利地，相对快地进行对根据本发明的实施例的存储器管理功能的实现，以便降低导致从在用户应用中执行的线程的观点来看过多的存储故障和延迟的风险。复制小的存储页面的内容比复制大的页面的内容要快。因此，当主体操作系统使用超过一个的页面尺寸时，图 5 的方法有利地提供了一种使用小的页面尺寸实现存储器复制操作的方式。

图 5 的方法包括将存储器管理中断从操作系统（432）导引（502）到管理体。管理体通过在处理器寄存器中设置比特来将存储器管理中断从操作系统导引到管理体，以便存储器管理中断被定向到管理体中断向量。当复制操作在页框上进行时，该机制允许管理体阻闭管理体中的处理器。由于使用管理体寄存器资源将中断呈现给管理体，因此存储故障对操作系统是透明的。

在图 5 的例子中，如果将小的页面尺寸取为 4KB，那么所示的操作系统（432）使用两个页面尺寸，4KB 和 16KB。这在页表（416）中进行了说明，其中 16KB 的虚拟页面（虚拟存储页面 346）被映射到四个 4KB 页框（页框 592、593、594 和 595）。其它的 4KB 虚拟页面 347、348、349 相对应地分别映射到 4KB 页框 596、597 和 598。图 5 的方法包括：将用于操作系统的存储器管理操作从用于操作系统的页表（416）切换（504）到临时可选页表(512)以便仅支持 4KB 页框中的复制操作，忽略页表(416)

中所呈现的来自操作系统的任何的大页面指示。在图 5 的方法中，复制（408）页框的内容包括复制（506）与被映射到 LMB 的页框的页面中的最小页面具有相同尺寸的分段中的页框的内容。也就是说，管理体仅实现 4KB 分段（4KB 页框 × 4KB 页框）中的复制操作。

当存储器管理中断出现时，管理体查找操作系统的真实页表以查看存储器管理中断是否在分区的真实页表在使用的情况下已经发生。如果是的话，则管理体对 OS 存储器管理中断向量给予控制。否则，将页框条目插入到临时可选页表中（如果复制操作不在进行中）。

在图 5 的方法中，复制（408）页框的内容还包括从临时可选页表（512）中删除（508）同样处在用于操作系统的页表中的页框。在图 5 的方法中，复制（408）页框的内容还包括在用于操作系统（432）的页表（416）中存储（510）这样的删除页框的状态比特。这样的删除页框的状态由参考比特（用于存储故障下的 LRU 操作）以及由变更比特（指示了当从高速缓存删除时页面已经被写入并且必须被保存回磁盘）来指示。

为了进一步解释，图 6 阐明了对另外的示例性方法进行说明的流程图，该示例性方法用于根据本发明的实施例在具有动态逻辑分区的计算机中管理计算机存储器，其中 LMB（402）的页框（406）中的至少一个被映射用于直接存储器存取（“DMA”）。在图 6 的方法中，复制（408）页框的内容包括在复制（660）被映射用于 DMA 的页框（423）的内容时，通过管理体（未示出）来阻闭（658）DMA 操作。

在图 6 的方法中，DMA 操作由含有这样的 I/O 适配器（未示出）的 I/O 槽（407）表示，即该 I/O 适配器实现了表示经由通过系统 RAM（168）中的页框的 DMA 通道（654）的数据存储（656）的磁盘 I/O。通过 DMA 映射（650）将系统 RAM 中的页框映射到 I/O 地址。在图 6 的方法中，复制（408）页框的内容包括将 DMA 映射的页框 550 复制（660）到 LMB（402）外部的页框（412），以及在 DMA 映射表（652）中为 LMB 的被映射用于 DMA 的每个页框存储（662）标识出向其复制了内容的页框的新的页框号。

DMA 映射（650、652）说明了根据图 6 的方法的存储器管理操作的效果。DMA 映射是数据结构，有时称为转换条目表或“TCE 表”，其中的每个条目将 I/O 地址空间中的地址映射到系统物理存储器中的页框。例如，I/O 地址空间中的地址可以是 I/O 适配器或 PCI（外设部件互连）总线适配器的地址空间中的地址。在图 6 中，DMA 映射（650、652）分别是在根据图 6 的方法的存储器管理操作之前（650）和之后（652）的同一 DMA 映射。在图 6 的例子中，最初将 I/O 地址（425）124 映射到页框 550。在为页面阻闭了 DMA 操作之后，复制 DMA 映射的页框，并且根据图 6 的方法，在映射中存储新的页框号，DMA 映射（652）示出了被映射到页框 725 的 I/O 地址 124。这有效地释放出 LMB（402）的页框 550 用于其它用途。可以将其列为空闲的，随其它页框或其它的 LMB 一起用于为新的 LPAR 安装大的页表、用于改善处理器－存储器亲和性，或者用于本领域的技术人员可以想到的其它方面。

页表通常是大型数据结构，常常大体上大于 LMB。当系统管理员尝试动态地创建新的 LPAR（而不重新引导）时，可能没有足够的相连存储器可用于新的 LPAR 的页表。有利地，根据本发明的实施例对具有动态逻辑分区的计算机中的计算机存储器的管理因此可以包括：创建一段空闲的相连存储器，其既大于 LMB 并且又大得足以容纳页表。

为了进一步解释，图 7 阐明了对创建一段空闲的相连存储器的示例性方法进行说明的流程图，该示例性方法包括：通过管理体，将处在用于 LPAR（450）中的操作系统（432）的页表（416）中的相连 LMB 的页框的内容，从相连 LMB（401，402）中的页框（406）复制（602）到相连 LMB 外部的页框（412）。图 7 的方法包括在页表（418）中存储（604）新的页框号，这包括通过管理体，为复制了其内容的每个页框存储标识出向其复制了内容的页框的新的页框号。

图 7 的方法还包括将 LMB 添加（606）到用于 LPAR（450）的空闲存储器的列表（608）。在图 7 的例子中，将 LMB 添加（606）到用于 LPAR 的空闲存储器的列表（608）是通过将释放的页框的页框号放置到空闲列表

（608）中来实现的。可选地，可以在空闲列表中列出 LMB 中的第一页框的页框号以指示整个 LMB 是空闲的。本领域的技术人员可以想到指示空闲存储器的其它方式，并且所有这样的方式同样在本发明的范围内。

通常必须释放超过两个的相连 LMB 以便为页表提供空间。因此，图 7 的方法有利地包括：参照预定的所要求的分段尺寸（610）来确定（609）存储器的空闲分段（freed segment）是否大得足以存储页表或满足空闲存储器的其它需求。如果空闲分段不够大，则通过重复（612）以下步骤继续处理，直到空闲分段足够大，即步骤：将相连 LMB 的页框的内容复制（602）到相连 LMB 外部的页框（412），在页表（418）中存储（604）新的页框号，以及将 LMB 添加（606）到用于 LPAR 的空闲存储器的列表（608）。

随着被访问存储器的亲和性相对于访问处理器而减少，整个系统性能降低。有利地，根据本发明的实施例对具有动态逻辑分区的计算机中的计算机存储器的管理因此可以包括：改善 LMB 对处理器的亲和性。为了进一步解释，图 8 阐明了对改善 LMB 对处理器的亲和性的示例性方法进行说明的流程图。图 8 的方法影响两个 LMB（402，403）的处理器－存储器亲和性。LMB（402，403）彼此远离，LMB（402）在 MCM 704 中而 LMB（403）在 MCM（705）中。如上所述，每个 MCM 均含有处理器和存储器。在管理体内实现图 8 的方法。通过管理体将来自各 MCM 的处理器和存储器分派给 LPAR 中的操作系统（图 8 中未示出）。

在图 8 的例子中，处理器（156）与位于相同的 MCM（704）上的 LMB（402）具有紧密的亲和性－而与位于不同的 MCM（705）上远离处理器（156）的 LMB（403）具有较少的亲和性。类似地，在图 8 的例子中，处理器（157）与位于相同的 MCM（705）上的 LMB（403）具有紧密的亲和性－而与位于不同的 MCM（704）上远离处理器（157）的 LMB（402）具有较少的亲和性。LMB（402）含有页框编号 600-699，并且 LMB（403）含有页框 800-899。LMB 中的页框分派仅用于进行解释，而非限制。读者可以认识到，在实际情况下 LMB 含有超过 100 的多个页框。所示的 MCM（705）和 MCM（704）通过系统总线（160）耦合，但是读者可以认识到，

该体系机构仅用于解释亲和性，而非对本发明的限制。事实上，可以通过分离的印刷电路板、连接，通过底板或子板，以及本领域的技术人员可以想到的其它方式来实现疏远的亲和性（remote affinity）。

在页表（416、418、417 和 419）中分别说明了用于 MCM（704，705）上的两个分区的页表条目。页表（416，418）分别示出了在亲和性改善操作之前（416）和之后（418）用于 MCM（705）的页表条目。类似地，页表（417，419）分别示出了在亲和性改善操作之前（417）和之后（419）用于 MCM（704）的页表条目。页表（416）示出了由在 MCM（705）上的处理器（157）上运行的线程使用的虚拟页面号 567、568 和 569 被映射到在物理上位于相对于处理器（157）具有疏远的亲和性的 MCM（704）上的 LMB（402）中的页框 666、667 和 668。类似地，页表（417）示出了由在 MCM（704）上的处理器（156）上运行的线程使用的虚拟页面号 444、445 和 446 被映射到在物理上位于相对于处理器（156）具有疏远的亲和性的 MCM（705）上的 LMB（403）中的页框 853、854 和 855。可以改善整个处理器－存储器亲和性和存储器管理效率，例如，在可以将被映射到处理器上正使用的虚拟页面的页框定位或移动到具有该处理器的相同 MCM 上的物理存储器的情况下。另外，可以利用多个 MCM 上的处理器来实现 LPAR，并且这样的 LPAR 还可以具有多个页表，例如，每个 MCM 具有一个页表。根据本发明的实施例改善 LMB 对处理器的亲和性对于在多个 MCM 上具有多个页表和处理器的这样的 LPAR 也是有用的。

图 8 的方法包括复制页框的内容（408），过程操作基本上如本说明书中上文所描述的。然而，为了改善亲和性，在图 8 的方法中，复制（408）LMB 的页框的内容有利地包括：将 LMB（402）的页框（406）的内容复制（802）到 LMB（402）外部的临时页框（702）。那么图 8 的方法中复制（408）页框的内容还包括：将 LMB（403）的页框（409）的内容复制（804）到 LMB（402）的页框（406），以及将临时页框（702）的内容复制（806）到 LMB（705）的页框（409）。图 8 的方法还包括存储（410）新的页框号，其一般如上所述进行操作，但在这里包括：对于 LMB（402）

的内容以及对于第二 LMB（403）的内容（409）这二者，存储（808）标识出向其复制了内容的页框的新的页框号。

页表（418，419）示出了这些亲和性改善操作的效果。页表（418）示出了由在 MCM（705）上的处理器（157）上运行的线程所使用的虚拟页面号 567、568 和 569 现在被映射到页框 853、854 和 855，页框 853、854 和 855 在物理上位于目前相对于相同的 MCM 上的处理器（157）具有紧密的亲和性的 MCM（705）上的 LMB（403）中。类似地，页表（419）示出了由在 MCM（704）上的处理器（156）上运行的线程所使用的虚拟页面号 444、445 和 446 现在被映射到页框 666、667 和 668，页框 666、667 和 668 在物理上位于相对于相同的 MCM 上的处理器（156）具有紧密的亲和性的 MCM（704）上的 LMB（402）中。

主要在用于管理具有动态逻辑分区的计算机中的计算机存储器的全功能计算机系统的环境下描述了本发明的示例性实施例。然而，本领域的技术人员可以认识到，本发明还可以体现在置于随任何适当的数据处理系统一起使用的信号承载介质上的计算机程序产品中。这样的信号承载介质可以是用于机器可读信息的传输介质或可记录介质，包括磁介质、光介质，或者其它合适的介质。可记录介质的例子包括磁带或硬盘驱动器中的磁盘、用于光驱动器的光盘、磁带，以及本领域的技术人员可以想到的其它方式。传输介质的例子包括用于话音通信的电话网络，以及举例来说，像 Ethernet$_{TM}$ 以及与网际协议和万维网通信的网络这样的数字数据通信网络。本领域的技术人员可以立即认识到，具有适当编程装置的任何计算机系统都将能够执行如程序产品中所体现的本发明的方法的步骤。本领域的技术人员可以立即认识到，尽管本说明书中所描述的一些示例性实施例是面向安装和执行于计算机硬件上的软件的，然而，实现为固件或硬件的可选实施例也属于本发明的范围之内。

根据前文的描述可以理解到，在本发明的范围内可以对本发明的上述说明性实施例进行各种修改。

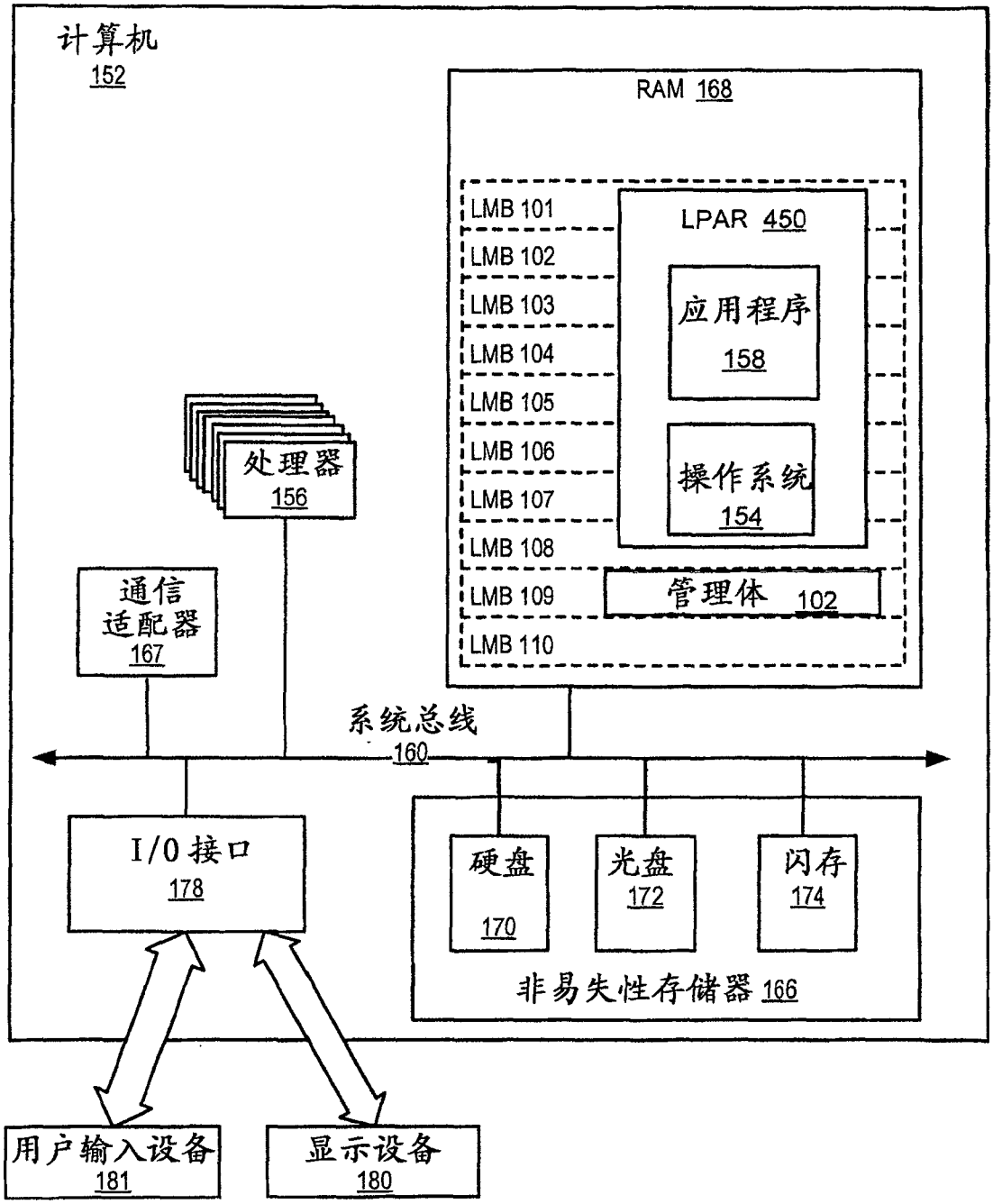

计算机
152
RAM 168
LMB 101
LMB 102
LMB 103
LMB 104
LMB 105
LMB 106
LMB 107
LMB 108
LMB 109
LMB 110
LPAR 450
应用程序
158
操作系统
154
管理体 102
处理器
156
通信
适配器
167
系统总线
160
I/O 接口
178
硬盘
170
光盘
172
闪存
174
非易失性存储器 166
用户输入设备
181
显示设备
180

图 1

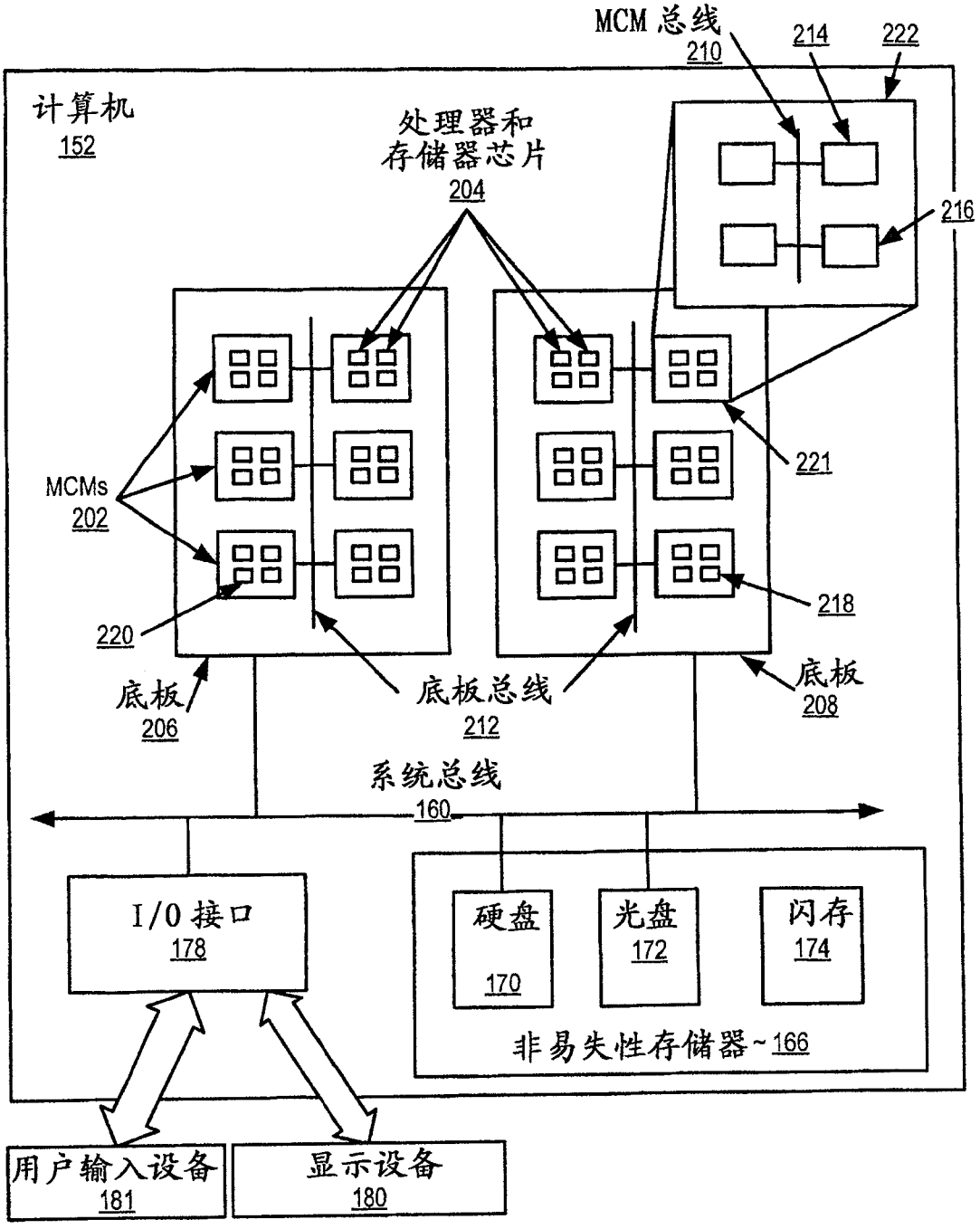

MCM 总线
210
214
222
计算机
152
处理器和
存储器芯片
204
216
MCMs
202
221
220
218
底板
206
底板总线
212
底板
208
系统总线
160
I/O 接口
178
硬盘
170
光盘
172
闪存
174
非易失性存储器~166
用户输入设备
181
显示设备
180

图 2

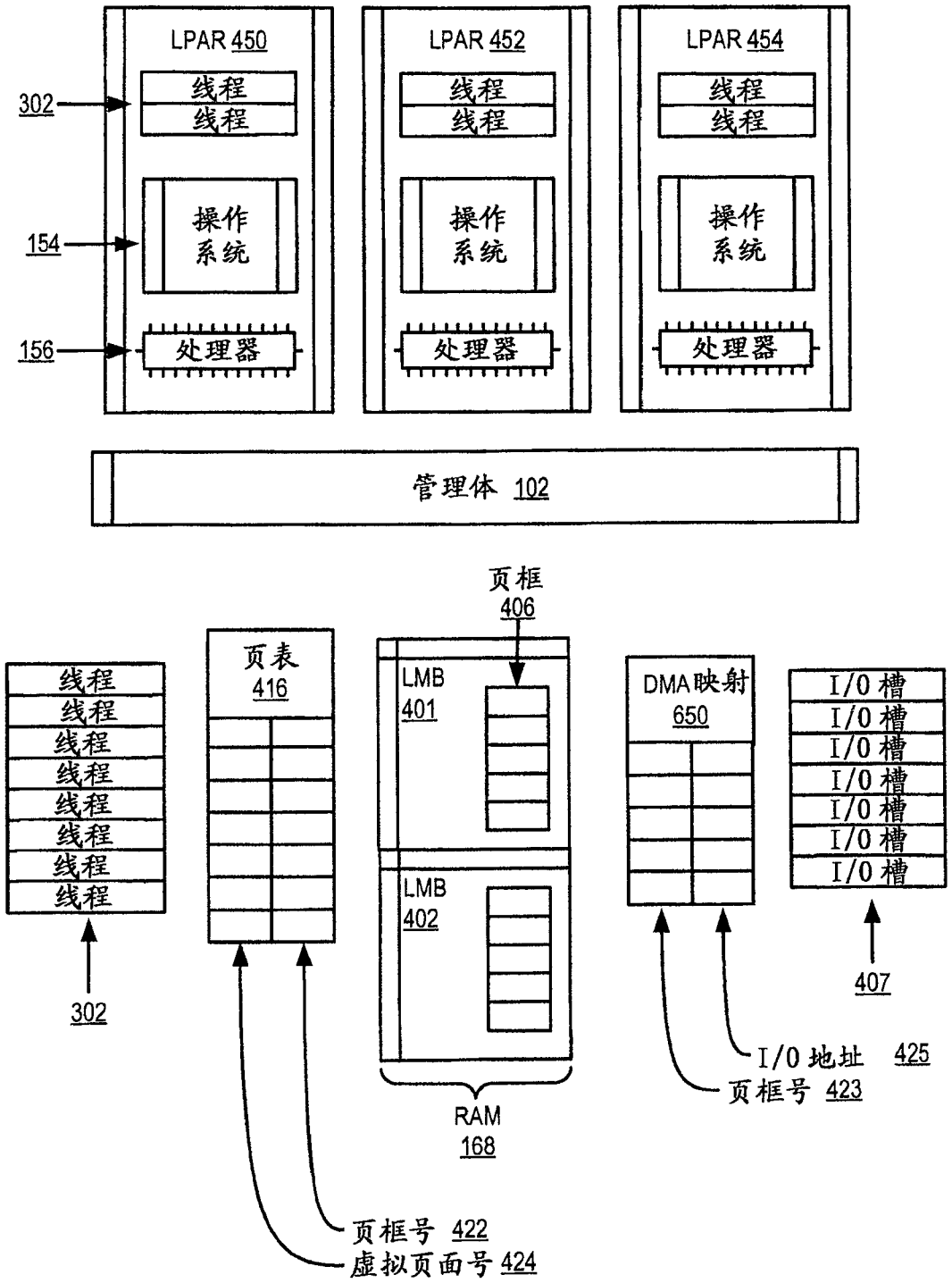

LPAR 450
LPAR 452
LPAR 454
线程
线程
操作系统
处理器
302
154
156
管理体 102
页框 406
页表 416
LMB 401
LMB 402
DMA映射 650
I/O 槽
I/O 槽
I/O 槽
I/O 槽
I/O 槽
I/O 槽
I/O 槽
线程
线程
线程
线程
线程
线程
线程
线程
302
407
RAM 168
I/O 地址 425
页框号 423
页框号 422
虚拟页面号 424

图 3

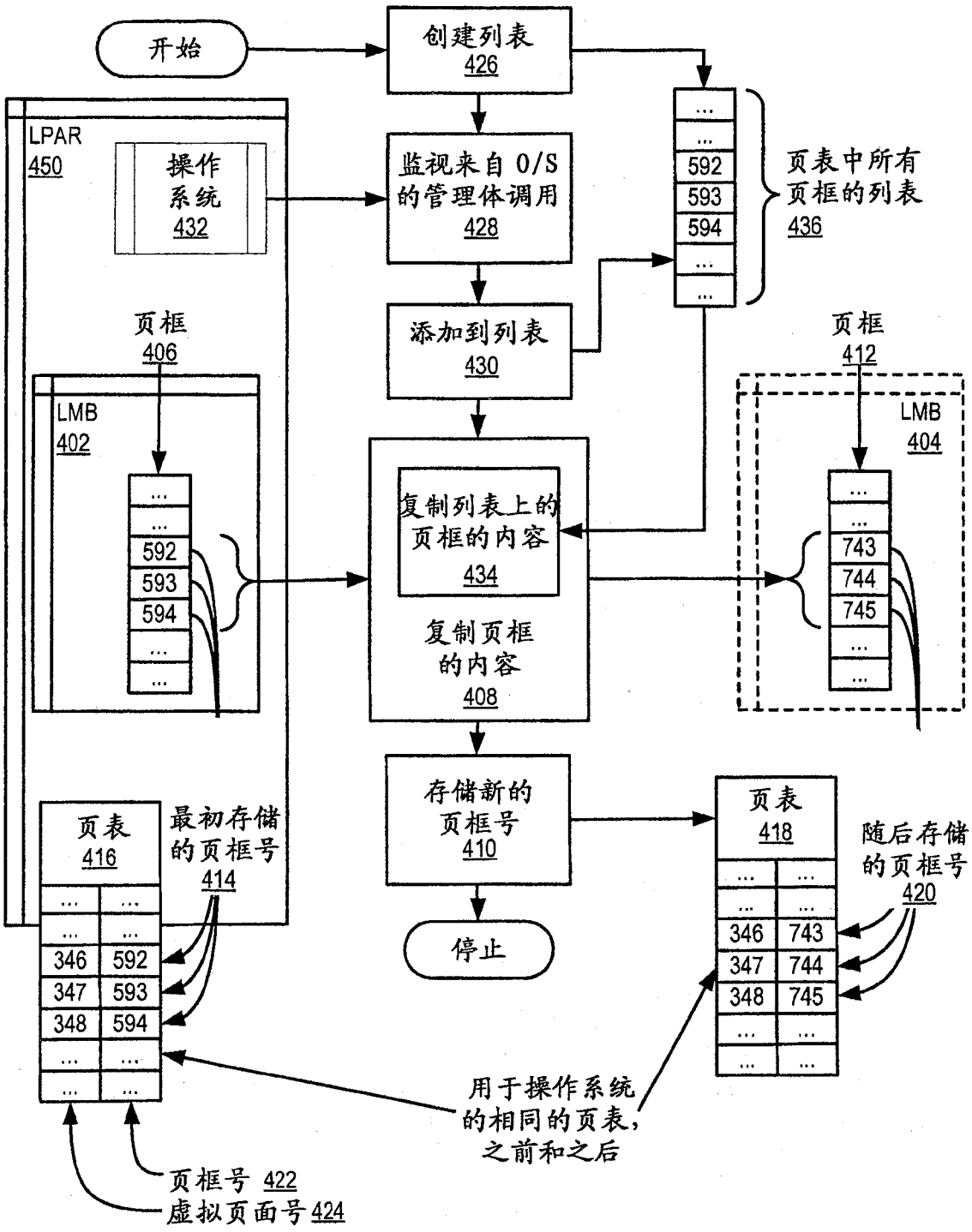

开始
创建列表
426
LPAR
450
操作
系统
432
监视来自 O/S
的管理体调用
428
添加到列表
430
页表中所有
页框的列表
436
页框
406
LMB
402
页框
412
LMB
404
复制列表上的
页框的内容
434
复制页框
的内容
408
存储新的
页框号
410
停止
页表
416
最初存储
的页框号
414
页表
418
随后存储
的页框号
420
用于操作系统
的相同的页表，
之前和之后
页框号 422
虚拟页面号 424

图 4

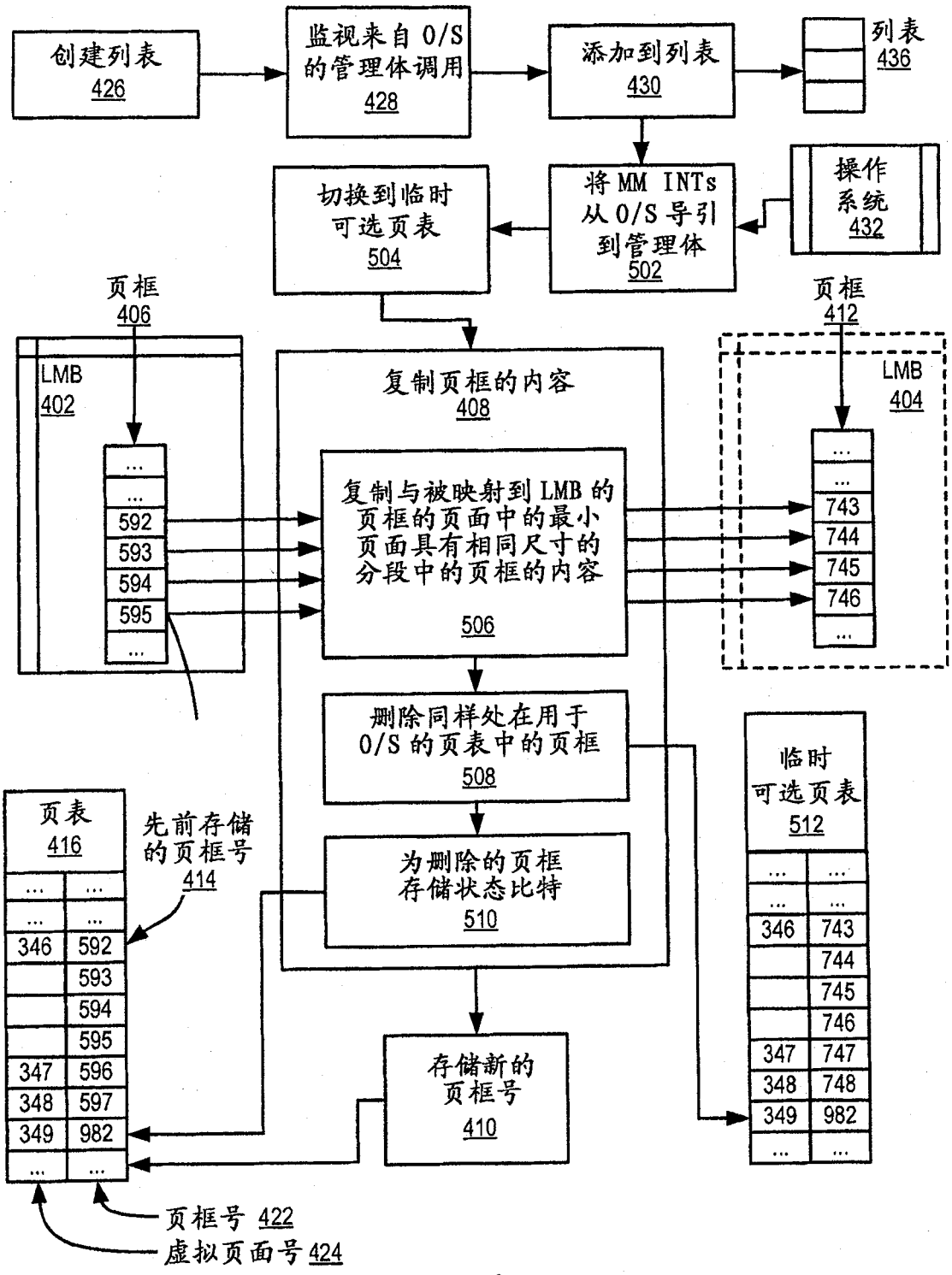

创建列表
426
监视来自 O/S
的管理体调用
428
添加到列表
430
列表
436
操作
系统
432
将 MM INTs
从 O/S 导引
到管理体
502
切换到临时
可选页表
504
页框
406
LMB
402
页框
412
LMB
404
复制页框的内容
408
复制与被映射到 LMB 的
页框的页面中的最小
页面具有相同尺寸的
分段中的页框的内容
506
删除同样处在用于
O/S 的页表中的页框
508
为删除的页框
存储状态比特
510
临时
可选页表
512
页表
416
先前存储
的页框号
414
存储新的
页框号
410
页框号 422
虚拟页面号 424

图 5

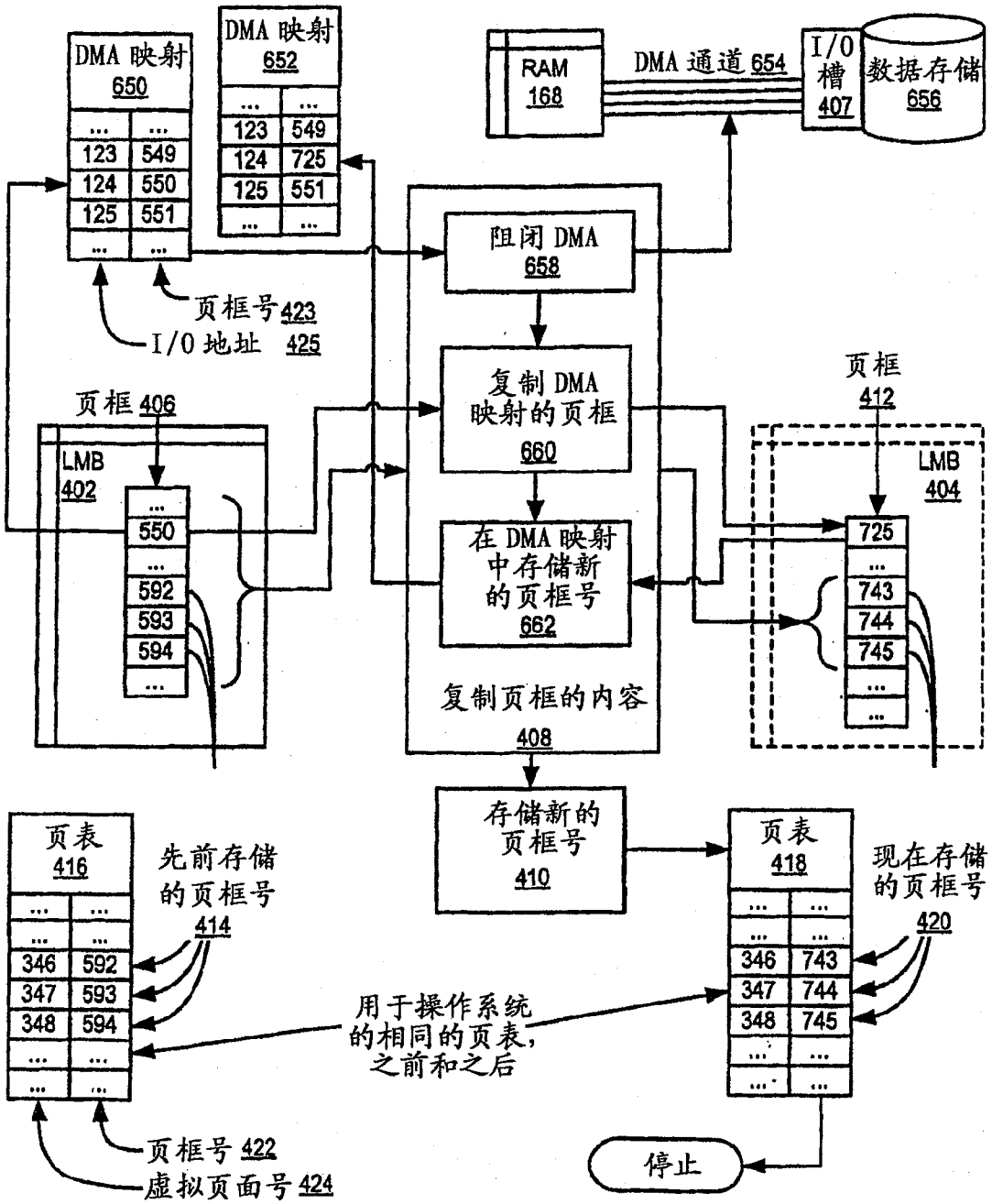

DMA 映射 650
DMA 映射 652
RAM 168
DMA 通道 654
I/O 槽 407
数据存储 656
阻闭 DMA 658
页框号 423
I/O 地址 425
复制 DMA 映射的页框 660
页框 406
LMB 402
在 DMA 映射中存储新的页框号 662
页框 412
LMB 404
复制页框的内容 408
存储新的页框号 410
页表 416
先前存储的页框号 414
页表 418
现在存储的页框号 420
用于操作系统的相同的页表，之前和之后
页框号 422
虚拟页面号 424
停止

图 6

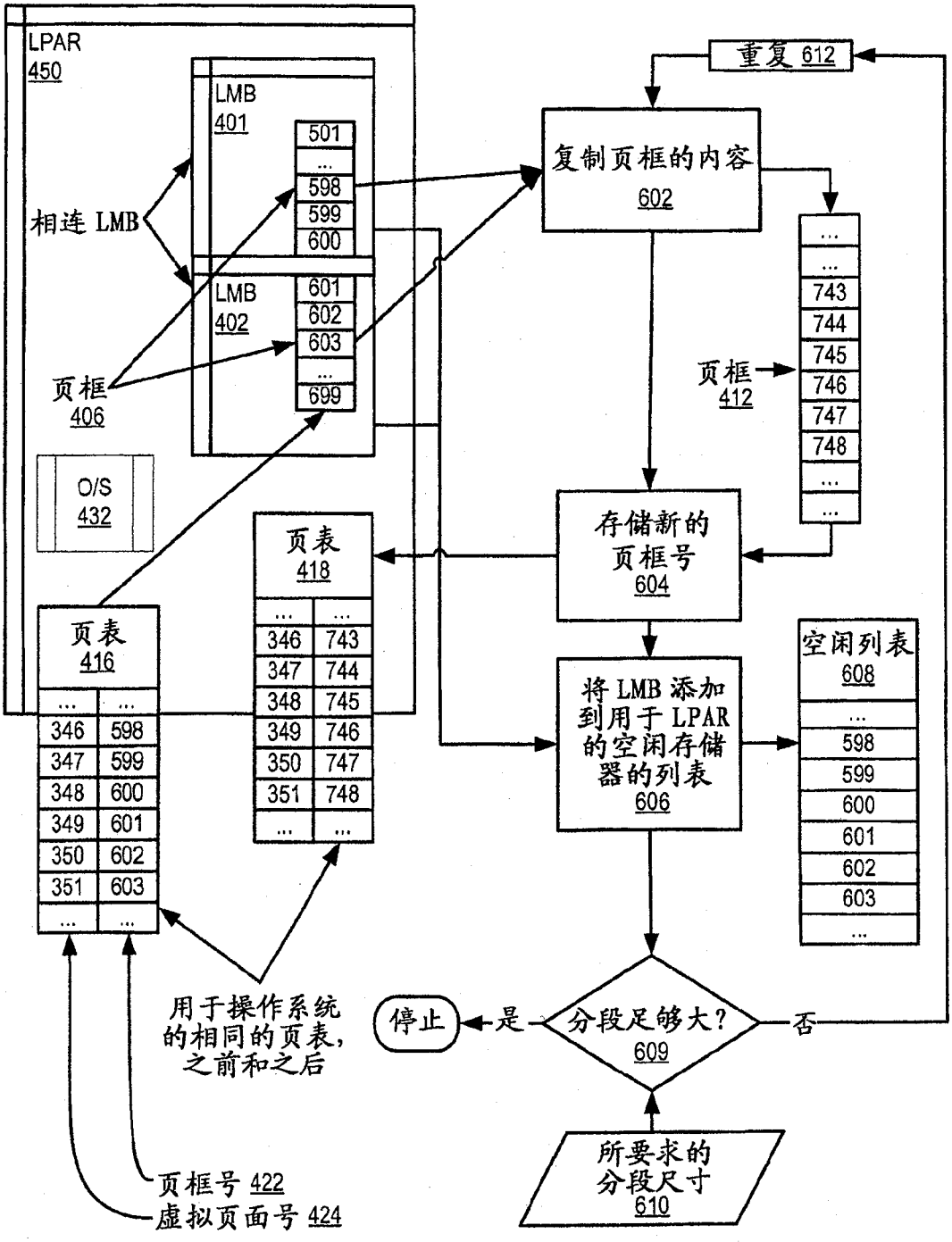

LPAR
450
LMB
401
501
...
598
599
600
LMB
402
601
602
603
...
699
相连 LMB
页框
406
O/S
432
页表
416
346 598
347 599
348 600
349 601
350 602
351 603
页表
418
346 743
347 744
348 745
349 746
350 747
351 748
用于操作系统的相同的页表，之前和之后
页框号 422
虚拟页面号 424
重复 612
复制页框的内容
602
页框
412
743
744
745
746
747
748
存储新的页框号
604
将 LMB 添加到用于 LPAR 的空闲存储器的列表
606
空闲列表
608
598
599
600
601
602
603
分段足够大？
609
是
否
停止
所要求的分段尺寸
610

图 7

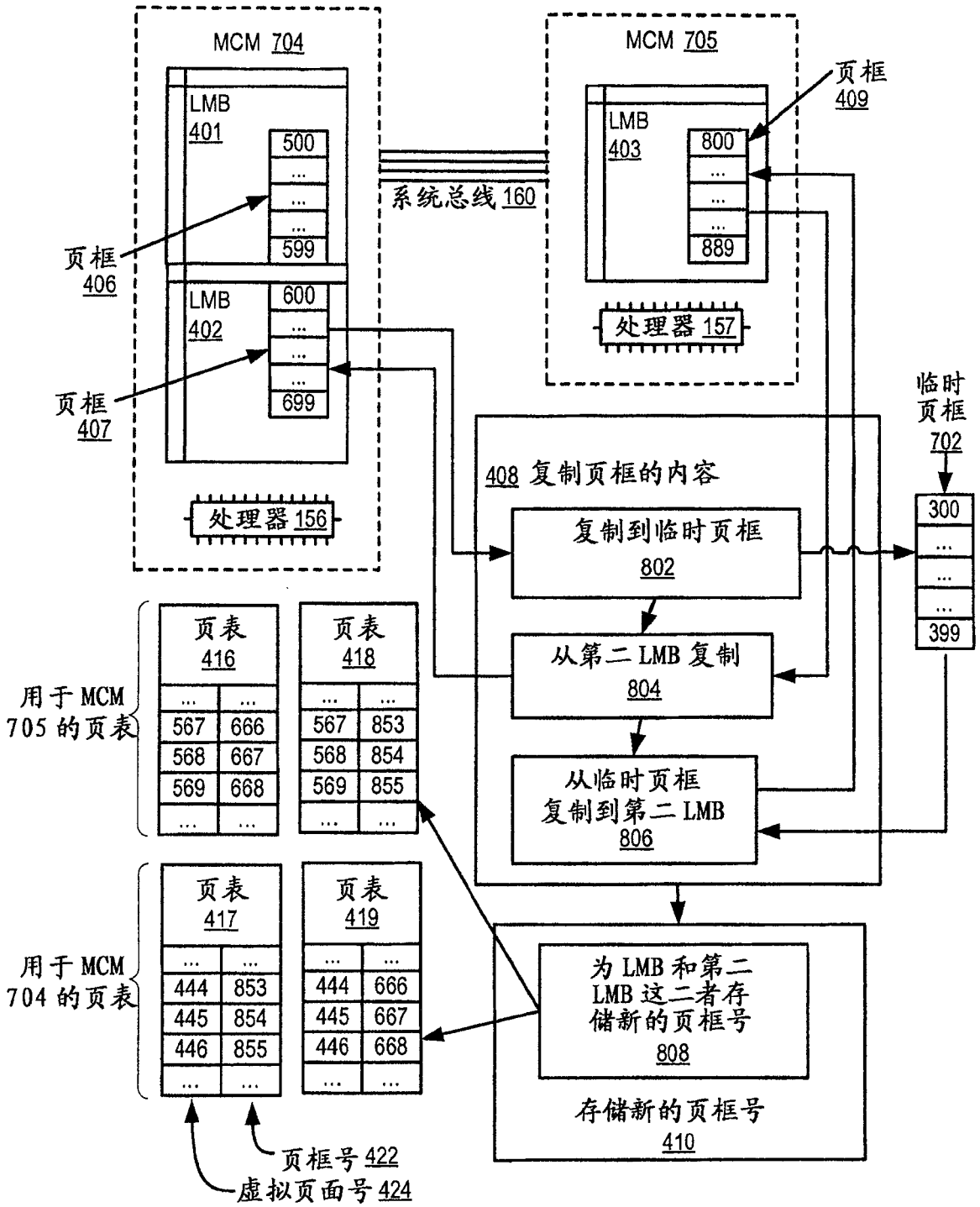

MCM 704
LMB 401
500
599
LMB 402
600
699
处理器 156
页框 406
页框 407
系统总线 160
MCM 705
LMB 403
800
889
处理器 157
页框 409
临时页框 702
300
399
408 复制页框的内容
复制到临时页框 802
从第二 LMB 复制 804
从临时页框复制到第二 LMB 806
为 LMB 和第二 LMB 这二者存储新的页框号 808
存储新的页框号 410
页表 416
页表 418
用于 MCM 705 的页表
567 666
568 667
569 668
567 853
568 854
569 855
页表 417
页表 419
用于 MCM 704 的页表
444 853
445 854
446 855
444 666
445 667
446 668
页框号 422
虚拟页面号 424

图 8