

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 9/38 (2006.01)



[12] 发明专利说明书

专利号 ZL 03826964.3

[45] 授权公告日 2008年7月2日

[11] 授权公告号 CN 100399262C

[22] 申请日 2003.8.26 [21] 申请号 03826964.3

[86] 国际申请 PCT/US2003/026531 2003.8.26

[87] 国际公布 WO2005/031565 英 2005.4.7

[85] 进入国家阶段日期 2006.2.24

[73] 专利权人 国际商业机器公司

地址 美国纽约

[72] 发明人 普拉迪普·伯斯

丹尼尔·M·西特朗

皮特·W·库克

菲利普·G·埃马

汉斯·M·雅克布森

普拉布哈卡尔·N·库德瓦

斯坦利·E·舒斯特

尤德·A·里弗斯

维克多·V·祖班

[56] 参考文献

US5987620A 1999.11.16

US2002/0169990A1 2002.11.14

CN1226700A 1999.8.25

EP0451661A2 1991.10.16

US6263448B1 2001.7.17

审查员 明 媚

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 康健忠

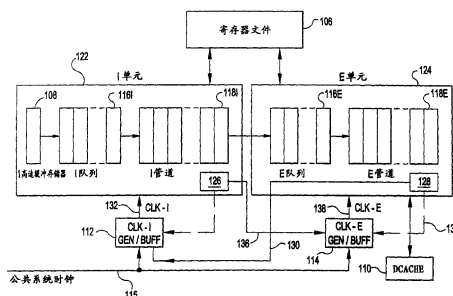
权利要求书 6 页 说明书 17 页 附图 9 页

[54] 发明名称

用于降低功率的按需驱动时钟调节的处理器

[57] 摘要

一种诸如标量处理器或者超标量处理器的同步集成电路。电路部件或者单元由公共系统时钟提供时钟并被同步到公共系统时钟。至少两个时钟驱动的单元包括多个寄存器级，例如，流水线级。每个时钟驱动的单元内的本地时钟发生器将公共系统时钟与来自一个或者多个其它单元的延迟状态组合，以上下调节寄存器的时钟频率。



1、一种同步集成电路，包括：

公共系统时钟；以及

多个时钟驱动的单元，与所述公共系统时钟同步，至少两个所述时钟驱动的单元分别包括：

多个寄存器级；以及

本地时钟发生器，接收所述公共系统时钟和延迟状态，而且响应所述延迟状态，调节所述多个寄存器的时钟频率。

2、根据权利要求1所述的同步集成电路，其中所述本地时钟发生器包括选通控制移位寄存器，该选通控制移位寄存器包括多个单位计数器。

3、根据权利要求2所述的同步集成电路，其中所述本地时钟发生器进一步包括多个本地时钟驱动器，每一本地时钟驱动器接收所述单位计数器之一的输出，而且将所述输出与所述系统时钟组合，以产生寄存器级时钟。

4、根据权利要求3所述的同步集成电路，其中所述多个寄存器级是寄存器流水线级，而且对于每个流水线级，所述选通控制移位寄存器包括所述多个单位计数器之一。

5、根据权利要求4所述的同步集成电路，其中所述集成电路是标量处理器，而且在所述标量处理器中，至少两个时钟驱动的单元包括指令单元和执行单元，所述指令单元对所述执行单元提供延迟状态，而所述执行单元对所述指令单元提供延迟状态。

6、根据权利要求5所述的同步集成电路，其中所述标量处理器进一步包括：

数据高速缓冲存储器，与所述执行单元通信；

寄存器文件，与所述指令单元和所述执行单元通信；

所述指令单元包括：指令高速缓冲存储器、从所述指令高速缓冲存储器接收数据的指令队列，以及从所述指令队列接收数据的指令管

道；以及

所述执行单元包括：从所述指令管道接收数据的执行队列，以及从所述执行队列接收数据的执行管道。

7、根据权利要求6所述的同步集成电路，其中在所述指令单元和所述执行单元的每一个中，在所述多个时钟驱动器的相应的一个，各所述选通控制移位寄存器的每个输出与所述系统时钟组合，每个所述多个时钟驱动器分别选通相应流水线级。

8、根据权利要求7所述的同步集成电路，其中延迟状态位被提供到所述选通控制移位寄存器的第一级。

9、根据权利要求7所述的同步集成电路，其中延迟状态位被提供到一位计数器，所述一位计数器保持置位，除非所述延迟状态位被断言，而且在所述延迟状态位被断言时进行计数，所述一位计数器的输出被提供到所述选通控制移位寄存器的第一级。

10、根据权利要求6所述的同步集成电路，其中在所述多个时钟驱动器的每一个中，所述选通控制移位寄存器的最后输出与所述系统时钟组合，延迟状态位被提供到一位计数器，所述一位计数器保持置位，除非所述延迟状态位被断言，而且在所述延迟状态位被断言时，进行计数，所述一位计数器的输出被提供到所述选通控制移位寄存器的第一级。

11、根据权利要求5所述的同步集成电路，其中所述标量处理器是超标量处理器，响应来自所述执行单元的所述延迟状态位，所述指令单元进一步调节指令高速缓冲存储器取带宽。

12、根据权利要求11所述的同步集成电路，其中所述执行单元包括：

定点单元，用于从所述指令单元接收指令，而且与通用寄存器/更名缓冲器单元通信；

负载存储单元，用于从所述指令单元接收指令，而且与所述通用寄存器/更名缓冲器单元以及数据高速缓冲存储器通信；

浮点单元，用于从所述指令单元接收指令，而且与浮点寄存器/更

名缓冲器单元通信,所述负载存储单元进一步与所述浮点寄存器/更名缓冲器单元通信,所述负载存储单元将延迟状态提供到所述浮点单元,而且所述浮点单元将延迟状态提供到所述负载存储单元;以及

矢量多媒体扩展单元,用于从所述指令单元接收指令,而且与所述指令单元内的完成单元通信。

13、根据权利要求 11 所述的同步集成电路,其中所述指令单元进一步包括:

指令高速缓冲存储器;

指令取单元/分支单元,用于从所述指令高速缓冲存储器接收指令;以及

调度单元,用于从所述指令取单元/分支单元接收指令,而且将收到的所述指令提供到所述执行单元。

14、根据权利要求 12 所述的同步集成电路,其中每个所述选通控制移位寄存器的每级的输出均与所述系统时钟组合,而且单独选通相应的流水线级。

15、根据权利要求 14 所述的同步集成电路,其中延迟状态位被提供到所述选通控制移位寄存器的第一级。

16、根据权利要求 14 所述的同步集成电路,其中在每个所述选通控制移位寄存器,延迟状态位被提供到一位计数器,所述一位计数器保持置位,除非所述延迟状态位被断言,而且在所述延迟状态位被断言时,进行计数,所述一位计数器的输出被提供到所述每个选通控制移位寄存器的第一级。

17、根据权利要求 14 所述的同步集成电路,其中在所述多个时钟驱动器的每一个中,所述选通控制移位寄存器的最后输出与所述系统时钟组合,延迟状态位被提供到一位计数器,所述一位计数器保持置位,除非所述延迟状态位被断言,而且在所述延迟状态位被断言时,进行计数,所述一位计数器的输出被提供到所述选通控制移位寄存器的第一级。

18、一种标量处理器,包括:

公共系统时钟；以及

指令单元，由所述系统时钟提供时钟；以及

执行单元，由所述系统时钟提供时钟，而且与所述指令单元通信；

所述指令单元和所述执行单元分别包括：

多个寄存器级；以及

本地时钟发生器，用于接收所述公共系统时钟和延迟状态，而且响应所述延迟状态，调节所述多个寄存器的时钟频率，所述指令单元将延迟状态提供到所述执行单元，而且所述执行单元将延迟状态提供到所述指令单元。

19、根据权利要求 18 所述的标量处理器，其中所述本地时钟发生器包括：

选通控制移位寄存器，包括多个单位计数器；以及

多个本地时钟驱动器，分别从所述多个单位计数器之一接收输出，而且将所述输出与所述系统时钟组合，以产生寄存器级时钟。

20、根据权利要求 19 所述的标量处理器，其中所述多个寄存器级是寄存器流水线级，而且对于每个流水线级，所述选通控制移位寄存器包括一个单位计数器。

21、根据权利要求 20 所述的标量处理器，进一步包括：

数据高速缓冲存储器，与所述执行单元通信；

寄存器文件，与所述指令单元和所述执行单元通信；

所述指令单元包括：指令高速缓冲存储器、从所述指令高速缓冲存储器接收数据的指令队列，以及从所述指令队列接收数据的指令管道；以及

所述执行单元包括：从所述指令管道接收数据的执行队列，以及从所述执行队列接收数据的执行管道。

22、根据权利要求 21 所述的标量处理器，其中在所述多个时钟驱动器的相应的一个，所述指令单元和所述执行单元将所述选通控制移位寄存器的输出与所述系统时钟组合，所述多个时钟驱动器的每一个分别选通相应流水线级。

23、根据权利要求 22 所述的标量处理器，其中延迟状态位被提供到所述选通控制移位寄存器的第一级。

24、根据权利要求 22 所述的标量处理器，其中延迟状态位被提供到一位计数器，所述一位计数器保持置位，除非所述延迟状态位被断言，而且在所述延迟状态位被断言时进行计数，所述一位计数器的输出提供到所述选通控制移位寄存器的第一级。

25、根据权利要求 21 所述的标量处理器，其中在所述多个时钟驱动器的每一个中，所述选通控制移位寄存器的最后输出与所述系统时钟组合，延迟状态位被提供到一位计数器，所述一位计数器保持置位，除非所述延迟状态位被断言，而且在所述延迟状态位被断言时，进行计数，所述一位计数器的输出被提供到所述选通控制移位寄存器的第一级。

26、一种超标量处理器，包括：

公共系统时钟；以及

指令单元，由所述系统时钟提供时钟，所述指令单元包括：

指令高速缓冲存储器，

指令取单元/分支单元，用于从所述指令高速缓冲存储器接收指令；以及

调度单元，用于从所述指令取单元/分支单元接收指令，而且转发收到的所述指令，以便执行；

执行单元，由所述系统时钟提供时钟，所述执行单元包括：

定点单元，用于从所述指令单元接收指令，而且与通用寄存器/更名缓冲器单元通信；

负载存储单元，用于从所述指令单元接收指令，而且与所述通用寄存器/更名缓冲器单元以及数据高速缓冲存储器通信；

浮点单元，用于从所述指令单元接收指令，而且与浮点寄存器/更名缓冲器单元通信，所述负载存储单元进一步与所述浮点寄存器/更名缓冲器单元通信，所述负载存储单元将延迟状态提供到所述浮点单元，而且所述浮点单元将延迟状态提供到所述负载存储单元；以及

矢量多媒体扩展单元，用于从所述指令单元接收指令，而且与所述指令单元内的完成单元通信；以及

本地时钟发生器，位于所述负载存储单元和所述浮点单元的每一个内，接收所述公共系统时钟和延迟状态，并根据所述延迟状态调节单元寄存器的时钟频率。

27、根据权利要求 26 所述的标量处理器，其中所述延迟状态指出单元的活动程度。

28、根据权利要求 27 所述的标量处理器，其中响应执行单元延迟状态，所述指令单元调节取带宽。

用于降低功率的按需驱动时钟调节的处理器

技术领域

本发明一般地涉及减少和控制由多个时钟驱动的部件或者单元构成的微处理器或者系统中的功率消耗。

背景技术

半导体技术和芯片制造业的进步使片上时钟频率、单芯片上的晶体管数量及其小片(die)大小持续提高,而且芯片电源电压也相应升高。通常,给定时钟驱动的(clocked)单元消耗的功率随其内的转换频率线性升高。因此,不但芯片电源电压的降低,而且还增加了芯片的功率消耗。芯片功率这样升高的当然结果是,在芯片级和系统级,散热和封装成本都提高。在系统(例如,手提式系统、便携式系统和移动式系统)的低端,电池寿命是关键,在不导致性能降低到不可接受程度的情况下,净能量的减少很重要。因此,微处理器功率耗散的增加变成未来性能提高的主要障碍。

标量处理器一次取并发出/执行一条指令。每条这种指令对标量数据操作数进行运算。每个这种操作数都是单数据值或者很小的(atomic)数据值或者数。在标量处理器内引入的流水线作业,该流水线作业被称为并发性处理,即,在给定的时钟周期内,处理多条指令,同时保存单发(single-issue)范型。

在给定的机器周期内,超标量处理器可以取、发出和执行多条指令。此外,通常流水线化每条指令的取、发出和执行通路,以进一步实现并行性。超标量处理器的例子包括:IBM公司的Power/PowerPC处理器、Intel公司的Pentium Pro(P6)处理器系列、Sun Microsystems的Ultrasparc处理器、Hewlett Packard公司(HP)的PA-RISC处理器以及先前的Compaq公司(现在与HP合并)的Alpha处理器系列。

通常流水线化矢量处理器，而且可以对单体系结构步骤或者指令中各数的整个阵列执行一次运算。例如，单条指令可以将阵列 A 的每项叠加到阵列 B 中的相应项，然后，将该结果存储在阵列 C 的相应项上。通常支持矢量指令，作为基本标量指令集的扩展。在矢量引擎上，仅执行那些在较大的应用中可以被矢量化的代码段。矢量引擎可以是单个流水线化执行单元；或者可以将其组织为阵列或者单指令多数据 (SIMD) 机，它具有同时对不同数据执行相同指令的多个相同执行单元。例如，Cray 巨型计算机通常是矢量处理器。

同步时钟处理器或者系统具有用于驱动构成该系统的所有单元或者部件的一个全局主时钟。有时，该时钟的比例从量 (ratioed derivative) 可以使特定单元的周期比主时钟频率或者总时钟频率快或者慢。通常，通过进行设计，利用统计方法，预定和预调这种时钟控制判定。例如，Intel Pentium 4 处理器以两倍于芯片主时钟的速度时钟驱动其整个流水线，就象使用本技术领域内所称的双抽运 (double-pumping) 或者波流水线。这种时钟加倍技术提高了处理器的执行速率并改善了处理器的性能。然而，总线和片外 (off-chip) 存储器速率不能与处理器计算逻辑核心保持同步。因此，在大多数情况下，该技术的处理器具有以主处理器时钟频率的整数倍量的频率工作的片外总线 and 高速缓冲存储器。通常，在进行系统设计期间，使这些时钟工作频率固定。这是因为当代处理器复杂，可以具有多个时钟速率。有时，将双抽运和波流水线用于更高端机器，以缓解该处理器与外部总线或者存储器之间的任何性能失配。

Rabaey, JanM. 和 Pedram, Massoud, ed., *Low Power Design Methodologies*, (Kluwer Academic Publishers, 1996) 描述了利用同步时钟选通 (gating) 降低功率，其中在再生点，即，在馈送到特定芯片区、部件或者锁存器的本地时钟缓冲器 (LCB) 内可以禁用该时钟。在粗程度控制下，沿着功能边界选通时钟。在较精确程度的控制下，在各锁存器选通时钟。例如，Gerosa 等人 "A 2.2W, 80 MHz, superscalar RISC microprocessor," *IEEE Journal of Solid State*

Circuit, vol. 29, no. 12, December 1994, pp. 1440-1454,教导了根据在每个周期发出和执行的指令,对不同执行单元选通时钟。

在处理器执行一系列某种功能类的指令,例如,仅整数或者仅浮点指令时的情况下,粗粒度(coarse-grain)单元级时钟选通是有益的。当输入工作负荷是这样的,使得处理器仅检查整数代码时,可以禁用到浮点型单元的时钟再生器。同样,在执行仅浮点运算期间,可以禁用到整数单元的时钟。这样可以节省大量的芯片功率。利用串行指令的软件或者使用硬件以检测空闲周期,可以在局部正常实现粗空闲控制。还可以在指令解码期间,通过避免不必要地传送无效或者无意义的数,在局部正常实现细空闲控制。将从其始端的初始点到下游级或者单元的选通控制信息因果流称为前馈流。这种流通道可以包括具有明显后馈流的回路,但是仍然认为因果信息流是前馈进程。因此,粗和细空闲控制都是自触发的、前馈控制。

另一方面,利用用于调节前馈流的下游流水线延迟(stall)信号构造反馈控制系统。在此,控制信息流是从下游的“果”到上游的“因”。粗和细粒度延迟控制主要用于防止覆盖写流水线处理器中的有效、延迟数据;但是这种机制还可以用于节省功率消耗。例如, Jacobson 等人在“Synchronous interlocked pipelines,”IEEE ASYNC-2002 conference, April 2002 中建议了一种在同步流水线中降低功率的细粒度延迟传播机制;这样弥补了使用如前所述的 Gerosa 等人描述的采用“有效”位进行时钟选通的更传统、细粒度前馈机制;此外,请参考 Gowan 等人的“Power considerations in the design of the Alpha 21264 microprocessor,” Proc. 1998 ACM/IEEE Design Automation Conference, pp. 726-731 (June 1998)。然而,象 Jacobson 等人描述那样,未将所公开的细粒度延迟选通(反馈)机制用于控制(通过进行时钟或者总线带宽调节(throttling))信息流速率,但是在我们的发明中控制信息流。

至少必须解决粗空闲控制引起的两个问题。第一,大瞬时电流降和增益可能使片上电源电压产生不可接受的感应(Ldi/dt)噪声。第

二，为了保持正确的功能操作，关断和选通过程需要开销周期。对于工作负荷中的较细粒度阶段变化，选通与启用方式之间转换的太频繁了，引起不可接受的性能冲击。

另外，该技术的细空闲控制的状态在局部取决于产生的选通信号或者流水线级 (stage) 级别 (level) 时钟选通的条件，例如，取决于数据无效或者无意义的操作数条件。该技术方法的这些状态不基于预测或者预计产生选通信号。因此，定时要求通常很重要，因为选通信号必须在断言之前可用，而且在无差错时钟选通操作的适当时长内被断言 (assert)。Gowan, M. K., Biro, L. L. 和 Jackson, D. B., "Power considerations in the design of the Alpha 21264 microprocessor," Proc. 1998 ACM/IEEE Design Automation Conference, pp. 726-731, (June 1998) 讨论了这些约束条件如何显著地使设计定时分析复杂化，甚至降低时钟频率性能的。

无论基本控制机制是前馈 (原因-结果流) 还是基于反馈 (结果-原因流)，无论粗还是细，该技术的时钟选通技术的状态还是仅空间控制。这是因为，在 (各) 起作用区，利用使用信息清除冗余时钟，而不考虑在该 (各) 区或者机器中的其它地方的瞬时活动或者历史。为了在非相邻区内 (例如，指令取或者传送单元) 调节上游 (产生方) 时钟或者信息流速率，不反馈下游 (消费方) 单元和级 (例如，执行流水线或者发出队列) 中的活动状态和事件。同样，为了调节下游消费方时钟或者信息流速率，不前馈上游产生方区中的活动状态和事件。此外，通常只能关断时钟信号，其中启动时钟信号或者不启动时钟信号。

因此，需要对可以以细粒空间和临时粒度操作的连接的流水线单元实现改善时钟控制，而不恶化性能 (开销)，而且不使大电流/电压波动到底层电路。

发明内容

本发明的目的是降低处理器的功率消耗，而没有显著的性能恶化。

本发明是诸如标量处理器或者超标量处理器的同步集成电路。电路部件或者单元由公共系统时钟提供时钟并被同步到公共系统时钟。至少两个时钟控制电路包括多个寄存器级，例如，流水线级。每个时钟驱动的单元中的本地时钟发生器将公共系统时钟和来自一个或者多个其它单元的延迟状态组合，以上下调节寄存器时钟频率。

附图说明

根据下面参考附图对本发明的说明性实施例所做的详细说明，可以更好地理解上述和其它目的、方面和优点，其中：

图 1A-B 示出该技术的流水线标量处理器的典型状态的例子和相应指令时序图；

图 2 示出根据本发明的优选实施例要求或者活动驱动功率控制的标量处理器的高级（high-level）例子；

图 3 示出活动监视与时钟控制逻辑暂停电路的第一个例子，其中选通控制移位寄存器（GCSR）使系统时钟传送到 I-PIPE 的各级；

图 4 示出在第二个例子中，用于请求式 I-CLK 调节的根据本发明优选实施例，可以代替图 3 所示暂停电路的，或者图 3 所示暂停电路包括的减速电路；

图 5 示出其中将各 GCSR 级输出传送到各相应“与”门的图 3 和图 4 所示实施例的变型；

图 6 进一步详细示出与图 3 所示截面对应的标量 I-管道的又一例子；

图 7 示出将本发明应用于流水线超标量处理器的例子；

图 8 示出如同图 7 所示的 E-UNIT 的 LSU 和 FPU 的更详细例子。

具体实施方式

现在，参考附图，更具体地说，图 1A-B 示出该技术的流水线标量处理器 100 的典型状态的高级方框图例子和相应指令时序图。将该主功能数据通路分割为通常称为指令单元（I-UNIT）102 和执行单元

(E-UNIT) 104 的两个主要部件或者单元。I-UNIT (I 单元) 102 和 E-UNIT (E 单元) 104 内的许多详细子单元和功能逻辑, 例如, 分支预测逻辑与本发明讨论的内容无关, 因此, 为了清楚地说明这种处理器 100 中的整个时钟控制过程, 省略该详细子单元和功能逻辑。该例子中的处理器 100 示出了流水线标量设计, 该设计不包括用于防止对单元、子单元或者包括在其内的辅助存储资源进行冗余时钟的时钟选通。

单元 102、104 之一或者二者访问的片上存储器包括: 寄存器文件 (REGFILE) 106; 指令高速缓冲存储器 (ICACHE) 108; 以及数据高速缓冲存储器 (DCACHE) 110。该 REGFILE 106 通常是可以两个单元 102、104 访问的共享资源, 这样, 可以作为独立实体对待它们。ICACHE (I 高速缓冲存储器) 108 是 I-UNIT 流水线的第一终端级, 这样, 将它看作 I-UNIT 102 的一部分。通常, 仅通过 E-UNIT 104 可以访问 DCACHE 110, 因此, 可以作为 E-UNIT 104 的一部分对待。两个独立的本地 (local) 时钟缓冲器 (LCB) 112、114 分别放大公共同步时钟 115, 然后, 将它分发到单元 102、104 中的相应之一。每个单元 102、104 包括输入队列 116I、116E 和 PIPE(管道) 118I、118E。可选地, 在 I-UNIT 102 和 E-UNIT 104 内, 可具有 LCB 112、114 的进一步分层结构, 以对公共系统时钟进行更细粒度分布、放大和控制。

关于该例子, 计算机程序指令包含在 ICACHE 108 内, 即, I-UNIT 管道第一级内。通常, 在可能引起 ICACHE 漏失 (miss) 的各种条件下, 对于可变处理器周期数量, ICACHE 108 可能延迟。这种延迟毫无疑问地包含了漏失对前面指令传送级的影响, 即, 在指令存储器分层结构的较低级的影响。

继续通过 LCB 112、114, 利用同步时钟驱动每个单元, 而不考虑流水线中的延迟。即使在很小的范围内, 开关电容调制和模式位变化使功率消耗发生变化。因此, 在从开始执行程序到结束执行程序的每个时钟周期内, 消耗了大约同等量的能量 (在此, 利用标准化能量单位表示)。因此, 可以应用根据本发明的功率节省技术, 特别是功率节省方法, 节省大量能量。

在代码生成期间，利用插入了包括在指令集体系中的特殊指令的编译器，可以合成粗空闲控制；或者，例如，在用于特殊中断或者在某一上下文切换时，可以利用操作系统交替动态发出这些指令。在最粗的控制级，可以发出专用休眠类指令或者命令；这种休眠命令可以产生用于停止将时钟送到芯片上的选择部分一段时间的禁用信号。这种同样的专用休眠命令可以用于禁用指令取处理。同样，当取消禁用信号时或者在休眠周期之后，开始隐式唤醒；或者，可以利用显式异步中断实现唤醒。在本技术领域内众所周知，可以提供各种省电模式（例如，打瞌睡、打盹或者睡眠），在 LCB 分层结构的各级选择地禁用时钟分布树。在下一个较细粒度级，每当编译器可以统计方法预测计算阶段时，编译器都可以插入专用指令，以开始关断（gate-off）送到给定单元，例如，浮点单元的时钟。

可以包括自检机制，每当单元发现自己空闲时，该机制可以使该单元禁用它自己的时钟一段时间。在硬件中，可以设计该逻辑，以检测处理器中的本地空闲时间。然后，对于某个空闲区或者所有空闲区，检测可以触发时钟禁用。同样，根据禁用或者休眠的单元收到的新工作，自启动唤醒。

对于较细的空闲控制，动态定义的信号逐个周期选通本地时钟。在指令解码过程中，例如，超标量机器、处理器确定在后续执行周期内，可以时钟选通哪个功能单元管道。在具有“顺序”发出机制的处理器中，它工作得很好，因此，在时间，即，解码或者传送的时间之前，可以无歧义地充分进行选通确定。如果以集中发出队列方式逐个项保护指令类信息，则甚至对于失序发出队列，在发出时，也可以生成这样的选通信号。

在任意流水线数据通路上，可以动态检测冗余时钟，并选择地防止冗余时钟，例如，Data Valid 标志或者位沿逻辑流水线传播；仅当在某个周期生成的数据有效时，设置该 Data Valid 标志。然后，可以将每个逻辑级的 Data Valid 标志用作用于设置该级的输出锁存器的时钟启用。因此，通过被称为细粒、基于有效位的流水线级级别时钟选

通的连续流水线级，没有不必要地对无效数据提供时钟。

2001年6月12号授予 Sproch 等人的标题为“Method 和 System for Pipe Stage Gating Within an Operating Pipelined Circuit for Power Savings”的第 6,247,1342 B1 号美国专利描述了一种处理器，该处理器具有将任何新收到的、在利用逻辑的第一级进行的先前周期的计算中在流水线中不发生变化的操作数识别为无意义的逻辑。被作为无意义的这种不变条件信号的检测可以用于禁用到第一级的时钟，然后到后续级的时钟。

Ohnishi, M., Yamada, A., Noda, H.和 Kambe, T. “A Method of Redundant Clocking Detection 和 Power Reduction at the Rt Level Design,” Proc. Int'l. Symp. On Low Power Electronics 和 Design(ISLPED), 1997, pp. 131-136 说明了另一个更详细的空闲检测机制，用于防止各种冗余锁存时钟驱动。

不管是粗还是细，该技术的空闲控制机制的这些状态是自触发的、空间或者前馈的，如上所述；即，通过检测单元的空闲或者无效状态，然后，避免不必要的逐个周期时钟并在存在这种无效或者空闲状态位时进行数据传播，在局部产生选通条件或者信号。单元可以是整个区或者功能单元；或者它可以是设置的流水线级锁存装置。

相反，在第一实施例、具有要求的驱动时钟调节的标量、流水线处理器中，可调节地与执行单元（E 单元）一起工作的指令单元（I 单元）在这两个单元之间建立产生方-消费方关系。产生方 I 单元将就绪和数据启用指令转发到执行单元，用于在其内以不高于执行单元可以接受的速率进行处理。每个单元分别使活动状态寄存器保持至少 1 位信息。在该实施例中，利用公共同步时钟，对 I-E 单元对提供时钟。然而，根据在两个单元之间通过的本地单元活动信息，在局部控制和调节每个单元的时钟。每个单元的本地时钟控制可以是本地单元和远程单元，即，两个单元输出的活动状态信息的函数。

图 2 示出根据本发明的与同样标记的图 1A 具有同样部件的优选实施例要求或者活动驱动功率控制的标量处理器 120 的高级例子。在

该实施例中，处理单元 122、124 分别包括活动监视与时钟控制逻辑 126、128，用于监视单元活动程度（level）。在简单实施例中，单活动状态位 130 是用于规定迟延/不迟延状态的迟延位。当 E 单元 124 读出迟延条件（或者当前的或者即将来临的）时，它断言迟延位 130。该迟延位 130 用于下调 I 单元时钟 CLK-I132 的时钟速度，以调低 I 单元 122，并有效地降低（或者截止）送到 E 单元 124 的指令速率。根据控制粒度，E 单元活动状态或者迟延位 130 可以调节它自己的时钟，例如，在 E 单元 124 内的 134。当 E 单元 124 迟延结束时，将 CLK-I 132 回调到它的正常时钟速率。同样，当 I-UNIT 122 经历迟延条件（例如，ICACHE108 漏失）时，因此，其后几个周期 I-PIPE 118I 为空，没有什么提供给 E 单元 124，I-PIPE 空位 136 用于下调 CLK-E138，以节省功率。可以利用许多不同的方法，下面提供了其几个例子，实现活动监视与时钟控制逻辑 126、128 的上下调节能力。

因此，例如，一次一个流水线级地使每个单元的时钟异相。同样，当选通条件结束时，一次一个流水线级地使每个单元的时钟重新同相。该控制逻辑使每个单元时钟及时异相或者同相，而不丢失有效的信息，而且无需增加接口逻辑、缓冲或者保持的费能的流水线以及循环。或者，可以减慢或者逐步降低单元的时钟频率，以节省功率。另外，如果/在需要时，可以减慢时钟，然后，使它停止。

图 3 示出活动监视与时钟控制逻辑暂停（suspend）电路的第一个例子，其中选通控制移位寄存器（GCSR）150 使系统时钟 115 传送到 I-PIPE 118I 的各级 152。GCSR 150 是时钟驱动的 1 位触发器 154 的 1 位线性移位寄存器，一个触发器 154 对应于每个 I-PIPE 级 152。在“与”门 156 之一，GCSR 150 的每位分别与系统时钟 115 进行“与”运算，以有选择地将系统时钟 115 作为“与”门输出传送到相应 I-PIPE 级 152，一起作为 I-CLK158。通常，在该例子中，不检测 E 单元迟延，迟延位 130 仍然为低，而且被反相器 160 反相。由于 GCSR 150 中都是 1，所以 I-PIPE 118I 处于完全调节（full-throttle）。因此，“与”门 156 将未调节系统时钟 115 作为 I-CLK158 传送到每个 I-PIPE 级

152, 对于系统时钟 115 每个等同。因此, 只要 E 单元延迟位 130 保持 0, 则 1 移位到 GCSR 150, 从而使 I-PIPE 118I 保持完全调节。

在该例子中, 当 E 单元活动监视与时钟控制逻辑 (图 2 分别示出的 124 和 128) 检测到即将来临的延迟条件时, 它断言 E 单元延迟位 130。被反相器 160 反相的延迟位 130 将 0 送到 GCSR 150, 在下一个后续系统时钟周期, 将其移位到 GCSR 150。在该周期, 禁用第一 I-PIPE 级, 即, 图 2 中的 ICACHE116 后面的级, 中止数据输入到 I-PIPE 118I, 同时, 使 0 同步移位到 GCSR 150 并通过 GCSR 150。只要断言延迟位 130, 则 0 移位到 GCSR 150, 每个系统时钟周期一次。每个 0 顺序地禁止使系统时钟通过“与”门 156 中的连续“与”门, 在该例子中, 它是从左到右脉动。因此, 从左到右一次禁用 I-PIPE 时钟 158 的一个级。因此, 在断言延迟位 130 之前, 有效 I-PIPE 项继续通过 I-PIPE 118I, 然后, 进入 E 单元。为了防止丢失 I-PIPE 118I 中的有效信息, 至少在可用 E 单元缓冲 (队列 116E) 空间等于 I-PIPE 118I 中的有效项数量时, 必须断言 E 单元延迟位 130。在极端情况下, 后者仅等于 I-PIPE 118I 的长度。因此, 在保守设计中, 每当 E 单元队列 116E 填充到可用 (自由) 队列项数量等于 I-PIPE 级数的程度时, E 单元 124 断言延迟位 130。与传统细粒度的级级别时钟选通相同, 假定利用每项有效位 (用于节省功率) 时钟选通 E 单元队列 116E 中的空项或者无效项。

同样, 当延迟位 130 返回 0 时, 即, 当检测到 E 单元活动回到预定程度以下时, 出现了反向升高操作或者反向向上调节操作。随着将有效输入数据移位到 I-PIPE 118I, 低延迟位 130 重新开始将 1 同时移位到 GCSR 150 内。因此, 在紧接着的后续系统时钟周期, 逐级启动到 I-PIPE 118I 的 I-CLK 158, 以便 I-PIPE 118I 重新开始正常操作, 以完全调节将数据传送到 E 单元 124。逐级关断/接通 I-CLK 158 防止大的电流波动, 从而将对电源电压的 Ldi/dt 噪声影响降低到最低。

图 4 示出在第二个例子中, 用于请求式 I-CLK 调节的根据本发明优选实施例, 可以代替图 3 所示暂停电路的, 或者图 3 所示暂停电路

包括的减速电路 170。该减速电路 170 基本上与具有利用相同编号标记的共享元件或者公共元件的图 3 所示暂停电路相同。将慢选择 172 提供到触发器/1 位双稳计数器 174 的倒置输入端和“与”门 176。当慢选择位 172 位是高，即被断言时，“与”门 176 选择地将系统时钟 115 传送到时钟 1 位计数器 174。除仅将最后级输出 180 传送到所有“与”门 182 之外，慢选择 GCSR 178 基本上与图 3 所示的 GCSR 150 相同。“与”门 182 可以是三输入“与”门，提供对图 3 的“与”门 156 的暂停选择功能。因此，“与”门 182 可以将最后级输出 180 与系统时钟 115 和来自 GCSR（在该例子中未示出的图 3 所示 130）的相应级输出组合。“与”门 182 的各输出对相应 I-PIPE 118I 级提供时钟。

在该实施例中，响应慢选择 172，可以上（下）调节 I 单元的时钟频率。通过断言（去断言）慢选择 172，E 单元使 I 单元注意 E 单元中的减慢（增加）需求。此外，对于一个或者多个周期，可以完全保持暂停 I-CLK 158 的上述特征；如上参考图 3 所述，当 E 单元队列几乎满时，触发暂停，从而禁止 GCSR 150 的输出到各相应“与”门 182。当 E 单元活动队列长度降低到低于如上所述的预定门限时，I-CLK 158 重新启动。

在正常操作条件下，慢选择 172 是低（“0”），这样使 1 位控制计数器输出的连续高移位到 GCSR 178。因此，通常，GCSR 178 全部是 1，而且系统时钟 115 不被调节传送到 I-PIPE 级 152。当例如因为延迟而慢选择 172 被断言以发送 E 单元内要求减慢的信号时，“与”门 176 防止 1 位控制计数器 174 触发。当送到其倒置（inverted set）输入端的慢选择 174 升高，而且“与”门 176 传送系统时钟 115 时，1 位控制计数器 174 被释放。1 位控制计数器 174 开始触发，将交替的 0 和 1 序列传送到 GCSR 178。一旦通过 GCSR 178 传播交替模式，则在交替时钟周期，启动和禁用送到“与”门 182 的 I-CLK 控制。实际上，这样将主系统时钟频率减半，其被提供为 I-CLK 158。

图 5 示出图 3 和图 4 所示实施例的变型，其中将各 GCSR 级的输出传送到各相应“与”门 182。运行过程基本上与图 4 相同，但是不同

之处可在于，最后 I-PIPE 时钟调节。在调节（减速）阶段期间，在稳态操作中，以实际上是系统时钟频率一半的给定系统时钟周期，对交替的 I-PIPE 级提供时钟。在仅作为例子，在每个 I-PIPE 级内，指出有效位（V）。该有效位通常出现在图 1 至 5 所示例子的 I-PIPE（I 管道）、E-PIPE（E 管道）、I-QUEUE（I 队列）和 E-QUEUE（E 队列）结构中。根据传统的前馈解决方案，上游 I-UNIT 有效位向下游传播到 E-UNIT，以启动细粒、级级别时钟选通，从而节省附加功率。在该例子中，每个流水线级的有效位对用于合成本地级级别时钟的“与”结构提供附加选通控制。在调节（减速）模式期间，该实施例的特定 I-PIPE 时钟设置要求底层电路（未示出），防止覆盖写 I-PIPE 级中的有效数据，例如，各级之间，或者每个锁存级的主部分与从部分之间的备份中间位置锁存器可以使信息存储加倍。

图 6A - B 进一步详细示出与图 3 所示截面对应的标量 I-管道 1181 的又一例子及其相应时序图。每个 I-PIPE 级 152 连接到寄存器级 192 的输入端和输出端。在该例子中，每个 GCSR 锁存器 154 都是两级锁存器，而且基本上是一位串行输入/并行输出寄存器。该两级锁存器 154 包括第一级锁存器 194 和第二级锁存器 196。第一级锁存器 194 与寄存器级 192 中的锁存器相同。利用与第一级锁存器 194 的时钟极性相反的时钟极性，启动第二级锁存器 196。因此，例如，如果在时钟上沿对第一级锁存器 194 时钟驱动，则时钟下沿对第二级锁存器 196 时钟驱动，以确保将有效输入供给 I 时钟驱动器 156。如上参考图 3 至 5 以及在此参考图 6 所述，I 时钟驱动器 156 是“与”门，因此，起“与”门的作用。然而，每个“与”门 156 可以包括时钟驱动器，如果需要，该时钟驱动器可以是双相时钟驱动器。输入块 198 对特定类型的、选择的暂停/减慢控制逻辑提供适当时钟控制逻辑。因此，在图 3 的例子中，输入模块 198 可以包括反相器 160 以及该第一级 GCSR 150 的锁存器 154 的第一级 194。同样，对于图 4 和图 5 所示的例子，输入块 130 可以包括锁存器 174 和“与”门 176。

图 7 示出将本发明应用于流水线超标量处理器 200 的例子。这种超标量处理器 200 包括 I 单元 202 和 E 单元 220。I 单元 202 包括：指

令高速缓冲存储器 (ICACHE) 204; 组合 IFU/BRU (206), 包括指令取单元 (IFU) 和分支单元 (BRU); 调度单元 (DPU) 208; 完成单元 (CMU) 210; 以及分支地址单元 214, 包括分支历史表 (BHT) 和分支目标地址高速缓冲存储器 (BTAC)。此外, I 单元 202 还包括监视与时钟控制逻辑 216。这些单元 204、206、208、210 和 214 的操作基本上与众所周知的这种相应单元相同, 但是根据本发明, 利用监视与时钟控制逻辑 216 对这些单元提供时钟 (clock), 如下所述。

通常, 在每周期, IFU/BRU 206 中的 IFU 从 ICACHE204 取指令。利用监视与时钟控制逻辑 216, 可以实时 (on the fly) 对在现有技术处理器中被固定为每周期取的最多指令数的取带宽(fetch_bw)进行调节。在具有可用自由空间的 IFU/BRU206 中, 该 IFU 将取的指令置于取队列(FETCH_Q)中。在每个周期开始时, IFU 中的指令取地址寄存器 (IFAR) 引导指令取, 并且提供下一个取地址。在每周期, IFU 将每个下一个取地址设置为下列之一: (a) 下一个顺序地址, 它是被递增从而足以对先前周期取到 FETCH_Q 内的指令数进行计数的先前周期的 IFAR 值; (b) 分支指令目标, 在先前周期要求解或者预测的; 或者 (c) 在确定先前预测错了之后, 正确求解的分支指令的取地址。分支与指令取地址预测硬件 206 包括分支历史表 (BHT) 和分支目标地址高速缓冲存储器 (BTAC), 而且引导指令取过程。在每个有效取 (或者调度) 周期, 正常取 (或者调度) 固定数量的指令, 如相应固定带宽参数 (fetch_bw 或者 disp_bw) 确定的。然而, 当 E 单元 220 指示有必要减慢/暂停时, 除了上述时钟调节之外, 优选实施例处理器 (该例子中的 200) 动态调节每个 fetch_bw 和/或者 disp_bw 的值。

作为在 E-UNIT 内产生和监视的状态信号的组合功能, 合成来自 E-UNIT 的减慢/暂停 (或者其反加速/连续) 信号。这种状态信号可以包括: (a) 发出队列 FXQ229, LSQ232, FPQ240 和 VXQ246 的满或者空的指示; (b) DCACHE 238 命中或者漏失事件; (c) E-UNIT 内部共享的总线通信阻塞或者缺少它 (例如, 可以共享 (和仲裁) 一条总线, 以将结束信息送到完成单元 240); 或者 (d) 由于分支错误

预测或者其它形式的错误推测，出现执行流水线清除或者重发条件。在该例子中，可在 FXU 管道中执行处理器分支指令。然而，作为一种选择，为了执行分支指令，可以存在分离并行 BRU 管道。

断言来自 E-UNIT 的减慢/暂停信号可以通过 I-CLK 调节和/或者限制相关 I-UNIT 总线带宽之一或者二者调节 I-UNIT 流水线流速率，而无需调节时钟，例如，通过禁用从 ICACHE 接收取数据的一半线路，可以使给定访问 `fetch_bw` 有效减半。因此，为了在调节带宽模式下节省功率，是正常数量一半的项在指令缓冲器（在 IFU 206 内）。通常，根据所指出的下游 E-UNIT 减慢/暂停的严重性，可以将取带宽调节到正常模式的任何部分，包括直到 0。同样，可以调节调度总线带宽（`disp_bw`），以节省功率，如果需要或者这样指示，则可以将较少指令调度到消耗 E-UNIT 执行管道。

E 单元 220 包括定点执行单元（FXU）222、负载存储单元（LSU）224、浮点执行单元（FPU）226 以及矢量多媒体扩展单元（VMXU）228。FXU 222 包括定点队列 229 和定点单元执行单元管道 230。LSU 224 包括负载存储队列 232 和负载存储单元管道 234。FXU 222 和 LSU 224 都与通用寄存器 236 通信。LSU 224 提供与数据高速缓冲存储器 238 通信。FPU 226 包括定点队列 240 和定点单元管道 242 以及定点寄存器和更名缓冲器 244。LSU 224 还与定点更名缓冲器 244 通信。VMXU 228 包括矢量扩展队列 246 和矢量多媒体扩展单元管道 248。

单元 229、230、232、234、236、238、240、242、246、248 的操作基本上都与众所周知的这种相应单元相同，但是根据本发明要对其提供时钟，如下所述。对于该技术超标量处理器的任意典型状态，在给定工作负荷执行阶段期间，FXU 222 和 FPU 226 中的活动经常互不相容。当 FPU 226 活动时，该优选实施例处理器 200 可以禁用或者减慢 FXU 222 的本地时钟，反之亦然。此外，该优选实施例处理器 200 使 LSU 224 和 FPU 226 互相暂停/减慢时钟速度。除了单元内较粗粒度模式之外，还对单元内、较细粒度按需驱动时钟调节模式进行了说明。

与上面描述的优选标量处理器例子不同，E 单元 220 内的这两个单元 224、226 没有直接数流通道产生方-消费方关系，即，在 LSU 224 与 FPU 226 之间不存在直接信息流。利用数据高速缓冲存储器/存储器和浮点寄存器文件 244，在这两个单元 224、226 之间间接进行通信。通常，FPU 流水线 242 具有几级（例如，在现代千兆赫兹范围的处理器中具有 6 至 8 级），而典型 LSU 执行管道 234 是 2 至 4 级。因为该原因，而且由于当前处理器具有大量寄存器更名缓冲器，所以在 DCACHE 238 命中阶段期间，LSU 管 234 基本上在 FPU 管道 242 之前运行。另一方面，在群集 DCACHE 238 漏失阶段期间，可以显著增加有效 LSU 通道等待时间。如果一系列快速漏失使 DCACHE 238 延迟，则 LSU 发出队列 232 填满，这样又可以使上游产生方延迟。利用 FPU 226 的上游资源或者本地时钟调节的活动驱动细粒度临时时钟选通，本发明采用该过程。

图 8 示出 E-UNIT 200 的 LSU 224 和 FPU 226 的更详细例子。在该实施例中，LSU 事件/活动状态监视逻辑 250 对使用各种 LSU 队列进行监视，然后，获得 LSU 226 的活动状态。在该例子中，LSU 队列包括负载存储发出队列（LSQ）232、未决负载队列（PLQ）252 和未决存储队列（PSQ）254 以及 DCACHE 236。监视 DCACHE 236，然后，记录高速缓冲存储器“漏失”事件。显然，仅作为例子，选择这四个单元 232、236、252、254，然而，可以对更少或者更多的队列和事件进行监视。对于该例子，LSU 事件/活动状态监视逻辑 250 断言在该例子中传送到 FPU 226 的输出延迟位 256。如果需要更细的控制，则可以使用延迟位集。为了对 I-UNIT 202 进行控制以及对 IFU 或者 DISPATCH 单元 206 进行时钟控制，可以传送延迟位 256。

例如，首先，去断言 LSU 活动状态监视逻辑输出延迟位 256 和 FPU 活动状态监视延迟 260 的输出 258，在 LSU224 和 FPU226 中引起正常的全调节操作。如果断言 FPU 活动状态监视延迟位 258，而仍未断言 LSU 活动状态延迟位 256，例如，因为 FPQ 内的高利用率。使 LSU 本地时钟减慢，以使 FPU226 赶上比 FPU226 超前的 LSU224，

因为高速缓冲存储器命中阶段。相反，在断言 LSU 活动状态延迟位 256，而仍未断言 FPU 活动状态监视延迟位 258 时，使 FPU 本地时钟减慢。如果同时断言/去断言 LSU 和 FPU 延迟位，则根据 E-UNIT 220 或者 I-UNIT 202 内的其它处的其它状态条件，使 LSU 和 FPU 本地时钟减慢或者加速到相同频率。

有利的是，响应其它处理器或者系统单元中的活动/不活动，本发明选择地使单元或者单元内的部件减速、加速或者禁用，即，本发明具有可变的时钟控制粒度。根据活动和相对于数据流方向的正向流动和反向流动的信息，得到每个单元的本地时钟控件。代替不是有就是无的现有技术时钟选通，该优选实施例的自适应时钟提供可以采用前馈和反馈控制，以对任选带宽调节提供更灵活的通用时钟调节机制。

因此，未决信息以适当大小被当前保持在单元队列中，但是在现有技术流水线单元中可能丢失该未决信息。通过与各种单元有关内活动的信息动态控制该单元。响应指出需要减少的其它单元中的（各）活动程度，以分阶方式，向下调节给定部件中的时钟速率，乃至降低到 0。当监视的活动程度指出单元需求返回正常时，再以相控方式，使特定部件中的时钟速率恢复其初始正常程度。在当在一个或者多个部件中向下调节本地时钟速率时的时间期间，按比例降低净系统功率消耗。以及时预计方式，调节每个部件的时钟速率，使得实现速率变化的硬件控制不因为例如开销延迟或者重新计算废弃的指令而引起任何显著性能降低。将现有技术处理器中因为这种损失导致的性能恶化几乎降到 0。

与传统的时钟选通方法相比，进行时钟速率调节以上下调节频率的相控方式可以确保优良（即，更良好）的电流波动（ di/dt ）特性。因此，利用相控降低或者提高时钟速率，可以将引起的（inductive）噪声降低到最小。因此，优选实施例系统消耗的功率相当少，而不显著降低性能。降低了平均功率，而没有显著的（体系）性能降低（例如，每周指令或者 IPC），而且不需要大量附加硬件。对于要求严格遵守最大功率消耗限制和最高温度限制，本发明成功地控制了功率

消耗，同时将性能降低限制到小预定时间窗，以保持正常工作条件，而且快速返回正常。

各系统部件的动态活动程度被监视，送到其它部件的时钟速率在通过芯片或者系统传播的单个同步时钟的框架内。此外，与具有本地时钟驱动的异步（或者自定时）单元的同步系统或处理器，或者在全局异步控制下，具有多个同步时钟域的多时钟同步系统不同，本发明不需要为了确保同步而在单独时钟驱动的部件之间设置“握手”协议。此外，本发明动态调节各种部件中的时钟速率，以使通常与传统粗粒度时钟选通方法有关的引起的噪声问题降低到最小。

尽管根据几个（例子）优选实施例，对本发明进行了说明，但是本技术领域的专业技术人员明白，在所附权利要求的精神实质范围内，可以利用修改例实现本发明。

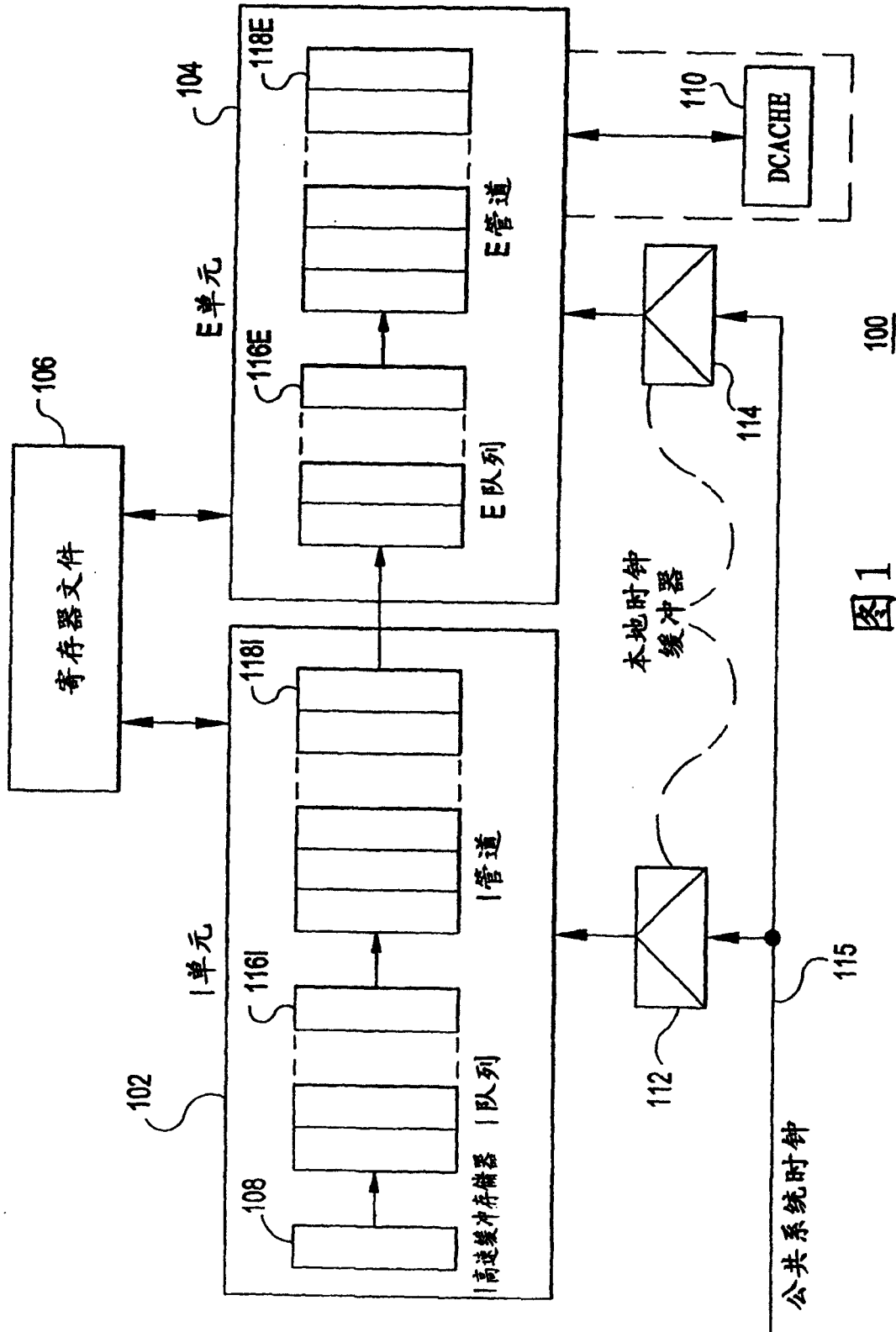


图1
(现有技术)

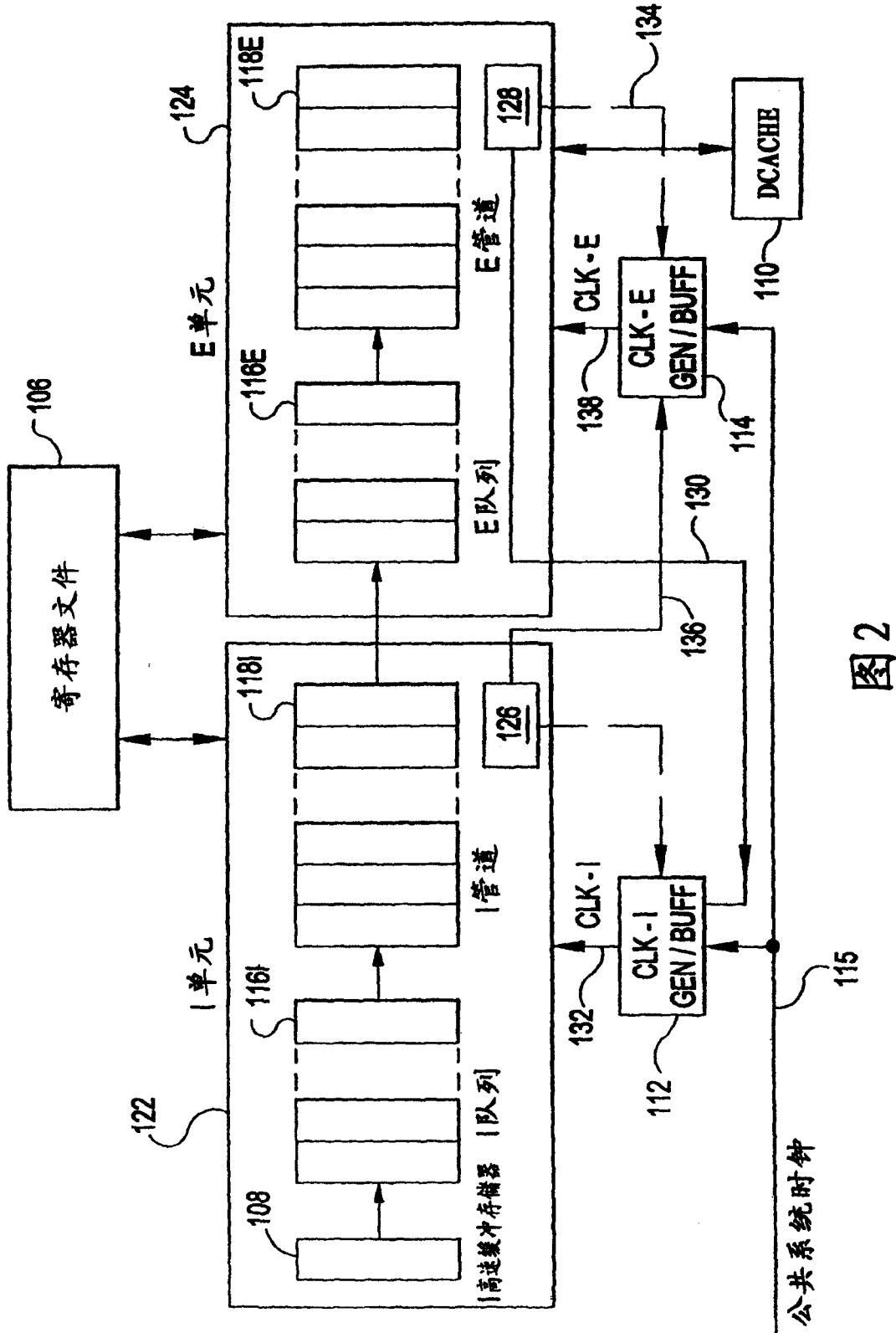


图 2

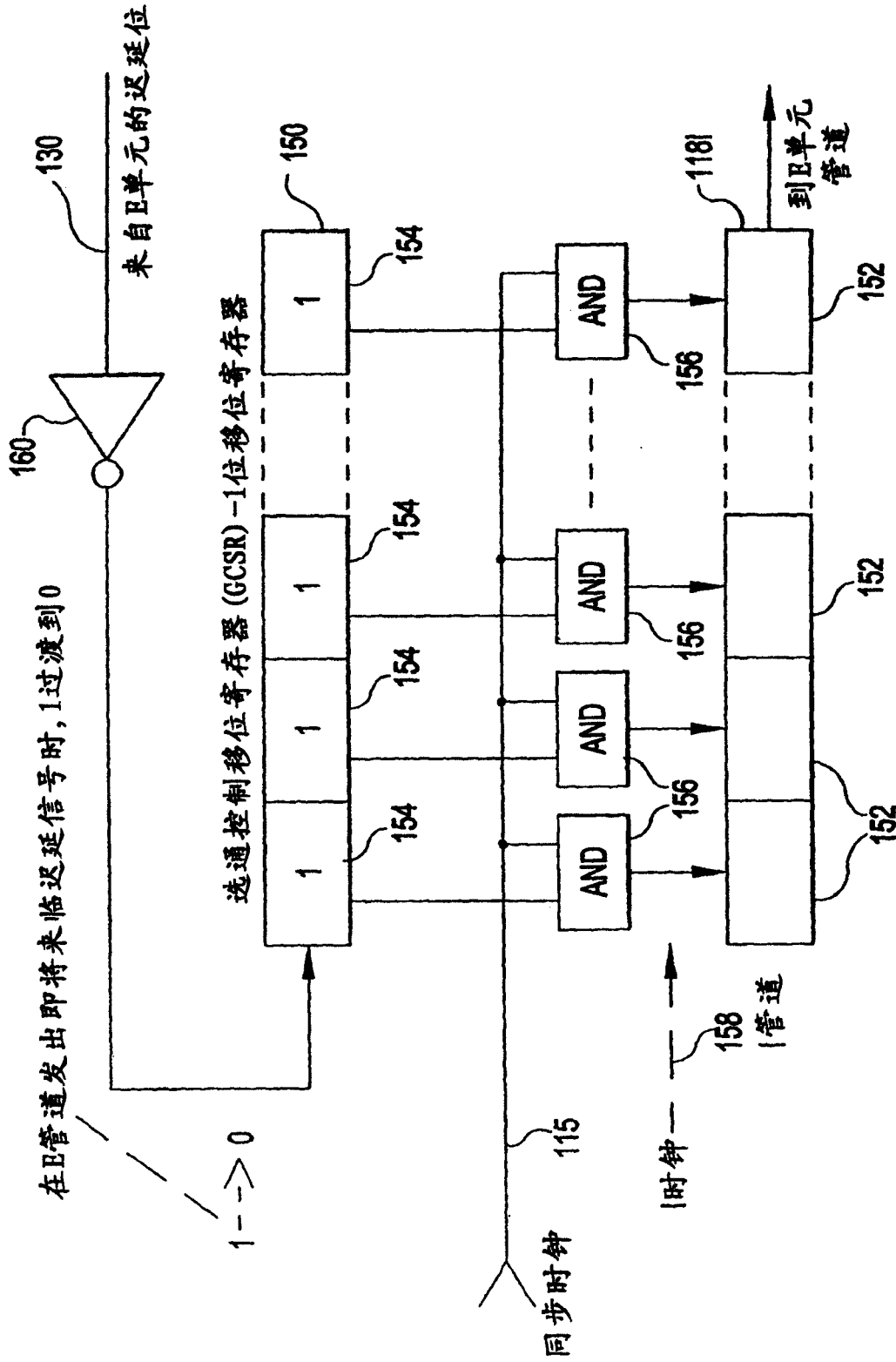


图3

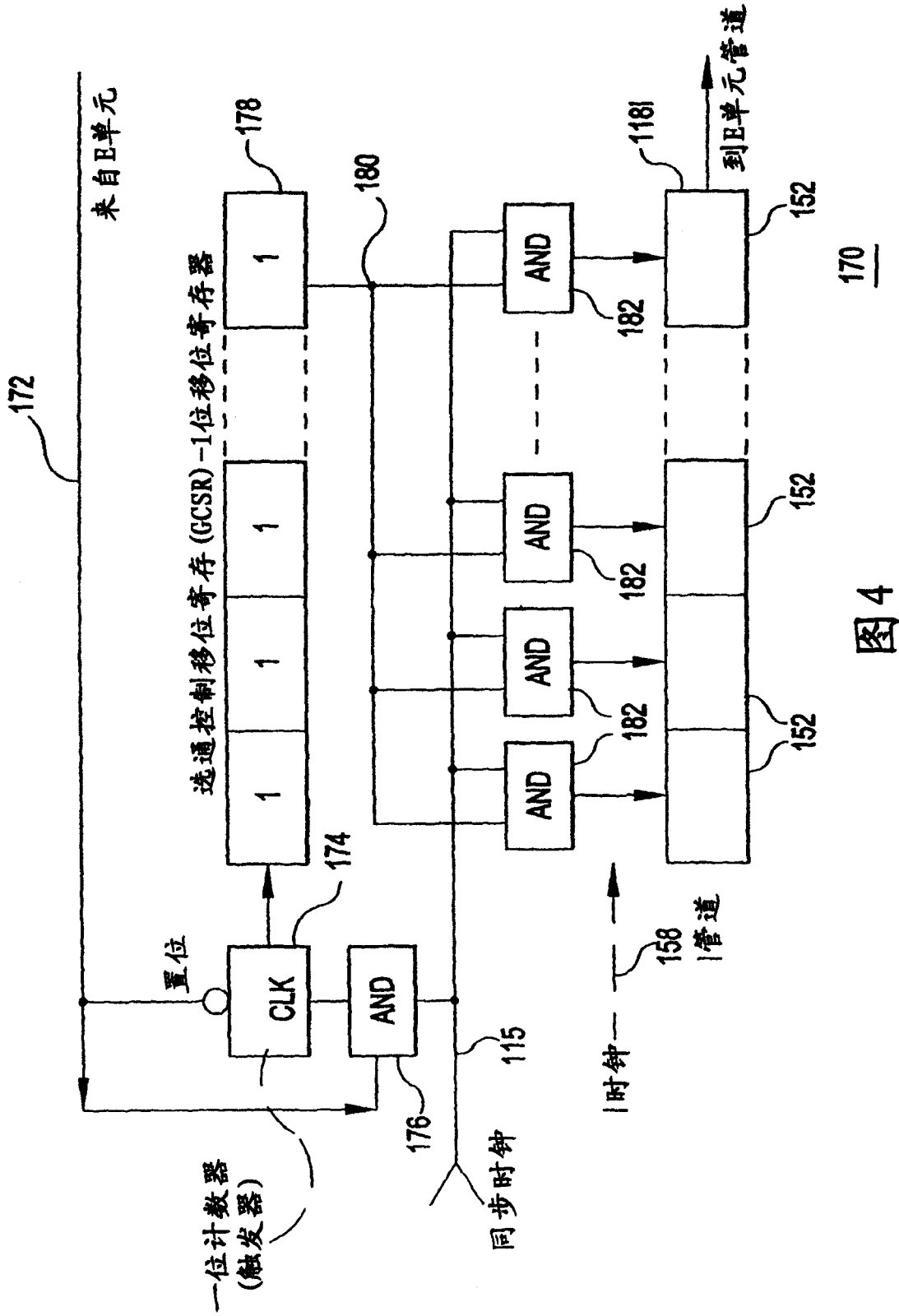


图 4

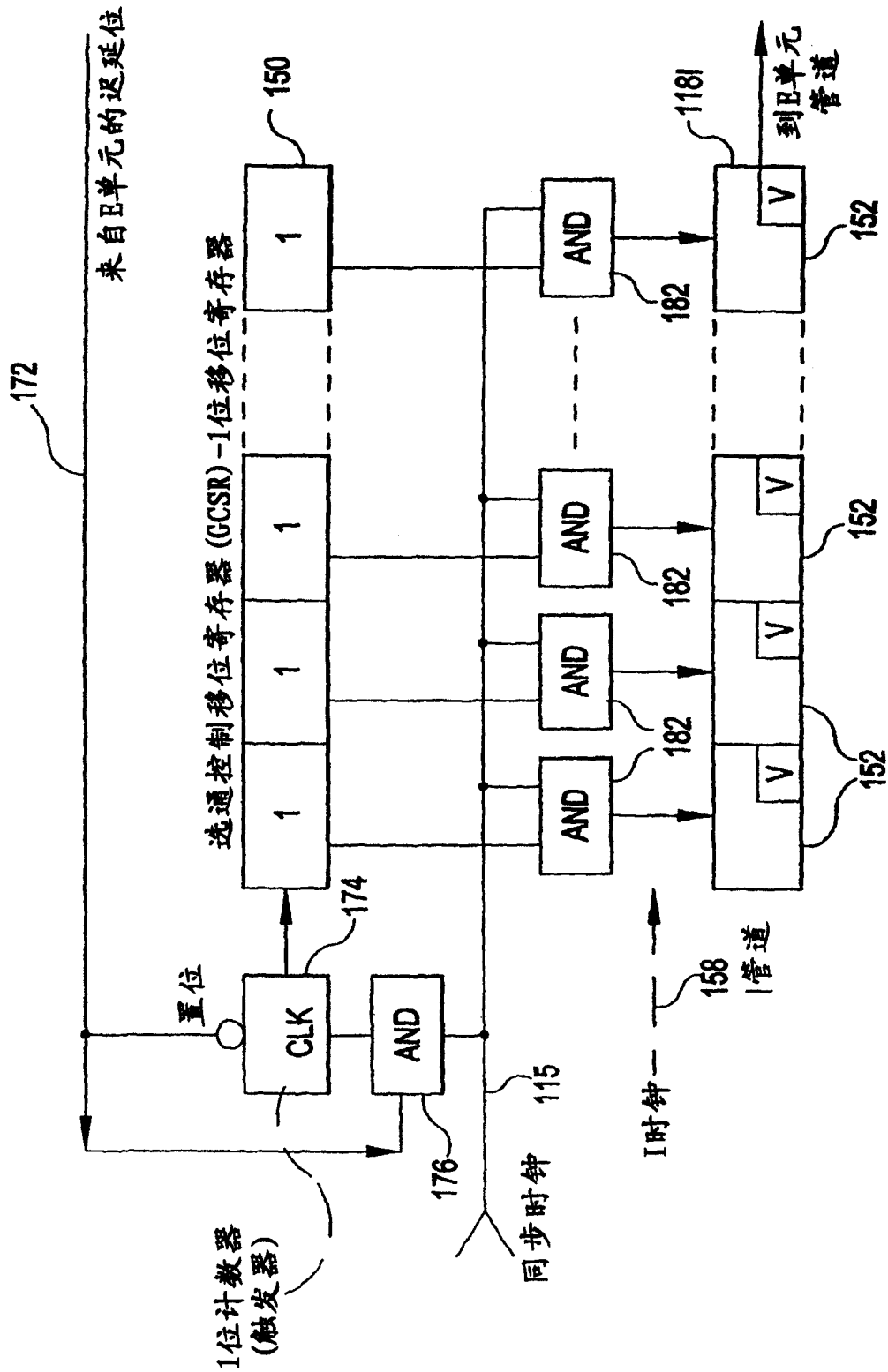


图5

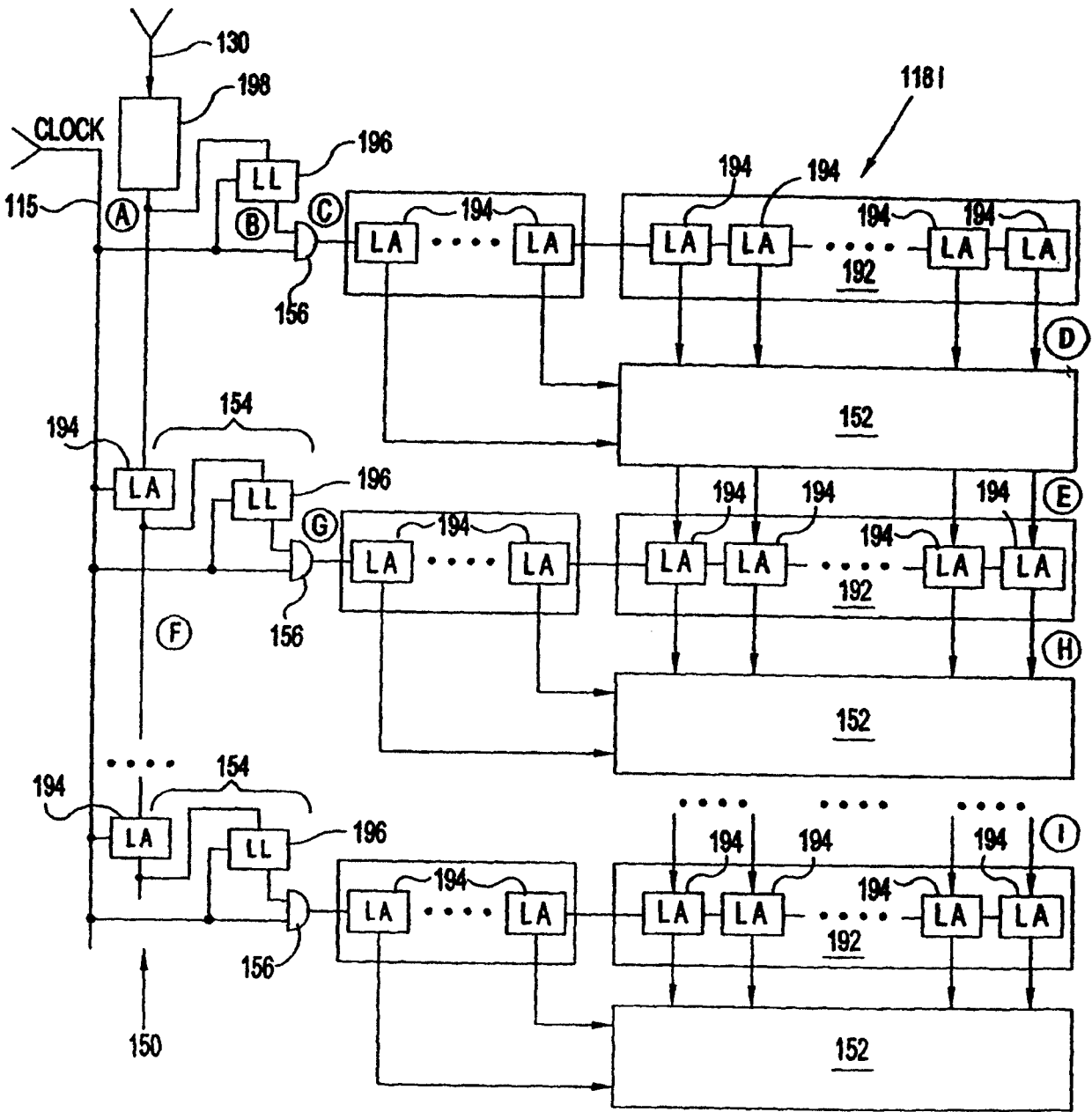
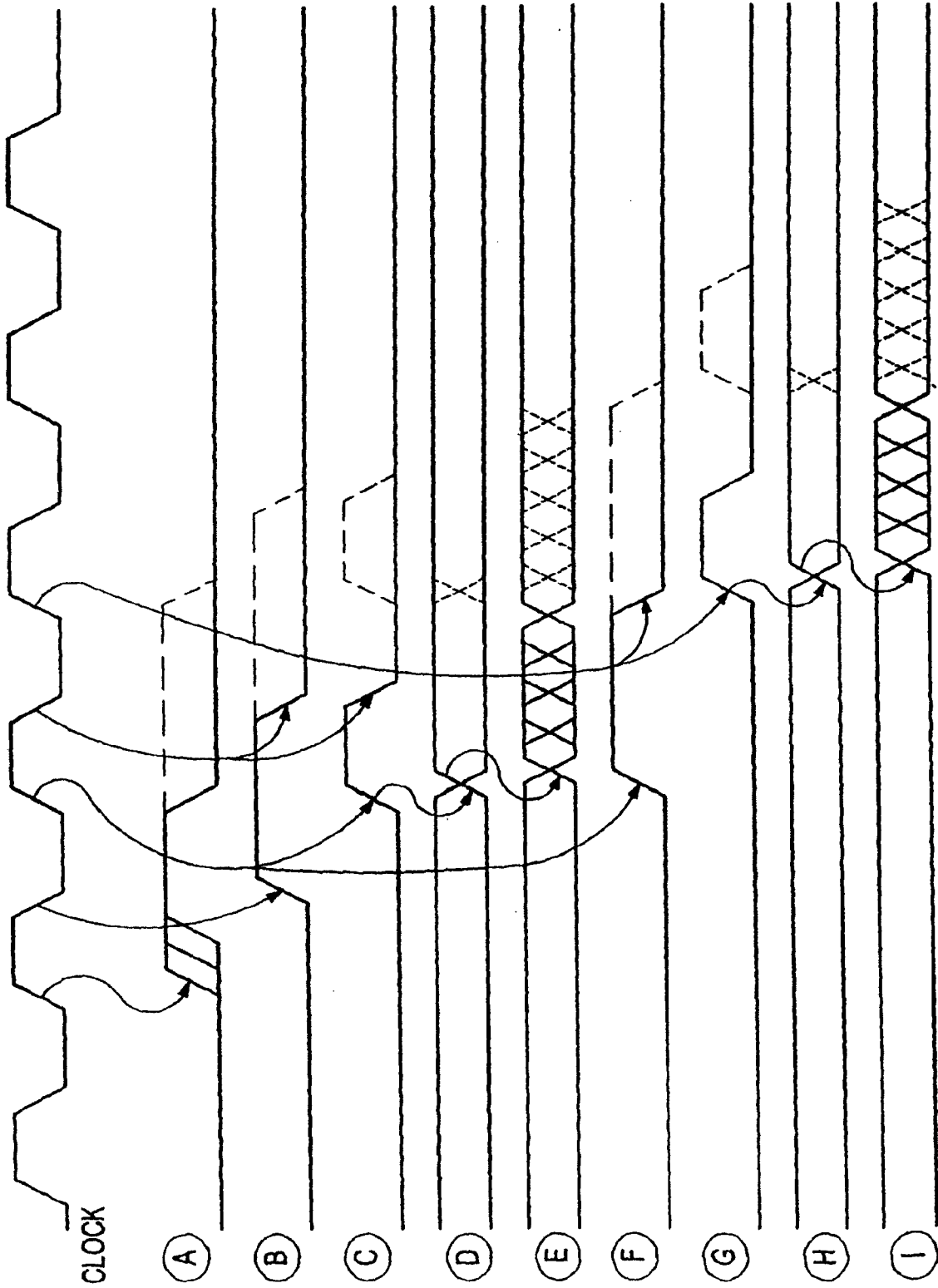
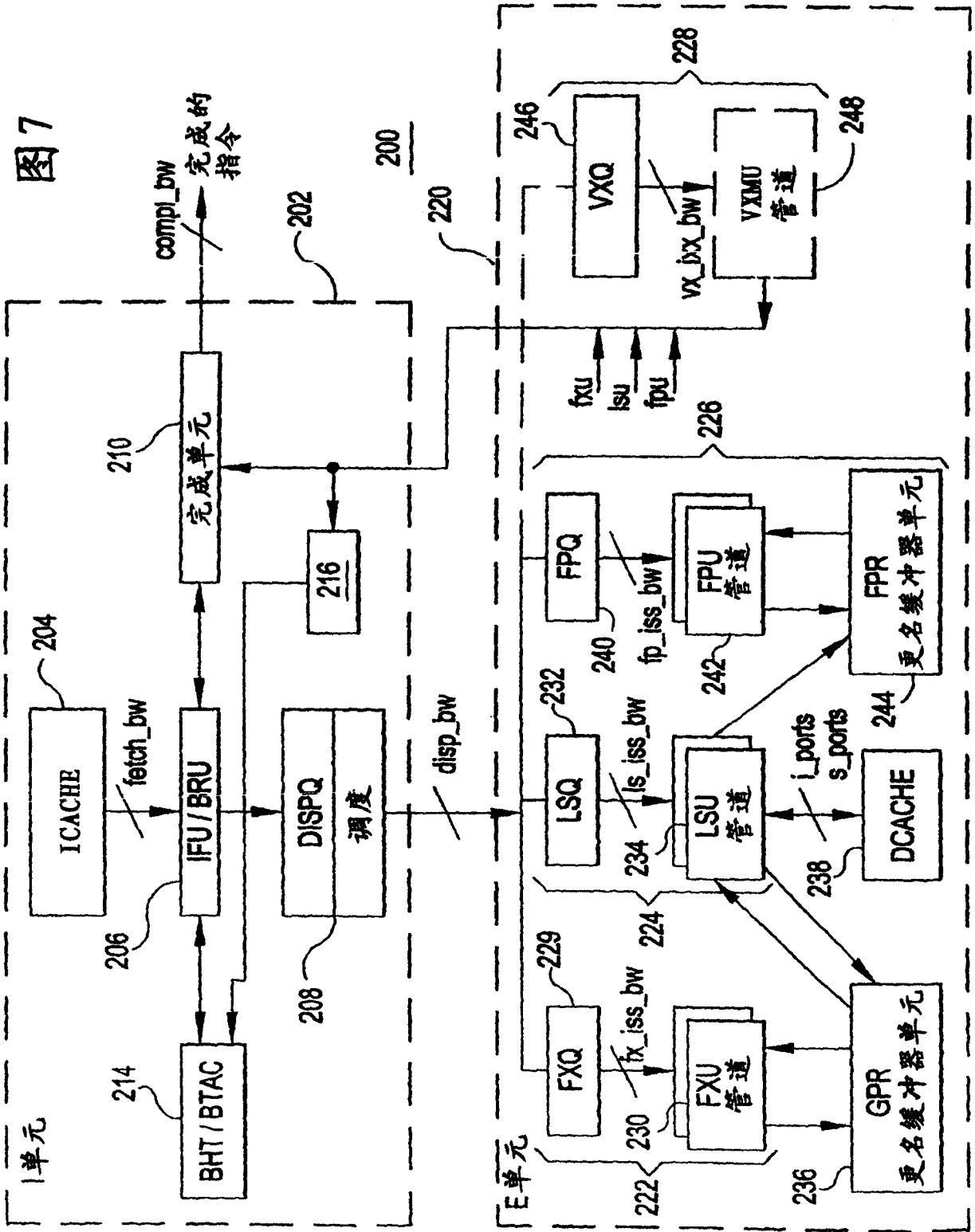


图 6A

图 6B





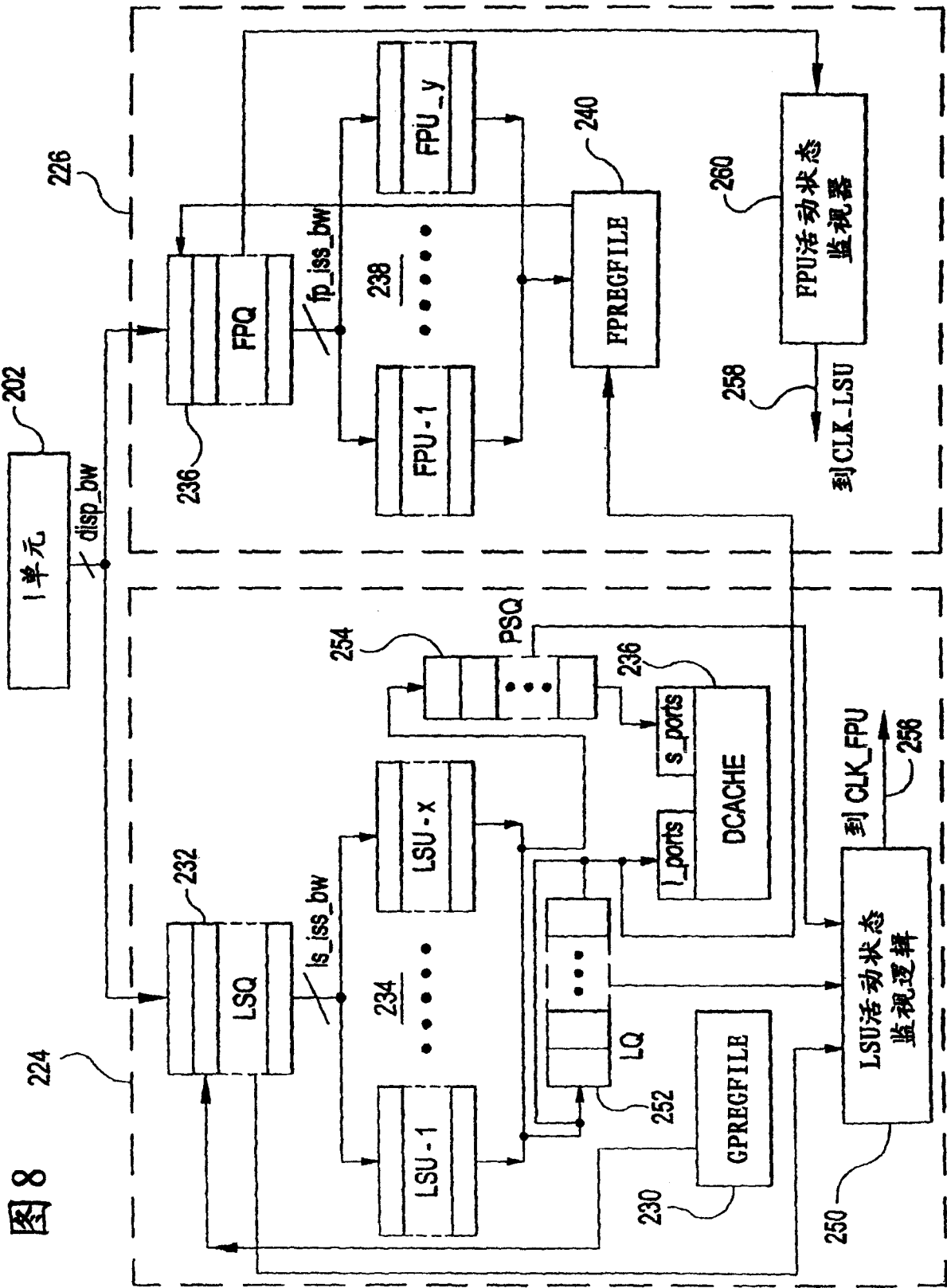


图 8