



(12)发明专利

(10)授权公告号 CN 106462219 B

(45)授权公告日 2019.05.14

(21)申请号 201580030639.7

(22)申请日 2015.06.04

(65)同一申请的已公布的文献号

申请公布号 CN 106462219 A

(43)申请公布日 2017.02.22

(30)优先权数据

14/300,457 2014.06.10 US

(85)PCT国际申请进入国家阶段日

2016.12.08

(86)PCT国际申请的申请数据

PCT/US2015/034169 2015.06.04

(87)PCT国际申请的公布数据

WO2015/191358 EN 2015.12.17

(73)专利权人 高通股份有限公司

地址 美国加利福尼亚

(72)发明人 K·K·普苏库里

G·C·卡斯卡瓦尔 M·韦伯

(74)专利代理机构 永新专利商标代理有限公司

72002

代理人 张扬 王英

(51)Int.Cl.

G06F 1/32(2006.01)

G06F 9/50(2006.01)

(56)对比文件

US 2009172424 A1,2009.07.02,

US 7818747 B1,2010.10.19,

US 2002065992 A1,2002.05.30,

US 2009217277 A1,2009.08.27,

US 2013024871 A1,2013.01.24,

US 2011093859 A1,2011.04.21,

CN 1938685 A,2007.03.28,

US 2009150893 A1,2009.06.11,

US 2013024871 A1,2013.01.24,

US 2012144218 A1,2012.06.07,

审查员 李妍君

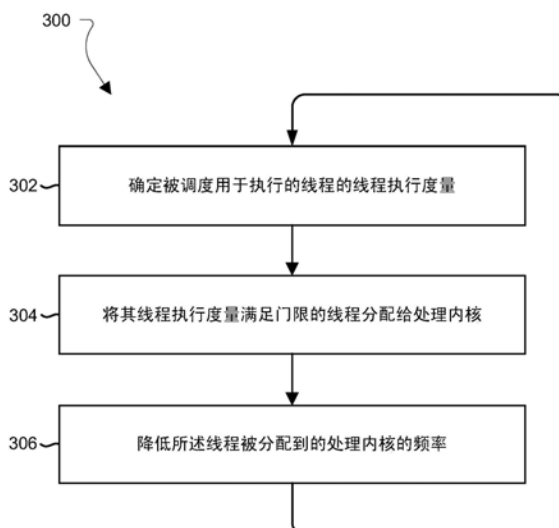
权利要求书7页 说明书13页 附图9页

(54)发明名称

管理处理器设备功耗的系统和方法

(57)摘要

本发明的方面包括管理处理器设备功耗的系统和方法。处理器可以确定被调度用于在包括多个处理内核的处理器中执行的多个线程中的每一个线程的线程执行度量。该处理器可以向选择的处理内核分配其线程执行度量满足门限的那些线程。该处理器可以降低所选择的处理内核的频率以降低功耗。



1. 一种管理处理器设备功耗的方法,包括:

确定被调度用于在包括多个处理内核的处理器中执行的多个线程中的每一个线程的线程执行度量,其中,所述线程执行度量基于通过执行所述多个线程中的相应线程引入的处理低效率;

向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程;以及

降低所述第一处理内核的频率以降低所述第一处理内核的功耗。

2. 如权利要求1所述的方法,其中,向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程包括:

将所述多个线程中其线程执行度量满足所述第一低效率门限的那些线程关联为第一组;以及

将所述第一组的线程分配给所述第一处理内核。

3. 如权利要求1所述的方法,还包括:

向所述多个处理内核中的第二处理内核分配所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程。

4. 如权利要求3所述的方法,还包括:

增加所述第二处理内核的频率,以增加处理所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的速率。

5. 如权利要求3所述的方法,其中,向所述多个处理内核中的第二处理内核分配其线程执行度量不超过所述第一低效率门限的那些线程包括:

将所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程关联为第二组;以及

将所述第二组分配给所述第二处理内核。

6. 如权利要求4所述的方法,还包括:

向所述多个处理内核中的第三处理内核分配其线程执行度量超过第二低效率门限但是不超过所述第一低效率门限的那些线程;以及

将所述第三处理内核的频率调整为所述第一处理内核的所述频率和所述第二处理内核的所述频率之间的频率。

7. 如权利要求1所述的方法,其中:

所述线程执行度量是缓存未命中率;并且

所述第一低效率门限包括缓存未命中率门限值。

8. 如权利要求7所述的方法,其中,降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述缓存未命中率确定所述第一处理内核的缓存未命中率特性;以及

基于所确定的缓存未命中率特性,来降低所述第一处理内核的所述频率。

9. 如权利要求8所述的方法,其中,所述缓存未命中率特性包括以下各项中的一项:

分配给所述第一处理内核的所述线程的平均缓存未命中率、分配给所述第一处理内核的所述线程的均值缓存未命中率、根据分配给所述第一处理内核的所述线程的所述均值缓存未命中率的标准偏差,和分配给所述第一处理内核的所述线程的聚合缓存未命中率。

10. 如权利要求1所述的方法, 其中:

所述线程执行度量还包括同步操作率; 以及

所述第一低效率门限还包括同步操作率门限值。

11. 如权利要求10所述的方法, 其中, 所述同步操作率包括以下各项中的至少一项:

自旋锁的数量、等待请求的共享资源的时间段、锁定指令的数量和同步指令执行率。

12. 如权利要求10所述的方法, 其中, 降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述同步操作率来确定所述第一处理内核的同步操作率特性; 以及

基于所确定的同步操作率特性, 来降低所述第一处理内核的所述频率。

13. 如权利要求1所述的方法, 其中:

所述线程执行度量还包括指令类型率; 以及

所述第一低效率门限还包括指令类型率门限值。

14. 如权利要求13所述的方法, 其中, 所述指令类型率包括以下各项中的至少一项:

浮点指令率、向量指令率和存储器指令率。

15. 如权利要求13所述的方法, 其中, 降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述指令类型率来确定所述第一处理内核的指令类型率特性; 以及

基于所确定的指令类型率特性, 来降低所述第一处理内核的所述频率。

16. 一种计算设备, 包括:

用于确定被调度用于在包括多个处理内核的处理器中执行的多个线程中的每一个线程的线程执行度量的单元, 其中, 所述线程执行度量基于通过执行所述多个线程中的相应线程引入的处理低效率;

用于向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程的单元; 以及

用于降低所述第一处理内核的频率以降低所述第一处理内核的功耗的单元。

17. 如权利要求16所述的计算设备, 其中, 所述用于向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程的单元包括:

用于将所述多个线程中其线程执行度量满足所述第一低效率门限的那些线程关联为第一组的单元; 以及

用于将所述第一组的线程分配给所述第一处理内核的单元。

18. 如权利要求16所述的计算设备, 还包括: 用于向所述多个处理内核中的第二处理内核分配所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的单元。

19. 如权利要求18所述的计算设备, 还包括:

用于增加所述第二处理内核的频率以增加处理所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的速率的单元。

20. 如权利要求18所述的计算设备, 其中, 用于向所述多个处理内核中的第二处理内核分配其线程执行度量不超过所述第一低效率门限的那些线程的单元包括:

用于将所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程关联为第二组的单元; 以及

用于将所述第二组分配给所述第二处理内核的单元。

21. 如权利要求19所述的计算设备,还包括:

用于向所述多个处理内核中的第三处理内核分配其线程执行度量超过第二低效率门限但是不超过所述第一低效率门限的那些线程的单元;以及

用于将所述第三处理内核的频率调整为所述第一处理内核的所述频率和所述第二处理内核的所述频率之间的频率的单元。

22. 如权利要求16所述的计算设备,其中:

所述线程执行度量是缓存未命中率;并且

所述第一低效率门限包括缓存未命中率门限值。

23. 如权利要求22所述的计算设备,其中,用于降低所述第一处理内核的频率的单元还包括:

用于基于分配给所述第一处理内核的所述线程的所述缓存未命中率来确定所述第一处理内核的缓存未命中率特性的单元;以及

用于基于所确定的缓存未命中率特性,来降低所述第一处理内核的所述频率的单元。

24. 如权利要求23所述的计算设备,其中,所述缓存未命中率特性包括以下各项中的一项:

分配给所述第一处理内核的所述线程的平均缓存未命中率、分配给所述第一处理内核的所述线程的均值缓存未命中率、根据分配给所述第一处理内核的所述线程的所述均值缓存未命中率的标准偏差,和分配给所述第一处理内核的所述线程的聚合缓存未命中率。

25. 如权利要求16所述的计算设备,其中:

所述线程执行度量还包括同步操作率;以及

所述第一低效率门限还包括同步操作率门限值。

26. 如权利要求25所述的计算设备,其中,所述同步操作率包括以下各项中的至少一项:

自旋锁的数量、等待请求的共享资源的时间段、锁定指令的数量和同步指令执行率。

27. 如权利要求25所述的计算设备,其中,用于降低所述第一处理内核的频率的单元还包括:

用于基于分配给所述第一处理内核的所述线程的所述同步操作率来确定所述第一处理内核的同步操作率特性的单元;以及

用于基于所确定的同步操作率特性,来降低所述第一处理内核的所述频率的单元。

28. 如权利要求16所述的计算设备,其中:

所述线程执行度量还包括指令类型率;以及

所述第一低效率门限还包括指令类型率门限值。

29. 如权利要求28所述的计算设备,其中,所述指令类型率包括以下各项中的至少一项:

浮点指令率、向量指令率和存储器指令率。

30. 如权利要求28所述的计算设备,其中,用于降低所述第一处理内核的频率的单元还包括:

用于基于分配给所述第一处理内核的所述线程的所述指令类型率来确定所述第一处

理内核的指令类型率特性的单元;以及

用于基于所确定的指令类型率特性,来降低所述第一处理内核的所述频率的单元。

31.一种计算设备,包括:

处理器,其被配置有用于执行包括以下步骤的操作的处理器可执行指令:

确定被调度用于在包括多个处理内核的处理器中执行的多个线程中的每一个线程的线程执行度量,其中,所述线程执行度量基于通过执行所述多个线程中的相应线程引入的处理低效率;

向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程;以及

降低所述第一处理内核的频率以降低所述第一处理内核的功耗。

32.如权利要求31所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作使得向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程包括:

将所述多个线程中其线程执行度量满足所述第一低效率门限的那些线程关联为第一组;以及

将所述第一组的线程分配给所述第一处理内核。

33.如权利要求31所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行还包括向所述多个处理内核中的第二处理内核分配所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的操作。

34.如权利要求33所述的计算设备,还包括:

增加所述第二处理内核的频率以增加处理所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的速率。

35.如权利要求33所述的计算设备,其中,向所述多个处理内核中的第二处理内核分配其线程执行度量不超过所述第一低效率门限的那些线程包括:

将所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程关联为第二组;以及

将所述第二组分配给所述第二处理内核。

36.如权利要求34所述的计算设备,其中,所述处理器被配置有用于执行还包括以下步骤的操作的处理器可执行指令:

向所述多个处理内核中的第三处理内核分配其线程执行度量超过第二低效率门限但是不超过所述第一低效率门限的那些线程;以及

将所述第三处理内核的频率调整为所述第一处理内核的所述频率和所述第二处理内核的所述频率之间的频率。

37.如权利要求31所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作,使得:

所述线程执行度量是缓存未命中率;并且

所述第一低效率门限包括缓存未命中率门限值。

38.如权利要求37所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作,使得降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述缓存未命中率来确定所述第一处理内核的缓存未命中率特性;以及

基于所确定的缓存未命中率特性,来降低所述第一处理内核的所述频率。

39.如权利要求38所述的计算设备,其中,所述缓存未命中率特性包括以下各项中的一项:

分配给所述第一处理内核的所述线程的平均缓存未命中率、分配给所述第一处理内核的所述线程的均值缓存未命中率、根据分配给所述第一处理内核的所述线程的所述均值缓存未命中率的标准偏差,和分配给所述第一处理内核的所述线程的聚合缓存未命中率。

40.如权利要求31所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作,使得:

所述线程执行度量包括同步操作率;以及

所述第一低效率门限包括同步操作率门限值。

41.如权利要求40所述的计算设备,其中,所述同步操作率包括以下各项中的至少一项:

自旋锁的数量、等待请求的共享资源的时间段、锁定指令的数量和同步指令执行率。

42.如权利要求40所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作,使得降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述同步操作率来确定所述第一处理内核的同步操作率特性;以及

基于所确定的同步操作率特性,来降低所述第一处理内核的所述频率。

43.如权利要求31所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作,使得:

所述线程执行度量包括指令类型率;以及

所述第一低效率门限包括指令类型率门限值。

44.如权利要求43所述的计算设备,其中,所述指令类型率包括以下各项中的至少一项:

浮点指令率、向量指令率和存储器指令率。

45.如权利要求43所述的计算设备,其中,所述处理器被配置有处理器可执行指令,所述处理器可执行指令用于执行操作,使得降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述指令类型率来确定所述第一处理内核的指令类型率特性;以及

基于所确定的指令类型率特性,来降低所述第一处理内核的所述频率。

46.一种其上存储有处理器可执行软件指令的非临时性处理器可读存储介质,所述处理器可执行软件指令被配置为使处理器执行用于管理处理器设备功耗的操作,所述操作包括:

确定被调度用于在包括多个处理内核的处理器中执行的多个线程中的每一个线程的线程执行度量,其中,所述线程执行度量基于通过执行所述多个线程中的相应线程引入的处理低效率;

向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第

一低效率门限的那些线程;以及

降低所述第一处理内核的频率以降低所述第一处理内核的功耗。

47. 如权利要求46所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得向所述多个处理内核中的第一处理内核分配所述多个线程中其线程执行度量超过第一低效率门限的那些线程包括:

将所述多个线程中其线程执行度量满足所述第一低效率门限的那些线程关联为第一组;以及

将所述第一组的线程分配给所述第一处理内核。

48. 如权利要求46所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行还包括向所述多个处理内核中的第二处理内核分配所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的操作。

49. 如权利要求48所述的非临时性处理器可读存储介质,还包括增加所述第二处理内核的频率以增加处理所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程的速率。

50. 如权利要求48所述的非临时性处理器可读存储介质,其中,向所述多个处理内核中的第二处理内核分配其线程执行度量不超过所述第一低效率门限的那些线程包括:

将所述多个线程中其线程执行度量不超过所述第一低效率门限的那些线程关联为第二组;以及

将所述第二组分配给所述第二处理内核。

51. 如权利要求49所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行还包括以下步骤的操作:

向所述多个处理内核中的第三处理内核分配其线程执行度量超过第二低效率门限但是不超过所述第一低效率门限的那些线程;以及

将所述第三处理内核的频率调整为所述第一处理内核的所述频率和所述第二处理内核的所述频率之间的频率。

52. 如权利要求46所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得:

所述线程执行度量是缓存未命中率;并且

所述第一低效率门限包括缓存未命中率门限值。

53. 如权利要求52所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述缓存未命中率来确定所述第一处理内核的缓存未命中率特性;以及

基于所确定的缓存未命中率特性,来降低所述第一处理内核的所述频率。

54. 如权利要求53所述的非临时性处理器可读存储介质,其中,所述缓存未命中率特性包括以下各项中的一项:

分配给所述第一处理内核的所述线程的平均缓存未命中率、分配给所述第一处理内核的所述线程的均值缓存未命中率、根据分配给所述第一处理内核的所述线程的所述均值缓存未命中率的标准偏差,和分配给所述第一处理内核的所述线程的聚合缓存未命中率。

55. 如权利要求46所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得:

所述线程执行度量包括同步操作率;以及

所述第一低效率门限包括同步操作率门限值。

56. 如权利要求55所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得所述同步操作率包括以下各项中的至少一项:

自旋锁的数量、等待请求的共享资源的时间段、锁定指令的数量和同步指令执行率。

57. 如权利要求55所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述同步操作率来确定所述第一处理内核的同步操作率特性;以及

基于所确定的同步操作率特性,来降低所述第一处理内核的所述频率。

58. 如权利要求46所述的非临时性处理器可读存储介质,其中,所存储的处理器可执行软件指令被配置为使处理器执行操作,使得:

所述线程执行度量包括指令类型率;以及

所述第一低效率门限包括指令类型率门限值。

59. 如权利要求58所述的非临时性处理器可读存储介质,其中,所述指令类型率包括以下各项中的至少一项:

浮点指令率、向量指令率和存储器指令率。

60. 如权利要求58所述的非临时性处理器可读存储介质,其中,降低所述第一处理内核的频率还包括:

基于分配给所述第一处理内核的所述线程的所述指令类型率确定所述第一处理内核的指令类型率特性;以及

基于所确定的指令类型率特性,来降低所述第一处理内核的所述频率。

管理处理器设备功耗的系统和方法

背景技术

[0001] 无线通信技术和移动电子设备(例如,蜂窝电话、平板电脑、膝上型电脑等等)已经越来越普及使用。为了与增长的消费者需求保持同步,移动设备已经变得更功能多样,并且现在一般包括多个处理设备、片上系统(SOC)和其它资源以用于允许移动设备用户在移动设备上执行复杂的且功率密集型软件应用(例如,视频和音频流和/或处理应用、网络游戏应用等等)。由于这些和其它改进,智能电话和平板计算机已经越来越普及,并且在替代膝上型和桌面型机器来作为很多用户的平台选择。但是,移动设备通常包括相对有限的电源。在一个移动设备中提供多个处理设备加剧了移动设备的功率存储局限性。

发明内容

[0002] 各个方面包括管理处理器设备功耗的方法,其可以包括确定被调度用于在包括多个处理内核的处理器中执行的多个线程中的每一个线程的线程执行度量;向所述多个处理内核中的第一处理内核分配所述多个线程中的其线程执行度量满足第一门限的那些线程;以及降低所述第一处理内核的频率以降低所选择的处理内核的功耗。在一个方面,向第一处理内核分配其线程执行度量满足第一门限的那些线程可以包括将其线程执行度量满足所述第一门限的那些线程关联为第一组;以及将所述第一组的线程分配给所述第一处理内核。

[0003] 在一个方面,该方法可以包括向所述多个处理内核中的第二处理内核分配所述多个线程中的其线程执行度量不满足所述第一门限的那些线程。在其它方面,该方法可以包括增加所述第二处理内核的频率以增加处理所述多个线程中的其线程执行度量不满足所述第一门限的那些线程的速率。在其它方面,向所述多个处理内核中的第二处理内核分配其线程执行度量不满足所述第一门限的那些线程可以包括将所述多个线程中的其线程执行度量不满足所述第一门限的那些线程关联为第二组,以及将所述第二组分配给所述第二处理内核。

[0004] 在一个其它方面,该方法可以包括向所述多个处理内核中的第三处理内核分配其线程执行度量满足第二门限但是不满足所述第一门限的那些线程,以及将所述第三处理内核的频率调整为所述第一处理内核的所述频率和所述第二处理内核的所述频率之间的频率。

[0005] 在一个其它方面,所述线程执行度量可以包括缓存未命中率,并且所述第一门限可以包括缓存未命中率门限值。在一些方面,所述缓存未命中率可以包括以下各项中的至少一项:在缓存中读取或写入数据的失败尝试的次数、基于所述缓存中读取或写入数据的失败尝试的主存储器访问次数,以及基于所述缓存中读取或写入数据的失败尝试的处理器等待状态次数。

[0006] 在一个其它方面,降低所述第一处理内核的频率还可以包括基于分配给所述第一处理内核的所述线程的所述缓存未命中率确定所述第一处理内核的缓存未命中率特性,以及基于所确定的缓存未命中率特性降低所选择的处理内核的所述频率。在其它方面,所述

缓存未命中率特性可以包括以下各项中的一项:分配给所选择的处理内核的所述线程的平均缓存未命中率、分配给所选择的处理内核的所述线程的均值缓存未命中率、根据分配给所选择的处理内核的所述线程的所述均值缓存未命中率的标准偏差,和分配给所选择的处理内核的所述线程的聚合缓存未命中率。

[0007] 在一个其它方面,所述线程执行度量可以包括同步操作率,并且所述第一门限可以包括同步操作率门限值。在一些方面,所述同步操作可以包括以下各项中的至少一项:自旋锁的数量、等待请求的共享资源的时间段、锁定指令的数量和同步指令执行率。

[0008] 在一个方面,降低所述第一处理内核的频率还可以包括基于分配给所述第一处理内核的所述线程的所述同步操作率确定所述第一处理内核的同步操作率特性,以及基于所确定的同步操作率特性降低所述第一处理内核的所述频率。

[0009] 在一个其它方面,所述线程执行度量可以包括指令类型率,并且所述第一门限可以包括指令类型率门限值。在其它方面,所述指令类型率可以包括以下各项中的至少一项:浮点指令率、向量指令率和存储器指令率。在其它方面,降低所述第一处理内核的频率还包括基于分配给所述第一处理内核的所述线程的所述指令类型率确定所述第一处理内核的指令类型率特性,以及基于所确定的指令类型率特性降低所述第一处理内核的频率。

[0010] 其它方面包括计算设备,其包括处理器,该处理器被配置有用于执行上述方面方法的操作的处理器可执行指令。其它方面包括其上存储有处理器可执行软件指令的非临时性处理器可读存储器介质,所述处理器可执行软件指令被配置为使处理器执行上述方面方法的操作。其它方面包括计算设备,其包括用于执行上述方面方法的操作的功能的单元。

附图说明

[0011] 合并到本文中并且组成本说明书的一部分的附图解释说明了本发明的示例性方面。与上面给出的概述和下面给出的具体实施方式一起,所述附图用于解释本发明的特征,而非限制所公开的方面。

[0012] 图1是示出可以用在实现各个方面的计算设备中的示例性片上系统(SOC)架构的组件框图。

[0013] 图2是示出可以用于实现各个方面的示例性多核处理器架构的功能框图。

[0014] 图3是示出管理处理器设备功耗的方面方法的过程流程图。

[0015] 图4是示出管理处理器设备功耗的另一方面方法的过程流程图。

[0016] 图5是示出管理处理器设备功耗的另一方面方法的过程流程图。

[0017] 图6是示出管理处理器设备功耗的另一方面方法的过程流程图。

[0018] 图7是适合与各个方面一起使用的示例性移动设备的组件框图。

[0019] 图8是适合与各个方面一起使用的示例性服务器的组件框图。

[0020] 图9是适合于实现各个方面的膝上型计算机的组件框图。

具体实施方式

[0021] 将参照附图详细描述各个方面。只要有可能,相同的附图标记在贯穿附图中将用于指代相同或相似的部分。针对特定示例和实现做出的引用是用于解释说明的目的,而并不意在限制权利要求的范围。

[0022] 本文中使用的词语“示例性”意味着“用作示例、实例或举例说明”。本文中被描述为“示例性”的任何实现不一定被解释为比其它实现更优选或更有优势。

[0023] 术语“移动设备”和“计算设备”在本文中交替使用以指代以下各项中的任意一项或所有项：蜂窝电话、智能电话、个人或移动多媒体播放器、个人数字助理 (PDA)、膝上型计算机、平板电脑、智能本、掌上计算机、无线电子邮件接收机、具备多媒体互联网功能的蜂窝电话、无线游戏控制器和类似的包括可编程处理器和存储器的电子设备。尽管各个方面具体地用于诸如蜂窝电话和其它便携式计算平台之类的移动设备 (其可能具有相对有限的处理功率和/或功率存储容量)，但是这些方面通常也可用于向处理设备或处理内核分配线程、进程或其它指令序列的任何计算设备。

[0024] 术语“片上系统” (SOC) 在本文中用于指代在单个基底上包含集成的多个资源和/或处理器的单个集成电路 (IC) 芯片。单个 SOC 可以包含用于数字、模拟、混合信号的电路和射频功能。单个 SOC 还可以包括任何数量的通用处理器和/或专用处理器 (数字信号处理器、调制解调器处理器、视频处理器等等)、存储器块 (例如, ROM、RAM、闪存等等) 和资源 (例如, 定时器、稳压器、振荡器等等)。SOC 还可以包括用于控制该集成资源和处理器以及用于控制外围设备的软件。

[0025] 术语“多核处理器”在本文中用于指代包含被配置为读取并执行程序指令的两个或更多独立处理设备或处理内核 (例如, CPU 内核) 的单个集成电路 (IC) 芯片或芯片包。SOC 可以包括多个多核处理器, 并且 SOC 中的每个处理器可以被称为一个内核或处理内核。术语“多处理器”在本文中用于指代包括被配置为读取并执行程序指令的两个或更多处理单元的系统或设备。术语“线程”或“进程”在本文中用于指代指令序列, 其可以被分配给处理内核。

[0026] 如本申请中所使用的, 术语“组件”、“模块”、“系统”等等意在包括计算机相关实体, 诸如但并不仅限于硬件、固件、硬件和软件的组合、软件或运行中的软件, 它们被配置为执行特定操作或功能。例如, 一个组件可以是, 但并不仅限于运行在处理器上的进程、处理器、对象、可执行程序、运行的线程、程序和/或计算机。举例来讲, 运行在计算设备上的应用和该计算设备二者可以被称为组件。一个或多个组件可以驻留在运行的进程和/或线程中, 并且一个组件可以位于一个处理器或内核上和/或分布在两个或更多处理器或内核之间。另外, 可以根据其上存储有各种指令和/或数据结构的各种非临时性处理器可读介质, 来执行这些组件。组件可以通过本地和/或远程过程、函数或程序调用、电子信号、数据分组、存储器读取/写入和其它公知的计算机、处理器和/或过程相关的通信方法的方式, 来进行通信。

[0027] 为了与增加的消费者需求保持同步, 移动设备已经变得越来越功能多样, 并且现在一般包括用于允许移动设备用户在移动设备上执行复杂的且功率密集型软件应用 (例如, 视频和音频流和/或处理应用、网络游戏应用等等) 的多个处理设备、多核处理器、片上系统 (SOC) 和其它资源。移动设备通常包括相对有限的电源。在一个移动设备中提供多个处理设备加剧了移动设备的功率存储局限。

[0028] 各个方面提供了用于通过基于确定线程执行度量来跨多核系统中的多个内核调度指令序列 (诸如进程、线程等等), 以及基于指派给每个处理内核的指令序列确定一个或多个处理内核的适当频率, 来改善移动设备的性能和功耗特性的方法、系统和设备。各个方

面可以实现为用户空间守护进程、(例如,移动设备的操作系统的)内核模块或操作系统调度器。在各个方面中,可以将所具有的线程执行度量满足门限线程执行度量的一个或多个线程分配给一个处理内核,并且可以降低该处理内核的操作频率以改善该处理内核的功耗和处理效率,同时提供该移动设备的门限水平的性能。

[0029] 各个方面可以具有多核系统中的特定应用,其中同时执行两个或更多应用。随着同时执行的应用的数量的增加,由该处理内核执行的指令序列的数量可能显著地增加,潜在地由于与可用的处理内核相比更多数量的线程而使该处理内核过载。当指令序列向处理器引入低效率时,诸如通过要求调用非缓存存储器,或通过包括更多地要求惊人的处理器资源或者当前由另一个进程或处理器使用的资源,该处理内核伴随的延迟(例如,等待时间)可能对应用性能有明显影响。

[0030] 将线程执行度量用作向特定内核分配指令序列的基础提供了一种用于针对一个或几个内核进程进行分配的低开销方法,所述内核进程涉及更严重的低效率,诸如在从非缓存存储器获取数据时的更多的等待时间、由同步操作造成的延迟和/或执行一种或多种类型的指令的速率。由于指令序列的处理时间可能主要由通过某些指令序列引入的处理低效率来决定,因此执行相对比较消耗处理器资源的指令序列的处理器内核可以以更低的频率并因此的更低的功率来操作,而不会明显地影响处理性能。因此,各个方面提供一种低开销方法,该方法用于将处理密集型指令序列分配给选择的能够以较高频率/较高功率来操作的处理内核以改善处理性能,并且将消耗更多资源/或时间的指令序列分配给能够以较低频率/较低功率来操作的一个或几个内核以便在不影响处理性能的前提下节省电池功率。

[0031] 可以在多个单处理器和多处理器计算机系统(包括片上系统(SOC))上来实现各个方面。图1示出可以用在实现各个方面的计算设备中的示例性片上系统(SOC) 100架构。SOC 100可以包括多个异构处理器,诸如数字信号处理器(DSP) 102、调制解调器处理器104、图形处理器106和应用处理器108。SOC 100还可以包括连接到异构处理器102、104、106、108中的一个或多个异构处理器的一个或多个协处理器110(例如,向量协处理器)。每个处理器102、104、106、108、110可以包括一个或多个内核(例如,应用处理器108中示出的处理内核108a、108b、108c和108d),并且每个处理器/内核可以独立于其它处理器/内核来执行操作。SOC 100可以包括用于执行操作系统(例如,FreeBSD、LINUX、OS X、Microsoft Windows 8等)的处理器,其可以包括被配置为将指令序列(诸如线程、进程或数据流)调度给一个或多个处理内核以用于执行的调度器。

[0032] 该SOC 100还可以包括模拟电路和定制电路114,其用于管理传感器数据、模拟到数字转换、无线数据传输,以及用于执行诸如处理要在网络浏览器中展示的经验码音频和视频信号之类的其它专门操作。该SOC 100还可以包括系统组件和资源116,诸如稳压器、振荡器、锁相环、外设桥接、数据控制器、存储器控制器、系统控制器、接入端口、定时器和其它类似的用于支持运行在计算设备上的处理器和软件程序的组件。

[0033] 该系统组件和资源116和/或定制电路114可以包括用于与外围设备(诸如摄像机、电子显示器、无线通信设备、外部存储器芯片等等)交互的电路。处理器102、104、106、108可以彼此之间进行通信,以及经由互连/总线模块124与一个或多个存储器元件112、系统组件和资源116和定制电路114通信,该互连/总线模块124可以包括可重新配置逻辑门阵列和/

或实现总线架构(例如,CoreConnect、AMBA等等)。可以由诸如高性能片上网络(NoC)之类的高级互连来提供通信。

[0034] SoC 100还可以包括输入/输出模块(未示出)用于与SoC外部的资源(诸如时钟118和稳压器120)通信。SOC外部的资源(例如,时钟118、稳压器120)可以由两个或更多内部SOC处理器/内核(例如,DSP 102、调制解调器处理器104、图形处理器106、应用处理器108等等)来共享。

[0035] 除了上面讨论的SOC 100,各个方面可以实现在很多不同计算系统中,该计算系统可以包括多个处理器、多核处理器或它们的任何组合。

[0036] 图2示出可以用于实现各个方面的示例性多核处理器架构。多核处理器202可以包括非常靠近(例如在单个基底、管芯(die)、集成芯片等上)的两个或更多独立处理内核204、206、230、232。处理内核204、206、230、232的接近性允许存储器以比如果信号不得经过芯片之外的可能频率/时钟速率高得多的频率/时钟速率来操作。处理内核204、206、230、232的接近性允许对片上存储器和资源(例如,电压轨)的共享,以及内核之间更协调的协作。尽管图2中示出了四个处理内核,但是应该了解的是,这并不是一个限制,多核处理器可以包括更多或更少的处理内核。

[0037] 该多核处理器202可以包括多级缓存,该多级缓存包括1级(L1)缓存212、214、238和240以及2级(L2)缓存216、226和242。多核处理器202还可以包括总线/互连接口218、主存储器220和输入/输出模块222。L2缓存216、226、242可以比L1缓存212、214、238、240更大(和更慢),但是比主存储器单元220更小(并且大体上更快)。每个处理内核204、206、230、232可以包括具有到L1缓存212、214、238、240的私有访问的处理单元208、210、234、236。处理内核204、206、230、232可以共享到L2缓存(例如,L2缓存242)的访问,或者可以具有到独立L2缓存(例如,L2缓存216、226)的访问。

[0038] L1和L2缓存可以用于存储被处理单元频繁访问的数据,而主存储器220可以用于存储被处理内核204、206、230、232访问的更大的文件和数据单元。多核处理器202可以被配置为使得处理内核204、206、230、232按照首先查询L1缓存,然后L2缓存以及然后如果信息没有被存储在缓存中则查询主存储器的顺序从存储器寻找数据。如果该信息没有被存储在缓存或主存储器220中,则多核处理器202可以从外部存储器和/或硬盘存储器224寻找信息。

[0039] 处理内核204、206、230、232可以经由总线/互连接口218彼此之间进行通信。每个处理内核204、206、230、232可以具有对一些资源的独有控制,也可以与其它内核共享其它资源。

[0040] 处理内核204、206、230、232可以彼此之间是相同的、是异构的和/或实现不同专用功能的。因此,处理内核204、206、230、232不需要对称,无论是从操作系统的角度(例如,可以执行不同的操作系统)还是从硬件角度(例如,可以实现不同的指令集/架构)。

[0041] 诸如上面参考图1和图2所讨论的那些多处理器硬件设计可以包括同一个包装内的多个不同能力的处理内核,通常也是在同一片硅片上的。对称多处理硬件包括连接到由单个操作系统控制的单个共享主存储器的两个或更多相同处理器。不对称或“松散耦合的”多处理硬件可以包括每一个可以是由独立操作系统控制并且连接到一个或多个共享存储器/资源的两个或更多异构处理器/内核。

[0042] 图3是示出了管理处理器设备功耗的方面方法300的过程流程图。在块302中,处理器可以确定被调度用于在多核处理器的各个处理内核(例如,图2中示出的处理内核204、206、230、232)中执行的多个线程中的每一个线程的线程执行度量。该线程执行度量可以包括,例如缓存未命中率、同步操作率、一种或多种类型的指令的执行率或其它线程执行度量。

[0043] 术语“缓存未命中”指的是从缓存存储器读取或写入数据单元的失败尝试,需要从系统存储器(例如,主存储器220)来访问数据,这比当数据处于缓存中时需要更多的时间来完成。例如,由处理内核204的处理单元208处理的线程可以尝试从L1缓存212取回数据或指令。当该数据或指令无法从L1缓存212获得时,可以从另一个缓存(诸如L2缓存216)或者从主存储器220取回该数据或指令。主存储器访问尤其大幅度地慢于缓存访问,因此从主存储器220取回数据或指令对该线程的处理引入时延。由于从主存储器220取回数据或指令的延迟,该处理内核可以进入等待状态同时该主存储器访问继续进行,在此期间不会执行额外的处理。如果一个线程包括很多非缓存读取或写入操作,则该处理内核会频繁地进入等待状态,因此减慢该内核的整体处理性能。

[0044] 在各个方面,由处理器在块302中针对每个线程计算出的缓存未命中率可以是在执行线程时将会发生的在缓存中读取或写入数据的失败尝试的数量。可以在块302中计算的缓存未命中率也可以是在基于在缓存中读取或写入数据的失败尝试将会在该线程执行期间发生的主存储器访问的次数。在块302中计算的缓存未命中率还可以是作为针对该线程的非缓存存储器执行而执行的数据或指令读取和写入操作的结果,处理器内核将会进入等待状态的次数。一个线程的缓存未命中率可以在块302中计算为每一单位时间的缓存未命中平均次数和/或每一单位时间的合计次数。该缓存未命中率还可以在块302中计算为每一单位时间的平均数量和/或每一单位时间的平均数量的标准偏差。其它缓存未命中计算也是可能的,包括上述的组合。

[0045] 术语“同步操作”指的是可以执行的使得两个或更多处理器能够,例如在公共数据结构集合上或者使用共享的应用代码片段来执行指令而不干扰另一处理器的处理操作,或者没有由于用于执行指令集合的处理或其它资源的临时不可用性而造成对该指令集合的执行的延迟的操作。同步操作可以是基于竞争的,使得对同步操作的执行率的上升可能向处理器性能引入时延。由处理器在块302中计算出的同步操作率可以包括自旋锁的数量、等待所请求的共享资源的时间段、锁定指令的数量和同步指令执行率中的至少一项。

[0046] 术语“指令类型”指的是可以根据指定的操作类型分类的指令,诸如浮点指令、向量指令、存储器指令率或另一种类型的指令。每个指令类型可能在用于执行的处理器和/或资源需求的程度上有差异,并且对某些指令类型的执行率的增加可能向处理器性能引入时延。处理器在块302中计算出的指令类型率可以包括浮点指令率、向量指令率和存储器指令率中的至少一项。

[0047] 在块304中,一个处理器可以将其线程执行度量满足门限的那些线程分配给选择的处理内核。可以满足该门限的该线程执行度量的值可以依赖于所考虑的线程执行度量的类型。在各个方面,根据特定度量,当线程执行度量低于门限、高于门限、小于或等于门限、大于或等于门限或者等于门限时,该线程执行度量可以满足该门限。换句话说,该门限可以是上边界或下边界。在一些方面,可以使用两个(或更多)门限,诸如以便确定线程执行度量是否高于第一门限、低于第二门限或者在第一门限和第二门限之间。

[0048] 举个例子,线程执行度量可以是缓存未命中率,并且可以将那些将执行多于门限数量的针对非缓存存储器的读/写操作的线程分配给一个或几个内核。以此方式,那些将在频繁地使处理内核置于等待状态时来从非缓存存储器访问数据或向其写入数据的线程可以被分配给一个或几个内核,于是这些内核可以以更低的频率来操作以节省功率。块304中的该过程还可以将剩余的主要向缓存存储器写入或从其读取并且因此会更少地将该处理内核置于等待状态的线程分配给其它处理内核,于是这些内核可以以更高的频率来操作以改善性能。

[0049] 再举一个例子,线程执行度量可以包括同步操作,并且当处理器执行门限率的同步操作时,可以将那些可能触发同步操作的执行的线程分配给一个或几个内核,于是这些内核可以以更低的频率来操作。

[0050] 再举一个例子,线程执行度量可以包括指令类型率,并且当该处理器执行门限率的一个或多个特定指令类型时,可以将那些一个或几个特定指令类型的线程分配给一个或几个内核,于是这些内核可以以更低的频率来操作。

[0051] 基于确定的线程执行度量,处理器可以在块304中将每个线程分配给(即,指派或调度)一个或一组处理内核。当多于一个线程的线程执行度量满足门限时,可以将这些线程中的每一个线程分配给相同的处理内核。

[0052] 在一个方面,线程可以被串行地分配给所选择的处理内核。换句话说,当确定一个线程的线程执行度量满足门限时,该线程可以被分配给所选择的处理内核。在另一个方面,线程可以在最小或门限数量的线程的线程执行度量满足该门限时被分配给所选择的处理内核。

[0053] 在块306中,处理器可以降低被选择用于执行具有满足门限的线程执行度量的线程的处理内核的频率,以便减少所选择的处理内核的功耗。例如,可以使用诸如动态电压和频率缩放、动态时钟和电压缩放和类似的功率减少技术之类的动态功率减少技术来降低所选择的处理内核的操作频率。在一个方面,所选择的处理内核的操作频率可以被降低到基于被分配给所选择的处理内核的线程的线程执行度量的水平。例如,线程执行度量越高,所选择的处理内核的频率可以越低。处理器可以周期性地重复方法300以便动态地重新确定一个或多个处理内核中的线程的线程执行度量。

[0054] 图4是示出通过基于线程执行度量向处理内核分配线程来管理处理器设备功耗的另一方面方法400的过程流程图。在块302中,处理器可以如上参考图3所描述的确定被调度用于在多个处理内核中的一个处理内核中执行的多个线程中的每一个线程的线程执行度量。

[0055] 在块404中,处理器可以将其线程执行度量满足门限的那些线程关联为一个组。例如,可以将每个线程的线程执行度量与门限线程执行度量进行比较。在一个方面,当一个线程的线程执行度量满足该门限时,该线程可以被分配到该组。当多于一个线程的线程执行度量满足门限时,这些线程中的每一个线程被分配到同一组。该线程组可以以高延迟或低效率(例如,高缓存未命中率),并且因此的高比率地将处理内核置于等待状态为特征。

[0056] 在块406中,处理器可以将其线程执行度量满足门限的线程组分配给所选择的处理内核。基于确定的线程执行度量,该组可以被分配(即,该组中的每个线程可以被指派或调度)给所选择的处理内核。在一个方面,当该组中的每个线程的线程执行度量满足门限时,该线程可以被分配给所选择的处理内核。该线程组的线程执行度量还可以基于被分配

到该组的线程的平均 (average) 线程执行度量 (例如, 平均缓存未命中率、平均同步操作率或者平均指令类型率)、被分配到该组的线程的均值 (mean) 线程执行度量 (例如, 均值缓存未命中率、均值同步操作率或均值指令类型率)、根据被分配到该组的线程的均值线程执行度量的标准偏差 (例如, 根据均值缓存未命中率、均值同步操作率或均值指令类型率的标准偏差), 以及被分配到该组的线程的聚合线程执行度量 (例如, 聚合的缓存未命中率、同步操作率或指令类型率) 来确定。其它示例也是可能的, 包括上述的组合。

[0057] 在块408中, 处理器可以降低所选择处理内核的频率, 以减少所选择处理内核的功耗。在一些方面, 可以使用诸如动态电压和频率缩放、动态时钟和电压缩放或另种类似的功率降低技术之类的动态功率降低技术来降低所选择处理内核的操作频率。在一个方面, 可以基于被分配给处理内核的那些线程的线程执行度量来确定针对所选择的处理内核的线程执行度量的特性, 并且可以基于该线程执行度量的特性来降低所选择的处理内核的操作频率。例如, 缓存未命中率的特性可以包括被分配给所选择的处理内核的线程 (或者被分配给该处理内核的该组) 的平均缓存未命中率、被分配给所选择的处理内核的线程的均值缓存未命中率、根据被分配给所选择的处理内核的线程的均值缓存未命中率的标准偏差和被分配给所选择的处理内核的线程的聚合缓存未命中率之一。在其它示例中, 除了缓存未命中率特性或替代于缓存未命中率特性, 可以类似地确定同步操作率特性和/或指令类型率特性。

[0058] 可以基于相对于确定的线程执行度量的特性的关系来降低处理内核的操作频率, 诸如通过将操作频率与确定的线程执行度量的特性成比例地降低、基于确定的线程执行度量的特性和操作频率之间的比率、基于确定的线程执行度量的特性来计算频率降低因子, 或者使用另一种基于确定的线程执行度量的特性来计算操作频率的降低的方法。处理器可以周期性地重复方法400以便动态地重新确定所选择的处理内核的和/或一个或多个处理内核中的线程的线程执行度量的特性。

[0059] 图5是示出管理处理器设备功耗的另一方面方法500的过程流程图。在块302中, 处理器可以如上参考图3描述地确定被调度用于在处理内核中执行的多个线程中的每一个线程的线程执行度量。

[0060] 在块404中, 处理器可以如上参考图4所描述地将其线程执行度量满足门限的那些线程关联为第一组。可以通过将每个线程的标识符与第一组标识符关联来将线程关联为该第一组。

[0061] 在块506中, 处理器可以将其线程执行度量不满足该门限的那些线程关联为第二组。可以通过将每个线程的标识符与第二组标识符关联来将这些线程关联为该第二组。

[0062] 在块508中, 处理器可以将第一组中的线程 (例如, 其线程执行度量满足缓存未命中率门限的那些线程) 分配给选择的第一处理内核。基于确定的线程执行度量, 该组可以被分配 (即, 该组中的每个线程可以被指派或调度) 给所选择的处理内核。在一个方面, 当该组中的每个线程的线程执行度量满足该门限时, 该组的线程可以被分配给所选择的处理内核。该线程组的线程执行度量还可以基于分配给该组的线程的平均线程执行度量、分配给该组的线程的均值线程执行度量、根据被分配给该组的线程的均值线程执行度量的标准偏差和被分配给该组的线程的聚合线程执行度量来确定。其它示例也是可能的, 包括上述的组合。可以基于线程标识符, 或基于用于指示一个线程已经被关联为第一组的组标识符来

将线程分配给所选择的处理内核。

[0063] 在块510中,处理器可以将第二组内的线程(即,其线程执行度量不满足门限线程执行度量的那些线程)分配给所选择的第二处理内核。该第二组的线程可以是基于线程标识符,或者基于用于指示一个线程已经被关联为该第二组的组标识符来分配给该第二处理内核的。

[0064] 在块512中,处理器可以基于被分配给第一组的那些线程的线程执行度量来确定该第一处理内核的线程执行度量的特性。该线程执行度量的特性可以包括缓存未命中率特性、同步操作率特性和指令类型率特性中的一项或多项。在各个方面,该缓存未命中率特性可以包括分配给所选择处理内核的线程(或分配给该处理内核的该组)的平均缓存未命中率、分配给所选择的处理内核的线程的均值缓存未命中率、根据被分配给所选择的处理内核的线程的均值缓存未命中率的标准偏差和分配给所选择的处理内核的线程的聚合缓存未命中率之一。额外或者替代地,该同步操作率特性可以包括被分配给所选择处理内核的线程(或被分配给该处理内核的该组)的平均同步操作率、分配给所选择的处理内核的线程的均值同步操作率、根据被分配给所选择的处理内核的线程的均值同步操作率的标准偏差和分配给所选择的处理内核的线程的聚合同步操作率之一。额外或者替代地,该指令类型率特性可以包括被分配给所选择处理内核的线程(或分配给该处理内核的该组)的平均指令类型率、分配给所选择的处理内核的线程的均值指令类型率、根据被分配给所选择的处理内核的线程的均值指令类型率的标准偏差和被分配给所选择的处理内核的线程的聚合指令类型率之一。

[0065] 在块514中,处理器可以降低该第一处理内核的操作频率以减少所选择处理内核的功耗。可以基于相对于确定的线程执行度量的特性的关系来降低该第一处理内核的操作频率。例如,可以基于确定的该线程执行度量的特性,基于确定的该线程执行度量的特性和操作频率之间的比率,通过基于确定的该线程执行度量的特性来计算频率降低因子,或者通过另一种基于确定的该线程执行度量的特性来计算操作频率的降低的方法,来确定第一处理内核的操作频率。在各个方面,可以使用诸如动态电压和频率缩放、动态时钟和电压缩放或另一种类似的功率降低技术之类的动态功率降低技术来降低所选择处理内核的操作频率。如上所讨论的,因为针对非缓存存储器的大量读/写操作的时延将会主要决定该内核的处理性能,所以可以在不显著影响整体性能的情况下来降低该第一处理内核的频率。

[0066] 在可选块516中,处理器可以增加第二处理内核的频率以便增加其线程执行度量不满足该门限的那些线程的处理速度。其线程执行度量不满足该门限的那些线程可以被确定更高效地使用第二处理器的处理资源,例如通过向第二处理器引入更少的延迟,或者通过要求更少的主存储器访问,从而使该第二处理器比该第一处理器相对更少地进入等待状态。尽管增加该第二处理器的操作频率可能增加其功耗,但是系统的整体效率可以通过第二处理器的更高效的使用而增加。在一些方面,可以使用诸如动态电压和频率缩放、动态时钟和电压缩放和类似技术之类的动态功率降低技术来增加所选择的处理内核的操作频率。

[0067] 线程还可以随着它们的线程执行度量的改变而被重新分类和/或重新关联为不同的组,诸如在系统存储器访问序列或操作序列之后。在可选的判定块518中,处理器可以判断初始与第一组相关联的线程的未决操作的线程执行度量是否不再满足门限缓存未命中率。这可以发生在,例如线程的未决操作将会在先前从系统存储器获得的数据上执行时。在

这种情况中,在具有高缓存未命中率的操作序列从系统存储器获取数据之后,一个线程可以转变为执行主要依赖于缓存读和写操作的操作,使得由线程要求的数据或指令可以在系统存储器调用序列之后的一段时间内在一个或多个缓存中可用,在该系统存储器调用序列期间该线程表现为高缓存未命中率并且该线程是由第一处理内核执行的。

[0068] 当处理器确定分配给第一处理内核的线程的线程执行度量不再满足该门限(即,可选判定块518=“是”)时,该处理器可以将该线程与第二组关联,并且因此可以在可选块520中将该线程分配给该第二处理内核,这样该未决操作可以被以更高的处理频率来执行。如果该第一组中没有线程不再满足该线程执行度量(即,可选判定块518=“否”),则该处理器可以不分配任何线程,并且可以周期性地重复方法500的操作。

[0069] 处理器可以周期性地重复方法500的操作以便动态地重新确定所选择处理内核和/或在一个或多个处理内核中的线程的线程执行度量特性。随着线程执行来动态地重新评估并向组或处理器内核重新分配线程可能是有益的,因为线程中的未决操作的线程执行度量可能随着时间改变。如上关于可选判定块518所讨论的,一个线程可能初始地具有满足门限的线程执行度量,因为其从系统存储器获取数据,但是可能由于从事于获得的数据的操作序列而具有不满足该门限的线程执行度量。类似地,初始地具有不满足门限的线程执行度量并且因此与第二组相关联的线程可以在向系统存储器或显示器缓存写入操作结果时转变为满足门限的线程执行度量。

[0070] 图6是示出管理处理器设备功耗的另一方面方法600的过程流程图。在块302中,处理器可以如上参考图3描述确定被调度用于在处理内核中执行的多个线程中的每一个线程的线程执行度量。

[0071] 在块604中,处理器可以将其线程执行度量满足第一门限的那些线程分配或指派给第一处理内核用于执行。在一个方面,每个线程的线程执行度量可以被独立地考虑。在一个方面,还可以基于被分配到该组的线程的平均线程执行度量、被分配到该组的线程的均值线程执行度量、根据被分配到该组的线程的均值线程执行度量的标准偏差和被分配到该组的线程的聚合线程执行度量,来确定线程组的线程执行度量。其它线程执行度量也是可能的,包括上述的组合。

[0072] 在块606中,处理器可以将其线程执行度量不满足该第一门限的那些线程分配给所选择的第二处理内核。在块608中,处理器可以将其线程执行度量满足第二门限但是不满足该第一门限的那些线程分配给多个处理内核中的第三处理内核。在一个方面,该第一门限可以对应于高于该第二门限的线程执行度量。在一个方面,多于两个门限可以被用于向一个或多个处理器分配线程,这可以提供更细粒度的线程的线程执行度量的评估。此外,尽管为了解释说明的清楚,块604、606和608是以特定序列描述的,但是可以以任何顺序来执行或大体上同时执行块604、606和608。

[0073] 在块610中,处理器可以降低该第一处理内核的操作频率以减少所选择的处理内核的功耗。该第一处理内核的操作频率可以是基于确定的与分配给该第一处理内核的线程相关联的线程执行度量的特性来降低的。如上所讨论的,因为针对非缓存存储器的大量读/写操作的时延将会主要决定该内核的处理性能,所以可以在不显著地影响整体性能的情况下降低该第一处理内核的频率。

[0074] 在可选块612中,处理器可以增加第二处理内核的频率以便增加其线程执行度量

不满足该门限的那些线程的处理速度。其线程执行度量不满足该第一门限的那些线程可以被确定用于更高效地使用第二处理器的处理资源,例如通过向第二处理器引入更少的延迟,或者通过要求更少的主存储器访问,从而使第二处理器比第一处理器相对更少地进入等待状态。在各个方面,可以使用诸如动态电压和频率缩放、动态时钟和电压缩放和类似技术之类的动态功率降低技术,来增加所选择的处理内核的操作频率。

[0075] 在可选块614中,处理器可以将第三处理内核的频率调整为第一处理内核和第二处理内核的频率之间的频率。可以基于确定的与分配给第三处理内核的线程相关联的线程执行度量的特性,以及还可以基于第一处理内核和第二处理内核的频率,来调整第一处理内核的操作频率。

[0076] 线程还可以随着它们的线程执行度量的改变而被重新分配给不同的处理内核,诸如在系统存储器访问序列或操作序列之后。在可选判定块618中,处理器可以判断被分配给第一处理内核、第二处理内核或第三处理内核的线程的线程执行度量是否已经关于第一门限和/或第二门限发生改变。当处理器确定被分配给处理内核的线程的线程执行度量已经关于第一门限和/或第二门限发生改变时(即,可选判定块618=“是”),处理器可以根据其改变的执行度量和第一门限和/或第二门限在可选块620中将该线程分配给不同的处理内核。如果所有线程都没有关于第一门限和/或第二门限发生改变(即,可选判定块618=“否”),则处理器可以不重新分配任何线程。

[0077] 处理器可以周期性地重复方法600的操作,以便通过动态地重新确定所选择处理内核和/或在一个或多个处理内核中的线程的线程执行度量特性,来重新评估并向组或处理器内核重新分配线程。随着线程执行来动态地重新评估并向组或处理器内核重新分配线程可能是有益的,因为线程中的未决操作的线程执行度量可能随着时间改变。如上关于可选判定块618所讨论的,一个线程可能初始地具有满足门限的线程执行度量(例如,由于其从系统存储器获取数据,或者需要某一水平的处理资源),但是可能由于例如从事于获取的数据,或者要求不同水平的处理资源的操作序列而具有不满足该门限的线程执行度量。

[0078] 各个方面可以实现在各种不同移动计算设备上,它的一个示例在图7中示出。更具体地,图7是适合与任何方面一起使用的以智能电话/蜂窝(cell)手机700的形式的移动收发机设备的系统框图。蜂窝手机700可以包括耦合到内部存储器702、显示器703和扬声器708的处理器701。另外,蜂窝手机700可以包括天线704,其用于发送和接收可以连接到无线数据链路和/或耦合到处理器701的蜂窝电话收发机705的电磁辐射。蜂窝手机700典型地还包括用于接收用户输入的菜单选择按钮或摇杆式开关706。

[0079] 典型的蜂窝手机700还包括声音编码/解码(CODEC)电路713,它将从麦克风接收到的声音数字化为适合于无线传输的数据分组并且解码接收到的声音数据分组以生成提供给扬声器708用于生成声音的模拟信号。并且,处理器701、无线收发机705和CODEC 713中的一个或多个可以包括数字信号处理器(DSP)电路(未单独示出)。蜂窝手机700还可以包括用于无线设备之间的低功率短距离通信的ZigBee收发机(即,IEEE 802.15.4收发机)713,或者其它类似的通信电路(例如,用于实现 Bluetooth®或WiFi协议的电路等等)。

[0080] 各个方面可以实现在各种不同商业可用服务器设备的任何一个上,诸如图8中示出的服务器800。该服务器800典型地包括耦合到易失性存储器802和大容量非易失性存储器(诸如硬盘驱动器)803的处理器801。该服务器800还可以包括耦合到处理器801的软盘驱

动器、压缩盘(CD)或DVD盘驱动器811。服务器800还可以包括耦合到处理器801用于与网络805(诸如耦合到其它通信系统计算机和服务器的局域网)建立数据连接的网络接入端口806。

[0081] 其它形式的计算设备也可以从各个方面受益。这些计算设备典型地包括图9中示出的组件,其显示为示例性个人膝上型计算机900。该个人计算机900一般包括耦合到易失性存储器902和大容量非易失性存储器(诸如硬盘驱动器)903的处理器901。该计算机900还可以包括耦合到处理器901的压缩盘(CD)和/或DVD驱动器904。该计算机设备900还可以包括耦合到处理器901用于建立数据连接或接收外部存储器设备的多个连接器端口,诸如用于将处理器901耦合到网络的网络连接电路905。该计算机900还可以耦合到计算机领域内众所周知的键盘908、诸如鼠标910之类的点击设备和显示器909。

[0082] 处理器701、801、901可以是任何能够由软件指令(应用)配置为执行各种不同功能(包括下面描述的各个方面的功能)的可编程微处理器、微计算机或多个处理器芯片。在一些移动设备中,可以提供多个处理器801,诸如专用于无线通信功能的一个处理器和专用于运行其它应用的一个处理器。典型的,软件应用可以在它们被访问或加载到处理器701、801、901之前被存储在内部存储器702、802、902中。该处理器701、801、901可以包括足够存储该应用软件指令的内部存储器。

[0083] 各个方面可以实现在任何数量的单个或多个处理器系统中。一般来讲,进程在很短时间片段内运行在处理器上,这样看起来多个进程同时运行在单个处理器上。当进程在一个时间片段结束时被从处理器移除时,关于该进程的当前操作状态的信息被保存在存储器中,这样当该进程返回在该处理器上执行时可以无缝地恢复其操作。该操作状态数据可以包括该进程的地址空间、栈空间、虚拟地址空间、寄存器集合镜像(例如,程序计数器、栈指针、指令寄存器、程序状态字等待)、账户信息、许可、访问限制和状态信息。

[0084] 一个进程可以再生其它进程,并且再生的进程(即,子进程)可以继承该再生进程(即,父进程)的一些许可和访问限制(即,上下文)。一个进程可以是包括多个轻量级进程或线程的重量级进程,它们是将其上下文(例如,地址空间、栈、许可和/或访问限制等等)的全部或一部分与其它进程/线程共享的进程。因此,单个进程可以包括共享、具有到单个上下文(即,处理器的上下文)的访问和/或在其内操作的多个轻量级进程或线程。

[0085] 上述方法描述和过程流程图仅仅作为示例性示例提供,而并不意在要求或暗示各个方面的块必须以示出的顺序执行。本领域的技术人员应该了解的是,上述方面中的块的顺序可以用任何顺序执行。诸如“之后”、“于是”、“接下来”等词语并不意在限制块的顺序;这些词语仅仅用于贯穿方法描述引导读者。此外,任何以单数形式对声明单元的引用,例如使用冠词“一(a)”、“一(an)”或“该(the)”并不是要解释为将该单元限制为单数。

[0086] 尽管上面描述为当一个值大于或等于一个门限时满足该门限,但是应该了解的是,这不是一个限制,并且在一些方面,当一个值超过该门限时可以满足该门限,而在该值小于或等于该门限时不满足该门限。

[0087] 结合本申请中公开的方面所描述的各种示例性逻辑框、模块、电路和算法块可以实现成电子硬件、计算机软件或其组合。为了清楚地示出硬件和软件之间的可交换性,上面对各种示例性组件、方框、模块、电路和块已经围绕其功能进行了总体描述。至于这种功能是实现成硬件还是实现成软件,取决于特定的应用和对整个系统所施加的设计约束。熟练

的技术人员可以针对每个特定应用,以变通的方式实现所描述的功能,但是,这种实现决策不应被解释为使得背离本发明的范围。

[0088] 可以用被设计用于执行本文所述功能的通用处理器、数字信号处理器 (DSP)、专用集成电路 (ASIC)、现场可编程门阵列 (FPGA) 或其它可编程逻辑设备、分立门或者晶体管逻辑器件、分立硬件组件或者其任意组合,来实现或执行用于实现结合本文公开方面描述的各种示例性逻辑、逻辑块、模块和电路的硬件。通用处理器可以是微处理器,或者,该处理器也可以是任何常规的处理器、控制器、微控制器或者状态机。处理器还可以实现为计算设备的组合,例如,DSP和微处理器的组合、多个微处理器、与DSP内核结合的一个或多个微处理器,或者任何其它此种结构。另外,可以由专用于给定功能的电路来执行一些块或方法。

[0089] 在一个或多个示例性方面,所描述的功能可以用硬件、软件、固件,或其任意结合来实现。如果在软件中实现,则功能可以作为一个或多个指令或代码存储在非临时性计算机可读介质或非临时性处理器可读介质上。本文中公开的方法或算法的步骤可以包含在处理器可执行软件模块中,其可以驻留在非临时性计算机可读或处理器可读存储介质上。非临时性计算机可读或处理器可读存储介质可以是计算机或处理器可访问的任何存储介质。举个例子,但是并不作为限制,该非临时性计算机可读或处理器可读介质可以包括RAM、ROM、EEPROM、闪存、CD-ROM或其它光盘存储器、磁盘存储器或其它磁存储设备,或可以用于以指令或数据结构的形式存储期望的程序代码并可以由计算机访问的任何其它介质。本文中所用的磁盘和光盘,包括压缩盘 (CD)、激光盘、光盘、数字多功能光盘 (DVD)、软盘和蓝光盘,其中,磁盘通常磁性地复制数据,而光盘则用激光光学地复制数据。上述的组合也包含在非临时性计算机可读和处理器可读介质的范围内。另外,方法或算法的操作可以作为代码和/或指令的一个或任何组合或集合存在于非临时性处理器可读介质和/或计算机可读介质上,其可以被合并入计算机程序产品中。

[0090] 提供前面对公开方面的描述以便使本领域的任何技术人员能够实施或使用本发明。对这些方面的修改对于本领域的技术人员来说是显而易见的,并且本文中定义的一般原则可以在不脱离本发明的精神或范围的前提下应用于其它方面。因此,本发明并不意在限制于本文中示出的方面,而是要符合与下面的权利要求和本文中公开的原则和新颖性特征相一致的最广范围。

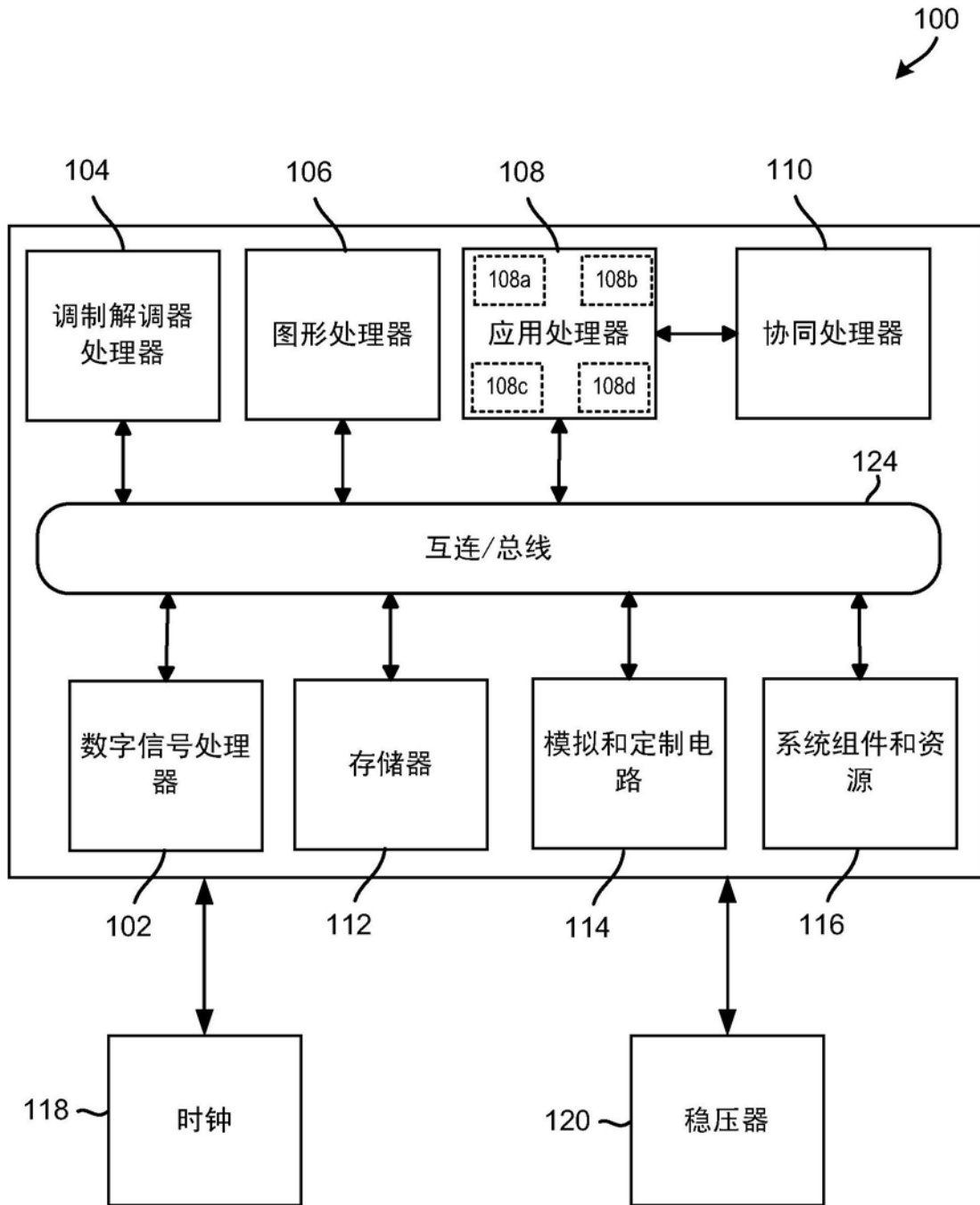


图1

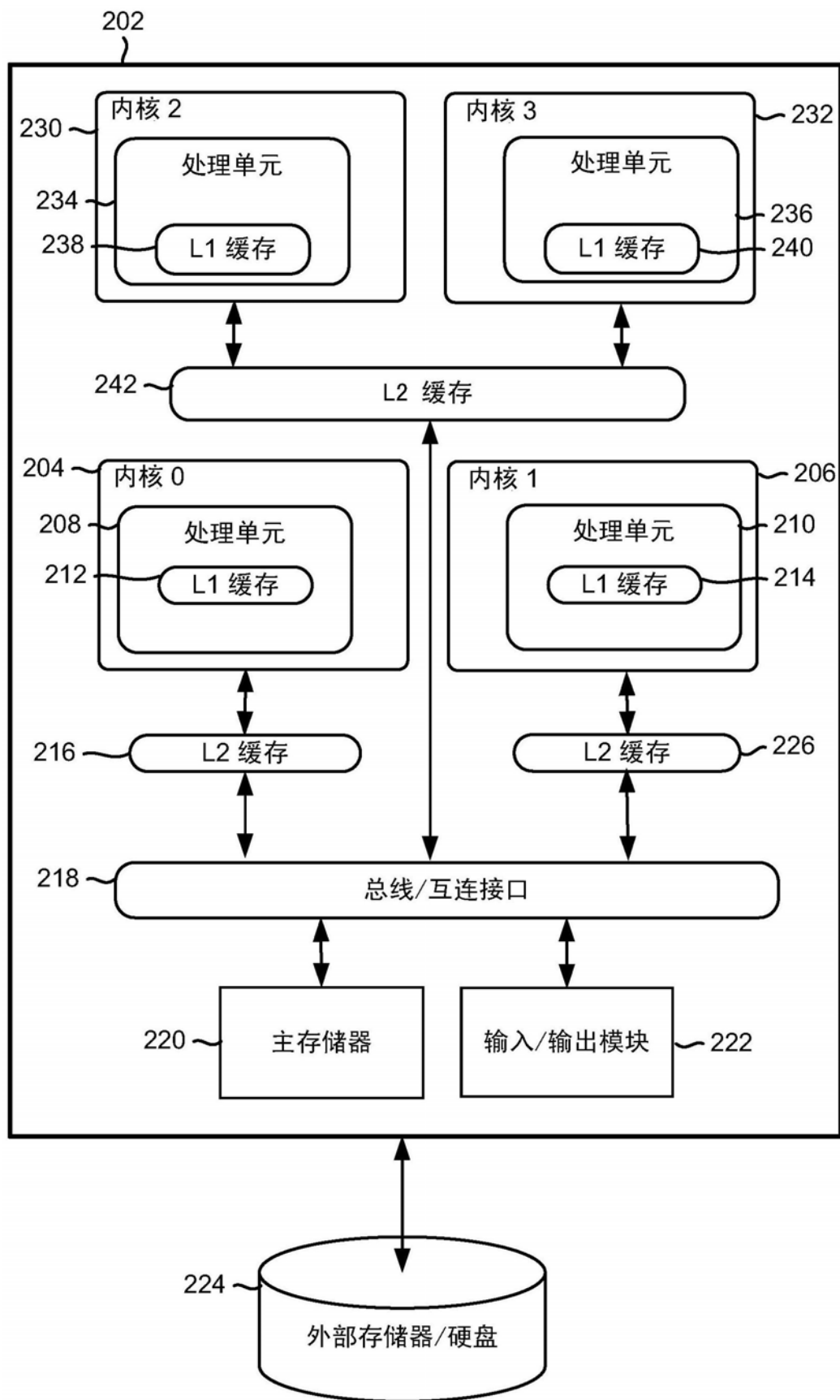


图2

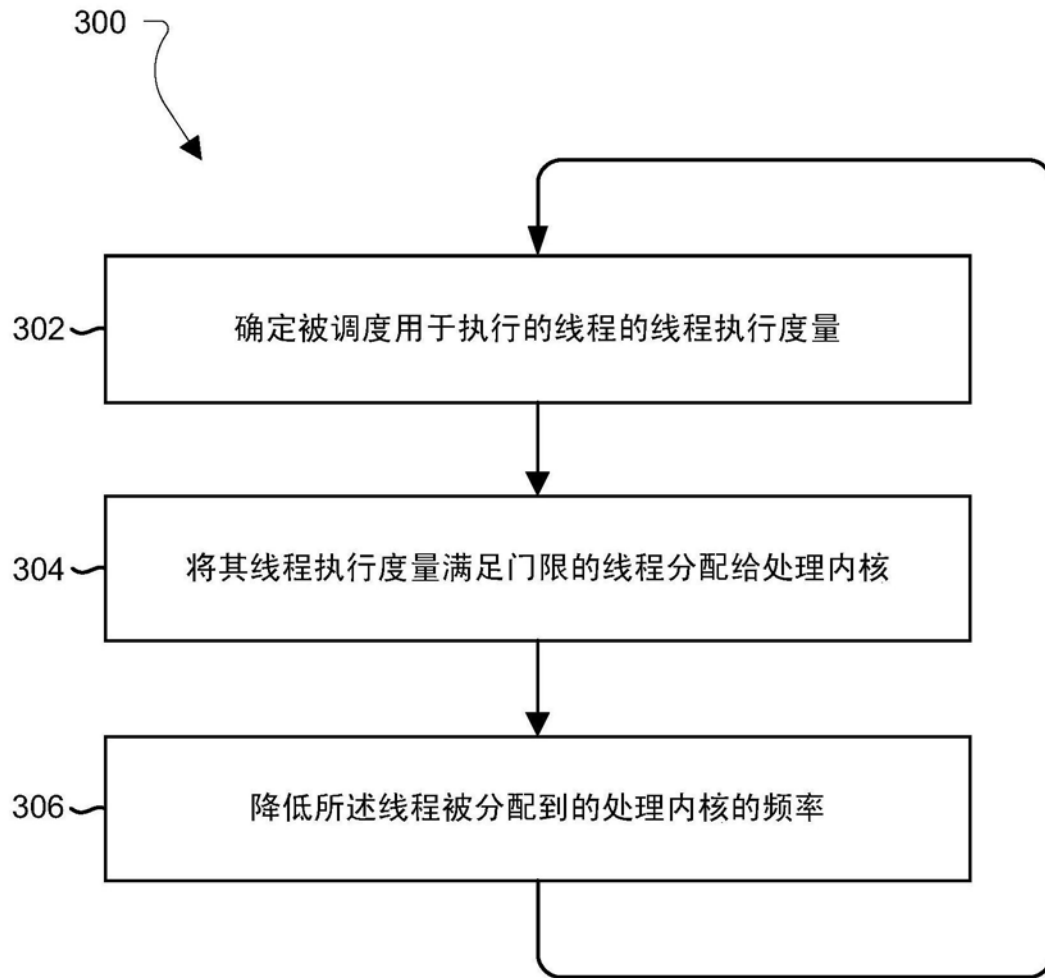


图3

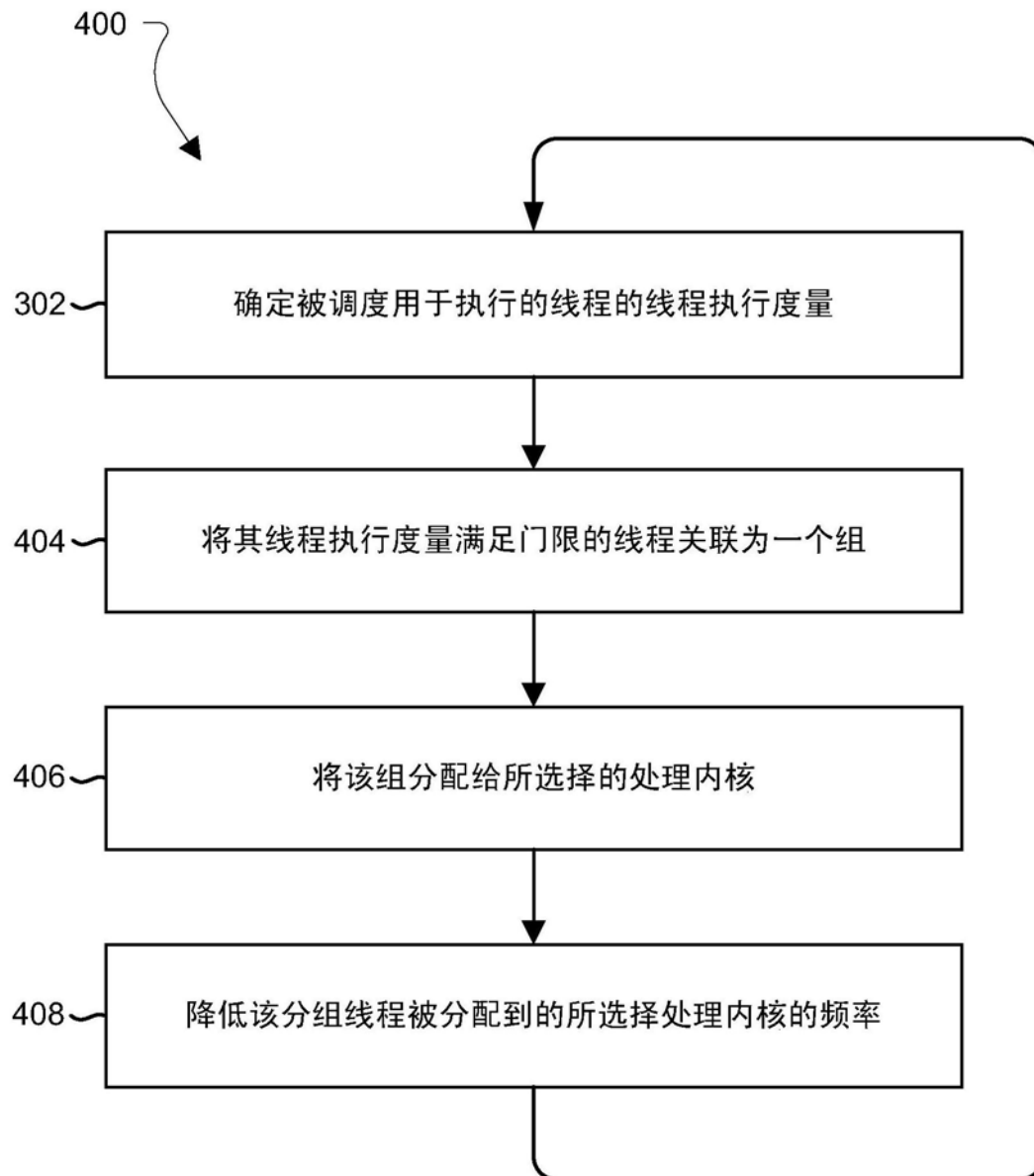


图4

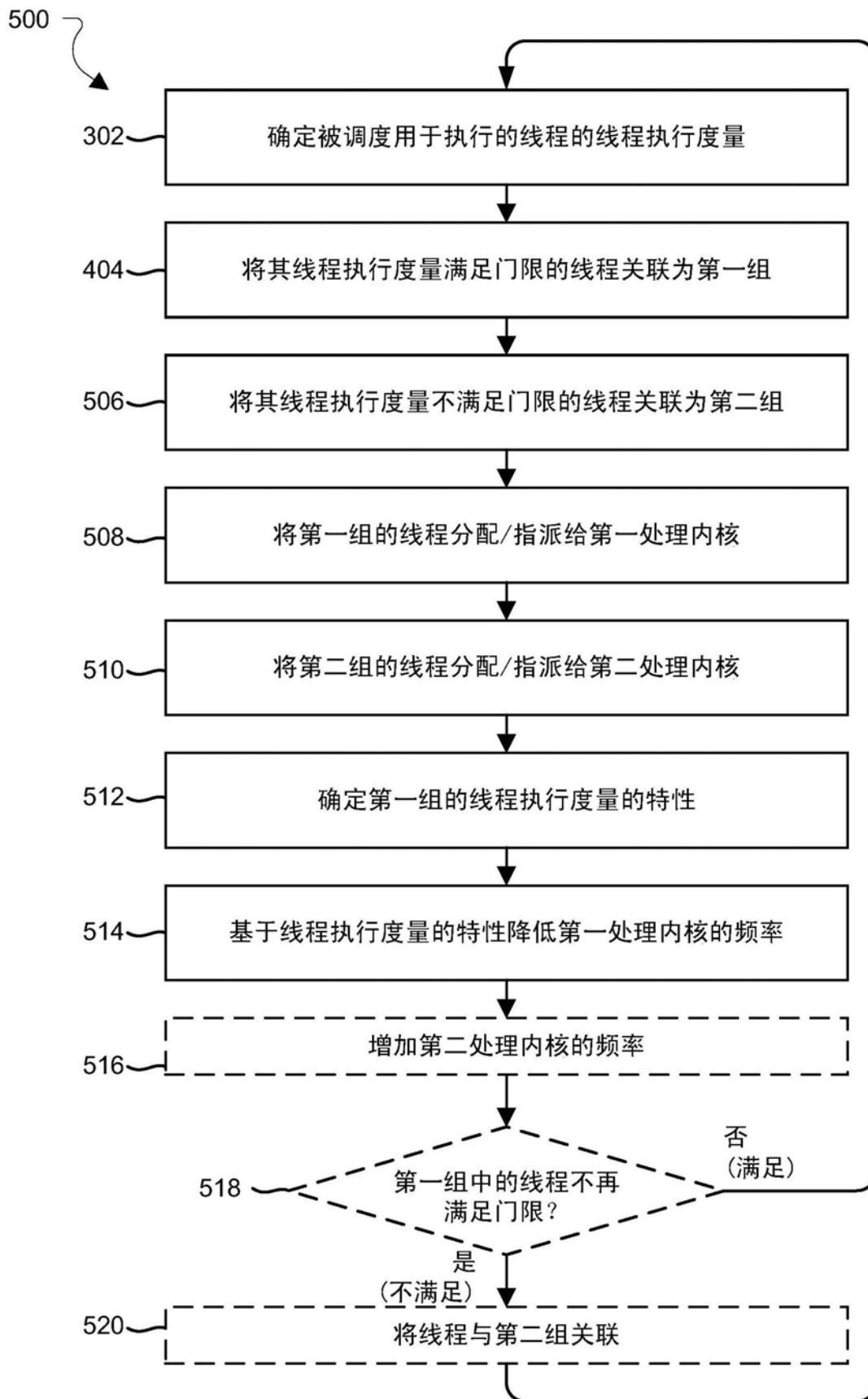


图5

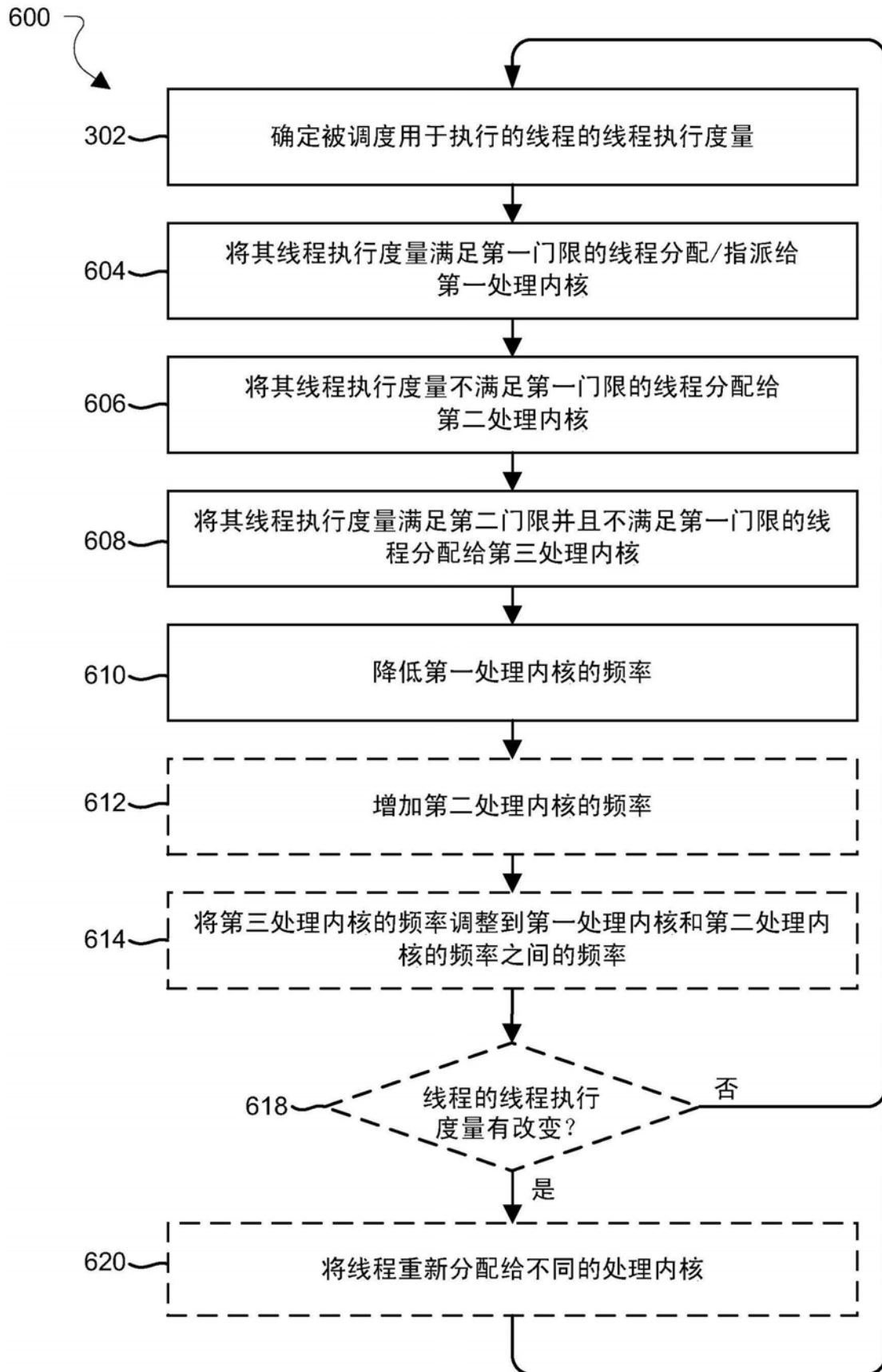


图6

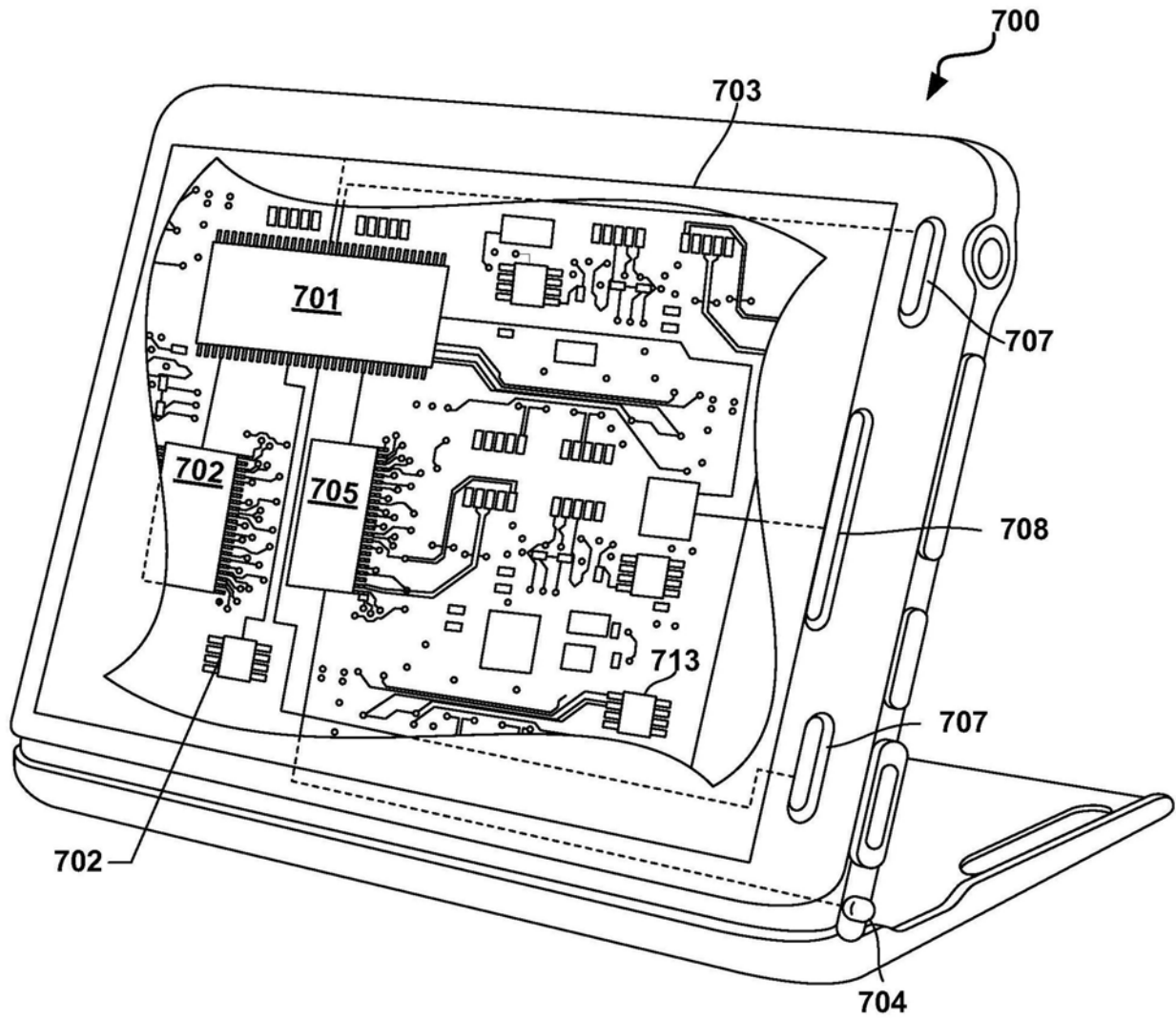


图7

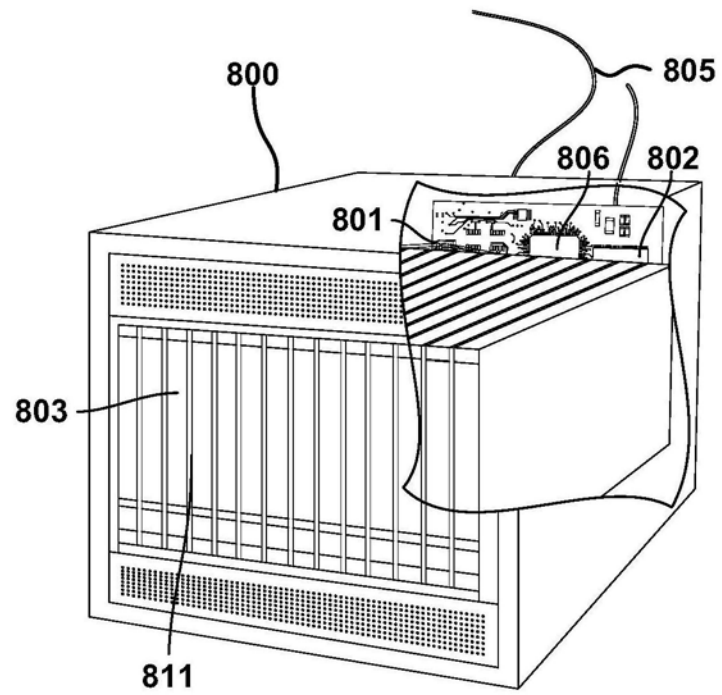


图8

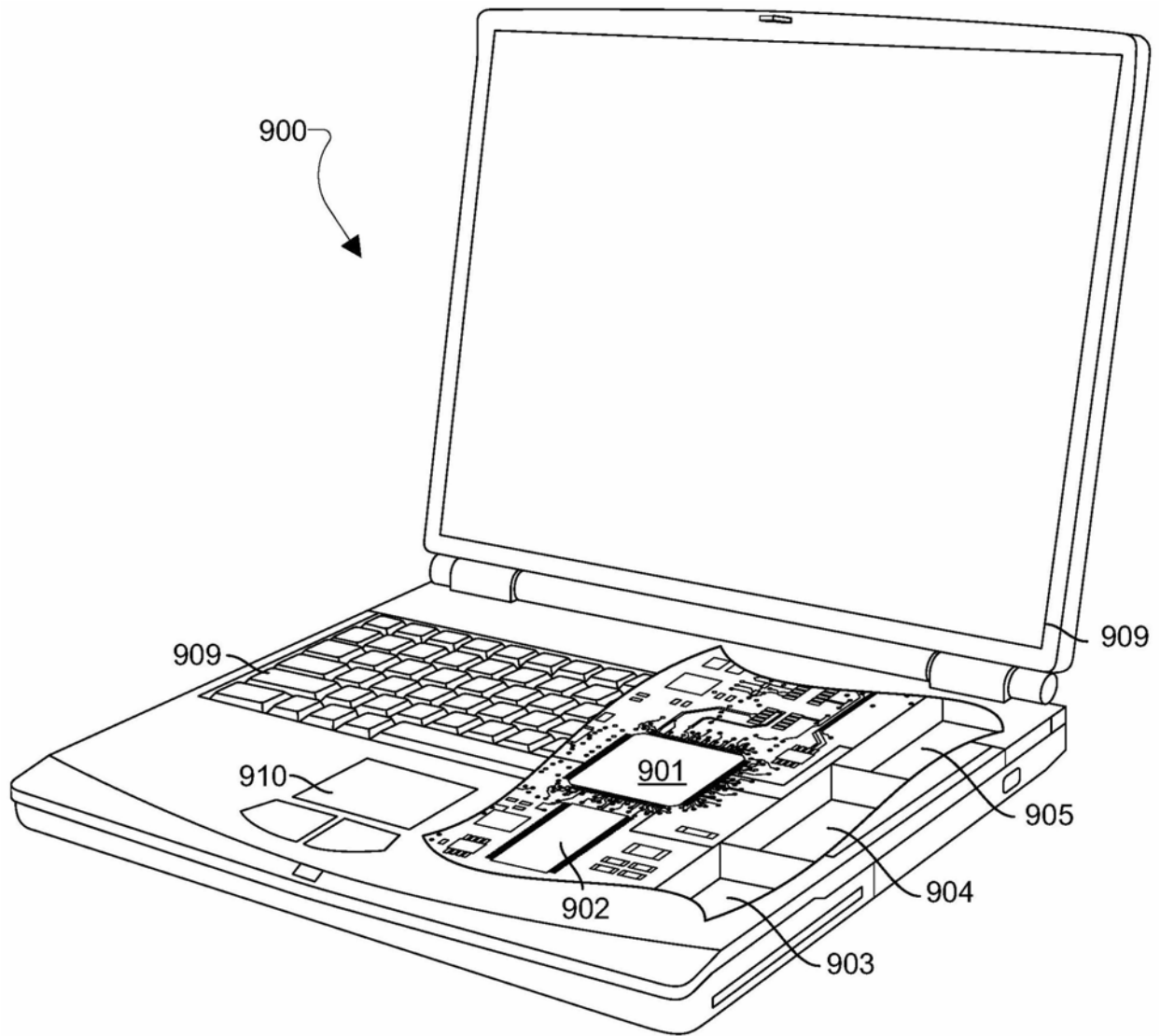


图9