

1. 一种用于操作具有多个人工神经元的人工神经系统的方法,包括:
同时操作所述人工神经系统中的多个超神经元处理单元以用于更新所述人工神经元的状态,其中所述人工神经元的子集被指派给每个超神经元处理单元;
将所述超神经元处理单元与存储器对接;以及
访问存储在所述存储器中的所述人工神经系统的突触权重和可塑性参数,其中所述存储器的组织允许毗连存储器访问。
2. 如权利要求1所述的方法,其特征在于,同时操作多个超神经元处理单元包括:
将所述超神经元处理单元时间复用以用于输入电流累加器和所述人工神经元的模型。
3. 如权利要求1所述的方法,其特征在于,进一步包括:
在每个定义的步阶更新指派给所述超神经元处理单元之一的任何人工神经元的状态。
4. 如权利要求1所述的方法,其特征在于,进一步包括:
在所述存储器中存储指向与任何所述人工神经元相关联的突触后扇出的指针。
5. 如权利要求1所述的方法,其特征在于,同时操作多个超神经元处理单元包括:
由每个所述超神经元处理单元实现神经输入累加器。
6. 如权利要求1所述的方法,其特征在于,每个超神经元处理单元中的计算单元的数目小于指派给该超神经元处理单元的所述人工神经元的数目。
7. 如权利要求1所述的方法,其特征在于:
所述存储器包括动态随机存取存储器DRAM,以及
所述DRAM的所述毗连存储器访问包括同时访问有关于多个突触的数据。
8. 如权利要求1所述的方法,其特征在于,所述存储器的组织包括用于实现可塑性的突触前和突触后神经元查找表。
9. 如权利要求1所述的方法,其特征在于,进一步包括:
用动态盖然性方式丢弃所述人工神经系统的尖峰。
10. 如权利要求1所述的方法,其特征在于,进一步包括:
在所述存储器中存储与所述人工神经系统的突触相关联的数据,其中所述数据包括目的地路由信息。
11. 如权利要求1所述的方法,其特征在于,进一步包括:
在所述人工神经系统中,通过利用所述超神经元处理单元和所述存储器的组织来集成外部和内部突触激活。
12. 一种用于操作具有多个人工神经元的人工神经系统的装置,包括:
第一电路,其配置成同时操作所述人工神经系统中的多个超神经元处理单元以用于更新所述人工神经元的状态,其中所述人工神经元的子集被指派给每个超神经元处理单元;
第二电路,其配置成将所述超神经元处理单元与存储器对接;以及
第三电路,用于访问存储在所述存储器中的所述人工神经系统的突触权重和可塑性参数,其中所述存储器的组织允许毗连存储器访问。
13. 如权利要求12所述的装置,其特征在于,所述第一电路还被配置成:
将所述超神经元处理单元时间复用以用于输入电流累加器和所述人工神经元的模型。
14. 如权利要求12所述的装置,其特征在于,进一步包括:
第四电路,其被配置成在每个定义的步阶更新指派给所述超神经元处理单元之一的任

何人工神经元的状态。

15. 如权利要求12所述的装置,其特征在于,进一步包括:

第四电路,其配置成在所述存储器中存储指向与任何所述人工神经元相关联的突触后扇出的指针。

16. 如权利要求12所述的装置,其特征在于,进一步包括:

第四电路,其配置成由每个所述超神经元处理单元实现神经输入累加器。

17. 如权利要求12所述的装置,其特征在于,每个超神经元处理单元中的计算单元的数目小于指派给该超神经元处理单元的所述人工神经元的数目。

18. 如权利要求12所述的装置,其特征在于:

所述存储器包括动态随机存取存储器DRAM,以及

所述DRAM的所述毗连存储器访问包括同时访问有关于多个突触的数据。

19. 如权利要求12所述的装置,其特征在于,所述存储器的组织包括用于实现可塑性的突触前和突触后神经元查找表。

20. 如权利要求12所述的装置,其特征在于,进一步包括:

第四电路,其配置成用动态盖然性方式丢弃所述人工神经系统的尖峰。

21. 如权利要求12所述的装置,其特征在于,进一步包括:

第四电路,其配置成在所述存储器中存储与所述人工神经系统的突触相关联的数据,其中所述数据包括目的地路由信息。

22. 如权利要求12所述的装置,其特征在于,进一步包括:

第四电路,其配置成在所述人工神经系统中,通过利用所述超神经元处理单元和所述存储器的组织来集成外部和内部突触激活。

23. 一种用于操作具有多个人工神经元的人工神经系统的设备,包括:

用于同时操作所述人工神经系统中的多个超神经元处理单元以用于更新所述人工神经元的状态的装置,其中所述人工神经元的子集被指派给每个超神经元处理单元;

用于将所述超神经元处理单元与存储器对接的装置;以及

用于访问存储在所述存储器中的所述人工神经系统的突触权重和可塑性参数的装置,其中所述存储器的组织允许毗连存储器访问。

24. 如权利要求23所述的设备,其特征在于,所述用于同时操作多个超神经元处理单元的装置进一步包括:

用于将所述超神经元处理单元时间复用以用于输入电流累加器和所述人工神经元的模型的装置。

25. 如权利要求23所述的设备,其特征在于,进一步包括:

用于在每个定义的步阶更新指派给所述超神经元处理单元之一的任何人工神经元的状态的装置。

26. 如权利要求23所述的设备,其特征在于,进一步包括:

用于在所述存储器中存储指向与任何所述人工神经元相关联的突触后扇出的指针的装置。

27. 如权利要求23所述的设备,其特征在于,所述用于同时操作多个超神经元处理单元的装置进一步包括:

用于由每个所述超神经元处理单元实现神经输入累加器的装置。

28. 如权利要求23所述的设备,其特征在于,每个超神经元处理单元中的计算单元的数目小于指派给该超神经元处理单元的所述人工神经元的数目。

29. 如权利要求23所述的设备,其特征在于:

所述存储器包括动态随机存取存储器DRAM,以及

所述DRAM的所述毗连存储器访问包括同时访问有关于多个突触的数据。

30. 如权利要求23所述的设备,其特征在于,所述存储器的组织包括用于实现可塑性的突触前和突触后神经元查找表。

31. 如权利要求23所述的设备,其特征在于,进一步包括:

用于用动态盖然性方式丢弃所述人工神经系统的尖峰的装置。

32. 如权利要求23所述的设备,其特征在于,进一步包括:

用于在所述存储器中存储与所述人工神经系统的突触相关联的数据的装置,其中所述数据包括目的地路由信息。

33. 如权利要求23所述的设备,其特征在于,进一步包括:

用于在所述人工神经系统中,通过利用所述超神经元处理单元和所述存储器的组织来集成外部和内部突触激活的装置。

34. 一种用于操作具有多个人工神经元的人工神经系统的非瞬态计算机可读介质,所述非瞬态计算机可读介质具有指令,所述指令能执行以用于:

同时操作所述人工神经系统中的多个超神经元处理单元以用于更新所述人工神经元的状态,其中所述人工神经元的子集被指派给每个超神经元处理单元;

将所述超神经元处理单元与存储器对接;以及

访问存储在所述存储器中的所述人工神经系统的突触权重和可塑性参数,其中所述存储器的组织允许毗连存储器访问。

35. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

将所述超神经元处理单元时间复用以用于输入电流累加器和所述人工神经元的模型。

36. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

在每个定义的步阶更新指派给所述超神经元处理单元之一的任何人工神经元的状态。

37. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

在所述存储器中存储指向与任何所述人工神经元相关联的突触后扇出的指针。

38. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

由每个所述超神经元处理单元实现神经输入累加器。

39. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,每个超神经元处理单元中的计算单元的数目小于指派给该超神经元处理单元的所述人工神经元的数目。

40. 如权利要求34所述的非瞬态计算机可读介质,其特征在于:

所述存储器包括动态随机存取存储器DRAM,以及

所述DRAM的所述毗连存储器访问包括同时访问有关于多个突触的数据。

41. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,所述存储器的组织包括用于实现可塑性的突触前和突触后神经元查找表。

42. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

用动态盖然性方式丢弃所述人工神经系统的尖峰。

43. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

在所述存储器中存储与所述人工神经系统的突触相关联的数据,其中所述数据包括目的地路由信息。

44. 如权利要求34所述的非瞬态计算机可读介质,其特征在于,还包括用于以下操作的代码:

在所述人工神经网络中,通过利用所述超神经元处理单元和所述存储器的组织来集成外部和内部突触激活。

尖峰网络的高效硬件实现

[0001] 根据35U.S.C.§119的优先权要求

[0002] 本申请要求以下申请的权益,于2013年5月21日递交的,题为“Spike Time Windowing for Implementing Spike-Timing Dependent Plasticity (STDP) (用于实现尖峰定时依赖可塑性 (STDP) 的尖峰时间加窗)”的美国临时专利申请序列号61/825,657,于2013年8月6日递交的,题为“Spike Time Windowing for Implementing Spike-Timing Dependent Plasticity (STDP) (用于实现尖峰定时依赖可塑性 (STDP) 的尖峰时间加窗)”的美国临时专利申请号61/862,714,于2013年8月6日递交的,题为“Computed Synapses for Neruomorphic System (用于神经元形态系统的计算突触)”的美国临时专利申请序列号61/862,741,于2013年8月6日递交的,题为“Implementing Delays between Neurons in an Artificial Nervous System (在人工神经网络中实现神经元之间的延迟)”的美国临时专利申请序列号61/862,734,于2014年1月28日递交的,题为“Efficient Hardware Implementation of Spiking Networks (尖峰网络的高效硬件实现)”的美国临时专利申请序列号61/932,364,上述每个申请通过引用被完全纳入于此。

背景技术

[0003] 领域

[0004] 本公开的某些方面一般涉及人工神经网络,尤其涉及用于尖峰神经网络的高效硬件实现的方法和装置。

背景技术

[0005] 可包括一群互连的人工神经元(即神经处理单元)的人工神经网络是一种计算设备或者表示将由计算设备执行的方法。人工神经网络可具有生物学神经网络中的对应的结构和/或功能。然而,人工神经网络可为其中传统计算技术是麻烦的、不切实际的、或不胜任的某些应用提供创新且有用的计算技术。由于人工神经网络能从观察中推断出功能,因此这样的网络在因任务或数据的复杂度使得通过常规技术来设计该功能较为麻烦的应用中是特别有用的。

[0006] 一种类型的人工神经网络是尖峰(spiking)神经网络,其将时间概念以及神经元状态和突触状态纳入到其工作模型中,由此提供了丰富的行为集,在神经网络中能从该行为集涌现出计算功能。尖峰神经网络基于以下概念:神经元基于该神经元的状态在一个或多个特定时间激发或“发放尖峰”,并且该时间对于神经元功能而言是重要的。当神经元激发时,它生成一尖峰,该尖峰行进至其他神经元,这些其他神经元继而可基于接收到该尖峰的时间来调整它们的状态。换言之,信息可被编码在神经网络中的尖峰的相对或绝对定时中。

[0007] 概述

[0008] 本公开的特定方面提供了用于以多个人工神经元来进行人工神经网络操作的方法。该方法一般包括同时操作人工神经网络中的多个超神经元处理单元来更新这些人工神经

元的状态,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及将这些超神经元处理单元与用于访问人工神经系统的突触权重和可塑性参数的存储器对接,其中存储器的组织允许毗连存储器访问。

[0009] 本公开的某些方面提供了一种用于操作人工神经系统的装置。该装置一般包括配置成同时操作人工神经网络中的多个超神经元处理单元来更新人工神经元的状态的第一电路,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及配置成将这些超神经元处理单元与用于访问人工神经系统的突触权重和可塑性参数的存储器对接的第二电路,其中存储器的组织允许毗连存储器访问。

[0010] 本公开的某些方面提供了一种用于操作人工神经系统的设备。该装置一般包括用于同时操作人工神经网络中的多个超神经元处理单元来更新人工神经元的状态的装置,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及用于将这些超神经元处理单元与用于访问人工神经系统的突触权重和可塑性参数的存储器对接的装置,其中存储器的组织允许毗连存储器访问。

[0011] 本公开的某些方面提供了一种用于操作人工神经系统的计算机程序产品。该计算机程序产品一般包括具有可执行以下操作的指令的非瞬态计算机可读介质(例如,存储设备):同时操作人工神经网络中的多个超神经元处理单元来更新人工神经元的状态,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及将这些超神经元处理单元与用于访问人工神经系统的突触权重和可塑性参数的存储器对接,其中存储器的组织允许毗连存储器访问。

[0012] 附图简述

[0013] 为了能详细理解本公开的以上陈述的特征所用的方式,可参照各方面来对以上简要概述的内容进行更具体的描述,其中一些方面在附图中解说。然而应该注意,附图仅解说了本公开的某些典型方面,故不应被认为限定其范围,因为本描述可允许有其他等同有效的方面。

[0014] 图1解说了根据本公开的某些方面的示例神经网络。

[0015] 图2解说了根据本公开的某些方面的计算网络(神经系统或神经网络)的示例处理单元(神经元)。

[0016] 图3解说了根据本公开的某些方面的示例尖峰定时依赖可塑性(STDP)曲线。

[0017] 图4是根据本公开的某些方面的用于人工神经元的状态的示例曲线图,其解说用于定义神经元的行为的正态相和负态相。

[0018] 图5解说了根据本公开的某些方面的人工神经系统的示例硬件实现。

[0019] 图6是根据本公开某些方面的用于操作人工神经系统的示例操作的流程图。

[0020] 图6A解说了能够执行图6中示出的操作的示例装置。

[0021] 图7解说了根据本公开的某些方面的用于使用通用处理器来操作人工神经系统的示例实现。

[0022] 图8解说了根据本公开的某些方面的用于操作人工神经系统的示例实现,其中存储器可与个体分布式处理单元对接。

[0023] 图9解说了根据本公开的某些方面的用于基于分布式存储器和分布式处理单元来操作人工神经系统的示例实现。

[0024] 图10解说根据本公开的某些方面的神经网络的示例实现。

[0025] 详细描述

[0026] 以下参照附图更全面地描述本公开的各个方面。然而，本公开可用许多不同形式来实施并且不应解释为被限定于本公开通篇给出的任何具体结构或功能。相反，提供这些方面是为了使得本公开将是透彻和完整的，并且其将向本领域技术人员完全传达本公开的范围。基于本文中的教导，本领域技术人员应领会，本公开的范围旨在覆盖本文中所披露的本公开的任何方面，不论其是与本公开的任何其他方面相独立地实现还是组合地实现的。例如，可以使用本文所阐述的任何数目的方面来实现装置或实践方法。另外，本公开的范围旨在覆盖使用作为本文中所阐述的本公开的各个方面的补充或者另外的其他结构、功能性、或者结构及功能性来实践的此类装置或方法。应当理解，本文中所披露的本公开的任何方面可由权利要求的一个或多个元素来实施。

[0027] 措辞“示例性”在本文中用于表示“用作示例、实例或解说”。本文中描述为“示例性”的任何方面不必被解释为优于或胜过其他方面。

[0028] 尽管本文描述了特定方面，但这些方面的众多变体和置换落在本公开的范围之内。虽然提到了优选方面的一些益处和优点，但本公开的范围并非旨在被限定于特定益处、用途或目标。相反，本公开的各方面旨在能宽泛地应用于不同的技术、系统配置、网络和协议，其中一些作为示例在附图以及以下对优选方面的描述中解说。详细描述和附图仅仅解说本公开而非限定本公开，本公开的范围由所附权利要求及其等效技术方案来定义。

[0029] 示例神经系统

[0030] 图1解说根据本公开的某些方面的具有多级神经元的示例神经系统100。神经系统100可包括一级神经元102，该级神经元102通过突触连接网络104（即，前馈连接）来连接到另一级神经元106。为简单起见，图1中仅解说了两级神经元，但在典型的神经系统中可存在更少或更多级神经元。应注意，一些神经元可通过侧向连接来连接至同层中的其他神经元。此外，一些神经元可通过反馈连接来后向连接至先前层中的神经元。

[0031] 如图1所解说的，级102中的每一神经元可接收输入信号108，输入信号108可以是由前一级（图1中未示出）的多个神经元所生成的。信号108可表示至级102的神经元的输入（例如，输入电流）。此类输入可在神经元膜上累积以对膜电位进行充电。当膜电位达到其阈值时，该神经元可激发并生成输出尖峰，该输出尖峰将被传递到下一级神经元（例如，级106）。此类行为可在硬件和/或软件（包括模拟和数字实现）中进行仿真或模拟。

[0032] 在生物学神经元中，在神经元激发时生成的输出尖峰被称为动作电位。该电信号是相对迅速、瞬态、全有或全无的神经脉冲，其具有约为100mV的振幅和约为1ms的历时。在具有一系列连通的神经元（例如，尖峰从图1中的一级神经元传递至另一级）的神经系统的特定方面，每个动作电位都具有基本上相同的振幅和历时，因此该信号中的信息仅由尖峰的频率和数目（或尖峰的时间）来表示，而不由振幅来表示。动作电位所携带的信息由尖峰、发放尖峰的神经元、以及该尖峰相对于一个或多个其他尖峰的时间来决定。

[0033] 尖峰从一级神经元向另一级神经元的传递可通过突触连接（或简称“突触”）网络104来达成，如图1所解说的。突触104可从级102的神经元（相对于突触104而言的突触前神经元）接收输出信号（即尖峰）。对于某些方面，这些信号可根据可调整的突触权重

$w_1^{(i,j+1)}, \dots, w_p^{(i,j+1)}$ (其中P是级102和106的神经元之间的突触连接的总数) 来缩放。对于其它方面, 突触104可以不应用任何突触权重。此外, (经缩放的) 信号可被组合以作为级106中每个神经元 (相对于突触104而言的突触后神经元) 的输入信号。级106中的每个神经元可基于对应的组合输入信号来生成输出尖峰110。随后可使用另一突触连接网络 (图1中未示出) 将这些输出尖峰110传递到另一级神经元。

[0034] 生物学突触可被分类为电的或化学的。电突触主要用于发送兴奋性信号, 而化学突触可调节突触后神经元中的兴奋性或抑制性 (超极化) 动作, 并且还可用于放大神经元信号。兴奋性信号通常使膜电位去极化 (即, 相对于静息电位增大膜电位)。如果在某个时段内接收到足够的兴奋性信号以使膜电位去极化到高于阈值, 则在突触后神经元中发生动作电位。相反, 抑制性信号一般使膜电位超极化 (即, 降低膜电位)。抑制性信号如果足够强则可抵消掉兴奋性信号之和并阻止膜电位到达阈值。除了抵消掉突触兴奋以外, 突触抑制还可对自发活跃神经元施加强力的控制。自发活动神经元是指在没有进一步输入的情况下 (例如, 由于其动态或反馈而) 发放尖峰的神经元。通过压制这些神经元中的动作电位的自发生成, 突触抑制可对神经元中的激发模式进行定形, 这一般被称为雕刻。取决于期望的行为, 各种突触104可充当兴奋性或抑制性突触的任何组合。

[0035] 神经系统100可由通用处理器、数字信号处理器 (DSP)、专用集成电路 (ASIC)、现场可编程门阵列 (FPGA) 或其他可编程逻辑器件 (PLD)、分立的门或晶体管逻辑、分立的硬件组件、由处理器执行的软件模块、或其任何组合来仿真。神经系统100可用在大范围的应用中, 诸如图像和模式识别、机器学习、电机控制、及类似应用等。神经系统100中的每一神经元可被实现为神经元电路。被充电至发起输出尖峰的阈值的神经元膜可被实现为例如对流经其的电流进行积分的电容器。

[0036] 在一方面, 电容器作为神经元电路的电流积分器件可被除去, 并且可使用较小的忆阻器元件来替代它。这种办法可应用于神经元电路中, 以及其中大容量电容器被用作电流积分器的各种其他应用中。另外, 每个突触104可基于忆阻器元件来实现, 其中突触权重改变可与忆阻器电阻的变化有关。使用纳米特征尺寸的忆阻器, 可显著地减小神经元电路和突触的面积, 这可使得实现超大规模神经系统硬件实现变得可行。

[0037] 对神经系统100进行仿真的神经处理器的功能性可取决于突触连接的权重, 这些权重可控制神经元之间的连接的强度。突触权重可存储在非易失性存储器中以在掉电之后保留该处理器的功能性。在一方面, 突触权重存储器可实现在与主神经处理器芯片分开的外部芯片上。突触权重存储器可与神经处理器芯片分开地封装成可更换的存储卡。这可向神经处理器提供多种多样的功能性, 其中特定功能性可基于当前附连至神经处理器的存储卡中所存储的突触权重。

[0038] 图2解说根据本公开某些方面的计算网络 (例如, 神经系统或神经网络) 的处理单元 (例如, 人工神经元202) 的示例200。例如, 神经元202可对应于来自图1的级102和106的任一个神经元。神经元202可接收多个输入信号204₁-204_N (x_1-x_N), 这些输入信号可以是该神经系统外部的信号、或是由同一神经系统的其他神经元所生成的信号、或这两者。输入信号可以是实数值或复数值的电流或电压。输入信号可包括具有定点或浮点表示的数值。可通过突触连接将这些输入信号递送到神经元202, 这些突触连接根据可调节突触权重206₁-

206_N (w₁-w_N) 对这些信号进行缩放,其中N可以是神经元202的输入连接的总数。

[0039] 神经元202可组合这些经缩放的输入信号,并且使用组合的经缩放的输入来生成输出信号208(即,信号y)。输出信号208可以是实数值或复数值的电流或电压。输出信号可包括具有定点或浮点表示的数值。随后该输出信号208可作为输入信号传递至同一神经系统的其他神经元、或作为输入信号传递至同一神经元202、或作为该神经系统的输出来传递。

[0040] 处理单元(神经元202)可由电路来仿真,并且其输入和输出连接可由具有突触电路的导线来仿真。处理单元、其输入和输出连接也可由软件代码来仿真。处理单元也可由电路来仿真,而其输入和输出连接可由软件代码来仿真。在一方面,计算网络中的处理单元可包括模拟电路。在另一方面,处理单元可包括数字电路。在又一方面,处理单元可包括具有模拟和数字组件两者的混合信号电路。计算网络可包括任何前述形式的处理单元。使用这样的处理单元的计算网络(神经系统或神经网络)可用在大范围的应用中,诸如图像和模式识别、机器学习、电机控制、及类似应用等。

[0041] 在神经网络的训练过程期间,突触权重(例如,来自图1的权重 $w_1^{(i,i+1)}$ 、...、 $w_p^{(i,i+1)}$ 和/或来自图2的权重206₁-206_N)可用随机值来初始化并根据学习规则而增大或减小。学习规则的某些示例是尖峰定时依赖型可塑性(STDP)学习规则、Hebb规则、Oja规则、Bienenstock-Copper-Munro(BCM)规则等。很多时候,这些权重可稳定至两个值(即,权重的双峰分布)之一。该效应可被用于减少每突触权重的位数、提高从/向存储突触权重的存储器读取和写入的速度、以及降低突触存储器的功耗。

[0042] 突触类型

[0043] 在神经网络的硬件和软件模型中,突触相关功能的处理可基于突触类型。突触类型可包括非可塑突触(对权重和延迟没有改变)、可塑突触(权重可改变)、结构化延迟可塑突触(权重和延迟可改变)、全可塑突触(权重、延迟和连通性可改变)、以及基于此的变型(例如,延迟可改变,但在权重或连通性方面没有改变)。此举的优点在于处理可以被细分。例如,非可塑突触不会要求执行可塑性功能(或等待此类功能完成)。类似地,延迟和权重可塑性可被细分成可一起或分开地、顺序地或并行地运作的操作。不同类型的突触对于适用的每一种不同的可塑性类型可具有不同的查找表或公式以及参数。因此,这些方法将针对该突触的类型来访问相关的表。

[0044] 还进一步牵涉到以下事实:尖峰定时依赖型结构化可塑性可独立于突触可塑性地来执行。结构化可塑性即使在权重幅值没有改变的情况下(例如,如果权重已达最小或最大值、或者其由于某种其他原因而不改变)也可执行,因为结构化可塑性(即,延迟改变的量)可以是pre-post(突触前-突触后)尖峰时间差的直接函数。替换地,结构化可塑性可被设为权重改变量的函数或者可基于与权重或权重改变的界限有关的条件来设置。例如,突触延迟可仅在发生权重改变时或者在权重到达0的情况下才改变,但在权重达到最大极限时不改变。然而,具有独立函数以使得这些过程能被并行化从而减少存储器访问的次数和交叠可能是有利的。

[0045] 突触可塑性的确定

[0046] 神经元可塑性(或简称“可塑性”)是大脑中的神经元和神经网络响应于新的信息、

感官刺激、发展、损坏、或机能障碍而改变其突触连接和行为的能力。可塑性对于生物学中的学习和记忆、以及对于计算神经元科学和神经网络是重要的。已经研究了各种形式的可塑性,诸如突触可塑性(例如,根据赫布理论)、尖峰定时依赖可塑性(STDP)、非突触可塑性、活动性依赖可塑性、结构化可塑性和自身稳态可塑性。

[0047] STDP是调节神经元(诸如大脑中的那些神经元)之间的突触连接的强度的学习过程。连接强度是基于特定神经元的输出与收到输入尖峰(即,动作电位)的相对定时来调节的。在STDP过程下,如果至某个神经元的输入尖峰平均而言倾向于紧挨在该神经元的输出尖峰之前发生,则可发生长期增强(LTP)。于是使得该特定输入在一定程度上更强。相反,如果输入尖峰平均而言倾向于紧接在输出尖峰之后发生,则可发生长期抑压(LTD)。于是使得该特定输入在一定程度上更弱,由此得名为“尖峰定时依赖可塑性”。因此,使得可能是突触后神经元兴奋原因的输入甚至更有可能在将来作出贡献,而使得不是突触后尖峰的原因的输入较不可能在将来作出贡献。该过程继续,直至初始连接集的子集保留,而所有其他连接的影响减轻至0或接近0。

[0048] 由于神经元一般在其许多输入都在一短时段内发生(即,足以累积到引起输出)时产生输出尖峰,因此通常保留下来的输入子集包括倾向于在时间上相关的那些输入。另外,由于在输出尖峰之前发生的输入被加强,因此提供对相关性的最早充分累积指示的输入将最终变成至该神经元的最后输入。

[0049] STDP学习规则可因变于突触前神经元的尖峰时间 t_{pre} 与突触后神经元的尖峰时间 t_{post} 之间的时间差(即, $t = t_{post} - t_{pre}$)来有效地适配将该突触前神经元连接到该突触后神经元的突触的突触权重。STDP的典型公式是若该时间差为正(突触前神经元在突触后神经元之前激发)则增大突触权重(即,增强该突触),以及若该时间差为负(突触后神经元在突触前神经元之前激发)则减小突触权重(即,抑压该突触)。

[0050] 在STDP过程中,突触权重随时间推移的改变可通常使用指数衰退来达成,如由下式给出的:

$$[0051] \quad \Delta w(t) = \begin{cases} a_+ e^{-t/k_+} + \mu, & t > 0 \\ a_- e^{t/k_-}, & t < 0 \end{cases}, \quad (1)$$

[0052] 其中, k_+ 和 k_- 分别是正和负时间差的时间常量, a_+ 和 a_- 是对应的缩放振幅,以及 μ 是可以被施加到正时间差和/或负时间差的偏移。

[0053] 图3解说根据STDP,突触权重因变于突触前尖峰(pre)和突触后尖峰(post)的相对定时而改变的示例曲线图300。如果突触前神经元在突触后神经元之前激发,则可使对应的突触权重增大,如曲线图300的部分302中所解说的。该权重增大可被称为该突触的LTP。从曲线图部分302可观察到,LTP的量可因变于突触前和突触后尖峰时间之差而大致呈指数地下降。相反的激发次序可减小突触权重,如曲线图300的部分304中所解说的,从而导致该突触的LTD。

[0054] 如图3中的曲线图300中所解说的,可向STDP曲线图的LTP(因果性)部分302应用负偏移 μ 。x轴的交越点306($y=0$)可被配置成与最大时间滞后重合以考虑到来自层 $i-1$ (突触前层)的各因果性输入的相关性。在基于帧的输入(即,输入是按包括尖峰或脉冲的特定历时的帧的形式)的情形中,可计算偏移值 μ 以反映帧边界。该帧中的第一输入尖峰(脉冲)可

被视为随时间衰退,要么如直接由突触后电位所建模地、要么以对神经状态的影响的形式而随时间衰退。如果该帧中的第二输入尖峰(脉冲)被视为与特定时间帧关联或相关,则该帧之前和之后的相关时间可通过偏移STDP曲线的一个或多个部分以使得相关时间中的值可以不同(例如,对于大于一个帧为负,而对于小于一个帧为正)来在该时间帧边界处被分开并在可塑性方面被不同地对待。例如,负偏移 μ 可被设为偏移LTP以使得曲线实际上在大于帧时间的pre-post时间处变得低于零并且它由此为LTD而非LTP的一部分。

[0055] 神经元模型及操作

[0056] 存在一些用于设计有用的尖峰发放神经元模型的一般原理。良好的神经元模型在以下两个计算态相(regime)方面可具有丰富的潜在行为:重合性检测和功能性计算。此外,良好的神经元模型应当具有允许时间编码的两个要素:输入的抵达时间影响输出时间,以及重合性检测能具有窄时间窗。最后,为了在计算上是有吸引力的,良好的神经元模型在连续时间上可具有闭合形式解,并且具有稳定的行为,包括在靠近吸引子和鞍点之处。换言之,有用的神经元模型是可实践且可被用于建模丰富的、现实的且生物学一致的行为并且可被用于对神经电路进行工程设计和反向工程两者的神经元模型。

[0057] 神经元模型可取决于事件,诸如输入抵达、输出尖峰或其他事件,无论这些事件是内部的还是外部的。为了达成丰富的行为库,能展现复杂行为的状态机可能是期望的。如果事件本身的发生在撇开输入贡献(若有)的情况下能影响状态机并约束在该事件之后的动态,则该系统的将来状态并非仅是状态和输入的函数,而是状态、事件和输入的函数。

[0058] 在一方面,神经元n可被建模为尖峰带漏泄积分激发神经元,其膜电压 $v_n(t)$ 由以下动态来支配:

$$[0059] \quad \frac{dv_n(t)}{dt} = \alpha v_n(t) + \beta \sum_m w_{m,n} y_m(t - \Delta t_{m,n}) \quad (2)$$

[0060] 其中 α 和 β 是参数, $w_{m,n}$ 是将突触前神经元m连接到突触后神经元n的突触的突触权重,而 $y_m(t)$ 是神经元m的尖峰输出,其可根据 $\Delta t_{m,n}$ 被延迟达树突或轴突延迟,直到神经元n的胞体抵达。

[0061] 应注意,从建立了对突触后神经元的充分输入的时间直至突触后神经元实际上激发的时间之间存在延迟。在动态尖峰发放神经元模型(诸如Izhikevich简单模型)中,如果在去极化阈值 v_t 与峰值尖峰电压 v_{peak} 之间有差量,则可引发时间延迟。例如,在该简单模型中,神经元胞体动态可由关于电压和恢复的微分方程对来支配,即:

$$[0062] \quad \frac{dv}{dt} = (k(v - v_t)(v - v_r) - u + I)/C \quad (3)$$

$$[0063] \quad \frac{du}{dt} = a(b(v - v_r) - u) \quad (4)$$

[0064] 其中 v 是膜电位, u 是膜恢复变量, k 是描述膜电位 v 的时间尺度的参数, a 是描述恢复变量 u 的时间尺度的参数, b 是描述恢复变量 u 对膜电位 v 的阈下波动的敏感度的参数, v_r 是膜静息电位, I 是突触电流,以及 C 是膜的电容。根据该模型,神经元被定义为在 $v > v_{peak}$ 时发放尖峰。

[0065] Hunzinger Cold模型

[0066] Hunzinger Cold神经元模型是能再现丰富多样的各种神经行为的最小双态相尖峰发放线性动态模型。该模型的一维或二维线性动态可具有两个态相,其中时间常数(以及耦合)可取决于态相。在阈下态相中,时间常数(按照惯例为负)表示漏泄通道动态,其一般作用于以生物学一致的线性方式使细胞返回到静息。阈上态相中的时间常数(按照惯例为正)反映抗漏泄通道动态,其一般驱动细胞发放尖峰,而同时在尖峰生成中引发等待时间。如图4中所示,该模型的动态可被划分成两个(或更多个)态相。这些态相可被称为负态相402(也可互换地称为带漏泄积分激发(LIF)态相,勿与LIF神经元模型混淆)以及正态相404(也可互换地称为抗漏泄积分激发(ALIF)态相,勿与ALIF神经元模型混淆)。在负态相402中,状态在将来事件的时间趋向于静息(v_-)。在该负态相中,该模型一般展现出时间输入检测性质及其他阈下行为。在正态相404中,状态趋向于尖峰发放事件(v_s)。在该正态相中,该模型展现出计算性质,诸如取决于后续输入事件而引发发放尖峰的等待时间。在事件方面对动态进行公式化以及将动态分成这两个态相是该模型的基础特性。

[0067] 线性双态相二维动态(对于状态 v 和 u)可按照惯例定义为:

$$[0068] \quad \tau_\rho \frac{dv}{dt} = v + q_\rho \quad (5)$$

$$[0069] \quad -\tau_u \frac{du}{dt} = u + r \quad (6)$$

[0070] 其中 q_ρ 和 r 是用于耦合的线性变换变量。符号 ρ 在本文中用于标示动态态相,在讨论或表达具体态相的关系时,按照惯例对于负态相和正态相分别用符号“-”或“+”来替换符号 ρ 。模型状态由膜电位(电压) v 和恢复电流 u 来定义。在基本形式中,态相在本质上是由模型状态来决定的。该精确和通用的定义存在一些细微却重要的方面,但目前考虑该模型在电压 v 高于阈值(v_+)的情况下处于正态相404中,否则处于负态相402中。

[0071] 态相相关时间常数包括负态相时间常数 τ_- 和正态相时间常数 τ_+ 。恢复电流时间常数 τ_u 通常是与态相无关的。出于方便起见,负态相时间常数 τ_- 通常被指定为反映衰退的负量,从而用于电压演变的相同表达式可用于正态相,在正态相中指数和 τ_+ 将一般为正,正如 τ_u 那样。

[0072] 这两个状态元素的动态可在发生事件之际通过使状态偏离其零倾线(null-cline)的变换来耦合,其中变换变量为:

$$[0073] \quad q_\rho = -\tau_\rho \beta u - v_\rho \quad (7)$$

$$[0074] \quad r = \delta (v + \epsilon) \quad (8)$$

[0075] 其中 δ 、 ϵ 、 β 和 v_- 、 v_+ 是参数。 v_ρ 的两个值是这两个态相的参考电压的基数。参数 v_- 是负态相的基电压,并且膜电位在负态相中一般将朝向 v_- 衰退。参数 v_+ 是正态相的基电压,并且膜电位在正态相中一般将趋向于背离 v_+ 。

[0076] v 和 u 的零倾线分别由变换变量 q_ρ 和 r 的负数给出。参数 δ 是控制 u 零倾线的斜率的比例缩放因子。参数 ϵ 通常被设为等于 $-v_-$ 。参数 β 是控制这两个态相中的 v 零倾线的斜率的电阻值。 τ_ρ 时间常数参数不仅控制指数式衰退,还单独地控制每个态相中的零倾线斜率。

[0077] 该模型被定义为在电压 v 达值 v_s 时发放尖峰。随后,状态通常在发生复位事件(其

在技术上可以与尖峰事件完全相同)时被复位:

$$[0078] \quad v = \hat{v}_- \quad (9)$$

$$[0079] \quad u = u + \Delta u \quad (10)$$

[0080] 其中 $\hat{v}_-$ 和 Δu 是参数。复位电压 $\hat{v}_-$ 通常被设为 v_- 。

[0081] 依照瞬时耦合的原理,闭合形式解不仅对于状态是可能的(且具有单个指数项),而且对于到达特定状态所需的时间也是可能的。闭合形式状态解为:

$$[0082] \quad v(t + \Delta t) = (v(t) + q_p) e^{\frac{\Delta t}{\tau_p}} - q_p \quad (11)$$

$$[0083] \quad u(t + \Delta t) = (u(t) + r) e^{\frac{\Delta t}{\tau_u}} - r \quad (12)$$

[0084] 因此,模型状态可仅在发生事件之际被更新,诸如基于输入(突触前尖峰)或输出(突触后尖峰)而被更新。还可在任何特定时间(无论是否有输入或输出)执行操作。

[0085] 而且,依照瞬时耦合原理,可以预计突触后尖峰的时间,因此到达特定状态的时间可提前被确定而无需迭代技术或数值方法(例如,欧拉数值方法)。给定了先前电压状态 v_0 ,直至到达电压状态 v_f 之前的时间延迟由下式给出:

$$[0086] \quad \Delta t = \tau_p \log \frac{v_f + q_p}{v_0 + q_p} \quad (13)$$

[0087] 如果尖峰被定义为发生在电压状态 v 到达 v_s 的时间,则从电压处于给定状态 v 的时间起测量的直至发生尖峰前的时间量或即相对延迟的闭合形式解为:

$$[0088] \quad \Delta t_s = \begin{cases} \tau + \log \frac{v_s + q_+}{v + q_+} & \text{若 } v > \hat{v}_+ \\ \infty & \text{其他} \end{cases} \quad (14)$$

[0089] 其中 $\hat{v}_+$ 通常被设为参数 v_+ ,但其他变型可以是可能的。

[0090] 模型动态的以上定义取决于该模型是在正态相还是负态相中。如所提及的,耦合和态相 p 可基于事件来计算。出于状态传播的目的,态相和耦合(变换)变量可基于在上一(先前)事件的时间的状态来定义。出于随后预计尖峰输出时间的目的,态相和耦合变量可基于在下一(当前)事件的时间的状态来定义。

[0091] 存在对该Cold模型、以及在时间上执行模拟、仿真、或建模的若干可能实现。这包括例如事件-更新、步阶-事件更新、以及步阶-更新模式。事件更新是其中基于事件或“事件更新”(在特定时刻)来更新状态的更新。步阶更新是以间隔(例如,1ms)来更新模型的更新。这不一定要迭代方法或数值方法。通过仅在事件发生于步阶处或步阶间的情况下才更新模型或即通过“步阶-事件”更新,基于事件的实现以有限的时间分辨率在基于步阶的模拟器中实现也是可能的。

[0092] 神经编码

[0093] 有用的神经网络模型(诸如包括图1的人工神经元102、106的神经网络模型)可经由各种合适的神经编码方案(诸如重合性编码、时间编码或速率编码)中的任一种来编码信息。在重合性编码中,信息被编码在神经元集群的动作电位(尖峰发放活动性)的重合性(或

时间邻近性)中。在时间编码中,神经元通过对动作电位(即,尖峰)的精确定时(无论是以绝对时间还是相对时间)来编码信息。信息由此可被编码在一群神经元间的相对尖峰定时中。相反,速率编码涉及将神经信息编码在激发率或集群激发率中。

[0094] 如果神经元模型能执行时间编码,则其也能执行速率编码(因为速率正好是定时或尖峰间间隔的函数)。为了提供时间编码,良好的神经元模型应当具有两个要素:(1)输入的抵达时间影响输出时间;以及(2)重合性检测能具有窄时间窗。连接延迟提供了将重合性检测扩展到时间模式解码的一种手段,因为通过恰适地延迟时间模式的元素,可使这些元素达成定时重合性。

[0095] 抵达时间

[0096] 在良好的神经元模型中,输入的抵达时间应当对输出时间有影响。突触输入——无论是狄拉克 δ 函数还是经定形的突触后电位(PSP)、无论是兴奋性的(EPSP)还是抑制性的(IPSP)——具有抵达时间(例如, δ 函数的时间或者阶跃或其他输入函数的开始或峰值的时间),其可被称为输入时间。神经元输出(即,尖峰)具有发生时间(无论其是在何处(例如在胞体处、在沿轴突的一点处、或在轴突末端处)测量的),其可被称为输出时间。该输出时间可以是尖峰的峰值时间、尖峰的开始、或与输出波形有关的任何其他时间。普适原理是输出时间取决于输入时间。

[0097] 乍看起来可能认为所有神经元模型都遵循该原理,但一般并不是这样。例如,基于速率的模型不具有此特征。许多尖峰模型一般也并不遵循这一点。带漏泄积分激发(LIF)模型在有额外输入(超过阈值)的情况下并不会更快一点地激发。此外,在以非常高的定时分辨率来建模的情况下或许遵循这一点的模型在定时分辨率受限(诸如限于1ms步长)时通常将不会遵循这一点。

[0098] 输入

[0099] 神经元模型的输入可包括狄拉克 δ 函数,诸如电流形式的输入、或基于电导率的输入。在后一种情形中,对神经元状态的贡献可以是连续的或状态依赖型的。

[0100] 尖峰网络的高效硬件实现

[0101] 本公开的特定方面提供了尖峰神经网络的高效硬件实现。基于尖峰的计算在用硬件实现时可能需要特殊的考量。非常细粒度的并行性通常在具有基于尖峰计算的人工神经网络中是固有的。例如,可能需要为尖峰神经网络(人工神经系统,例如图1的系统100)中的大量人工神经元来高效地计算神经元“扇出”。对该问题的解决方案可以涉及利用动态随机存取存储器(DRAM),其代表了具有最高密度的当前技术,并且要求毗连接入以便实现最高存储器吞吐量。此外,基于尖峰的计算可能需要对人工神经系统的突触前和突触后人工神经元的可塑性查询。

[0102] 图5解说了根据本公开的某些方面的人工神经系统的示例硬件实现500。如以上所描述的基于STDP学习规则的突触权重的更新(学习)可在‘实施可塑性更新以及重组块’502中发生。对于本公开的特定方面,经更新的突触权重和可塑性参数可以经由高速缓存行接口504被存储在片外存储器(例如,DRAM 506)中。

[0103] 在典型的人工神经系统中,有远多于人工神经元的突触(突触连接)。由此,对于大神经网络(人工神经系统),总是期望用高效的方式处理突触权重更新。大量突触可使得想到将突触权重和其他参数存储在存储器(例如,DRAM506)中。当人工神经系统的人工神经元

在所谓的“超神经元(SN)” (例如,图5中的SN 508) 中生成尖峰时,这些人工神经元可通过DRAM查找以确定突触后神经元和对应神经权重的方式来将那些尖峰转发给突触后人工神经元。

[0104] 为了实现快速和高效的查找,突触排序可例如基于来自人工神经元的扇出被连贯地保持在存储器中。稍后,当在图5中的‘实施可塑性更新以及重组块’ 502中处理STDP更新时,在给定此存储器布局的情况下效率可规定基于转发扇出来处理STDP更新,因为不需要搜索DRAM或大的查找表来确定针对LTP更新的反向映射。图5中示出的办法促成了这一点。‘实施可塑性更新以及重组’ 块502可查询超神经元508以力图获得突触前和突触后尖峰时间,由此再次减少所涉及的状态存储器量。

[0105] 图5进一步解说了基于经时间复用的硬件来用于输入电流累加器和神经模型的超神经元的概念。如图5中所解说的,多个超神经元508可以并行操作从而改善效率,即,性能速度。根据本公开的特定方面,超神经元508中的计算单元的数目可以少于指派给该超神经元的人工神经元的数目,即,一个计算单元可以与多个人工神经元相关联。

[0106] 在本公开的一方面,人工神经元的状态可以每个定义的时段被更新,即“Tau”步阶。在一个方面,人工神经元可以将指向突触后扇出的指针存储在存储器中,例如,在DRAM 506中。在一方面,每个超神经元508可以实现神经输入的累加。

[0107] 根据本公开的特定方面,基于扇出的突触存储器组织可以导致高吞吐量和长“突发性”读取,这适合于DRAM。例如,单个512位“高速缓存行”(例如,图5中的高速缓存行接口 504) 可包括多个突触连接。应当注意,512位访问当前代表了DRAM的最高效率。

[0108] 根据本公开的特定方面,突触前和突触后神经元查找可以被用于实现可塑性学习规则。进一步,本公开的特定方面支持尖峰的动态盖然性丢弃,其中常规处理器可以被紧集成到该织构中。在本公开的一方面,涉及存储在DRAM存储器506中的突触(例如,权重和其他参数)的信息可以包括目的地路由信息。由于存储器映射和组织,这可以被实现。此外,外部和内部突触激活可以被集成到示例硬件实现500以及与其他处理节点的同步中。本公开中提供的存储器组织支持每个突触前神经元可以将其所有突触包括在一起。

[0109] 根据本公开的特定方面,可塑性查询可以向突触后神经元提供路由请求。消息可以被从突触前(攻击者)神经元发送到突触后(受攻击者)神经元,该突触后神经元可以随后路由回信息(包括返回地址)。单独的突触块可以促成这一点。

[0110] 图6是根据本公开的某些方面的用于操作具有多个人工神经元的人工神经系统的示例操作600的流程图。操作600可以硬件(例如由一个或多个神经处理单元,诸如神经元形态处理器)、以软件或以固件来执行。该人工神经系统可建模在各种生物或虚构神经系统中的任一者上,诸如视觉神经系统、听觉神经系统、海马体等。

[0111] 操作600可以在602始于同时操作人工神经系统中的多个超神经元处理单元以用于更新人工神经元的状态,其中这些人工神经元的子集可以被指派给每个超神经元处理单元。在604,超神经元处理单元可以与用于访问人工神经系统的突触权重和可塑性参数的存储器对接,其中存储器的组织允许毗连存储器访问。

[0112] 在本公开的一方面,同时操作多个超神经元处理单元可包括将这些超神经元处理单元时间复用以用于输入电流累加器和这些人工神经元的模型。指派给超神经元处理单元之一的人工神经元的状态可以在每个所定义的步阶(例如,“Tau”步阶)被更新。在一方面,

指向与任何人工神经元相关联的突触后扇出的指针可以被存储在存储器中。在一方面,同时操作多个超神经元处理单元可以包括由每个超神经元处理单元实现神经输入累加器。

[0113] 在本公开的一方面,每个超神经元处理单元中的计算单元的数目可以少于指派给该超神经元处理单元的人工神经元的数目。该存储器可包括,例如,DRAM,并且该DRAM的毗连存储器访问可包括对于涉及多个突触(例如,至多512位)的数据的同时访问。

[0114] 在一方面,存储器的组织可包括突触前和突触后查找以实现可塑性。在一方面,人工神经系统的尖峰的丢弃可以用动态盖然性方式来执行。在一方面,与人工神经系统的突触相关联的数据可以被存储在存储器中,其中该数据可包括目的地路由信息。在一方面,外部和内部突触激活可以通过利用超神经元处理单元和存储器的组织来被集成在人工神经网络中。

[0115] 图7解说了根据本公开的某些方面的使用通用处理器702来操作具有多个人工神经元的人工神经网络的前述方法的示例框图700。与计算网络(神经网络)相关联的变量(神经信号)、突触权重和/或系统参数可被存储在存储器块704中,而在通用处理器702处执行的有关指令可从程序存储器706中加载。在本公开的一方面,加载到通用处理器702中的指令可以包括,用于执行以下步骤的代码,同时操作人工神经网络中的多个超神经元处理单元来更新人工神经元的状态,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及将这些超神经元处理单元与用于访问人工神经网络的突触权重和可塑性参数的存储器对接,其中存储器的组织允许毗连存储器访问。

[0116] 图8解说了根据本公开的某些方面的用于操作具有多个人工神经元的人工神经网络的前述方法的示例框图800,其中存储器802可经由互连网络804与计算网络(神经网络)的个体(分布式)处理单元(神经处理器)806对接。与计算网络(神经网络)相关联的变量(神经信号)、突触权重和/或系统参数可被存储在存储器802中,并且可从存储器802经由互连网络804的连接被加载到每个处理单元(神经处理器)806中。在本公开的一方面,处理单元806可以配置成同时操作人工神经网络中的多个超神经元处理单元来更新人工神经元的状态,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及将这些超神经元处理单元与用于访问人工神经网络的突触权重和可塑性参数的存储器对接,其中存储器的组织允许毗连存储器访问。

[0117] 图9解说了根据本公开的某些方面的基于分布式权重存储器902和分布式处理单元(神经处理器)904来操作具有多个人工神经元的人工神经网络的前述方法的示例框图900。如图9中所解说的,一个存储器组902可直接与计算网络(神经网络)的一个处理单元904对接,其中该存储器组902可存储与该处理单元(神经处理器)904相关联的变量(神经信号)、突触权重和/或系统参数。在本公开的一方面,(诸)处理单元904可以配置成同时操作人工神经网络中的多个超神经元处理单元来更新人工神经元的状态,其中这些人工神经元的子集被指派给每个超神经元处理单元,以及将这些超神经元处理单元与用于访问人工神经网络的突触权重和可塑性参数的存储器对接,其中存储器的组织允许毗连存储器访问。

[0118] 图10解说根据本公开的某些方面的神经网络1000的示例实现。如图10中所解说的,神经网络1000可包括多个局部处理单元1002,它们可执行以上描述的方法的各种操作。每个处理单元1002可包括存储该神经网络的参数的局部状态存储器1004和局部参数存储器1006。另外,处理单元1002可包括具有局部(神经元)模型程序的存储器1008、具有局部学

习程序的存储器1010、以及局部连接存储器1012。此外,如图10中所解说的,每个局部处理单元1002可与用于配置处理的单元1014对接并且与路由连接处理元件1016对接,单元1014可提供对局部处理单元的局部存储器的配置,元件1016提供局部处理单元1002之间的路由。

[0119] 根据本公开的某些方面,每个局部处理单元1002可被配置成基于神经网络的一个或多个期望功能性特征来确定神经网络的参数,以及随着所确定的参数被进一步适配、调谐和更新来使这一个或多个功能性特征朝着期望的功能性特征发展。

[0120] 以上所描述的方法的各种操作可由能够执行相应功能的任何合适的装置来执行。这些装置可包括各种硬件和/或软件组件和/或模块,包括但不限于电路、专用集成电路(ASIC)、或处理器。例如,各种操作可由图5、7-10中所示的各种处理器中的一个或多个处理器来执行。一般而言,在存在附图中解说的操作的场合,这些操作可具有带相似编号的相应配对装置加功能组件。例如,图6中所解说的操作600对应于图6A中所解说的装置600A。

[0121] 例如,用于显示的装置可包括显示器(例如,监视器、平面屏幕、触摸屏等)、打印机、或任何其他用于输出数据以供视觉描绘(例如表、图表或图形)的合适装置。用于处理的装置、用于接收的装置、用于跟踪的装置、用于调节的装置、用于更新的装置、或用于确定的装置可包括处理系统,该处理系统可包括一个或多个处理器或处理单元。用于感测的装置可包括传感器。用于存储的装置可包括可由处理系统访问的存储器或任何其它合适的存储设备(例如,RAM)。

[0122] 如本文所使用的,术语“确定”涵盖各种各样的动作。例如,“确定”可包括演算、计算、处理、推导、研究、查找(例如,在表、数据库或其他数据结构中查找)、查明、及类似动作。而且,“确定”可包括接收(例如接收信息)、访问(例如访问存储器中的数据)、及类似动作。同样,“确定”还可包括解析、选择、选取、建立、及类似动作。

[0123] 如本文中所使用的,引述一系列项目中的“至少一个”的短语是指这些项目的任何组合,包括单个成员。作为示例,“a、b或c中的至少一者”旨在涵盖:a、b、c、a-b、a-c、b-c、以及a-b-c。

[0124] 结合本公开描述的各种解说性逻辑框、模块、以及电路可用设计成执行本文中描述的功能的通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其他可编程逻辑器件(PLD)、分立的门或晶体管逻辑、分立的硬件组件、或其任何组合来实现或执行。通用处理器可以是微处理器,但在替换方案中,处理器可以是任何市售的处理器、控制器、微控制器、或状态机。处理器还可以被实现为计算设备的组合,例如DSP与微处理器的组合、多个微处理器、与DSP核心协同的一个或多个微处理器、或任何其它此类配置。

[0125] 结合本公开所描述的方法或算法的步骤可直接在硬件中、在由处理器执行的软件模块中、或在这两者的组合中体现。软件模块可驻留在本领域所知的任何形式的存储介质中。可使用的存储介质的一些示例包括随机存取存储器(RAM)、只读存储器(ROM)、闪存、EPROM存储器、EEPROM存储器、寄存器、硬盘、可移动盘、CD-ROM,等等。软件模块可包括单条指令、或许多条指令,且可分布在若干不同的代码段上,分布在不同的程序间以及跨多个存储介质分布。存储介质可被耦合到处理器以使得该处理器能从/向该存储介质读写信息。替换地,存储介质可以被整合到处理器。

[0126] 本文所公开的方法包括用于实现所描述的方法的一个或多个步骤或动作。这些方法步骤和/或动作可以彼此互换而不会脱离权利要求的范围。换言之,除非指定了步骤或动作的特定次序,否则具体步骤和/或动作的次序和/或使用可以改动而不会脱离权利要求的范围。

[0127] 所描述的功能可在硬件、软件、固件或其任何组合中实现。如果以硬件实现,则示例硬件配置可包括设备中的处理系统。处理系统可以用总线架构来实现。取决于处理系统的具体应用和整体设计约束,总线可包括任何数目的互连总线和桥接器。总线可将包括处理器、机器可读介质、以及总线接口的各种电路链接在一起。总线接口可用于尤其将网络适配器等经由总线连接至处理系统。网络适配器可用于实现信号处理功能。对于某些方面,用户接口(例如,按键板、显示器、鼠标、操纵杆,等等)也可以被连接到总线。总线还可以链接各种其他电路,诸如定时源、外围设备、稳压器、功率管理电路以及类似电路,它们在本领域中是众所周知的,因此将不再进一步描述。

[0128] 处理器可负责管理总线和一般处理,包括执行存储在机器可读介质上的软件。处理器可用一个或多个通用和/或专用处理器来实现。示例包括微处理器、微控制器、DSP处理器、以及其他能执行软件的电路系统。软件应当被宽泛地解释成意指指令、数据、或其任何组合,无论是被称作软件、固件、中间件、微代码、硬件描述语言、或其他。作为示例,机器可读介质可包括RAM(随机存取存储器)、闪存、ROM(只读存储器)、PROM(可编程只读存储器)、EPROM(可擦式可编程只读存储器)、EEPROM(电可擦式可编程只读存储器)、寄存器、磁盘、光盘、硬驱动器、或者任何其他合适的存储介质、或其任何组合。机器可读介质可被实施在计算机程序产品中。该计算机程序产品可以包括包装材料。

[0129] 在硬件实现中,机器可读介质可以是处理系统中与处理器分开的一部分。然而,如本领域技术人员将容易领会的,机器可读介质或其任何部分可在处理系统外部。作为示例,机器可读介质可包括传输线、由数据调制的载波、和/或与设备分开的计算机产品,所有这些都可由处理器通过总线接口来访问。替换地或补充地,机器可读介质或其任何部分可被集成到处理器中,诸如高速缓存和/或通用寄存器文件可能就是这种情形。

[0130] 处理系统可以被配置为通用处理系统,该通用处理系统具有一个或多个提供处理器功能性的微处理器、以及提供机器可读介质中的至少一部分的外部存储器,它们都通过外部总线架构与其他支持电路系统链接在一起。替换地,处理系统可以用带有集成在单块芯片中的处理器、总线接口、用户接口、支持电路系统、和至少一部分机器可读介质的ASIC(专用集成电路)来实现,或者用一个或多个FPGA(现场可编程门阵列)、PLD(可编程逻辑器件)、控制器、状态机、门控逻辑、分立硬件组件、或者任何其他合适的电路系统、或者能执行本公开通篇所描述的各种功能性的电路的任何组合来实现。取决于具体应用和加诸于整体系统上的总设计约束,本领域技术人员将认识到如何最佳地实现关于处理系统所描述的功能性。

[0131] 机器可读介质可包括数个软件模块。这些软件模块包括当由处理器执行时使处理系统执行各种功能的指令。这些软件模块可包括传送模块和接收模块。每个软件模块可以驻留在单个存储设备中或者跨多个存储设备分布。作为示例,当触发事件发生时,可以从硬驱动器中将软件模块加载到RAM中。在软件模块执行期间,处理器可以将一些指令加载到高速缓存中以提高访问速度。随后可将一个或多个高速缓存行加载到通用寄存器文件中以供

处理器执行。在以下述及软件模块的功能性时,将理解此类功能性是在处理器执行来自该软件模块的指令时由该处理器来实现的。

[0132] 如果以软件实现,则各功能可作为一条或多条指令或代码存储在计算机可读介质上或藉其进行传送。计算机可读介质包括计算机存储介质和通信介质两者,这些介质包括促成计算机程序从一地向另一地转移的任何介质。存储介质可以是能被计算机访问的任何可用介质。作为示例而非限定,这样的计算机可读介质可包括RAM、ROM、EEPROM、CD-ROM或其他光盘存储、磁盘存储或其他磁存储设备、或能用于携带或存储指令或数据结构形式的期望程序代码且能被计算机访问的任何其他介质。任何连接也被正当地称为计算机可读介质。例如,如果软件是使用同轴电缆、光纤电缆、双绞线、数字订户线(DSL)、或无线技术(诸如红外(IR)、无线电、以及微波)从web网站、服务器、或其他远程源传送而来,则该同轴电缆、光纤电缆、双绞线、DSL或无线技术(诸如红外、无线电、以及微波)就被包括在介质的定义之中。如本文中所使用的盘(disk)和碟(disc)包括压缩碟(CD)、激光碟、光碟、数字多用碟(DVD)、软盘、和蓝光[®]碟,其中盘(disk)常常磁性地再现数据,而碟(disc)用激光来光学地再现数据。因此,在一些方面,计算机可读介质可包括非瞬态计算机可读介质(例如,有形介质)。另外,对于其他方面,计算机可读介质可包括瞬态计算机可读介质(例如,信号)。上述的组合应当也被包括在计算机可读介质的范围内。

[0133] 因此,一些方面可包括用于执行本文中给出的操作的计算机程序产品。例如,此种计算机程序产品可包括其上存储(和/或编码)有指令的计算机可读介质,这些指令能由一个或多个处理器执行以执行本文中所描述的操作。对于一些方面,计算机程序产品可包括包装材料。

[0134] 另外,应领会,用于执行本文中所描述的方法和技术的模块和/或其它恰适装置能由设备在适用的场合下载和/或以其他方式获得。例如,此类设备能被耦合至服务器以促成用于执行本文中所描述的方法的装置的转移。替换地,本文中所描述的各种方法能经由存储装置(例如,RAM、ROM、诸如压缩碟(CD)或软盘之类的物理存储介质等)来提供,以使得一旦将该存储装置耦合到或提供给设备,该设备就能获得各种方法。此外,可利用适于向设备提供本文中所描述的方法和技术的任何其他合适的技术。

[0135] 将理解,权利要求并不被限定于以上所解说的精确配置和组件。可在以上所描述的方法和装置的布局、操作和细节上作出各种改动、更换和变形而不会脱离权利要求的范围。

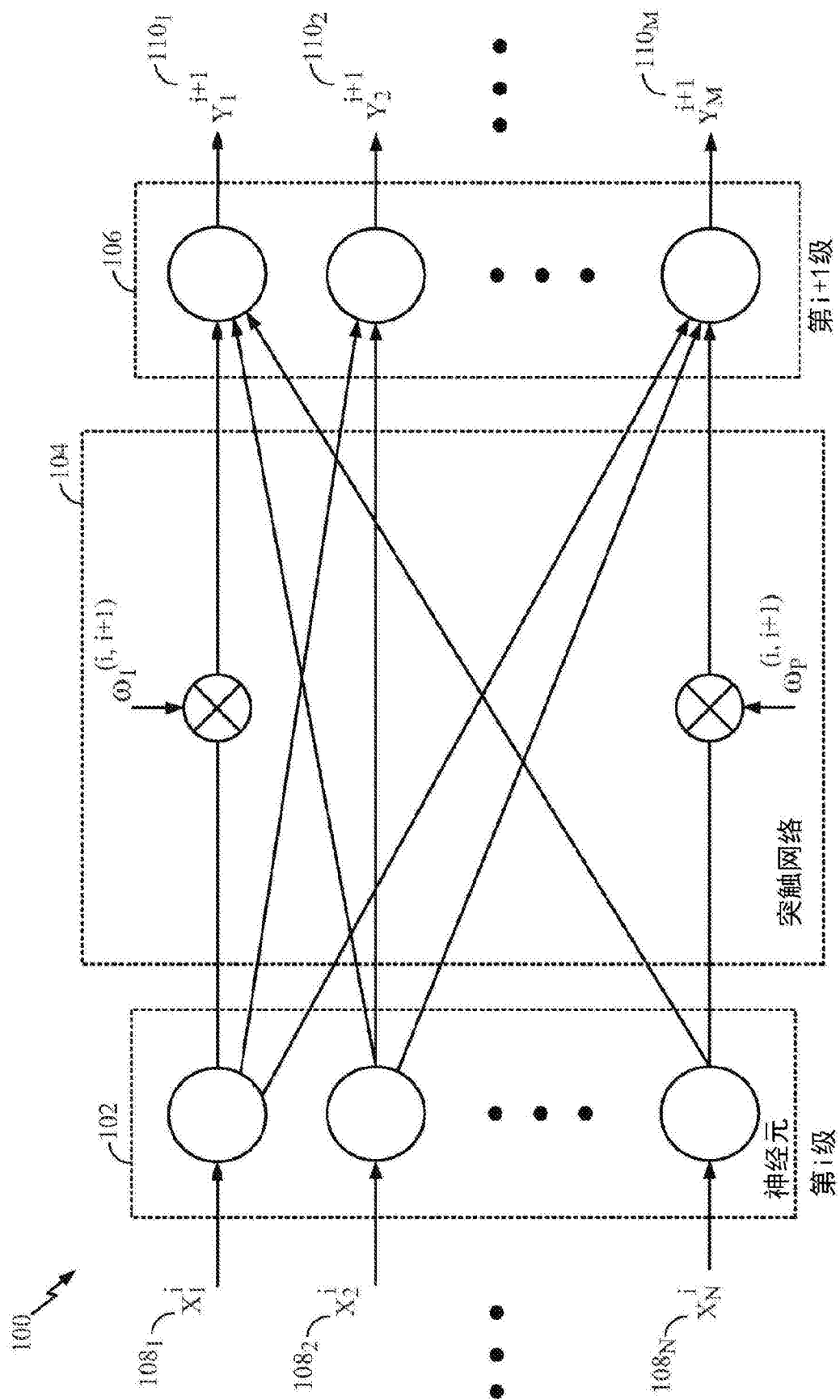


图1

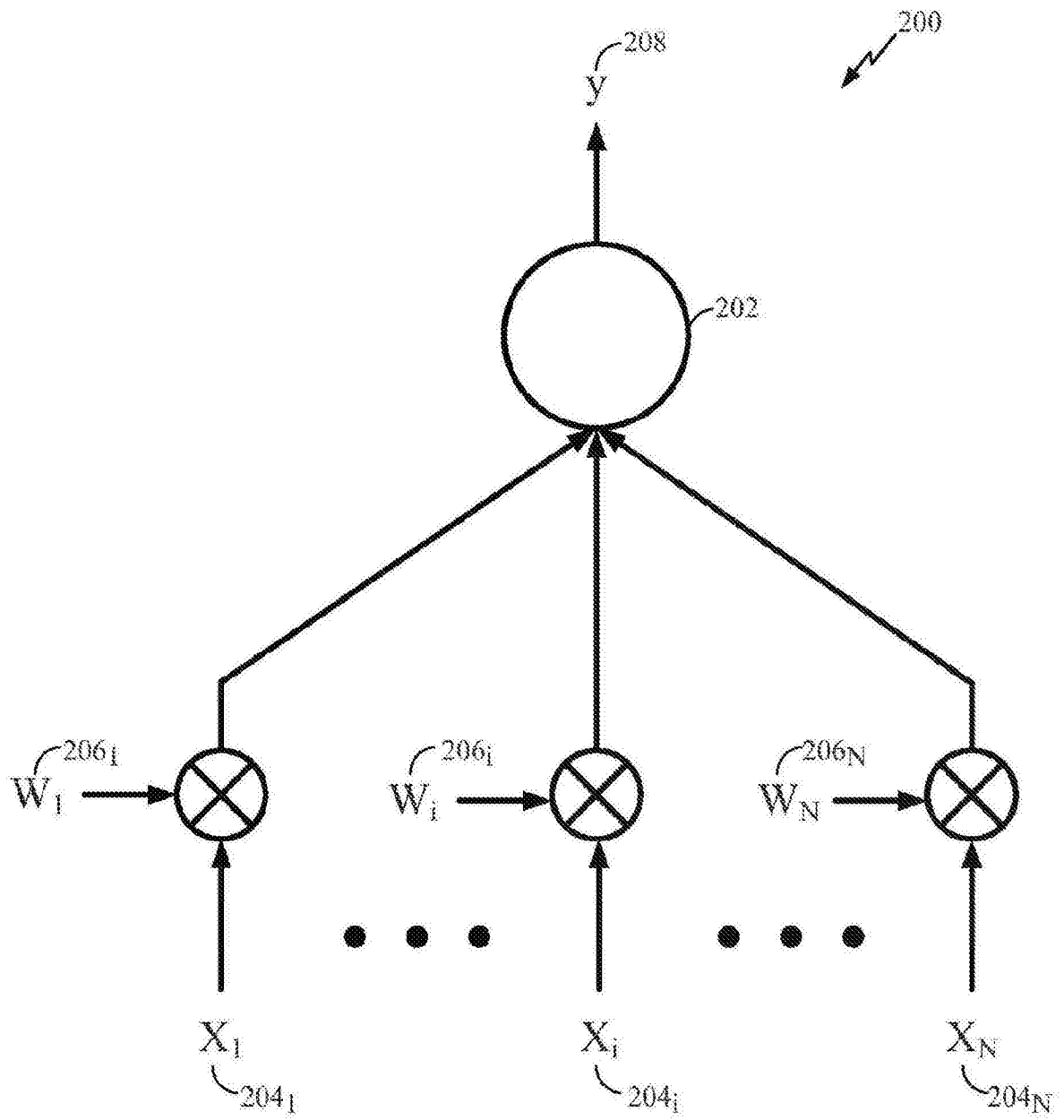


图2

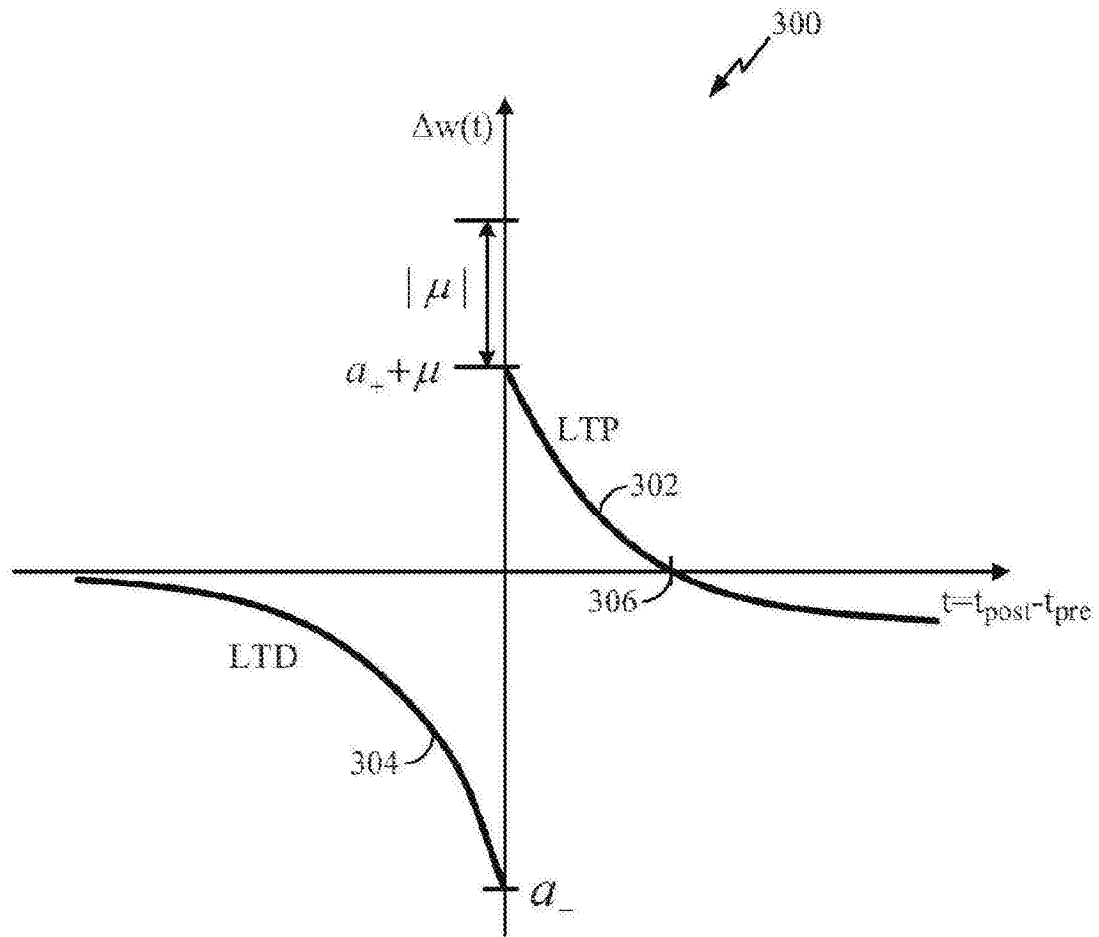


图3

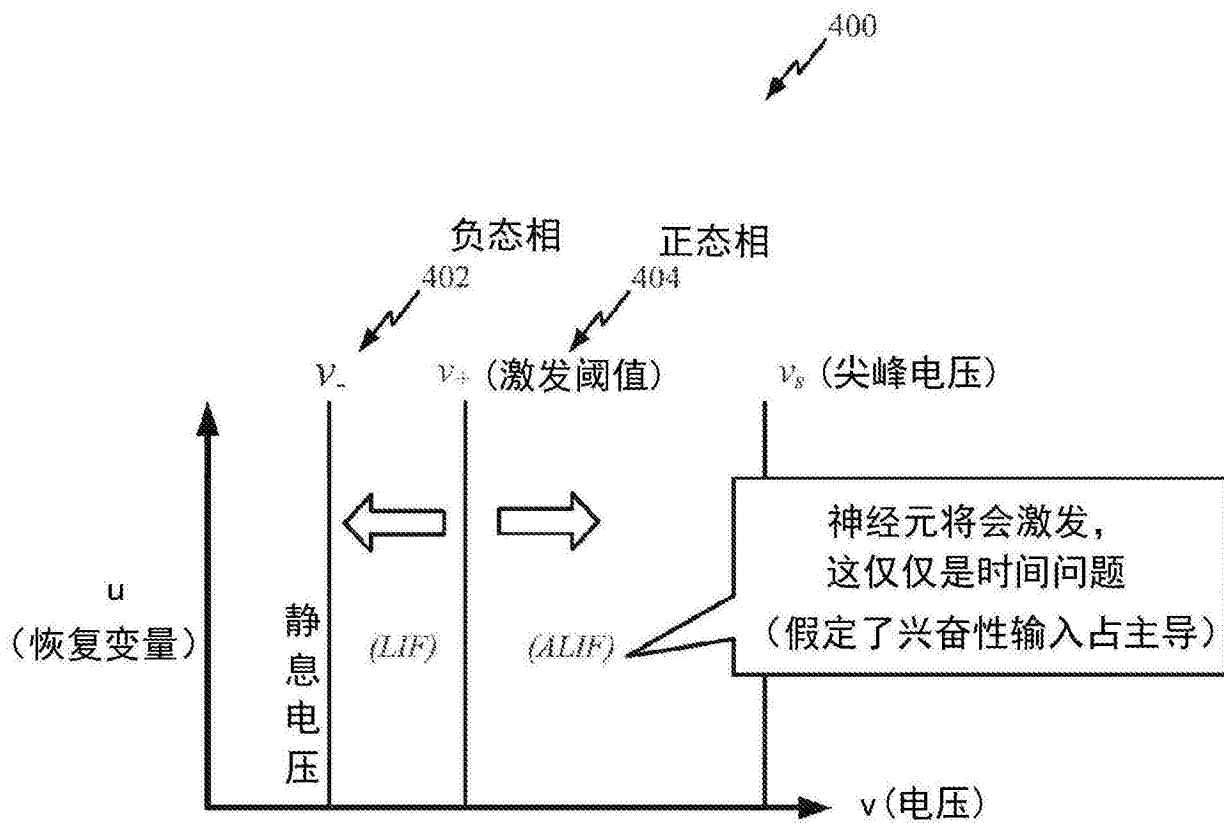


图4

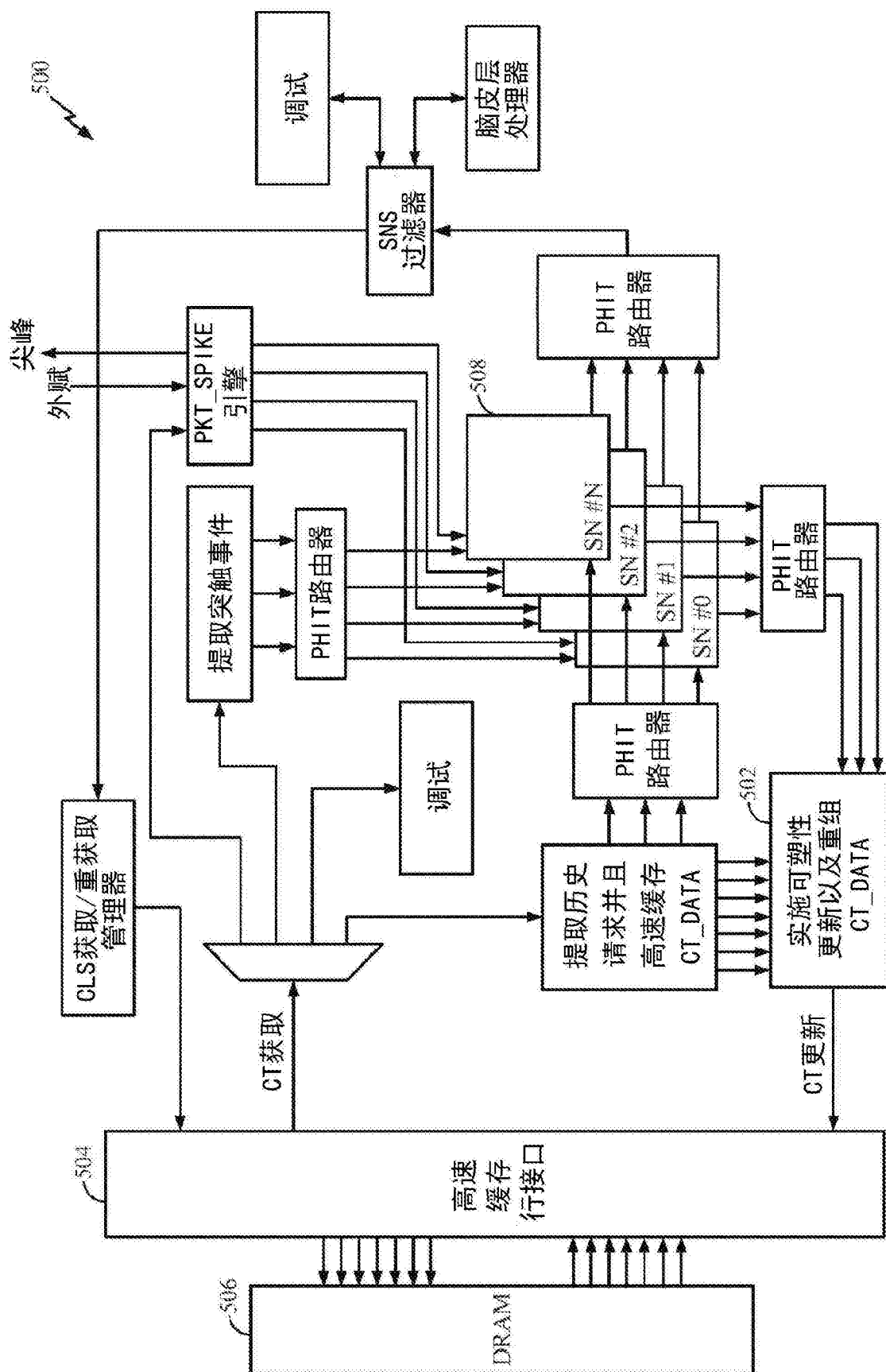


图5

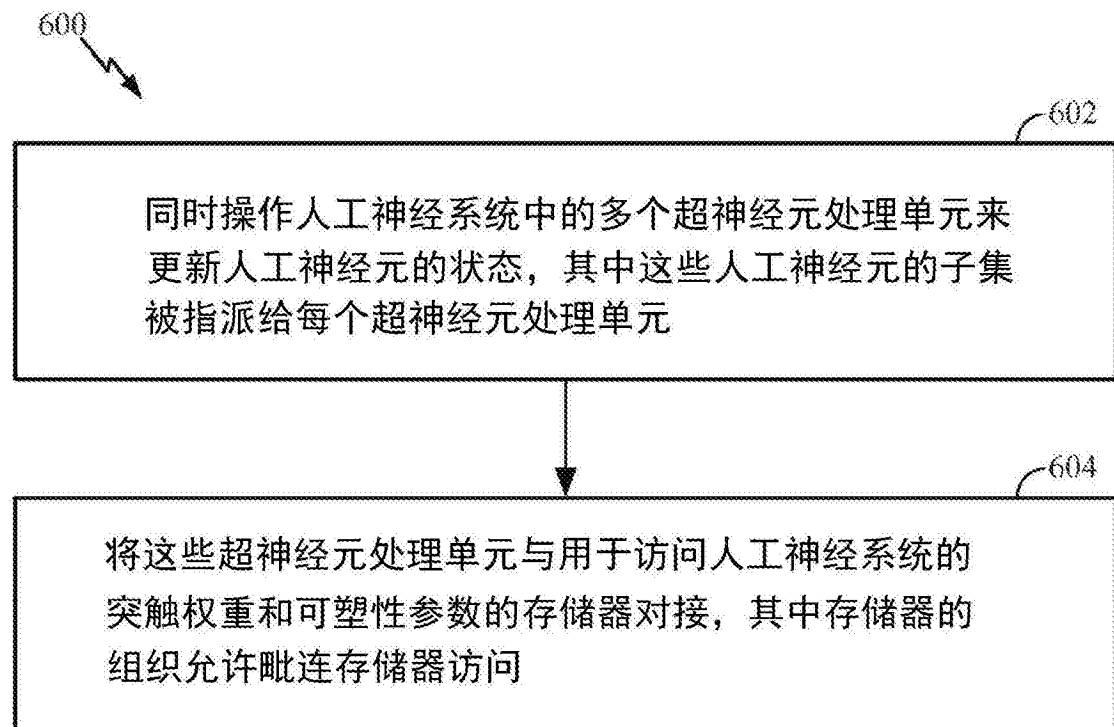


图6

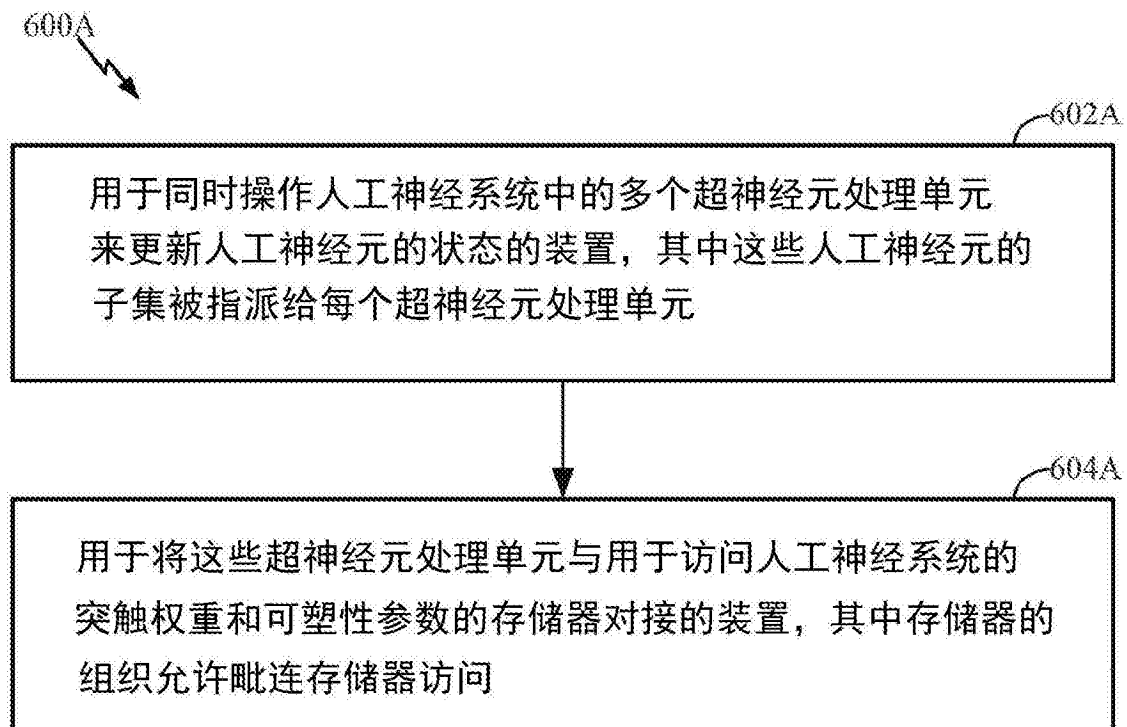


图6A

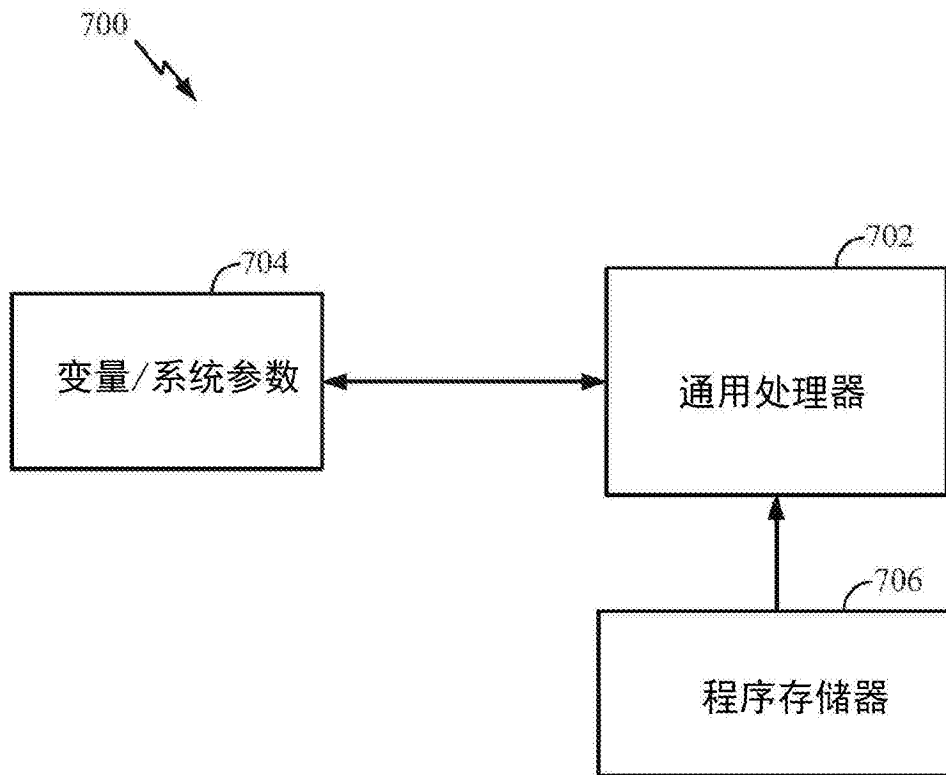


图7

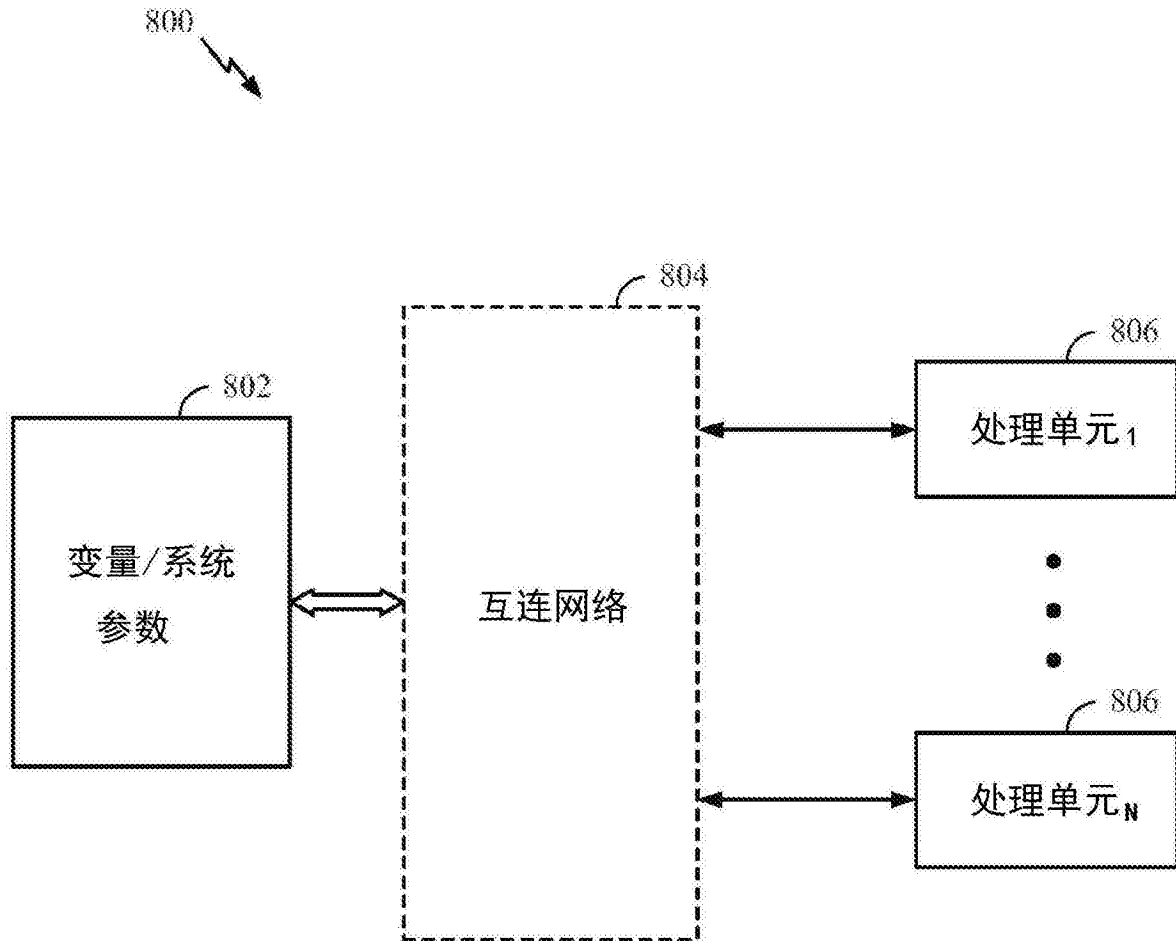


图8

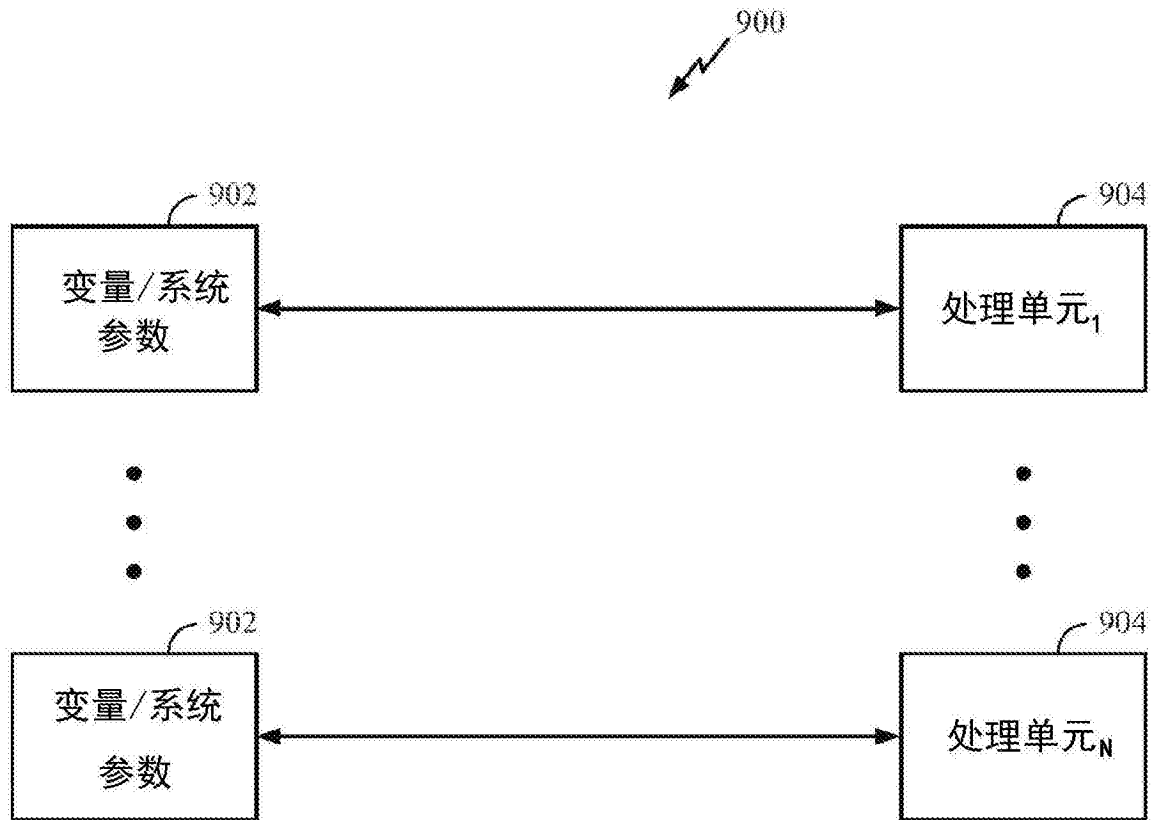


图9

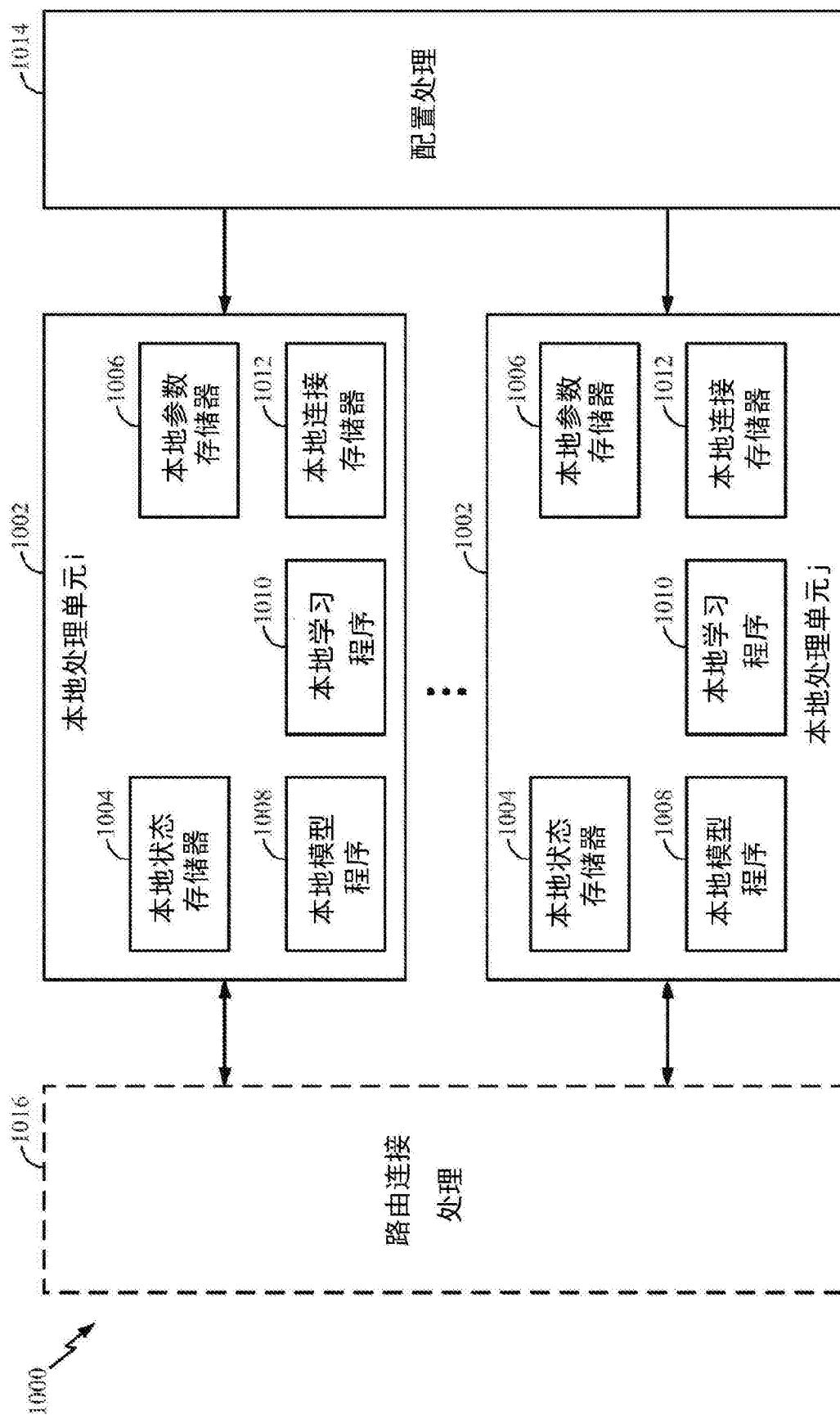


图10