



(12)发明专利

(10)授权公告号 CN 104781836 B

(45)授权公告日 2018.05.08

(21)申请号 201380059005.5

(22)申请日 2013.11.05

(65)同一申请的已公布的文献号
申请公布号 CN 104781836 A

(43)申请公布日 2015.07.15

(30)优先权数据

61/728,360 2012.11.20 US

61/734,716 2012.12.07 US

61/740,633 2012.12.21 US

61/756,889 2013.01.25 US

14/070,659 2013.11.04 US

(85)PCT国际申请进入国家阶段日
2015.05.12

(86)PCT国际申请的申请数据
PCT/US2013/068531 2013.11.05

(87)PCT国际申请的公布数据
W02014/081561 EN 2014.05.30

(73)专利权人 高通股份有限公司
地址 美国加利福尼亚州

(72)发明人 R·派多瓦尼 Y·C·尹

(74)专利代理机构 上海专利商标事务所有限公司 31100

代理人 李小芳

(51)Int.Cl.
G06N 3/04(2006.01)

(56)对比文件

CN 1417742 A,2003.05.14,

CN 1846218 A,2006.10.11,

EP 1172763 A2,2002.01.16,

US 4979126 A,1990.12.18,

Arnaud Tonnelier et al..“Piecewise linear differential equations and integrate-and-fire neurons: Insights from”.《The American Physical Society》.2003,

审查员 张丽娜

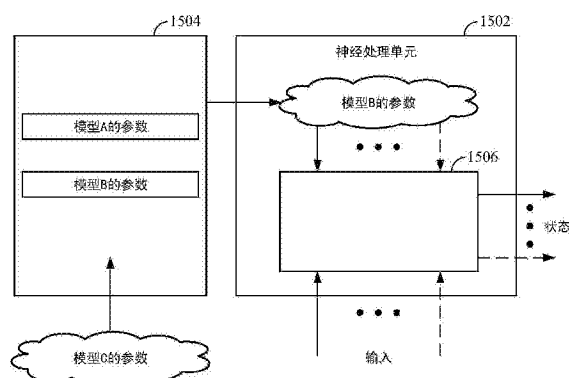
权利要求书3页 说明书46页 附图27页

(54)发明名称

分段线性神经元建模

(57)摘要

用于基于一个或多个线性化神经元模型在人工神经网络中进行分段线性神经元建模以及实现一个或多个个人工神经元的方法和装置。一种示例方法(用于在神经处理单元系统中实现多个神经元模型的组合)一般包括:将从多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元中;至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态;以及至少部分地基于第一神经元模型的参数并且基于第一状态来确定第一神经处理单元的第二状态。该方法还可以包括更新这多个神经元模型(例如,通过添加、删除或调整第一神经元模型或另一个神经元模型的参数)。



1. 一种用于在神经处理单元系统中实现多个神经元模型的组合的方法,包括:
将从所述多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元中;
至少部分地基于所述第一神经元模型的所述参数来确定所述第一神经处理单元的第一状态;以及
至少部分地基于所述第一神经元模型的所述参数并且基于所述第一状态来确定所述第一神经处理单元的第二状态;
将第二神经元模型的参数加载到所述第一神经处理单元中;以及
至少部分地基于所述第二神经元模型的所述参数和所述第二状态来确定所述第一神经处理单元的第三状态,
其中所述第一神经处理单元的所述第一状态、第二状态和第三状态是通过膜电位和恢复电流来定义的。
2. 根据权利要求1所述的方法,还包括更新所述多个神经元模型。
3. 根据权利要求1所述的方法,其中加载所述第二神经元模型的所述参数是响应于配置事件的。
4. 根据权利要求3所述的方法,其中所述配置事件包括对所述神经处理单元系统或所述第一神经处理单元中的至少一者上电。
5. 根据权利要求1所述的方法,其中所述第二神经元模型至少部分地被配置成产生振荡。
6. 根据权利要求1所述的方法,其中加载所述第二神经元模型的所述参数是至少部分地基于以下至少一者的:状态条件、时间条件或触发,其中所述触发位于所述第一神经处理单元的外部。
7. 根据权利要求1所述的方法,还包括:
将从所述多个神经元模型中选择的第二神经元模型的参数加载到第二神经处理单元中;以及
至少部分地基于所述第二神经元模型的所述参数来确定所述第二神经处理单元的状态。
8. 根据权利要求7所述的方法,其中所述第一神经元模型和所述第二神经元模型表示不同类型的神经元。
9. 根据权利要求1所述的方法,其中加载所述第一神经元模型的所述参数包括:从存储器获取所述第一神经元模型的所述参数的至少一部分。
10. 根据权利要求9所述的方法,其中所述存储器的至少一部分位于所述第一神经处理单元的本地。
11. 根据权利要求9所述的方法,其中所述存储器的至少一部分是高速缓存存储器。
12. 根据权利要求9所述的方法,其中加载所述第一神经元模型的所述参数还包括:在所述存储器中访问与所述第一神经元模型相关联的存储器块。
13. 根据权利要求12所述的方法,其中所述存储器块是由指向所述存储器块中的存储器地址的指针来指定的。
14. 根据权利要求1所述的方法,其中所述多个神经元模型包括以下至少一者:

Izhikevich简单模型、指数积分激发(EIF)模型、FitzHugh-Nagumo模型、四次模型或者固有电导模型。

15.一种用于在神经处理单元系统中实现多个神经元模型的组合的装置,包括:

处理系统,所述处理系统被配置成:

将从所述多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元中;

至少部分地基于所述第一神经元模型的所述参数来确定所述第一神经处理单元的第一状态;以及

至少部分地基于所述第一神经元模型的所述参数并且基于所述第一状态来确定所述第一神经处理单元的第二状态;

将第二神经元模型的参数加载到所述第一神经处理单元中;以及

至少部分地基于所述第二神经元模型的所述参数和所述第二状态来确定所述第一神经处理单元的第三状态,

其中所述第一神经处理单元的所述第一状态、第二状态和第三状态是通过膜电位和恢复电流来定义的;以及

耦合到所述处理系统的存储器。

16.根据权利要求15所述的装置,其中所述处理系统还被配置成更新所述多个神经元模型。

17.根据权利要求15所述的装置,其中所述处理系统被配置成:响应于配置事件而加载所述第二神经元模型的所述参数。

18.根据权利要求17所述的装置,其中所述配置事件包括对所述神经处理单元系统或所述第一神经处理单元中的至少一者上电。

19.根据权利要求15所述的装置,其中所述第二神经元模型至少部分地被配置成产生振荡。

20.根据权利要求15所述的装置,其中所述处理系统被配置成:至少部分地基于状态条件、时间条件或触发中的至少一者来加载所述第二神经元模型的所述参数,其中所述触发位于所述第一神经处理单元的外部。

21.根据权利要求15所述的装置,其中所述处理系统还被配置成:

将从所述多个神经元模型中选择的第二神经元模型的参数加载到第二神经处理单元中;以及

至少部分地基于所述第二神经元模型的所述参数来确定所述第二神经处理单元的状态。

22.根据权利要求21所述的装置,其中所述第一神经元模型和所述第二神经元模型表示不同类型的神经元。

23.根据权利要求15所述的装置,其中所述处理系统被配置成:通过从所述存储器获取所述第一神经元模型的所述参数的至少一部分来加载所述第一神经元模型的所述参数。

24.根据权利要求23所述的装置,其中所述存储器的至少一部分位于所述第一神经处理单元的本地。

25.根据权利要求23所述的装置,其中所述存储器的至少一部分是高速缓存存储器。

26. 根据权利要求23所述的装置, 其中所述处理系统被配置成: 通过在所述存储器中访问与所述第一神经元模型相关联的存储器块来加载所述第一神经元模型的所述参数。

27. 根据权利要求26所述的装置, 其中所述存储器块是由指向所述存储器块中的存储器地址的指针来指定的。

28. 根据权利要求15所述的装置, 其中所述多个神经元模型包括以下至少一者: Izhikevich简单模型、指数积分激发 (EIF) 模型、FitzHugh-Nagumo模型、四次模型或者固有电导模型。

29. 一种用于在神经处理单元系统中实现多个神经元模型的组合的装置, 包括:

用于将从所述多个神经元模型中选择的第一神经元模型的参数加载到神经处理单元中的装置;

用于至少部分地基于所述第一神经元模型的所述参数来确定所述神经处理单元的第一状态的装置; 以及

用于至少部分地基于所述第一神经元模型的所述参数并且基于所述第一状态来确定所述神经处理单元的第二状态的装置;

用于将第二神经元模型的参数加载到所述神经处理单元中的装置; 以及

用于至少部分地基于所述第二神经元模型的所述参数和所述第二状态来确定所述神经处理单元的第三状态的装置,

其中所述神经处理单元的所述第一状态、第二状态和第三状态是通过膜电位和恢复电流来定义的。

30. 一种用于在神经处理单元系统中实现多个神经元模型的组合的计算机程序产品, 包括计算机可读存储设备, 所述计算机可读存储设备具有可执行以进行以下操作的指令:

将从所述多个神经元模型中选择的神经元模型的参数加载到神经处理单元中;

至少部分地基于所述神经元模型的所述参数来确定所述神经处理单元的第一状态; 以及

至少部分地基于所述神经元模型的所述参数并且基于所述第一状态来确定所述神经处理单元的第二状态;

将第二神经元模型的参数加载到所述神经处理单元中; 以及

至少部分地基于所述第二神经元模型的所述参数和所述第二状态来确定所述神经处理单元的第三状态,

其中所述神经处理单元的所述第一状态、第二状态和第三状态是通过膜电位和恢复电流来定义的。

分段线性神经元建模

[0001] 根据35 U.S.C. §119的优先权要求

[0002] 本申请要求以下美国临时专利申请的权益：于2012年11月20日提交的题为“Piecewise Linear Neuron Modeling (分段线性神经元建模)”的美国临时专利申请S/N.61/728,360、于2012年12月7日提交的题为“Piecewise Linear Neuron Modeling (分段线性神经元建模)”的美国临时专利申请S/N.61/734,716、于2012年12月21日提交的题为“Piecewise Linear Neuron Modeling (分段线性神经元建模)”的美国临时专利申请S/N.61/740,633、以及于2013年1月25日提交的题为“Piecewise Linear Neuron Modeling (分段线性神经元建模)”的美国临时专利申请S/N.61/756,889,上述所有申请的全部内容通过援引纳入于此。

[0003] 背景

技术领域

[0004] 本公开的某些方面一般涉及人工神经系统,尤其涉及将神经元模型的非线性函数的至少一部分近似为分段线性函数,以及涉及在一个或多个人工神经元中使用结果所得的线性化神经元模型。

背景技术

[0005] 可包括一群互连的人工神经元(即神经元模型)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] 本公开的某些方面一般涉及将神经元模型的非线性函数的至少一部分近似为分段线性函数。例如,还提供了用于在一个或多个人工神经元中实现结果所得的线性化神经元模型的方法和装置。本公开的某些方面一般涉及用于实现神经元模型的动态的共用且灵活的架构。设计目标包括对动态的低复杂度、准确的建模,以及实现任何神经元模型(一维、二维或更多维)的能力。分段线性近似提供了在这种架构内仅通过替代与各种神经元模型

相关联的不同参数来改变神经元模型的简单方式。

[0009] 本公开的某些方面提供了一种用于操作人工神经元的方法。该方法一般包括：确定人工神经元的第一状态位于第一区域内；至少部分地基于第一组线性方程来确定人工神经元的第二状态，其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数；确定人工神经元的第二状态位于第二区域内；以及至少部分地基于第二组线性方程来确定人工神经元的第三状态，其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0010] 本公开的某些方面提供了一种用于操作人工神经元的装置。该装置一般包括处理系统和耦合到处理系统的存储器。处理系统一般被配置成：确定人工神经元的第一状态位于第一区域内；至少部分地基于第一组线性方程来确定人工神经元的第二状态，其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数；确定人工神经元的第二状态位于第二区域内；以及至少部分地基于第二组线性方程来确定人工神经元的第三状态，其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0011] 本公开的某些方面提供了一种用于操作人工神经元的设备。该设备一般包括：用于确定人工神经元的第一状态位于第一区域内的装置；用于至少部分地基于第一组线性方程来确定人工神经元的第二状态的装置，其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数；用于确定人工神经元的第二状态位于第二区域内的装置；以及用于至少部分地基于第二组线性方程来确定人工神经元的第三状态的装置，其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0012] 本公开的某些方面提供了一种用于操作人工神经元的计算机程序产品。该计算机程序产品一般包括计算机可读介质（例如，存储设备），该计算机可读介质具有可执行以进行以下操作的指令：确定人工神经元的第一状态位于第一区域内；至少部分地基于第一组线性方程来确定人工神经元的第二状态，其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数；确定人工神经元的第二状态位于第二区域内；以及至少部分地基于第二组线性方程来确定人工神经元的第三状态，其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0013] 本公开的某些方面提供了一种用于在神经处理单元系统中实现多个神经元模型的组合的方法。该方法一般包括：将从多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元中；至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态；以及至少部分地基于第一神经元模型的参数且基于第一状态来确定第一神经处理单元的第二状态。

[0014] 本公开的某些方面提供了一种用于在神经处理单元系统中实现多个神经元模型的组合的装置。该装置一般包括处理系统和耦合到处理系统的存储器。该处理系统一般被配置成：将从多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元中；至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态；以及至少部分地基于第一神经元模型的参数并且基于第一状态来确定第一神经处理单元的第二状态。

[0015] 本公开的某些方面提供了一种用于在神经处理单元系统中实现多个神经元模型的组合的装置。该装置一般包括：用于将从多个神经元模型中选择的第一神经元模型的参

数加载到第一神经处理单元中的装置;用于至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态的装置;以及用于至少部分地基于第一神经元模型的参数并且基于第一状态来确定第一神经处理单元的第二状态的装置。

[0016] 本公开的某些方面提供了一种用于在神经处理单元系统中实现多个神经元模型的组合的计算机程序产品。该计算机程序产品一般包括(非瞬态)计算机可读介质,该计算机可读介质具有可执行以进行以下操作的指令:将从多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元中;至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态;以及至少部分地基于第一神经元模型的参数并且基于第一状态来确定第一神经处理单元的第二状态。

[0017] 本公开的某些方面提供了一种用于操作人工神经元的方法。该方法一般包括:确定人工神经元的第一状态位于第一区域内;至少部分地基于第一组线性方程来确定人工神经元的第二状态,其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数;确定人工神经元的第二状态位于第二区域内,其中第一区域或第二区域中的至少一者是由两个或更多个维度来定义的;以及至少部分地基于第二组线性方程来确定人工神经元的第三状态,其中第二组线性方程是至少部分地基于与第二区域相对应的第二组参数。

[0018] 本公开的某些方面提供了一种用于操作人工神经元的装置。该装置一般包括处理系统和耦合到处理系统的存储器。该处理系统一般被配置成:确定人工神经元的第一状态位于第一区域内;至少部分地基于第一组线性方程来确定人工神经元的第二状态,其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数;确定人工神经元的第二状态位于第二区域内,其中第一区域或第二区域中的至少一者是由两个或更多个维度来定义的;以及至少部分地基于第二组线性方程来确定人工神经元的第三状态,其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0019] 本公开的某些方面提供了一种用于操作人工神经元的装置。该装置一般包括:用于确定人工神经元的第一状态位于第一区域内的装置;用于至少部分地基于第一组线性方程来确定人工神经元的第二状态的装置,其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数;用于确定人工神经元的第二状态位于第二区域内的装置,其中第一区域或第二区域中的至少一个区域是由两个或更多个维度来定义的;以及用于至少部分地基于第二组线性方程来确定人工神经元的第三状态的装置,其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0020] 本公开的某些方面提供了一种用于操作人工神经元的计算机程序产品。该计算机程序产品一般包括计算机可读介质(例如,存储设备或其它非瞬态介质),该计算机可读介质具有可执行的以进行以下操作的指令:确定人工神经元的第一状态位于第一区域内;至少部分地基于第一组线性方程来确定人工神经元的第二状态,其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数;确定人工神经元的第二状态位于第二区域内,其中第一区域或第二区域中的至少一个区域是由两个或更多个维度来定义的;以及至少部分地基于第二组线性方程来确定人工神经元的第三状态,其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。

[0021] 附图简述

[0022] 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。

[0023] 图1解说了根据本公开的某些方面的示例神经元网络。

[0024] 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)。

[0025] 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。

[0026] 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。

[0027] 图5A和图5B根据本公开的某些方面分别解说了用于将非线性的时变简单模型与基于泰勒展开法的示例性线性化进行比较的膜电压 v 和恢复电流 u 相对于时间的示例曲线图。

[0028] 图6A和图6B根据本公开的某些方面分别解说了Hunzinger Cold模型的阈下动态的膜电压 v 和恢复电流 u 相对于时间的示例曲线图。

[0029] 图7根据本公开的某些方面解说了对分段线性神经元建模的各种方法的概述。

[0030] 图8根据本公开的某些方面解说了具有就单个维度进行划分的分段线性神经元建模的示例。

[0031] 图9根据本公开的某些方面解说了具有由2个维度定义的3个矩形区域的广义线性神经元建模的示例。

[0032] 图10根据本公开的某些方面解说了具有由2个维度定义的4个区域的广义线性神经元建模的示例,其可以用于展示衰退、保持以及增长的阈下振荡行为。

[0033] 图11根据本公开的某些方面解说了具有不同形状的5个区域的广义线性神经元建模的示例。

[0034] 图12A-C根据本公开的某些方面分别解说了衰退、保持以及增长的阈下振荡的示例标绘。

[0035] 图13根据本公开的某些方面解说了具有由2个维度定义的6个区域的广义线性神经元建模的示例,其可以用于多级衰退、多级保持及/或多级增长区域来支持逐渐衰退的、多个保持的振荡及/或阈下振荡的逐渐增加。

[0036] 图14是根据本公开的某些方面的用于操作人工神经元的示例性操作的流程图。

[0037] 图14A解说了能够执行图14中所示的操作的示例装置。

[0038] 图15A-D根据本公开的某些方面解说了单个神经处理单元的共用且灵活的神经架构的实现,其中可以选择、加载、访问、添加、删除和/或更新神经元模型的参数。

[0039] 图16是根据本公开的某些方面的用于在神经处理单元系统中实现多个神经元模型的组合的示例操作的流程图。

[0040] 图16A解说了能够执行图16中示出的操作的示例装置。

[0041] 图17解说了根据本公开的某些方面的用于使用通用处理器来确定人工神经元的状态的示例实现。

[0042] 图18解说了根据本公开的某些方面的用于确定人工神经元的状态的示例实现,其中存储器可与个体分布式处理单元对接。

[0043] 图19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确定人工神经元的状态的示例实现。

[0044] 图20解说了根据本公开的某些方面的神经网络的示例实现。

[0045] 图21是根据本公开的某些方面的分段线性神经元建模的示例实现的框图,其中根据当前状态的量化从存储器获取用于更新人工神经元状态的参数。

[0046] 详细描述

[0047] 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于本公开通篇给出的任何具体结构或功能。确切而言,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种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。

[0048] 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。

[0049] 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。

[0050] 示例神经系统

[0051] 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。

[0052] 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。

[0053] 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。

[0054] 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节突触权重 $w_1^{(i,i+1)}$ 、...、 $w_P^{(i,i+1)}$ (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。

[0055] 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,而化学突触可调停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动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整形,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。

[0056] 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中的每个神经元(或神经元模型)都可以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。

[0057] 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。这种方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。

[0058] 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。

[0059] 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(x₁-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 w_{1-206N} (w_1-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。

[0060] 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。

[0061] 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202、其输入和输出连接也可由软件代码来仿真。处理单元202还可由电路来仿真,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202可包括模拟电路。在另一方面,处理单元202可以包括数字电路。在又一方面,处理单元202可以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。

[0062] 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 $w_1^{(i,j+1)}$ 、...、 $w_p^{(i,j+1)}$ 和/或来自图2的权重 w_{1-206N})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。

[0063] 突触类型

[0064] 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。

[0065] 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。

[0066] 突触可塑性的确定

[0067] 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。

[0068] 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。

[0069] 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。

[0070] 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)。

[0071] 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:

$$[0072] \quad \Delta w(t) = \begin{cases} a_+ e^{-t/k_+} + \mu, & t > 0 \\ a_- e^{t/k_-}, & t < 0 \end{cases}, \quad (1)$$

[0073] 其中 k_+ 和 k_- 分别是针对正和负时间差的时间常数, a_+ 和 a_- 是对应的比例缩放幅值,以及 μ 是可应用于正时间差和/或负时间差的偏移。

[0074] 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示图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。

[0075] 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 $i-1$ 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。

[0076] 神经元模型及操作

[0077] 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。

[0078] 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技能集合,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。

[0079] 在一方面,神经元 n 可被建模为尖峰带漏泄积分激发(LIF)神经元,其膜电压 $v_n(t)$ 由以下动态来支配:

$$[0080] \quad \frac{dv_n(t)}{dt} = \alpha v_n(t) + \beta \sum_m w_{m,n} y_m(t - \Delta t_{m,n}), \quad (2)$$

[0081] 其中 α 和 β 是参数, $w_{m,n}$ 是将突触前神经元 m 连接至突触后神经元 n 的突触的突触权重,以及 $y_m(t)$ 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才抵达神经元 n 的胞体。

[0082] 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来支配,即:

$$[0083] \quad \frac{dv}{dt} = (k(v - v_r)(v - v_t) - u + I)/C, \quad (3)$$

$$[0084] \quad \frac{du}{dt} = a(b(v - v_r) - u). \quad (4)$$

[0085] 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。

[0086] Hunzinger Cold模型

[0087] 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。

[0088] 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。

[0089] 线性双态相二维动态(对于状态 v 和 u)可按照惯例定义为:

$$[0090] \quad \tau_\rho \frac{dv}{dt} = v + q_\rho \quad (5)$$

$$[0091] \quad -\tau_u \frac{du}{dt} = u + r \quad (6)$$

[0092] 其中 q_ρ 和 r 是用于耦合的线性变换变量。

[0093] 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 。

[0094] 模型状态通过膜电位(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。

[0095] 态相依赖型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。

[0096] 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:

$$[0097] \quad q_\rho = -\tau_\rho \beta u - v_\rho \quad (7)$$

$$[0098] \quad r = \delta (v + \epsilon) \quad (8)$$

[0099] 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_+ 。

[0100] 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

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_p 时间常数参数不仅控制指数衰退,还单独地控制每个态相中的零倾线斜率。

[0101] 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:

$$[0102] \quad v = \hat{v}_- \quad (9)$$

$$[0103] \quad u = u + \Delta u \quad (10)$$

[0104] 其中 $\hat{v}_-$ 和 Δu 是参数。复位电压 $\hat{v}_-$ 通常被设为 v_- 。

[0105] 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:

$$[0106] \quad v(t + \Delta t) = (v(t) + q_p) e^{\frac{\Delta t}{\tau_p}} - q_p \quad (11)$$

$$[0107] \quad u(t + \Delta t) = (u(t) + r) e^{\frac{\Delta t}{\tau_u}} - r \quad (12)$$

[0108] 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。

[0109] 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:

$$[0110] \quad \Delta t = \tau_p \log \frac{v_f + q_p}{v_0 + q_p} \quad (13)$$

[0111] 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:

$$[0112] \quad \Delta t_s = \begin{cases} \tau_+ \log \frac{v_s + q_+}{v + q_+} & \text{若 } v > \hat{v}_+ \\ \infty & \text{否则} \end{cases} \quad (14)$$

[0113] 其中 $\hat{v}_+$ 通常被设为参数 v_+ ,但其他变型可以是可能的。

[0114] 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)变量可基于在上一个(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个(当前)事件的时间的状态来定义。

[0115] 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、步点-事件更新、以及步点-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。步点更新是以间隔(例如,1ms)来更新模型的情况下的更新。这不一定要求迭代方法或数值方法。通过仅在事件发生于步点处或步点间的情况下才更新模型或即通过“步点-事件”更新,基于事件的实现以有限的时间分辨率在基于步点的模拟器中实现也是可能的。

[0116] 神经编码

[0117] 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。

[0118] 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。

[0119] 抵达时间

[0120] 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。

[0121] 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。

[0122] 输入

[0123] 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。

[0124] 示例分段线性神经元建模

[0125] 数十年来都在寻找和研究神经元动态的数学模型。已经提出了各种神经元模型,其中这些模型的复杂度及与其生物对应物匹配的准确性各有不同。从根本上说,所有的神经元模型试图捕捉由于多种离子通道的相互作用而引起的细胞膜电压的非线性行为,并且具有共同的起点,即Hodgkin-Huxley在20世纪50年代的突破性工作所提供的数学描述。

[0126] 多年来,神经学家们在很大程度上向二维神经元模型集中,该二维神经元模型似乎在对他们寻求建模的生物细胞的测得行为进行复制的能力与可以对其进行分析和仿真的容易程度和速度之间提供了良好折衷。下面描述了全部经由一对微分方程来构建的最常见的二维模型。

[0127] 然而,常规神经元模型实现缺乏灵活性。通常,人工神经元的神经元模型的实现例如是基于预先选择特定的神经元模型。这种方法的缺点在于:实现完全不同的或者甚至稍

微修改的神经元模型的难度。

[0128] 本公开的某些方面利用了以下事实：各种动态神经元模型的微分方程可以被视为等效的，但是如上所述，就非线性函数 $F(v)$ 而言，试图捕捉细胞膜电压的非线性行为。在这种认识的情况下，一种方法包括：以有限量化区间对用于对神经元动态进行建模的微分方程中的非线性函数进行线性化，以提供分段线性近似。这种方法的优点包括能够相对容易地并且使用一般数学框架来获得连续时间或离散时间的动态的解，从而可以对任何神经元模型进行分析和仿真。这些解提供了与给定神经元模型的每个量化区间相对应的参数，从而使得本公开的某些方面的不同神经元模型的实现可能涉及对参数的简单替换。

[0129] 本公开的某些方面将该分段线性化方法应用到包括突触电流的函数。这种更广义的方法得到系统矩阵，该系统矩阵是时变突触电导（以及其它变量）的函数。本公开首先研究结果所得的分段线性时变系统的一些近似解。接着，本公开研究通过在给定时间区间内由常数近似时变电导来从先前的时变系统获得的分段线性时不变（LTI）系统。这种更精确的方法（具有包括突触电流的函数）导致具有大动态范围的系统矩阵将被量化过程覆盖，并且因此导致用于存储各种预先计算出的矩阵的相对较大存储器需求。该方法的好处在于对原始非线性时变系统的真正动态的更准确的近似。还提供了通过在动态范围的一部分上计算相关矩阵而不是存储相关矩阵来减轻存储器需求的方法。

[0130] 本公开的某些方面提供了一种用于实现支持任何一维、二维或更高维的神经元模型的通用架构的手段。使用这种灵活架构，可以根据需要来执行和替代各种适合的神经元模型中的任一种。例如，神经元模型可以包括以下模型中的至少一种：Izhikevich简单模型、指数积分激发（EIF）模型、FitzHugh-Nagumo模型、四次模型、或如上所述并且在2012年5月30日提交的题为“Dynamical Event Neuron and Synapse Models for Learning Spiking Neural Networks（用于学习尖峰神经网络的动态事件神经元和突触模型）”的美国专利申请S/N.13/483,811[代理人案卷号122024]（其通过援引纳入于此）中描述的Hunzinger Cold模型。可以使用本文中描述的分段线性近似来实现这些神经元模型。

[0131] 引言

[0132] 以描述对神经元动态进行建模的微分方程开始。虽然本描述关注二维神经元模型，但是也可以将该方法扩展到更高维的模型或者应用于一维模型。

$$[0133] \quad C \frac{d}{dt} v = F(v) - u + I \quad (15)$$

$$[0134] \quad \frac{d}{dt} u = a(b \cdot (v - v_r) - u) \quad (16)$$

[0135] 上面的等式定性地描述了人工神经元的动态（为了标示简化，省略了变量的时间依赖性）。这些等式是Hodgkin-Huxley四维模型到由变量 v 和 u 表示的二维模型的简化结果。变量 v 捕捉神经元膜电压和钠激活的行为，而 u 表示试图捕捉钾激活和钠失活的较慢行为的“适应”或“恢复”变量，从而将Hodgkin-Huxley的四个变量减少到两个变量。等式（15）中的变量 I 表示输入电流。更通用的二维模型可以是以下形式：

$$[0136] \quad C \frac{d}{dt} v = F(v, u) + I$$

[0137] $\frac{d}{dt}u = G(v, u)$

[0138] 其中这两个微分方程皆可以包含非线性项。这里,重点在于诸如等式(15)和(16)描述的那些模型,但是贯穿其中所开发的相同方法也可以应用于上面的描述。

[0139] 从根本上来讲,文献中已建议的最流行的神经元模型在等式(15)中的函数F(v)的选择上有所不同。一些示例包括由Izhikevich(也被称为简单模型)所提出的二次函数:

[0140] $F(v) = k(v - v_r)(v - v_t) \quad (17)$

[0141] 线性加指数函数,按照Brette和Gerstner:

[0142] $F(v) = g_L \cdot \Delta \cdot e^{\frac{v - v_L}{\Delta}} + g_L \cdot (E_L - v) \quad (18)$

[0143] 表示FitzHugh-Nagumo模型的线性加三次函数

[0144] $F(v) = v - \frac{1}{3}v^3 \quad (19)$

[0145] 按照Touboul和Gerstner,线性加四次项,被称为四次模型

[0146] $F(v) = 2a \cdot v + v^4 \quad (20)$

[0147] 以及最后可以被成为“固有电导”模型,被定义为:

[0148] $F(v) = G(v) \cdot v + p(v) \quad (21)$

[0149] 其中G(v)是分段常数函数(以电导为单位)并且p(v)也是分段常数函数(以电流为单位)。当G(v)和p(v)仅在两个区间上是分段常数时(如在上述Hunzinger Cold模型中),获得固有电导模型的最简单形式,对此,F(v)呈现以下形式:

[0150]
$$F(v) = \begin{cases} \frac{C}{\tau_-}(v - v_r) & \text{若 } v \leq v_t \Rightarrow G(v) = \frac{C}{\tau_-} \text{ 且 } p(v) = -\frac{C}{\tau_-}v_r \\ \frac{C}{\tau_+}(v - v_t) & \text{若 } v > v_t \Rightarrow G(v) = \frac{C}{\tau_+} \text{ 且 } p(v) = -\frac{C}{\tau_+}v_t \end{cases} \quad (22)$$

[0151] 因为除了等式(15)和(16)给出的描述之外还给出了重置条件,所以上面的所有模型(除了FitzHugh-Nagumo模型之外)被称为二维混合模型。例如,因为一旦跨越了电压阈值,变量v就将增长到无穷,所以在这些模型中提供了重置条件。因而可以使用以下重置条件:

[0152] 若 $v > v_{\text{peak}}$, 则 $v \rightarrow v_r$ 且 $u \rightarrow u + d \quad (23)$

[0153] 换句话说,当电压v超过尖峰发放阈值 v_{peak} (或者,对于某些方面来说,确定电压将超过 v_{peak})时,则将电压重置为静息值 v_r ,并且将恢复变量u重置为与当前值加常数d相等的值。对于某些方面来说,可以将u重置(reset)为预定的常数值(u_{reset}),而不是 $u + d$ 。根据某些方面,重置条件可以响应于控制信号的激活或接收而发生。在混合模型中的重置条件丰富了二维动态系统的可能行为。

[0154] 更一般地,通过突触电流 $I_{\text{SYN}}(t)$ 和通用电流 $I_{\text{ext}}(t)$ 的组合来对输入电流I(t)进行建模。突触电流具有以下形式:

[0155] $I_{\text{SYN}}(t) = \sum_i g_i(t) \cdot [E_i - v(t)] \quad (24)$

[0156] 在等式(24)中, $g_i(t)$ 指示特定信道(第i个信道)的依赖于时间的电导,并且 E_i 指示

针对该通道的逆转电位。虽然上述形式的突触电流足以描述实质上线性的电流-电压关系，但是在某些情况下(例如，N-甲基-D-天冬氨酸(NMDA)通道)，电导还是突触后膜电压的函数。在这种情况下，等式(24)呈现更加复杂的形式，即

$$[0157] \quad I_{\text{NMDA}}(t) = g_{\text{NMDA}}(t) \cdot h(v) [E_{\text{NMDA}} - v(t)] \quad (25)$$

[0158] 其中函数 $h(v)$ 捕捉对突触后电压的依赖性。将函数 $h(v)$ 建模成：

$$[0159] \quad h(v) \equiv 1 / (1 + \beta e^{-\alpha \cdot v}) \quad (26)$$

[0160] 其中参数 α 和 β 取以下值，例如： $\alpha = 0.062$ 且 $\beta = 1/3.57$ 。

[0161] 因此，最一般地，可以将多达 L 个不同的突触通道建模成：

$$[0162] \quad I_{\text{SYN}}(t) = \sum_{i=1}^L g_i(t) h_i(v(t)) (E_i - v(t)) \quad (27)$$

[0163] 其中

$$[0164] \quad h_i(v(t)) = \frac{1}{1 + \beta_i e^{-\alpha \cdot v(t)}} \quad \text{其中} \quad \begin{cases} \beta_i = \beta & \text{若 } i = \text{NMDA} \\ \beta_i = 0 & \text{否则} \end{cases} \quad (28)$$

[0165] 此外，可以通过简单指数函数、阿尔法(α)函数或者通过差分指数函数来对依赖于时间的电导 $g_i(t)$ 进行建模。在具有时间常数 τ 的简单衰退指数(其中 $\theta(t)$ 是Heaviside阶跃函数)的情况下，得到

$$[0166] \quad g_i(t) = g_i e^{-\frac{t}{\tau_i}} \theta(t) \quad (29)$$

[0167] 在阿尔法函数的情况下，得到

$$[0168] \quad g_i(t) = g_i \frac{t}{\tau_i} e^{-\frac{t}{\tau_i}} \theta(t) \quad (30)$$

[0169] 在差分指数函数的情况下，指数具有不同的上升(rise)和衰退(decay)时间常数，得到

$$[0170] \quad g_i(t) = \bar{g}_i \left(e^{-\frac{t}{\tau_{\text{decay}}}} - e^{-\frac{t}{\tau_{\text{rise}}}} \right) \theta(t) \quad (31)$$

[0171] 常数 $\bar{g}_i$ 包含正规化因子，从而使得峰值等于如下 g_i ：

$$[0172] \quad \bar{g}_i = \frac{g_i}{e^{-\frac{t_{\text{peak}}}{\tau_{\text{decay}}}} - e^{-\frac{t_{\text{peak}}}{\tau_{\text{rise}}}}} \quad (32)$$

[0173] 其中

$$[0174] \quad t_{\text{peak}} = \frac{\tau_{\text{decay}} \cdot \tau_{\text{rise}}}{\tau_{\text{decay}} - \tau_{\text{rise}}} \cdot \ln \left(\frac{\tau_{\text{decay}}}{\tau_{\text{rise}}} \right) \quad (33)$$

[0175] 在接下来的部分中，探讨了对动态系统的解的近似。由于系统是非线性的，所以不能找到精确的解。因此，当提到“精确的”解或者与“精确的”解相比较时(如示例1中)，意思是获得的数值解(例如，Runge-Kutta)，但具有高精度。

[0176] 公式化并推导

[0177] 在进行推导之前，介绍了一些定义和符号约定。此后，黑体大写字母表示矩阵，而

黑体小写字母表示向量。二维状态向量 $\mathbf{x}$ 可以被定义为

$$[0178] \quad \mathbf{x} \equiv \begin{bmatrix} v \\ u \end{bmatrix} \quad (34)$$

[0179] 第一状态变量是膜电压 v ，而第二状态变量是恢复变量 u 。现在针对最一般的情况，如下明确地重写等式(15)和(16)：

$$[0180] \quad C \frac{d}{dt} v(t) = F(v(t)) - u(t) + \sum_{i=1}^L g_i(t) h_i(v(t)) (E_i - v(t)) + I_{\text{ext}}(t) \quad (35)$$

$$[0181] \quad \frac{d}{dt} u(t) = abv(t) - au(t) - abv_r \quad (36)$$

[0182] 因此，在最一般的情况下，处理的是非线性时变(NLTV)二维动态系统，如图7中的702处所示。在下文中，在706处通过分段线性时变(pLTV)系统首先对此类系统进行近似，并且相继将其进一步近似为：(1)在710处的具有依赖于时间区间的常系数的分段线性时不变系统(pLTI)或(2)在714处的具有独立于时间区间的常系数的pLTI。

[0183] 分段线性时变系统

[0184] 首先将时间轴 $t \in [0, \infty]$ 划分为任意非重迭的时间区间，即

$$[0185] \quad \{t \in [T_n, T_{n+1}], n=0, 1, 2, 3, \dots\} \quad (37)$$

[0186] 在每个时间区间(例如，第 n 个区间)内，神经元行为由线性时变(LTV)系统来近似，该线性时变系统的参数是在区间的开始处(例如，在时刻 $t = T_n$ 处)从人工神经元的状态获得的。为了开发这种近似，首先定义

$$[0187] \quad \Gamma(v, t) \equiv \frac{1}{C} \left\{ F(v(t)) + \sum_{i=1}^L g_i(t) h_i(v(t)) (E_i - v(t)) + I_{\text{ext}}(t) \right\} \quad (38)$$

[0188] 并且等式(35)和(36)的系统变成

$$[0189] \quad \frac{d}{dt} v(t) = \Gamma(v, t) + a_{12} u(t) \quad (39)$$

$$[0190] \quad \frac{d}{dt} u(t) = a_{21} v(t) + a_{22} u(t) + b_2 \quad (40)$$

[0191] 下一步，使用在区间 $t \in [T_n, T_{n+1}]$ 上的近似仿射线性表达式来对函数 $\Gamma(v, t)$ 进行近似。该近似中的系数是电压在时刻 T_n 的函数。为了在明确时简化符号，表示 $v(T_n) \leftrightarrow v_n$ ，从而使得

$$[0192] \quad \Gamma(v, t) \cong a_{11}[v_n, t] v(t) + b_1[v_n, t] \quad (41)$$

[0193] 这些近似的一些示例包括泰勒展开法、平均斜率法、一阶线性内插法以及最小化 L_p 近似误差的最优线性内插法。对于泰勒展开法：

$$[0194] \quad \Gamma(v, t) \cong \Gamma(v(T_n), t) + \left. \frac{\partial \Gamma(v, t)}{\partial v} \right|_{v(T_n)} (v(t) - v(T_n))$$

$$[0195] \quad a_{11}[v_n, t] \equiv \left. \frac{\partial \Gamma(v, t)}{\partial v} \right|_{v_n} \quad \text{且} \quad (41a)$$

$$[0196] \quad b_1[v_n, t] \equiv \Gamma(v_n, t) - a_{11}[v_n, t]v_n$$

[0197] 在平均斜率法中,对于给定的电压步长 Δv_n ,平均斜率被计算如下:

$$[0198] \quad a_{11}[v(T_n), t] \equiv \frac{\Gamma(v(T_n) + \Delta v_n, t) - \Gamma(v(T_n), t)}{\Delta v_n} \text{ 且} \quad (41b)$$

$$[0199] \quad b_1[v_n, t] \equiv \Gamma(v_n, t) - a_{11}[v_n, t]v_n$$

[0200] 在一阶线性内插法中,将电压轴划分成区间, $v \in [V_k, V_{k+1}]$, $k=0, 1, 2, \dots$, 其中 $V_k \leq v(T_n) \leq V_{k+1}$ 。则

$$[0201] \quad a_{11}[v(T_n), t] \equiv \frac{\Gamma(V_{k+1}, t) - \Gamma(V_k, t)}{V_{k+1} - V_k} \text{ 且} \quad (41c)$$

$$[0202] \quad b_1[v_n, t] \equiv \Gamma(v_n, t) - a_{11}[v_n, t]v_n$$

[0203] 在最小化 L_p 近似误差的最优线性内插法中,如之前那样,将电压轴划分为区间, $v \in [V_k, V_{k+1}]$, $k=0, 1, 2, \dots$ 。然而,线性近似

$$[0204] \quad \hat{\Gamma}(v, t) = a_{11}[v_n, t]v(t) + b_1[v_n, t]$$

[0205] 被设计为使关于基于 L_p 平均数的原始函数的线性近似误差最小化。在区间 $v \in [V_k, V_{k+1}]$ 上的近似误差可以被定义为

$$[0206] \quad \varepsilon_{p,n,t}(a_{11}[v_n, t], b_1[v_n, t]) = \left[\int_{V_k}^{V_{k+1}} |\Gamma(v, t) - \hat{\Gamma}(v, t)|^p dv \right]^{\frac{1}{p}}$$

[0207] 其中

$$[0208] \quad \hat{\Gamma}(v, t) = a_{11}[v_n, t]v(t) + b_1[v_n, t]$$

[0209] 基于等式 (41), 用等式 (39) 和 (40) 描述的LTV系统现在可以被表示如下:

$$[0210] \quad \frac{d}{dt}v(t) = a_{11}[v_n, t]v(t) + a_{12}u(t) + b_1[v_n, t] \quad (42)$$

$$[0211] \quad \frac{d}{dt}u(t) = a_{21}v(t) + a_{22}u(t) + b_2 \quad (43)$$

[0212] 更简洁地,有如下矩阵方程:

$$[0213] \quad \dot{\mathbf{x}}(t) = \mathbf{A}(v_n, t)\mathbf{x}(t) + \mathbf{b}(v_n, t) \quad (44)$$

$$[0214] \quad \dot{\mathbf{x}}(t) = \begin{bmatrix} a_{11}[v_n, t] & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_1[v_n, t] \\ b_2 \end{bmatrix} \quad (45)$$

[0215] 其中 $a_{12} = -1C$, $a_{21} = ab$, $a_{22} = -a$, 并且 $b_2 = -abv_r$ 。

[0216] 对于 $t \in [T_n, T_{n+1}]$, 上面LTV系统的解可以被表示为

$$[0217] \quad \mathbf{x}(t) = \Phi(t, T_n)\mathbf{x}(T_n) + \int_{T_n}^t \Phi(t, \tau)\mathbf{b}(v_n, \tau) d\tau \quad (46)$$

[0218] 其中转换矩阵 $\Phi(t, T_n)$ 是由Peano-Baker方程给出的, 即

$$\begin{aligned}
\Phi(t, T_n) = & \mathbf{I} + \int_{T_n}^t \mathbf{A}(v_n, t_0) dt_0 + \int_{T_n}^t \mathbf{A}(v_n, t_0) \left[\int_{T_n}^{t_0} \mathbf{A}(v_n, t_1) dt_1 \right] dt_0 + \\
& \int_{T_n}^t \mathbf{A}(v_n, t_0) \left[\int_{T_n}^{t_0} \mathbf{A}(v_n, t_1) \left[\int_{T_n}^{t_1} \mathbf{A}(v_n, t_2) dt_2 \right] dt_1 \right] dt_0 + \dots
\end{aligned} \quad (47)$$

[0220] 针对一些特定的情况,对等式(47)中给出的级数进行了简化,例如:

[0221] (i) $\mathbf{A}(v_n, t)$ 是常数并且等于 $\mathbf{A}_n$ 。则 $\Phi(t, T_n)$ 采用矩阵指数的常见形式,即

$$[0222] \quad \Phi(t, T_n) = e^{\mathbf{A}_n(t-T_n)}$$

[0223] 其中矩阵 $\mathbf{A}$ 的矩阵指数被定义为

$$[0224] \quad e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \dots$$

[0225] (ii) 更一般的情况是当 $\mathbf{A}(v_n, t)$ 和 $\int_{T_n}^t \mathbf{A}(v_n, \tau) d\tau$ 针对任何 t 进行互换时。则

$$[0226] \quad \Phi(t, T_n) = \exp \left[\int_{T_n}^t \mathbf{A}(v_n, \tau) d\tau \right]$$

[0227] 以下条件中的任一个条件保证满足上面的互换性质: (a) $\mathbf{A}(v_n, t)$ 是常数; (b) $\mathbf{A}(v_n, t) = \mathbf{a}(t) \mathbf{M}$, 其中 $\mathbf{a}(t)$ 是标量函数并且 $\mathbf{M}$ 是常数矩阵; 或者 (c) $\mathbf{A}(v_n, t) = \sum \mathbf{a}_i(t) \mathbf{M}_i$, 其中 $\mathbf{a}_i(t)$ 是标量函数, 并且 $\mathbf{M}_i \mathbf{M}_j = \mathbf{M}_j \mathbf{M}_i$ 是对于任何 i, j 进行互换的常数矩阵。

[0228] (iii) 稍微更加普遍的情况是, 当针对所有 $t \in [T_n, T_{n+1}]$ 时, 对于某些整数 $k \geq 0$ 和 $s \geq 1$, 某些固定的时刻 $t_c \leq \tau \leq t$ 以及某些固定的矩阵 $\mathbf{A}_k(v_n, \tau, t_c)$ 和 $\mathbf{A}_{k+s}(v_n, \tau, t_c)$, 可以将矩阵 $\mathbf{A}(v_n, t)$ 表示为两个部分之和:

$$[0229] \quad \mathbf{A}(v_n, t) = \mathbf{A}_k(v_n, \tau, t_c) (t - t_c)^k + \mathbf{A}_{k+s}(v_n, \tau, t_c) \left[\frac{(t - t_c)^{k+1} - (\tau - t_c)^{k+1}}{k+1} \right]^s。$$

[0230] 时刻和矩阵在以下意义上是“固定的”: 其独立于时间参数 $t \in [T_n, T_{n+1}]$ 。可能最有用的示例是当 $k=0$ 并且 $s=1$ 或 2 时。在任何情况下, 若上面表达式中的第二项 (即, 涉及矩阵 $\mathbf{A}_{k+s}(v_n, \tau, t_c)$ 的项) 比针对所有 $t \in [T_n, T_{n+1}]$ 的第一项小得多, 则证明矩阵 $\Phi(t, T_n)$ 可以由以下无穷级数来近似:

$$\begin{aligned}
\Phi(t, T_n) = & \exp \left[\mathbf{A}_k(v_n, \tau, t_c) \frac{(t - t_c)^{k+1} - (\tau - t_c)^{k+1}}{k+1} \right] + \\
& \sum_{m=0}^{\infty} \mathbf{A}_k^m(v_n, \tau, t_c) \mathbf{A}_{k+s}(v_n, \tau, t_c) G_{m,k,s}(\mathbf{A}_k(v_n, \tau, t_c), (t - t_c), (\tau - t_c))
\end{aligned} \quad (48)$$

[0232] 其中定义了复数变量 z 和实数值参数 t 和 τ 的函数 $G_{m,k,s}(z, t, \tau)$ 。函数 $G_{m,k,s}(z, t, \tau)$ 在 $z=0$ 处是解析的, 并且由下式给出:

$$[0233] \quad G_{m,k,s}(z, t, \tau) \equiv \left\langle \frac{\partial^{k+s}}{\partial z^{k+s}} \left(z^{-(m+1)} \exp \left[z \frac{t^{k+1} - \tau^{k+1}}{k+1} \right] \right) \right\rangle$$

[0234] 其中符号 $\langle f(z) \rangle$ 表示 $f(z)$ 的解析分量。在上面的等式中,通过用指数项的幂级数表示来代替指数项、计算乘积和导数、随后丢掉与 z 的负幂相关联的项,来获得解析分量。

[0235] 针对时变矩阵 $A(v_n, t)$ 的所有其它情况,可以找到对等式(47)的转换矩阵的近似,随后可以获得等式(46)的近似解。

[0236] 分段线性时不变系统

[0237] 通过将等式(44)描述的线性时变(LTV)系统变换成线性时不变(LTI)系统,可以获得进一步的简化。为了做到这一点,系数 $a_{11}[v_n, t]$ 在区间 $t \in [T_n, T_{n+1}]$ 上保持恒定。可以通过多种方式获得该结果,下面描述其中几种方式。针对上面描述的仿射线性近似中的每一个,对于任何区间 $t \in [T_n, T_{n+1}]$,常系数可以被定义如下:

$$[0238] \quad a_{11}[v_n, t] \equiv a_{11}[v_n, T_n]$$

[0239] 或者,若下一时间步长 T_{n+1} 的值在时刻 $t = T_n$ 是已知的,则针对任何区间 $t \in [T_n, T_{n+1}]$,系数 $a_{11}[v_n, t]$ 的平均值可以被计算如下:

$$[0240] \quad a_{11}[v_n, t] \equiv \frac{1}{T_{n+1} - T_n} \int_{T_n}^{T_{n+1}} a_{11}[v_n, \tau] d\tau$$

[0241] 为了阐明泰勒展开法的示例,通过使用下式来获得LTI系统:

$$[0242] \quad a_{11}[v_n, T_n] \equiv \left. \frac{\partial \Gamma(v, T_n)}{\partial v} \right|_{v_n}$$

[0243] 或者

$$[0244] \quad a_{11}[v_n, T_n] \equiv \frac{1}{T_{n+1} - T_n} \int_{T_n}^{T_{n+1}} \left. \frac{\partial \Gamma(v, t)}{\partial v} \right|_{v_n} dt$$

[0245] 可以对上述方法中的任一种方法应用相同的求平均值方法。现在通过下面的矩阵方程来描述LTI系统:

$$[0246] \quad \dot{\mathbf{x}}(t) = \begin{bmatrix} a_{11}[v_n, T_n] & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} b_1[v_n, t] \\ b_2 \end{bmatrix} \quad (48)$$

[0247] 解可以被表示为

$$[0248] \quad \mathbf{x}(t) = \Phi(t, T_n) \mathbf{x}(T_n) + \int_{T_n}^t \Phi(T_n + t - \tau, T_n) \mathbf{b}(v_n, \tau) d\tau \quad (49)$$

[0249] 其中转换矩阵 $\Phi(t, T_n)$ 现在是矩阵指数

$$[0250] \quad \Phi(t, T_n) = e^{A(v_n, T_n)(t - T_n)} \quad (50)$$

[0251] 对于其中 $T_n = nT$ (即,长度为 T 的固定并均匀的时间区间)的情况,等式(49)变成

$$[0252] \quad \mathbf{x}(nT + T) = e^{A(v_n, nT)T} \mathbf{x}(nT) + \int_{nT}^{nT+T} e^{A(v_n, nT)(nT+T-\tau)} \mathbf{b}(v_n, \tau) d\tau \quad (51)$$

[0253] 总之,初始的通用模型是非线性的时变二维系统。通过在给定的时间区间上将线性近似应用到非线性函数 $\Gamma(v, t)$,首先将该非线性系统变换成线性时变系统。通过采用时变系统矩阵 $A(v_n, t)$ 并且在相同的给定区间上用常数矩阵 A_n 来逼近时变系统矩阵 $A(v_n, t)$,

将时变系统进一步转换成线性时不变(LTI)系统。

[0254] 解

[0255] 由于已经阐述了问题并且已经提出了对非线性函数的多个可能的仿射线性近似,则现在关注的是由等式(49)和(51)给出的在区间 $t \in [T_n, T_{n+1}]$ 上的LTI系统的解及其实现。进一步简化等式(49)中的符号并且表示 $A(v_n, T_n) = A_n$,

$$[0256] \quad \mathbf{x}(t) = e^{A_n(t-T_n)} \mathbf{x}(T_n) + \int_{T_n}^t e^{A_n(t-\tau)} \mathbf{b}(v_n, \tau) d\tau \quad (52)$$

[0257] 并且矩阵 $A(v_n, T_n)$ 和向量 $\mathbf{b}(v_n, t)$ 由下式提供:

$$[0258] \quad A_n = \begin{bmatrix} a_{11}[v_n, T_n] & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{b}(v_n, t) = \begin{bmatrix} b_1[v_n, t] \\ b_2 \end{bmatrix}$$

[0259] 其中通过上面描述的方法中的任何一种方法来计算系数 $a_{11}[v_n, T_n]$ 和 $b_1[v_n, t]$ 。

[0260] 下一步,推导出对泰勒展开法的解(相应地,后面其它方法中的任一种方法的解)。在这种情况下,具有

$$[0261] \quad a_{11}[v_n, T_n] \equiv \left. \frac{\partial \Gamma(v, T_n)}{\partial v} \right|_{v_n} \text{ 且 } b_1[v_n, t] \equiv \Gamma(v_n, t) - a_{11}[v_n, T_n] v_n \quad (53)$$

[0262] 并且

$$[0263] \quad \Gamma(v_n, t) \equiv \frac{1}{C} \left\{ F(v_n) + \sum_{i=1}^L g_i(t) h_i(v_n) (E_i - v_n) + I_{\text{ext}}(t) \right\} \quad (54)$$

$$[0264] \quad a_{11}[v_n, T_n] \equiv \left. \frac{\partial \Gamma(v, T_n)}{\partial v} \right|_{v_n} = \frac{1}{C} \left. \frac{\partial F(v)}{\partial v} \right|_{v_n} + \frac{1}{C} \sum_{i=1}^L g_i(T_n) \left[\left. \frac{\partial h_i(v)}{\partial v} \right|_{v_n} (E_i - v_n) - h_i(v_n) \right] \quad (55)$$

[0265] 简化表示:

$$[0266] \quad a_{11}[v_n, T_n] = \frac{1}{C} F'(v_n) + \frac{1}{C} \sum_{i=1}^L g_i(T_n) [h'_i(v_n) (E_i - v_n) - h_i(v_n)] \quad (56)$$

[0267] 将向量 $\mathbf{b}(v_n, t)$ 如下表示是方便的:

$$[0268] \quad \mathbf{b}(v_n, t) = \begin{bmatrix} b_1[v_n, t] \\ b_2 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} + \begin{bmatrix} b_1(t) \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{C} I_{\text{ext}}(t) \\ 0 \end{bmatrix}$$

[0269] 其中

$$[0270] \quad \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{C} \{ F(v_n) - a_{11}[v_n, T_n] v_n \} \\ b_2 \end{bmatrix} \text{ 且 } \begin{bmatrix} b_1(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{C} \left\{ \sum_{i=1}^L g_i(t) h_i(v_n) (E_i - v_n) \right\} \\ 0 \end{bmatrix}$$

[0271] 随后可以将等式(52)表示为:

$$[0272] \quad \mathbf{x}(t) = e^{A_n(t-T_n)} \mathbf{x}(T_n) + \int_{T_n}^t e^{A_n(t-\tau)} d\tau \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} + \int_{T_n}^t e^{A_n(t-\tau)} \begin{bmatrix} b_1(\tau) \\ 0 \end{bmatrix} d\tau + \int_{T_n}^t e^{A_n(t-\tau)} \begin{bmatrix} \frac{1}{C} I_{\text{ext}}(\tau) \\ 0 \end{bmatrix} d\tau \quad (57)$$

[0273] 可以用闭合形式来求解等式(57)的第一积分,从而得到

$$[0274] \quad \int_{T_n}^t e^{A_n(t-\tau)} d\tau \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} = A_n^{-1} (e^{A_n(t-T_n)} - I) \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} \quad (58)$$

[0275] 其中I是2x2单位矩阵。

[0276] 现在可以将等式(52)表示如下：

[0277]

$$x(t) = e^{A_n(t-T_n)} x(T_n) + A_n^{-1} (e^{A_n(t-T_n)} - I) \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} + \int_{T_n}^t e^{A_n(t-\tau)} b_1(\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \int_{T_n}^t e^{A_n(t-\tau)} \frac{1}{C} I_{\text{ext}}(\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (59)$$

[0278] 若外部电流是使得无法获得等式(59)的最后一个积分的闭合形式的表达式，则可能不得不对其进行近似。近似的一个示例是零保持，即

[0279]

$$\int_{T_n}^t e^{A_n(t-\tau)} \frac{1}{C} I_{\text{ext}}(\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cong \frac{1}{C} I_{\text{ext}}(T_n) \int_{T_n}^t e^{A_n(t-\tau)} d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{C} I_{\text{ext}}(T_n) A_n^{-1} (e^{A_n(t-T_n)} - I) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (60)$$

[0280] 在这种情况下，针对任意区间 $t \in [T_n, T_{n+1}]$ 的闭合形式解采用以下形式：

[0281]

$$x(t) = e^{A_n(t-T_n)} x(T_n) + A_n^{-1} (e^{A_n(t-T_n)} - I) \begin{bmatrix} b_0 + \frac{1}{C} I_{\text{ext}}(T_n) \\ b_2 \end{bmatrix} + \int_{T_n}^t e^{A_n(t-\tau)} b_1(\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (61)$$

[0282] 等式(61)中的关键部分是矩阵指数 $e^{A_n t}$ ，其可以被表示如下。令 λ_{n1} 和 λ_{n2} 是2x2矩阵 A_n 的特征值。则

$$[0283] \quad e^{A_n t} = e^{\lambda_{n1} t} I + \frac{e^{\lambda_{n1} t} - e^{\lambda_{n2} t}}{\lambda_{n1} - \lambda_{n2}} (A_n - \lambda_{n1} I) \quad \text{对于 } \lambda_{n1} \neq \lambda_{n2} \text{ 实数} \quad (62)$$

$$[0284] \quad e^{A_n t} = e^{\lambda_{n1} t} I + t e^{\lambda_{n1} t} (A_n - \lambda_{n1} I) \quad \text{对于 } \lambda_{n1} = \lambda_{n2} \text{ 实数} \quad (63)$$

$$[0285] \quad e^{A_n t} = e^{\alpha t} \cos \beta t I + \frac{e^{\alpha t} \sin \beta t}{\beta} (A_n - \alpha I) \quad \text{对于 } \lambda_{n1} = \lambda_{n2}^* = \alpha + i\beta \quad \beta > 0 \quad (64)$$

[0286] 其中 λ_{n1} 和 λ_{n2} 是特征方程 $\det(A_n - \lambda I) = 0$ 的解。

[0287] 等式(62) - (64) 提供了针对矩阵指数的明确表达式。或者，大量的系数数值算法可用于计算矩阵指数。因此，通过任何一种方法来确定系统在时刻 $t = T_{n+1}$ 的状态是可行的。但在硬件实现中，使用如接下来描述的不同方法可能是有利的。

[0288] 首先，注意在矩阵 A_n 中，仅有一个系数 $a_{11}[v_n, T_n]$ 随着系统演变而发生变化；其余的三个系数始终保持不变。例如，系数由等式(56)给出。示例的高效程序如下：(1) 每次对第一状态变量（即，膜电压 v_n ）进行量化从而变到有限数量的值，并且(2) 使用预先计算出的函数 $F'(v_n)$ ， $h'_i(v_n)$ 以及 $h_i(v_n)$ 的值来为查找表编索引。类似地，也可以对时刻进行量化，并且可以将 $g_i(T_n)$ 的预先计算出的值存储在查找表中。现在可以使用检索到的值经由等式(56)来求 $a_{11}[v_n, T_n]$ 的值，并且将结果量化为有限数量的值 $Q(a_{11}[v_n, T_n])$ 。

[0289] 类似地，可以对 ΔT_n 的多个选择进行预选择，并且可以用 $Q(a_{11}[v_n, T_n])$ 和 ΔT_n 来对等式(62)中出现的矩阵的矩阵指数和逆矩阵进行预计算和处理。

[0290] 自适应和固定的时间步长

[0291] 在前面的部分中,推导出了在任意时刻的针对LTI系统的状态的表达式。下面更详细地描述了其中自适应地选择时间步长的情形以及同步系统的较简单的情形(其中时刻是均匀的,并且具有固定并预先确定的步长大小 $\Delta T_n = T$)。

[0292] 自适应策略可以有利于关于复杂度和精度两个方面。例如,当状态正在缓慢地演变时,可以选择间隔更远(即,较大 ΔT_n)的时间步长,从而提高计算的速度。相反,当例如神经元模型靠近尖峰发放事件时,可以使时间步长较小,从而提高计算尖峰时间的精度。接着,描述了用于选择时间步长的可能算法。

[0293] 如等式(63)所示,系统演变的时间常数与矩阵 A_n 的特征值成反比。特别地,若特征值接近于零,则时间常数较大,并且系统缓慢地演变。相反,若特征值中的至少一个特征值变大(并且可能为正),则系统非常迅速地演变。完成上述任务的用于选择时间步长的一个示例算法如下:

$$[0294] \quad T_{n+1} = T + \min\left(\frac{\ln \mu_n}{\lambda_{n,\max}}, \Delta T_{\max}\right) \quad (65)$$

[0295] 在等式(65)中, $\lambda_{n,\max}$ 表示矩阵 A_n 的最大特征值(就幅度而言),而 $\Delta T_{\max}$ 和 μ_n 是可配置参数。这保证以下关系成立:

$$[0296] \quad \lambda_{n,\max} \leq |a_{11}[v(T_n), T_n]| + |a_{22}|, T_{n+1} - T_n \leq \Delta T_{\max} \quad (66)$$

[0297] 并且对于任意区间 $t \in [T_n, T_{n+1}]$ 来说

$$[0298] \quad ||x(t)|| \leq \mu_n ||x(T_n)|| \quad (67)$$

[0299] 接下来,针对大小为 T 的固定并均匀的时间步长的重要情形,对前面部分的通用解进行专门化。以等式(61)开始,得到

[0300]

$$x(nT + T) = e^{A_n T} x(nT) + A_n^{-1} (e^{A_n T} - I) \left[\begin{array}{c} b_0 + \frac{1}{C} I_{\text{ext}}(nT) \\ b_2 \end{array} \right] + \int_0^T e^{A_n t} b_1 (nT + T - t) dt \left[\begin{array}{c} 1 \\ 0 \end{array} \right] \quad (68)$$

[0301] 示例1:Izhikevich(简单)模型的泰勒展开

[0302] 在该示例中,基于泰勒展开法,针对简单模型和线性化求等式(68)的值。为了使公式有所简化,假定不存在外部电流。

[0303] 首先,从等式(56)开始推导出简单模型的泰勒展开(即, $F(v) = k(v - v_r)(v - v_t)$)的系数。假如

$$[0304] \quad \left. \frac{1}{C} \frac{\partial F(v)}{\partial v} \right|_{v_n} = \frac{2k}{C} v_n - \frac{k}{C} (v_r + v_t) \quad (69)$$

[0305] 则得到

$$[0306] \quad a_{11}[v_n, nT] = \frac{2k}{C} v_n - \frac{k}{C} (v_r + v_t) + \frac{1}{C} \sum_{i=1}^L g_i(nT) [h'_i(v_n)(E_i - v_n) - h_i(v_n)] \quad (70)$$

[0307]

$$b_1[v_n, t] = \Gamma(v_n, t) - a_{11}[v_n, nT]v_n = \frac{1}{C} \left\{ k(v_n - v_r)(v_n - v_t) + \sum_{i=1}^L g_i(t)h_i(v_n)(E_i - v_n) \right\} - \left\{ \frac{2k}{C}v_n^2 - \frac{k}{C}(v_r + v_t)v_n + \frac{1}{C} \sum_{i=1}^L g_i(nT)[h'_i(v_n)(E_i - v_n) - h_i(v_n)]v_n \right\} \quad (71)$$

[0308] 对于非NMDA突触通道来说,等式(70)和(71)简化为

$$a_{11}[v_n, nT] = \frac{2k}{C}v_n - \frac{k}{C}(v_r + v_t) - \frac{1}{C} \sum_{i=1}^L g_i(nT) \quad (72)$$

$$b_1[v_n, nT] = \frac{1}{C} \left\{ -kv_n^2 + kv_rv_t + \sum_{i=1}^L g_i(nT)v_n \right\} + \frac{1}{C} \sum_{i=1}^L g_i(t)(E_i - v_n) \quad (73)$$

[0311] 若

$$b_0 = \frac{1}{C} \left\{ -kv_n^2 + kv_rv_t + \sum_{i=1}^L g_i(nT)[h'_i(v_n)(E_i - v_n) - h_i(v_n)]v_n \right\}$$

[0313] 则现在等式(68)变成

[0314]

$$\mathbf{x}(nT+T) = e^{\mathbf{A}_n T} \mathbf{x}(nT) + \mathbf{A}_n^{-1} (e^{\mathbf{A}_n T} - \mathbf{I}) \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} + \sum_{i=1}^L [h_i(v_n)(E_i - v_n)] \int_0^T e^{\mathbf{A}_n \tau} g_i(nT+T-\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (74)$$

[0315] 其中

$$\mathbf{A}_n = \begin{bmatrix} \frac{2k}{C}v_n - \frac{k}{C}(v_r + v_t) + \frac{1}{C} \sum_{i=1}^L g_i(nT)[h'_i(v_n)(E_i - v_n) - h_i(v_n)] & -\frac{1}{C} \\ ab & -a \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{C} \left\{ -kv_n^2 + kv_rv_t + \sum_{i=1}^L g_i(nT)[h'_i(v_n)(E_i - v_n) - h_i(v_n)]v_n \right\} \\ -abv_r \end{bmatrix}$$

[0318] 再一次对于非NMDA突触电流来说,等式(74)简化为

$$\mathbf{x}(nT+T) = e^{\mathbf{A}_n T} \mathbf{x}(nT) + \mathbf{A}_n^{-1} (e^{\mathbf{A}_n T} - \mathbf{I}) \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} + \sum_{i=1}^L E_i \int_0^T e^{\mathbf{A}_n \tau} g_i(nT+T-\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (75)$$

[0320] 其中

$$\mathbf{A}_n = \begin{bmatrix} \frac{2k}{C}v_n - \frac{k}{C}(v_r + v_t) - \frac{1}{C} \sum_{i=1}^L g_i(nT) & -\frac{1}{C} \\ ab & -a \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{C} \left\{ -kv_n^2 + kv_rv_t + \sum_{i=1}^L g_i(nT)v_n \right\} \\ -abv_r \end{bmatrix}$$

[0323] 此外,可以以封闭形式来求解等式(74)和(75)中的积分。例如,若

[0324] $g_i(t) = g_i e^{-t/\tau_i} \theta(t)$

[0325] 则

[0326]
$$\int_0^T e^{A_n \tau} g_i(nT+T-\tau) d\tau = g_i(nT+T) (e^{A_n} + \tau_i^{-1} \mathbf{I})^{-1} (e^{(A_n + \tau_i^{-1} \mathbf{I})T} - \mathbf{I}) \quad (76)$$

[0327] 为了完成该示例,针对具有参数 $k=0.7$ 、 $C=100$ 、 $a=0.03$ 、 $b=-2$ 、 $v_r=-60\text{mV}$ 以及 $v_t=-40\text{mV}$ 的慢兴奋神经元模型以及具有参数 $\tau_{\text{AMPA}}=5\text{ms}$ 、 $E_{\text{AMPA}}=0\text{mV}$ 和 $g_{\text{AMPA}}=5$ 的单个指数AMPA通道,用 $T=1\text{ms}$ 对等式(75)的准确性进行测试。如图5A和图5B的膜电位标绘500和恢复电流标绘520中分别所示,基于等式(75)的线性化准确地跟踪了基于非线性的时变模型的标绘。

[0328] 示例2:Hunzinger Cold模型的阈下动态

[0329] 开发了另一个示例以致力于检查被称为Hunzinger Cold模型的简单固有电导模型的阈下动态。在这个示例中,假定没有突触电流,并且假定简单但有趣的脉冲外部电流的情况。

[0330] 在Hunzinger Cold模型中(见上面的等式(22)),当膜电压低于阈值时,矩阵 A_n 为常数并且等于 A_- :

[0331]
$$A_- \equiv \begin{bmatrix} -1/\tau_- & -1/C \\ ab & -a \end{bmatrix}$$

[0332] 若第一状态变量被定义为膜电压减去参考电压,则推导可以被进一步简化。在这种定义的情况下,所有常数项等于0,并且等式(61)被简化为

[0333]
$$\mathbf{x}(t) = e^{A_-(t-T_n)} \mathbf{x}(T_n) + \frac{1}{C} \int_{T_n}^t e^{A_-(t-\tau)} I_{\text{ext}}(\tau) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (77)$$

[0334] 此外,若假定外部电流是在时刻 T_n 、具有幅度 I 的狄拉克 δ 函数,即, $I_{\text{ext}}(t) = I\delta(t-T_n)$, 则

[0335]
$$\mathbf{x}(t) = e^{A_-(t-T_n)} \left(\mathbf{x}(T_n) + \frac{1}{C} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \quad (78)$$

[0336] 注意脉冲输入与初始条件对系统具有相同的效果。若特征值是实数并且 $\lambda_1 \neq \lambda_2$, 则可以将等式(62)代入等式(78)以获得下面的针对 $v < v_t$ 的连续时间解:

[0337]
$$\mathbf{x}(t) = \left[e^{\lambda_1(t-T_n)} \mathbf{I} + \frac{e^{\lambda_1(t-T_n)} - e^{\lambda_2(t-T_n)}}{\lambda_1 - \lambda_2} (A_- - \lambda_1 \mathbf{I}) \right] \left(\mathbf{x}(T_n) + \frac{I}{C} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \quad (79)$$

[0338] 通过定义新的矩阵

[0339]
$$\overline{A}_- = \frac{(A_- - \lambda_1 \mathbf{I})}{\lambda_1 - \lambda_2}$$

[0340] 并且表示

[0341]
$$\mathbf{x}(T_n) + \frac{I}{C} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} v_n \\ u_n \end{bmatrix}$$

[0342] 可以将等式 (72) 写为

$$[0343] \quad \mathbf{x}(t) = \left[e^{\lambda_1(t-T_n)} (\mathbf{I} + \overline{\mathbf{A}}_-) - e^{\lambda_2(t-T_n)} \overline{\mathbf{A}}_- \right] \begin{bmatrix} v_n \\ u_n \end{bmatrix} \quad (80)$$

[0344] 尤其针对 $T_n=0$, 得到

$$[0345] \quad \mathbf{x}(t) = \left[e^{\lambda_1 t} (\mathbf{I} + \overline{\mathbf{A}}_-) - e^{\lambda_2 t} \overline{\mathbf{A}}_- \right] \begin{bmatrix} v_0 + I/C \\ u_0 \end{bmatrix} \quad (81)$$

[0346] 图6A和图6B的膜电位标绘600和恢复电流标绘620分别针对以下参数值示出等式 (81) 的时间演变: $\tau_r=14\text{ms}$ 、 $C=100$ 、 $a=0.03$ 、 $b=-2$ 、 $I_{\text{ext}}(t)=100\delta(t)$, 并且 $v_0=u_0=0$ 。由以下函数给出膜电压和恢复电流:

$$[0347] \quad v(t) - v_r = 0.8228e^{-0.0827t} + 0.1771e^{-0.0186t} \quad (82)$$

$$[0348] \quad u(t) = 0.9351e^{-0.0827t} - 0.9351e^{-0.0186t} \quad (83)$$

[0349] 由于等式 (80) 的计算中涉及的矩阵和特征值是先验知道的, 因此可以通过以给定的时间分辨率预先计算并 (例如, 在查找表中) 存储两个指数函数来很容易地、并且以期望的精度计算出系统的时间过程。或者, 可以使用针对指数函数的计算 (近似) 的快速算法, 并且可以实时地计算等式 (80)。

[0350] 上面的构建在基于事件的实现中特别有关系, 在基于事件的实现中, 关心的是计算在事件之时的状态向量。例如, 给定等式 (80) 以及在时刻 T_{n+1} 发生的事件, 可能希望计算 T_{n+1}^- 。在 $\Delta_n \equiv T_{n+1} - T_n$ 的情况下, 有

$$[0351] \quad \mathbf{x}(T_{n+1}^-) = \left[e^{\lambda_1 \Delta_n} (\mathbf{I} + \overline{\mathbf{A}}_-) - e^{\lambda_2 \Delta_n} \overline{\mathbf{A}}_- \right] \begin{bmatrix} v_n \\ u_n \end{bmatrix} \quad (84)$$

[0352] 一个重要的事件是尖峰时间 (即, 如上所述, 人工神经元的膜电压超过峰值阈值的时间)。当人工神经元的模型接近尖峰时间时, 这些特征值中的至少一个特征值为正并且占主导地位。因此, 可以通过仅保留电压表达式中的主导项来达到良好近似。换句话说, 由于函数 $v(t)$ 是等式 (82) 中所示的形式, 即

$$[0353] \quad v(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \quad (85)$$

[0354] $v(t)$ 可以由其主导项来近似, 即,

$$[0355] \quad v(t) \cong c_1 e^{\lambda_1 t} + c_2 \quad (86)$$

[0356] 可以容易地根据下式来获得尖峰时间:

$$[0357] \quad t_{\text{spike}} \cong \frac{1}{\lambda_1} \ln((v_{\text{peak}} - c_2)/c_1) \quad (87)$$

[0358] 额外的简化

[0359] 题为“解”的小节的结尾提到: 为了预先计算矩阵指数和逆矩阵, 应该将矩阵 A_n 的第一个元素 $a_{11}[v_n, T_n]$ 量化成覆盖动态范围的有限数量的值。如根据前一小节的推导并且尤其从示例1可见, 系数 $a_{11}[v_n, T_n]$ 是 v_n 和在时间 T_n 处的电导的函数。虽然膜电压的动态范围相当有限, 但是电导的动态范围可以非常大。因此, 在这种情况下, 可能极有可能不得不对大量预先计算出的矩阵进行存储。接下来, 为了降低存储器需求, 探讨进一步简化。一种方

法是,对有限范围内的期望矩阵进行预先计算并存储,并且针对超出该范围的值执行计算以对期望矩阵进行近似。

[0360] 例如,当系数 $a_{11}[v_n, T_n]$ 变得比矩阵 A_n 的其它三个固定系数大得多(就幅度而言)时,针对诸如 $\psi(x) = e^x$ 和 $\psi(x) = x^{-1}$ 等任何复数值函数 $\psi(\cdot)$ 可以使用以下近似:

$$[0361] \quad \psi(A) \cong \frac{1}{(a_{11} - a_{22})^2 + a_{12}a_{21}} \begin{bmatrix} a_{11} - a_{22} & -a_{12} \\ a_{21} & a_{11} - a_{22} \end{bmatrix} \begin{bmatrix} \psi(\gamma_1) & 0 \\ 0 & \psi(\gamma_2) \end{bmatrix} \begin{bmatrix} a_{11} - a_{22} & a_{12} \\ -a_{21} & a_{11} - a_{22} \end{bmatrix}$$

$$[0362] \quad \gamma_1 = a_{11} + \frac{a_{12}a_{21}}{a_{11} - a_{22}} \text{ 且 } \gamma_2 = a_{22} - \frac{a_{12}a_{21}}{a_{11} - a_{22}} \quad (88)$$

[0363] 使用该方法,可以对存储器需求和计算复杂度进行权衡。

[0364] 可以经由某种近似来实现对存储器需求的进一步降低。例如,可以将仿射线性近似应用到经修订的函数 $\Gamma(v, t) = \Gamma(v) \equiv \frac{1}{C} F(v(t))$,而不是由等式(38)定义的函数。此外,可

以将突触电流视为外部电流,并且结果所得的积分可以由后向矩形规则(backward rectangular rule)来近似。在这种情况下,对矩阵指数和逆矩阵进行预先计算和存储所涉及的动态范围只是膜电压的动态范围,因此动态范围很小。权衡是在存储器需求与解的准确性之间的权衡。

[0365] 接下来,通过在该附加简化的情况下重新推导出解来更详细地阐述上面的段落。像以前一样,可以从等式(35)和(36)开始,在这里为了方便起见对等式(35)和(36)重复如下:

$$[0366] \quad C \frac{d}{dt} v(t) = F(v(t)) - u(t) + \sum_{i=1}^L g_i(t) h_i(v(t)) (E_i - v(t)) + I_{\text{ext}}(t) \quad (89)$$

$$[0367] \quad \frac{d}{dt} u(t) = abv(t) - au(t) - abv_r \quad (90)$$

[0368] 然而,现在将等式(38)中的函数修改如下:

$$[0369] \quad \Gamma(v, t) = \Gamma(v) \equiv \frac{1}{C} F(v(t)) \quad (91)$$

[0370] 并且将方程组重新用公式表示为

$$[0371] \quad \frac{d}{dt} v(t) = \Gamma(v) + a_{12}u(t) + \frac{1}{C} \left\{ \sum_{i=1}^L g_i(t) h_i(v(t)) (E_i - v(t)) + I_{\text{ext}}(t) \right\} \quad (92)$$

$$[0372] \quad \frac{d}{dt} u(t) = a_{21}v(t) + a_{22}u(t) + b_2 \quad (93)$$

[0373] 系统矩阵 A_n 现在简化为

$$[0374] \quad A_n \equiv \begin{bmatrix} a_{11}[v_n] & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (94)$$

[0375] 并且现在将向量 $b(v_n, t)$ 定义为

$$[0376] \quad \mathbf{b}(v_n, t) = \begin{bmatrix} b_1[v_n, t] \\ b_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{C} \left\{ \sum_{i=1}^L g_i(t) h_i(v_n) (E_i - v_n) + I_{\text{ext}}(t) \right\} \\ b_2 \end{bmatrix} \quad (95)$$

[0377] 使用上面的近似,避免了系统矩阵(系数 $\mathbf{a}_{11}$)对电导值的依赖,从而极大地降低了在计算矩阵指数时所覆盖的动态范围。可以通过在该时间段期间将膜电压保持在常数值来将突触电流视为外部电流。则在固定时间步长的实现中,解将是

$$[0378] \quad \mathbf{x}(nT+T) = e^{\mathbf{A}_n T} \mathbf{x}(nT) + \int_{nT}^{nT+T} e^{\mathbf{A}_n(nT+T-\tau)} \mathbf{b}(v_n, \tau) d\tau \quad (96)$$

[0379] 等式(96)中向量积分的第一分量是

$$[0380] \quad \int_{nT}^{nT+T} e^{\mathbf{A}_n(nT+T-\tau)} \frac{1}{C} \left\{ \sum_{i=1}^L g_i(\tau) h_i(v_n) (E_i - v_n) + I_{\text{ext}}(\tau) \right\} d\tau \quad (97)$$

[0381] 可以在大多数感兴趣的情况下对其求解。已经通过使用后向矩形规则(即,使用等式(97)中的值 $v_n = v(nT+T)$)获得了上面近似的最佳结果。

[0382] 示例3:使用后向矩形规则的近似

[0383] 为了简化阐述和符号,该示例假定突触电流不包含依赖于电压的电导通道(例如,NMDA通道)并且没有外部电流 $I_{\text{ext}}(t)$ 。针对长度为T的固定步长大小,获得推导。开始,可以将等式(68)简化如下:

[0384]

$$\mathbf{x}(nT+T) = e^{\mathbf{A}_n T} \mathbf{x}(nT) + \mathbf{q} + \frac{1}{C} \int_0^T e^{\mathbf{A}_n \tau} \sum_{i=1}^L g_i(nT+T-\tau) (E_i - v(nT+T)) d\tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (98)$$

[0385] 其中向量 $\mathbf{q}$ 包含常数项,即

$$[0386] \quad \mathbf{q} = \mathbf{A}_n^{-1} (e^{\mathbf{A}_n T} - \mathbf{I}) \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} \quad (99)$$

[0387] 为了标示方便起见,也可以定义以下向量和矩阵:

$$[0388] \quad \mathbf{b} \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{B} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (100)$$

[0389] 从而可以将等式(98)重写为

$$[0390] \quad \mathbf{x}(nT+T) = e^{\mathbf{A}_n T} \mathbf{x}(nT) + \frac{1}{C} \int_0^T e^{\mathbf{A}_n \tau} \sum_{i=1}^L g_i(nT+T-\tau) E_i d\tau \mathbf{b} - \frac{1}{C} \int_0^T e^{\mathbf{A}_n \tau} \sum_{i=1}^L g_i(nT+T-\tau) d\tau \mathbf{B} \cdot \mathbf{x}(nT+T) + \mathbf{q} \quad (101)$$

[0391] 现在,根据等式(29),针对被建模为简单指数的 $g_i(t)$ 求解。在该情况下,有

$$[0392] \quad \mathbf{x}(nT+T) = e^{\mathbf{A}_n T} \mathbf{x}(nT) + \frac{1}{C} \sum_{i=1}^L g_i(nT) E_i \int_0^T e^{\mathbf{A}_n \tau} e^{-(T-\tau)/\tau_i} d\tau \cdot \mathbf{b} - \frac{1}{C} \sum_{i=1}^L g_i(nT) \int_0^T e^{\mathbf{A}_n \tau} e^{-(T-\tau)/\tau_i} d\tau \mathbf{B} \cdot \mathbf{x}(nT+T) + \mathbf{q} \quad (102)$$

[0393] 可以容易地求解积分,并且若进一步定义

$$[0394] \quad \mathbf{H}_i \equiv \int_0^T e^{\Lambda \sigma} e^{\frac{\sigma-T}{\tau_i}} d\sigma = \begin{bmatrix} h_{11}^i & h_{12}^i \\ h_{21}^i & h_{22}^i \end{bmatrix}, \quad \mathbf{h}_i \equiv \mathbf{H}_i \cdot \mathbf{b} = \begin{bmatrix} h_{11}^i \\ h_{21}^i \end{bmatrix} \quad (103)$$

[0395] 则,可以将等式(102) 写为

[0396]

$$\mathbf{x}(nT+T) = e^{\Lambda T} \mathbf{x}(nT) + \sum_i \frac{E_i g_i(nT)}{C} \cdot \mathbf{h}_i - \sum_i \frac{g_i(nT)}{C} \cdot \mathbf{H}_i \cdot \mathbf{B} \cdot \mathbf{x}(nT+T) + \mathbf{q} \quad (104)$$

[0397] 聚集在左侧的、包含在时刻nT+T的状态的项,得到

$$[0398] \quad \left[\mathbf{I} + \sum_i \frac{g_i(nT)}{C} \cdot \mathbf{H}_i \cdot \mathbf{B} \right] \cdot \mathbf{x}(nT+T) = e^{\Lambda T} \mathbf{x}(nT) + \sum_i \frac{E_i g_i(nT)}{C} \cdot \mathbf{h}_i + \mathbf{q} \quad (105)$$

[0399] 且最后

$$[0400] \quad \cdot \mathbf{x}(nT+T) = \left[\mathbf{I} + \sum_i \frac{g_i(nT)}{C} \cdot \mathbf{H}_i \cdot \mathbf{B} \right]^{-1} \cdot \left\{ e^{\Lambda T} \mathbf{x}(nT) + \sum_i \frac{E_i g_i(nT)}{C} \cdot \mathbf{h}_i + \mathbf{q} \right\} \quad (106)$$

[0401] 注意,要求逆的矩阵是

[0402]

$$\left[\mathbf{I} + \sum_i \frac{g_i(nT)}{C} \cdot \mathbf{H}_i \cdot \mathbf{B} \right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sum_i \frac{g_i(nT)}{C} \cdot \begin{bmatrix} h_{11}^i & 0 \\ h_{21}^i & 0 \end{bmatrix} = \begin{bmatrix} 1 + \sum_i h_{11}^i \frac{g_i(nT)}{C} & 0 \\ \sum_i h_{21}^i \frac{g_i(nT)}{C} & 1 \end{bmatrix} \quad (107)$$

[0403] 并且其逆矩阵为

$$[0404] \quad \left[\mathbf{I} + \sum_i \frac{g_i(nT)}{C} \cdot \mathbf{H}_i \cdot \mathbf{B} \right]^{-1} = \frac{1}{\Delta} \begin{bmatrix} 1 & 0 \\ -\sum_i h_{21}^i \frac{g_i(nT)}{C} & \Delta \end{bmatrix}, \text{ 其中 } \Delta = 1 + \sum_i h_{11}^i \frac{g_i(nT)}{C} \quad (108)$$

[0405] 对状态更新的计算可以如下继续(其中 $e^{\Lambda \cdot T} \equiv G$):

[0406] I. 对期望量化进行预先计算并存储所有的系数(G 、 $\mathbf{h}_i$ 、和 $\mathbf{q}$)。

[0407] II. 针对每个通道,计算 $h_{11}^i g_i(nT)/C$ 和 $h_{21}^i g_i(nT)/C$ 。

[0408] III. 计算总和以及经 E_i 缩放的总和

$$[0409] \quad \sum_i \frac{E_i h_{11}^i g_i(nT)}{C}, \sum_i \frac{h_{11}^i g_i(nT)}{C}, \sum_i \frac{E_i h_{21}^i g_i(nT)}{C}, \sum_i \frac{h_{21}^i g_i(nT)}{C}$$

[0410] IV. 对等式(105)的右侧进行求值,并且存储在辅助(aux)变量中,即,

$$[0411] \quad \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} v(nT) \\ u(nT) \end{bmatrix} + \begin{bmatrix} \sum_i \frac{E_i h_{11}^i g_i(nT)}{C} \\ \sum_i \frac{E_i h_{21}^i g_i(nT)}{C} \end{bmatrix} + \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$$

$$[0412] \quad auxv = g_{11}v(nT) + g_{12}u(nT) + \sum_i \frac{E_i h_{11}^i g_i(nT)}{C} + q_1$$

$$[0413] \quad auxu = g_{21}v(nT) + g_{22}u(nT) + \sum_i \frac{E_i h_{21}^i g_i(nT)}{C} + q_2$$

[0414] V. 利用逆矩阵执行乘法

$$[0415] \quad v(nT+T) = \frac{auxv}{\Delta} = \frac{auxv}{1 + \sum_i \frac{h_{11}^i g_i(nT)}{C}}$$

$$[0416] \quad u(nT+T) = -\frac{auxv}{\Delta} \sum_i \frac{h_{21}^i g_i(nT)}{C} + auxu = -v(nT+T) \sum_i \frac{h_{21}^i g_i(nT)}{C} + auxu$$

[0417] 注意,对于具有双指数的NMDA情况,很有可能单独地对每个指数进行处理。这导致对两个向量进行预先计算并存储,其中一个向量针对衰退的指数项,一个向量针对上升的指数项。

[0418] 图21是解说等式(106)的实现的框图2100。为了简单起见,表示等式(106)的第一部分的矩阵求逆块2102没有被展开。根据目前的状态向量 $x(nT)$,在量化块2104中对来自状态向量的膜电位 $v(nT)$ 进行量化,以确定存储器索引。存储器索引用于选择与膜电位的量化区间相对应的、并且被存储在存储器查找表2106中的用于神经元模型的分段线性近似的预定参数(例如,系数)。基于存储器索引,加载与量化区间相关联的参数,这些参数包括指数矩阵 G 2108和来自上面的等式(99)的所定义向量 q 2110的系数。这些参数还包括输入电流向量 h 2112和 N 个突触电流输入向量 h_i 2114的系数。求和块2116对应用到向量 h 2112的输入电流 $I(nT)/C$ 和由应用到 N 个突触电流输入向量 h_i 2114的反向电位 $E_i g_i(nT)/C$ 缩放电导值进行求和。根据等式(106),用矩阵求逆块2102对结果进行处理,以产生经更新的状态向量 $x(nT+T)$ 。可以添加延迟块2118,从而使得在时间上按照每个步长大小 T 而不是连续地对状态进行更新。

[0419] 可以使用对等式(101)中的积分的其它近似。例如,不是计算精确值并将其存储在 h_i 向量中,而是可以使用应用到 $g(t)$ 的梯形规则,从而避免对 h_i 向量的存储并以稍差近似的代价来进一步降低存储器需求。

[0420] 小结

[0421] 在上面的小节中,描述了对神经元模型进行分段线性化的一般方法。图7对各种步骤和线性化方法进行了小结。在702处以非线性时变系统开始,可以在704处在区间 $T_n \leq t \leq T_{n+1}$ 上执行分段线性化,从而在706处以矩阵 $A(v_n, t)$ 的方式来产生分段线性时变系统。若在708处,将常数(即,分段常函数)用于每个区间 $T_n \leq t \leq T_{n+1}$,则在710处,以矩阵 $A(v_n, T_n)$ 的方式来形成分段LTI系统。若在712处此线性化是独立于时间区间而被执行的,则在714处以矩阵 $A(v_n)$ 的方式来创建分段LTI系统。

[0422] 示例依赖于状态-空间区域的线性神经元建模

[0423] 本公开的上述某些方面主要涵盖了对局限于一个状态变量(例如,膜电位 v)的二维状态-空间的划分。如上所述,使用单个状态将 (v, u) 状态-空间划分成垂直块,并且每一块(其表示电压区间)与不同的线性微分方程组相关联。

[0424] 如下所述,本公开的某些方面对该构思进行了扩展。可以将状态空间划分成不同的区域(不局限于垂直块),其中每个区域与不同的线性微分方程组相关联。这些区域可以

是不重迭的。下面的描述主要关注二状态(或者等效地,二维)模型,并且理解可以容易地将二状态模型扩展到三个或更多个状态,其中划分可以分别是容积和多维区域。

[0425] 作为 v 的函数的分段线性公式化的综述

[0426] 可以将二维分段线性微分方程表示为一组 K 对微分方程,其中第 n 对由下式定义:

$$[0427] \quad C \frac{d}{dt} v = m_k v + d_k - u + I, \quad v_k \leq v < v_{k+1} \quad (109)$$

$$[0428] \quad \frac{d}{dt} u = a b v - a u - a b v_r \quad (110)$$

[0429] 其中 $k \in [1, K]$ 并且 $F(v) = m_k v + d_k$ 。在该方法中,将电压空间划分为 K 个不同的区域,并且第 k 个区域与第 k 对微分方程相关联。 K 个区域的联合覆盖了 (v, u) 值的空间。图8解说在给定 $K=4$ 情况下的划分,其中将电压空间划分成4个区域: $v < v_1$ 、 $v_1 \leq v < v_2$ 、 $v_2 \leq v < v_3$ 以及 $v_3 \leq v$ 。对于 $K=2$ 来说,这简化成Hunzinger Cold神经元模型(具有切换的封闭和开放区间,从而使得 $v_k < v \leq v_{k+1}$)。

[0430] 广义公式化

[0431] 基本思想是将分段线性构思扩展到仅根据 v 定义的区域之外(即,允许等式(110)以及等式(109)随每个区间而变化)。这可以通过在 (v, u) 空间中定义 K 个二维区域 Ω_k 来实现,其中 $k \in [1, K]$,并且其中这 K 个区域的联合覆盖了 (v, u) 值的整个空间。此外,除了 m_k 和 d_k 之外,还允许参数 a 和 b 对于每个区域是不同的,从而使得这些参数变成 k 的函数。还可以使参数 C 成为针对各区域的不同 C_k ,但自此以后对于每个区域是固定的,并且理解 C 可以成为因区域而异的。这些扩展可能不一定具有生物上的等效物。然而,从工程设计和计算的角度来看,这些扩展实现了对更丰富的神经元模型的设计以及用于开发神经元模型的更简单的平台。

[0432] 前面提到的两个扩展产生:

$$[0433] \quad C \frac{d}{dt} v = m_k v + d_k - u + I, \quad (v, u) \in \Omega_k \quad (111)$$

$$[0434] \quad \frac{d}{dt} u = a_k b_k v - a_k u - a_k b_k v_r, \quad (v, u) \in \Omega_k \quad (112)$$

[0435] 使用上述状态-空间符号可以将上面的表达式更加简洁地表达为

$$[0436] \quad \dot{\mathbf{x}} = \mathbf{A}_k \mathbf{x} + \mathbf{b} \left(\frac{I}{C} \right) + \mathbf{p}_k, \quad \mathbf{x} \in \Omega_k \quad (113)$$

[0437] 其中

$$[0438] \quad \mathbf{x} \equiv \begin{bmatrix} v \\ u \end{bmatrix} \quad (114)$$

$$[0439] \quad \mathbf{A}_k \equiv \begin{bmatrix} \frac{m_k}{C} & -\frac{1}{C} \\ a_k b_k & -a_k \end{bmatrix} \quad (115)$$

$$[0440] \quad \mathbf{p}_k \equiv \begin{bmatrix} d_k / C \\ -a_k b_k v_r \end{bmatrix} \quad (116)$$

$$[0441] \quad \mathbf{b} \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (117)$$

接下来,对依赖于M维状态-空间区域的情形和状态-空间解进行描述。

[0442] M维状态-空间情形

[0443] 在存在M个状态变量的M维情况下,状态-空间方程变成

$$[0444] \quad \dot{\mathbf{x}} = \mathbf{A}_k \mathbf{x} + \mathbf{b} \left(\frac{I}{C} \right) + \mathbf{p}_k, \quad \mathbf{x} \in \Omega_k \quad (118)$$

[0445] 其中

$$[0446] \quad \mathbf{x} \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix} \quad (119)$$

$$[0447] \quad \mathbf{A}_k \equiv \begin{bmatrix} a_{11}^{(k)} & \cdots & a_{1M}^{(k)} \\ \vdots & \ddots & \vdots \\ a_{M1}^{(k)} & \cdots & a_{MM}^{(k)} \end{bmatrix} \quad (120)$$

$$[0448] \quad \mathbf{p}_k \equiv \begin{bmatrix} d_1^{(k)} \\ \vdots \\ d_M^{(k)} \end{bmatrix} \quad (121)$$

$$[0449] \quad \mathbf{b} \equiv \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (122)$$

[0450] 依赖于区域的状态-空间解

[0451] 提供了方程 (113) 在连续时域和离散时域中的状态-空间解。它们的推导自然使用与上述方法相同的方法。连续时间解可以表示为:

$$[0452] \quad \mathbf{x}(t) = e^{\mathbf{A}_k t} \mathbf{x}(0) + \int_0^t e^{\mathbf{A}_k(t-\tau)} \cdot \mathbf{b} \cdot I(\tau)/C d\tau + \mathbf{q}_k, \quad \mathbf{x}(0) \in \Omega_k \quad (123)$$

[0453] 其中x(0)表示x(t)的初始值,

$$[0454] \quad \mathbf{q}_k \equiv \mathbf{A}_k^{-1} (e^{\mathbf{A}_k t} - \mathbf{I}) \cdot \mathbf{p}_k \quad (124)$$

[0455] 并且 $e^{\mathbf{x}}$ 表示矩阵X的矩阵指数。

[0456] 相应的离散时间解(在采样周期为T的情况下)可以表示为:

$$[0457] \quad \mathbf{x}(nT + T) = e^{\mathbf{A}_k T} \cdot \mathbf{x}(nT) + \mathbf{h}_k \cdot I(nT)/C + \mathbf{q}_k, \quad \mathbf{x}(nT) \in \Omega_k \quad (125)$$

[0458] 其中

$$[0459] \quad \mathbf{h}_k \equiv \int_0^T e^{\mathbf{A}_k \tau} d\tau \mathbf{b} \quad (126a)$$

$$[0460] \quad = \mathbf{A}_k^{-1} (e^{\mathbf{A}_k t} - \mathbf{I}) \quad (126b)$$

[0461] 当前的状态-空间向量 $\mathbf{x}(nT)$ 标识了区域 Ω_k 以及根据等式 (120) 来计算 $\mathbf{x}(nT+T)$ 所要使用的哪一组 $\{e^{\mathbf{A}_k t}, \mathbf{h}_k, \mathbf{q}_k\}$ 矩阵和向量。

[0462] 此外,上述针对非突触和突触电流类型的相同解容易应用。这包括线性时变 (LTV) 解。这将涉及使用这些更通用的状态-空间区域来代替依赖于电压的区域 (垂直划分)。

[0463] 解说性示例

[0464] 图9解说具有三个区域 (即, Ω_k , 其中 $k \in [1, 3]$) 的示例。图9中的三个区域皆是矩形区域。

[0465] 图10解说具有四个区域的示例,解说了为何这些区域不需要是矩形。可以使用该特定设计来创建阈下振荡神经元,其中 Ω_1 、 Ω_2 、 Ω_3 和 Ω_4 分别表示以下区域:其中 (v, u) 以增长的半径向外螺旋的区域、以圆形 (没有衰退也没有增长) 的方式螺旋的区域、以衰退半径向内螺旋的区域、以及趋于发放尖峰的区域 (等效于Hunzinger Cold模型的ALIF区域)。下面对该设计进行更加详细的描述。

[0466] 图11解说更加理论性的示例,该示例具有5个区域以解说理论上可以创建具有任何形状的区域 (例如,六边形、三角形或星形区域)。

[0467] 依赖于状态-空间区域的线性神经元建模的实现

[0468] 目前,区域由下式定义:

$$[0469] \quad \Omega_k = \{ (v, u) \mid v_k \leq v < v_{k+1} \}$$

[0470] 其中 v_k 和 v_{k+1} 表示电压的左边界和右边界。因此,在每个时间步长处,通过取当前状态 $(v(nT), u(nT))$ 并针对 K 个区域: $\Omega_k, k \in [1, K]$ 中的每一个区域对当前状态进行核查来确定索引 k 。索引 k (用 k_F 表示) 可以经由仅 v 以及量化区间、最小可能电压和区间数量的函数来直接计算:

$$[0471] \quad k = \left\lfloor \frac{v(nT) - v_{\min, F}}{\Delta v_F} \right\rfloor + 1$$

[0472] 对于某些方面来说,这些区域 (或者其至少一部分) 可以是稍微重迭的。在这种情况下,可以使用滞后来标识 K 个区域 ($\Omega_k, k \in [1, K]$) 中的哪个区域包含当前状态 $(v(nT), u(nT))$ 。

[0473] 下面提供了广义区域的两个示例。也可以开发其它区域。

[0474] 针对依赖于 v 和 u 二者的矩形区域

[0475] 针对广义矩形区域,第 k 个区域可以由下式定义:

$$[0476] \quad \Omega_k = \{ (v, u) \mid v_k \leq v < v_{k+1}, u_k \leq u < u_{k+1} \}$$

[0477] 其中 v_k 和 v_{k+1} 分别表示针对电压的左边界和右边界,并且其中 u_k 和 u_{k+1} 分别表示恢复变量的下边界和上边界。因此,在每个时间步长处,可以通过标识 K 个区域 ($\Omega_k, k \in [1, K]$) 中的哪个区域包含当前状态 $(v(nT), u(nT))$ 来确定索引 k 。

[0478] 对于椭圆区域

[0479] 为了易于阐述,假定只有皆以大约静息状态 $(v_r, 0)$ 为中心的椭圆区域。则第 k 个区域可以由下式定义

[0480] $\Omega_k = \{ (v, u) \mid \rho_k \leq \rho(v, u) < \rho_{k+1} \}$

[0481] 其中 ρ_k 是沿 v 轴到内椭圆边界的“径向距离”， ρ_k 是沿 v 轴到外椭圆边界的“径向距离”，“径向距离”可以被定义为

[0482] $\rho(v, u) \equiv \sqrt{(v - v_r)^2 + (u/\beta)^2}$

[0483] 并且 β 是与椭圆相关联的参数($\beta=1$ 导致同心区域)。与之前一样,在每个时间步长处,通过取当前状态 $(v(nT), u(nT))$ 并且针对 K 个区域: $\Omega_k, k \in [1, K]$ 中的每一个区域对当前状态进行核查来确定索引 k 。

[0484] 依赖于状态-空间区域的线性神经元建模的优点

[0485] 对神经元建模的状态-空间的这种广义划分可有助于支持神经元的综合并且以系统方式来实现这点。其两个示例是实现:(1)发动检测和事件计数,以及(2)阈下振荡行为,其中在下一部分中对后者进行描述。

[0486] 可以针对发动检测和事件计数来开发基于Hunzinger Cold模型的人工神经元(其具有两个划分)。将该人工神经元(起初处于静息状态)设计为若其接收到用于发动检测的尖峰则立即激发。此后,该神经元被设计为开始对传入尖峰进行计数,并且在某个数量的尖峰之后激发。当尖峰到达时,该动作继续自我重复。由于神经元在达到平衡的静息状态之前会花大约500ms,所以在没有更多尖峰到达之后该设计的潜在缺点出现。在神经元可以执行发动检测之前,这可能是一段太长的时间。该设计中的部分约束是:缩短达到静息状态的时间如何通过使其行为更像LIF神经元来影响尖峰计数的能力。

[0487] 根据本公开的某些方面,可以通过划分将这两种行为分开来消除该约束,这类似于可以如何通过用Hunzinger Cold模型划分两个区域来分解Izhikevich神经元模型的LIF和ALIF行为的方式。可以通过注意 u 状态变量在LIF区域中的阈值约0.3之上和之下如何行为来缩短将神经元重置以执行发动检测的时间。在该值以下,系统需要长时间来达到静息状态。这一点可以通过根据图9的三区域状态-空间中所示的以 u 和 v 划分状态-空间来加快,从而实现改进的发动和事件计数神经元。在 $v < v_t$ 的LIF区域中,若 $u > u_0$,则人工神经元现在的行为如用于事件计数的积分和激发(IF)神经元(即,区域 Ω_2),否则如达到 v_t 的LIF神经元(即,区域 Ω_1)。一旦状态进入区域 Ω_1 ,可以将系统设计为根据需要以较短的时间常数尽快移动到静息状态。通过作为 u 的函数将LIF区域划分为两个期望区域(针对IF的 Ω_2 和针对LIF的 Ω_1),其中每个区域使用两对微分方程,可以消除大的延迟。

[0488] 基于依赖于状态-空间区域的线性神经元模型的示例阈下振荡神经元设计

[0489] 可以在人工神经系统中使用阈下振荡以增加在膜电位从静息电位的上升(即,使人工神经元“准备”发放尖峰)中发放尖峰的可能性并且降低在下降中的该可能性。前一部分描述了通过允许不仅基于一个状态变量(例如,电压)而且基于状态变量的组合和区域(或函数)对二维(或多维)状态-空间进行划分,来对分段线性神经元建模进行扩展。本部分克服了现有的阈下振荡神经元设计的缺点。例如,现有的神经元模型可以支持展示三种振荡类型(例如,如图12A的曲线图1200中所示的衰退振荡、如图12B的曲线图1210中所示的持续振荡、或者图12C的曲线图1220中所示的增长振荡)中的最多一种振荡类型的阈下振荡神经元的设计,但是这些模型无法实现展示这三种类型中的两种或更多种类型的人工神经元的设计。此外,现有的神经元模型不能支持具有特定振荡频率、衰退振荡的衰退速率、增长

振荡的增长速率、和/或持续振荡的幅度的阈下振荡神经元的系统设计。

[0490] 相应地,需要具有能够展示全部三种类型的振荡并且支持对所期望的振荡行为的系统实现的改进型神经元模型的人工神经元。

[0491] 能够进行阈下振荡的神经元模型之间的比较

[0492] 有四种现有的神经元模型能够产生阈下振荡:Izhikevich (简单) 模型、自适应指数 (AdEx) 神经元模型 (其是上述Brette和Gerstner的线性加指数函数)、四次模型以及Hunzinger Cold模型。可以通过四种方式来表征这些神经元模型的特征:(1) 其可以产生三种可能的阈下振荡行为(衰退振荡、持续振荡或增长振荡)中的哪一种;(2) 单个神经元可以展示这三种行为中的几种;(3) 设计具有所期望的振荡频率的神经元的容易程度;以及(4) 对振荡设计进行微调的容易程度(例如,定义持续振荡的幅度,或者定义衰退或增长的多个速率)。下表对每种神经元模型针对首次测量如何行为进行了概括(其中这些神经元模型有可能能够支持这三种振荡类型中的一种到所有类型):

[0493]

神经元模型	衰退振荡	持续振荡	增长振荡
Izhikevich 模型和 AdEx 模型	是		
四次模型	是	是	是
Hunzinger Cold 模型	是	是	是
依赖于状态-空间区域的模型	是	是	是

[0494] 下一个表对每种神经元模型针对其余特性测量如何行为进行了概括:

[0495]

模型	一种振荡类型	两种振荡类型	三种振荡类型	系统化、简单设计	微调能力
Izhikevich 模型和 AdEx 模型	是				

[0496]

型					
四次模型	是				
Hunzinger Cold 模型	是			是	
依赖于状态- 空间区域的模 型	是	是	是	是	是

[0497] 每个神经元模型(除了下述依赖于状态-空间区域的方法)可以再现仅能够展示单个振荡类型的神经元。且仅有Hunzinger Cold模型(由于其线性设计)能够提供对目标进行设计的系统方式。然而,这些模型中没有一种(除了下述依赖于状态-空间区域的方法)提供了微调的能力。

[0498] 基于依赖于状态-空间区域的线性神经元建模的阈下振荡

[0499] 如上面的表所示,并且与现有的神经元模型形成对比,本公开的某些方面可以:
(1) 产生可以在阈下态相中展示这三种振荡类型中的所有或任何子集的人工神经元,(2) 被设计为不仅是像Hunzinger Cold模型一样系统化,而且还可以支持设计的微调。

[0500] 设计

[0501] 可以通过使用上述依赖于状态-空间区域的方法来创建所有三种振荡类型的展示。例如,可以定义如图10中所示的四区域、二维神经元模型。

[0502] 每个状态-空间区域 Ω_k 与一组线性微分方程相关联,使得

$$[0503] \quad C \frac{d}{dt} v = m_k v + d_k - u + I, \quad (v, u) \in \Omega_k \quad (127)$$

$$[0504] \quad \frac{d}{dt} u = a_k b_k v - a_k u - a_k b_k v_r, \quad (v, u) \in \Omega_k \quad (128)$$

[0505] 或者,以状态-空间向量符号表示,

$$[0506] \quad \dot{\mathbf{x}} = \mathbf{A}_k \mathbf{x} + \mathbf{b} \left(\frac{I}{C} \right) + \mathbf{p}_k, \quad \mathbf{x} \in \Omega_k \quad (129)$$

[0507] 并且

$$[0508] \quad \mathbf{A}_k \equiv \begin{bmatrix} \frac{m_k}{C} & -\frac{1}{C} \\ a_k b_k & -a_k \end{bmatrix} \quad (130)$$

$$[0509] \quad \mathbf{p}_k \equiv \begin{bmatrix} d_k / C \\ -a_k b_k v_r \end{bmatrix} \quad (131)$$

[0510] 区域 Ω_1 (显示为圆形,但通常为椭圆)、 Ω_2 (也通常是椭圆)、 Ω_3 和 Ω_4 可以被设计为分别再现增长振荡、持续振荡、衰退振荡和ALIF行为。

[0511] 根据某些方面,可以由不止两个维度来定义状态-空间。在这种情况下,替代地,可以使用下面的替代形式:

$$[0512] \quad \mathbf{A}_k \equiv \begin{bmatrix} a_{11}^{(k)} & \cdots & a_{1M}^{(k)} \\ \vdots & \ddots & \vdots \\ a_{M1}^{(k)} & \cdots & a_{MM}^{(k)} \end{bmatrix} \quad (132)$$

$$[0513] \quad \mathbf{p}_k \equiv \begin{bmatrix} d_1^{(k)} \\ \vdots \\ d_M^{(k)} \end{bmatrix} \quad (133)$$

[0514] 若期望持续振荡,则在时间步长方法的情况下,可以保证一些厚度来处理 $v(nT+T)$ 可能无意中绕过持续振荡区域的情形。

[0515] 重置修改

[0516] 如上所述,重置机制可以涉及将 u 重置为 $u+d$,这相当于,例如,在尖峰事件之后或在控制信号被激活之后递增 u 。对于某些方面,为了确保 $u+d$ 值不会如此之大以至于会导致神经元从刚开始就立即发放尖峰或进入持续振荡,可能期望具有将 u 重置为目标常数(即,重置为 $u+d$ 以外的某个值)的选项。这样,设计者具有对神经元在状态-空间中何处进行重置的控制。

[0517] 定义状态-空间区域

[0518] 一般而言,持续振荡中的 (v, u) 轨迹可以呈现椭圆的形式,该椭圆的长轴和短轴可能不与 v, u 笛卡尔轴对准(即,椭圆可以有倾斜)。解可以呈现以下形式:

$$[0519] \quad v(t) = \mathbf{A}_c^{(v)} \cos(\omega t) + \mathbf{A}_s^{(v)} \sin(\omega t) \quad (134a)$$

$$[0520] \quad u(t) = \mathbf{A}_c^{(u)} \cos(\omega t) + \mathbf{A}_s^{(u)} \sin(\omega t) \quad (134b)$$

[0521] 其中四个系数是 $\mathbf{A}$ 中的矩阵元素(其中去掉了索引 k)和初始条件: $(v(0), u(0))$ 的函数。

[0522] 为了便于阐述和设计,此处的描述关注长轴和短轴相等的圆形轨迹。这些要求矩阵元素满足下列约束条件:

$$[0523] \quad a_{11} = a_{22} = \lambda_R = \frac{1}{\tau} \quad (135)$$

$$[0524] \quad a_{12} = -a_{21} = \lambda_I = 2\pi f_c \quad (136)$$

[0525] 其中 λ_R 和 λ_I 分别表示复数特征值 $\lambda_{1,2} = \lambda_R \pm \lambda_I$ 的实部和虚部, τ 表示(衰退或上升)时间常数, f_c 表示振荡频率,并且已经去掉了索引 k 。这得到圆形区域和明确形式的矩阵:

$$[0526] \quad \mathbf{A} \equiv \begin{bmatrix} \frac{1}{\tau} & 2\pi f_c \\ -2\pi f_c & \frac{1}{\tau} \end{bmatrix} \quad (137)$$

[0527] 使用具有圆形区域的这种更简单的形式,以平衡点 $(v_r, 0)$ 为中心的圆的半径为

$$[0528] \quad \rho(v, u) \equiv \sqrt{(v - v_r)^2 + u^2} \quad (138)$$

[0529] 其表示从 (v, u) 到平衡点的距离。例如, $\rho(v_r + 2mV, 0) = 2$ 。从而, 第 k 个区域可以由下式来定义:

$$[0530] \quad \Omega_k = \{(v, u) \mid \rho_{k, \text{in}} < \rho(v, u) \leq \rho_{k, \text{out}} \text{ 并且 } v \leq v_t\} \quad (139)$$

[0531] 其中 $\rho_{k, \text{in}}$ 表示 Ω_k 的内圆边界的半径, 并且 $\rho_{k, \text{out}}$ 表示 Ω_k 的外圆边界的半径。

[0532] 针对图10中的 $K=4$ 的示例, 持续振荡区域 Ω_2 可以由下式来定义:

$$[0533] \quad \Omega_2 = \{(v, u) \mid \rho_g < \rho(v, u) \leq \rho_d \text{ 并且 } v \leq v_t\} \quad (140)$$

[0534] 其中 ρ_g 表示分开 Ω_1 和 Ω_2 的圆边界的半径, 并且其中 ρ_d 表示 Ω_2 和 Ω_3 之间的圆边界的半径。

[0535] 针对增长振荡区域 Ω_1 :

$$[0536] \quad \Omega_1 = \{(v, u) \mid \rho(v, u) \leq \rho_g \text{ 并且 } v \leq v_t\} \quad (141)$$

[0537] 针对衰退振荡区域 Ω_3 :

$$[0538] \quad \Omega_3 = \{(v, u) \mid \rho_d < \rho(v, u) \text{ 并且 } v \leq v_t\} \quad (142)$$

[0539] 图10中的ALIF区域 Ω_4 由下式定义

$$[0540] \quad \Omega_4 = \{(v, u) \mid v > v_t\} \quad (143)$$

[0541] 多级衰退和增长区域的额外区域

[0542] 以上思想可以被扩展为支持振荡的更逐渐的衰退或逐级增加。例如, 为了具有不同速率的衰退轨道和不同速率的增长轨道, 如图13中所示, 可以使用多个环和6个区域。在该示例中, Ω_3 可以被视为持续振荡区域。 Ω_5 可以具有较快的衰退轨道, 而 Ω_4 可以具有较慢的衰退轨道。这将允许向持续振荡区域 Ω_3 进行逐级并更缓和的衰退。类似地, Ω_1 可以具有较快的增长轨道, 而 Ω_2 可以具有较慢的增长轨道。可以将该构思扩展到更多环, 以实现甚至更逐渐的变化。此外, 若期望不同的持续振荡参数 (例如, 不同频率), 则可以使用多级持续振荡区域。

[0543] 用于实现人工神经元的示例操作

[0544] 图14是根据本公开的某些方面用于对人工神经网络中的人工神经元的状态进行更新的示例操作1400的流程图。操作1400可以用硬件 (例如, 通过一个或多个神经处理单元 (例如, 人工神经元), 诸如神经形态处理器)、软件或固件来实现。人工神经网络可建模在各种生物或想像神经网络中的任一者上, 诸如视觉神经系统、听觉神经系统、海马体等。

[0545] 操作1400可始于在1402确定人工神经元的第一状态位于第一区域内。在1404, 至少部分地基于第一组线性动态方程来确定人工神经元的第二状态, 其中第一组线性动态方程至少部分地基于与第一区域相对应的第一组参数 (例如, 系数)。在1406, 确定人工神经元的第二状态位于第二区域内。在1408, 至少部分地基于第二组线性动态方程来确定人工神经元的第三状态。第二组线性动态方程可以至少部分地基于与第二区域相对应的第二组参数。

[0546] 根据某些方面, 第一和第二组线性方程包括线性时不变 (LTI) 状态空间方程的离散时间解。对于其它方面, 离散时间解可以至少部分地基于连续时间状态-空间方程的闭合形式解。

[0547] 根据某些方面, 操作1400还包括从存储器中获取第一或第二组参数中的至少一组参数。如本文中所使用的, 存储器可以指用于存储数据的各种合适装置中的任一种, 无论是永久地还是临时地、在处理单元本地还是远程、片上或片外, 包括随机存取存储器 (RAM)、

高速缓存存储器、寄存器、锁存器、触发器等。这种获取可以包括从人工神经元本地的存储器获取第一组参数或第二组参数中的至少一组参数。对于某些方面,操作1400还包括计算第一组参数或第二组参数中的该至少一组参数的至少一部分。可以使用从存储器获取的一个或多个值来计算第一组参数或第二组参数中的该至少一组参数的至少一部分。

[0548] 根据某些方面,可以通过用分段线性函数近似与人工神经元相关联的神经元模型中的至少一部分非线性函数来获得第一组参数或第二组参数中的至少一组参数。对于某些方面,非线性函数包括膜电位(v)乘以依赖于电压的电导($g(v)$)。依赖于电压的电导可以用分段常数函数来近似。对于其它方面,非线性函数包括依赖于电压的函数($F(v)$),其中 v 是人工神经元的膜电位。针对第一区域和第二区域中的每一者,分段线性函数可以具有斜率(m)和截距(d)。对于某些方面,在分段线性函数中,第一区域具有与第二区域不同的宽度。分段线性函数中的第一区域或第二区域的宽度可以依赖于非线性函数。对于某些方面,分段线性近似至少部分地基于下列方法中的至少一种方法:泰勒展开法、一阶线性内插法、最优线性内插法、或平均斜率法。

[0549] 根据某些方面,第一组参数或第二组参数中的至少一组参数可以至少部分地设计为在人工神经元中实现特定行为。例如,可以生成第一和/或第二组参数(例如,由神经元模型设计者手动选择)以有效地创建特定函数(例如,可调谐的曲线),其对所期望的行为进行近似。通过这种方式,人工神经元可以基于全新的神经元模型或者对现有神经元模型的修改来进行操作。

[0550] 人工神经元的第一、第二和第三状态可以由膜电位(v)和恢复电流(u)来定义。这是人工神经元状态变量的两个示例。根据某些方面,操作1400还可以包括:至少部分地基于确定尖峰事件已经发生或将要发生而重置人工神经元的膜电位或恢复电流中的至少一者。可以将膜电位重置为静息电位($v(t) \rightarrow v_r$)。可以将恢复电流重置为恢复电流的当前值与偏移之和($u(t) \rightarrow u(t) + d$)。

[0551] 根据某些方面,第一组或第二组线性方程中的至少一组至少部分地基于人工神经元的神经元模型。对于某些方面,神经元模型至少部分地基于以下至少一者:Izhikevich简单模型、指数积分激发(EIF)模型、FitzHugh-Nagumo模型、四次模型、或者固有电导模型(例如,表示为膜电位 v 乘以依赖于电压的电导 $g(v)$ 的神经元模型)。对于某些方面,神经元模型包括至少两个维度(即,至少两个状态变量)。神经元模型可以至少部分地基于一个或多个一阶常微分方程(ODE),其可以是线性的。

[0552] 根据某些方面,神经元模型的时间步长大小至少部分地基于正建模的人工神经元的类型。神经元模型的时间步长大小可以是非均匀的。对于某些方面,特定时间步长的时间步长大小是至少部分地基于人工神经元的当前状态或过去状态中的至少一者、或者基于特定的参数集(其可以与当前状态或过去状态相关联)来确定的。对于某些方面,确定第一和第二状态之间的第一时间步长大小不同于确定第二和第三状态之间的第二时间步长大小。根据某些方面,在1404确定人工神经元的第二状态可以在第一时间步长中执行,而在1408确定人工神经元的第三状态可以在第一时间步长之后的第二时间步长中执行。

[0553] 根据某些方面,确定人工神经元的第二或第三状态中的至少一者是至少部分地基于输入人工神经元的电流。对于某些方面,输入电流包括突触电流或一般外部电流中的至少一者。突触电流可以至少部分地基于一个或多个通道中的每一者的依赖于时间的电导以

及一个或多个信道中的每一者的逆转电位。对于某些方面,突触电流是至少部分地基于人工神经元的针对N-甲基-D-天冬氨酸(NMDA)通道的突触后膜电位。依赖于时间的电导可以通过指数函数、阿尔法函数或差分指数函数来建模。

[0554] 对于某些方面,第一区域和第二区域是同一区域。

[0555] 根据某些方面,操作1400还可包括将第一状态、第二状态或第三状态中的至少一者的指示输出到显示器。

[0556] 根据某些方面,第一区域或第二区域中的至少一者是由两个或更多个维度定义的。第一区域或第二区域中的至少一者是M维子空间,其中 $M \geq 2$ 。M维子空间可以具有任何形状或者由任何边界(例如, $v > 10\text{mV}$)定义。对于某些方面,两个或更多个维度是二维,并且M维子空间具有二维形状,例如椭圆形、圆形、多边形、矩形或正方形。对于其它方面,两个或更多个维度是三维,并且M维子空间具有三维形状,例如球形、椭球形、多边形棱柱、矩形棱柱或立方体。对于某些方面,第一区域关联于与第二区域不同的线性方程组。第一区域和第二区域可以是部分重迭或不重迭的区域。第一区域和第二区域可以具有不同形状或不同大小中的至少一者。对于某些方面,第一区域与人工神经元的第一行为相关联,而第二区域是与人工神经元的第二行为相关联,第二行为不同于第一行为。对于某些方面,第一区域和第二区域具有分层形状,例如同心环、管、矩形框架等。

[0557] 根据某些方面,两个或更多个维度是由膜电位(v)和恢复电流(u)定义的两个维度组成的。对于某些方面,操作1400还可以包括:至少部分地基于对控制信号的接收或者对尖峰事件已经发生或将要发生的确定中的至少一者而重置人工神经元的膜电位或恢复电流中的至少一者。在这种情况下,可以将膜电位重置为静息电位,并且可以将恢复电流重置为常数。

[0558] 根据某些方面,第二区域的至少一部分与第一区域重迭。在这种情况下,在1406确定人工神经元的第二状态位于第二区域内可以至少部分地基于滞后。

[0559] 根据某些方面,人工神经元被配置成用于发动检测和事件计数。在这种情况下,第一区域可以与用于事件计数的积分激发(IF)行为相关联,并且第二区域可以与带漏泄积分激发(LIF)行为相关联。

[0560] 根据某些方面,人工神经元被配置成用于阈下振荡行为。阈下振荡行为可以包括衰退振荡行为、持续振荡行为或增长振荡行为中的至少一种。对于某些方面,第一区域可以与增长振荡行为相关联,第二区域可以与持续振荡行为相关联,并且第三区域可以与衰退振荡行为相关联。此外,第四区域与ALIF行为相关联。对于某些方面,第一区域可以具有椭圆形状,而第二区域可以具有环形形状。

[0561] 示例共用且灵活的神经架构

[0562] 本公开的某些方面一般涉及开发支持动态神经元模型的共用且灵活的架构。设计目标包括对神经元动态进行低复杂度、准确的建模,以及实现任何神经元模型的能力。使用该灵活架构,可以根据需要来执行和取代各种合适的神经元模型中的任何一种。例如,神经元模型可以包括以下至少一种:Izhikevich简单模型、指数积分激发(EIF)模型、FitzHugh-Nagumo模型、四次模型或者Hunzinger Cold模型。如上所述,可以使用分段线性近似以及与之相关联的参数(例如,系数)来实现这些神经元模型。此外,这种灵活的架构允许人工神经网络系统中的不同人工神经元与不同的神经元模型同时操作。

[0563] 图15解说根据本公开的某些方面针对单个神经处理单元1502的这种共用且灵活的神经架构的实现,其中可以对神经元模型的参数进行选择、加载、访问、添加、删除、调整和/或更新。可以将本文中描述的构思容易地扩展到神经处理单元1502的系统(例如,人工神经网络),但是为了便于描述,图15A中仅示出单个神经处理单元。如图所示,神经处理单元1502可以实现状态机1506,该状态机1506接收神经元模型的一个或多个输入和一组参数,并且输出一个或多个状态变量。状态机1506可以实现人工神经元的任何合适的处理算法。

[0564] 可以从用于多个神经元模型的一组参数1504中选择给定神经元模型的参数。对于某些方面,例如,操作者可能希望在给定的神经处理单元中使用某种神经元模型,并从可用神经元模型列表中选择该模型。对于其它方面,可以由神经处理单元系统基于学习或训练操作来选择某种神经元模型。给定神经元模型的参数可以是基于上述分段线性近似的参数。对于某些方面,可以将该组参数1504存储在与神经处理单元1502相关联和/或在神经处理单元1502本地的存储器中。对于其它方面,可以将该组参数1504存储在可由多个神经处理单元全局访问的存储器中、或者存储在神经处理单元1502内部的高速缓存存储器中。可以将神经元模型A的参数存储在第一存储器位置1508(例如,可寻址的存储器块),而可以将神经元模型B的参数存储在第二存储器位置1510。

[0565] 在图15A中,已经将神经元模型A的参数加载到神经处理单元1502中。如本文中所使用的,可以广义地定义术语“加载”,并且“加载”可以包括:从神经处理单元(或神经处理单元系统)可访问的存储器获取给定神经元模型的参数;将参数存储在神经处理单元本地的存储器中;或者在存储器中访问与该神经处理单元相关联的(例如,被指定用于该神经处理单元的)一个或多个存储器区域。根据某些方面,将某个神经元模型的参数加载到特定神经处理单元可以响应于配置事件而发生。例如,配置事件可以包括对特定神经处理单元上电,对神经处理单元的一个或多个区域(其可以包括或不包括该特定处理单元)上电,或者对神经处理单元的整个系统上电。

[0566] 在图15B中,已经将神经元模型B的参数加载到了神经处理单元1502中。从那时起,状态机1506可以至少部分地基于这些最新加载的参数来进行操作。通过这种方式,神经处理单元1502可以仅通过加载不同的参数来根据不同的神经元模型起作用。此外,可以在任何时候更新或删除特定神经元模型的参数。

[0567] 在图15C中,可以将神经元模型C的参数添加到该组参数1504中。例如,神经元模型C可以是最新开发、购买或获许可的神经元模型,其在神经处理单元1502变得可操作时是不可用的。在其它示例情景中,设计师或系统设计者可能也没有考虑神经元模型C或可能起初以为该模型不符合所期望的应用,但现在期望添加该模型。

[0568] 在图15D中,可以将神经元模型C的参数存储在第三存储器位置1512。可以将神经元模型C的参数加载到神经处理单元1502中,从而使得状态机1506可以至少部分地基于这些最新加载的参数来进行操作。对于神经处理单元仅通过在其中加载不同参数来有效地改变神经元模型的能力提供了可以根据需要而更新和改变的非常灵活的架构。

[0569] 图16是根据本公开的某些方面用于在神经处理单元系统(例如,人工神经网络)中实现多个神经元模型的组合的示例操作1600的流程图。操作1600可以用硬件(例如,通过一个或多个处理单元,诸如神经形态处理器)、软件或固件来实现。人工神经网络可建模在各

种生物或想像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。

[0570] 多个神经元模型可以包括上述神经元模型的任何组合、以及现有或尚未开发或未公开的任何合适的神经元模型。例如,多个神经元模型可以包括以下至少一种:Izhikevich简单模型、指数积分激发(EIF)模型、FitzHugh-Nagumo模型、四次模型、Hunzinger Cold模型或固有电导模型。如本文中所使用的,短语“神经元模型的组合”一般是指多个神经元模型的集合,其中该集合可包括单个成员。换句话说,多个神经元模型的组合包括这多个神经元模型中的一个以及这多个神经元模型的任何子集的任何组合。

[0571] 操作1600可以始于在1602将从多个神经元模型中选择的第一神经元模型的参数加载到第一神经处理单元(例如,人工神经元)中。给定神经元模型的参数可以是如上所述地从对神经元模型的分段线性近似获得的参数。在1604,可以至少部分地基于加载的第一神经元模型的参数来确定第一神经处理单元的第一状态。在1606,可至少部分地基于第一神经元模型的参数以及第一状态来确定第一神经处理单元的第二状态。

[0572] 根据某些方面,在1608,可以更新多个神经元模型。如上所述,更新多个神经元模型可以包括删除或调整现有神经元模型的某些参数,或者添加另一个神经元模型(例如,在这多个神经元模型对系统可用时尚不可用的新神经元模型,或者对现有神经元模型的期望修改以有效地创建新神经元模型)的参数。对于某些方面,更新多个神经元模型可以在加载(或重新加载)第一神经元模型的参数并确定第一神经处理单元的下一状态之前发生。

[0573] 在1610,可选地,可以将第二神经元模型的参数加载到第一神经处理单元中。第二神经元模型的这些参数可以取代在1602加载的第一神经元模型的参数。在1610加载第二神经元模型的参数之前,不需要在1608更新多个神经元模型。在1612,可以至少部分地基于第二神经元模型的参数来确定第一神经处理单元的第三状态(例如,在第二状态之后)。

[0574] 根据某些方面,在1602加载第一模型的参数(或者在1610加载第二神经元模型的参数)是响应于配置事件。对于某些方面,配置事件是对神经处理单元系统或第一神经处理单元中的至少一者上电。

[0575] 根据某些方面,在1612确定第三状态还至少部分地基于第二状态。若基于后续神经元模型的参数来确定后续状态是基于根据先前神经元模型(不同于后续神经元模型)的参数而确定的先前状态,则这可以被称为“热插拔”。对于某些方面,第二神经元模型至少部分被配置成在该“热插拔”条件下产生阈下振荡。该振荡可以从第二状态或者在另一时刻开始。对于某些方面,在1610加载第二神经元模型的参数是至少部分地基于以下至少一者:状态条件(例如,状态进入特定区域或展示某种行为,诸如变化速率等)、时间条件(例如,基于定时器)或触发。例如,触发可以在第一神经处理单元外部。

[0576] 根据某些方面,操作1600还可以包括将从多个神经元模型中选择的第二神经元模型的参数加载到第二神经处理单元。随后,可以至少部分地基于第二神经元模型的参数来确定第二神经处理单元的状态。对于某些方面,第一和第二神经处理单元位于系统的不同区域中。对于某些方面,第一和第二神经元模型表示不同类型的神经元。换句话说,神经处理单元系统中的不同神经元模型可以用于模拟神经系统的不同区域中的神经元。例如,视觉、听觉或电机控制系统中的神经元可以由不同的神经元模型来表示。

[0577] 加载第一、第二或任何其它神经元模型的参数可以涉及各种合适操作中的任一种。为了便于描述,下面的描述包括加载第一神经元模型的参数,但是这种加载也适用于任

何其它神经元模型的参数。根据某些方面,加载第一神经元模型的参数包括将第一神经元模型的参数加载到包括第一神经处理单元在内的多个神经处理单元。对于某些方面,加载第一神经元模型的参数包括从存储器获取第一神经元模型的参数的至少一部分。对于某些方面,此存储器可由神经处理单元系统访问。对于某些方面,该存储器的至少一部分可以是第一神经处理单元本地的。该存储器的至少一部分可以是高速缓冲存储器。对于某些方面,加载第一神经元模型的参数还包括在存储器中访问与第一神经元模型相关联的存储器块。在这种情况下,存储器块可以由指向存储器块中的存储器地址的指针来指定。

[0578] 图17根据本公开的某些方面解说用于使用通用处理器1702来实现前述操作人工神经元的方法的组件的示例框图1700。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1704中,而在通用处理器1702处执行的有关指令可从程序存储器1706中加载。在本公开的一个方面,加载到通用处理器1702中的指令可以包括:用于确定人工神经元的第一状态位于第一区域内的代码;用于至少部分地基于第一组线性方程来确定人工神经元的第二状态的代码,其中第一组线性方程是至少部分地基于与第一区域相对应的第一组参数;用于确定人工神经元的第二状态位于第二区域内的代码;以及用于至少部分地基于第二组线性方程来确定人工神经元的第三状态的代码,其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。对于某些方面,第一区域或第二区域中的至少一者是由两个或更多个维度来定义的。

[0579] 在本公开的另一个方面,加载到通用处理器1702的指令可以包括:用于将从多个神经元模型中选择的第一神经元模型的参数加载到(位于神经处理单元系统中的)第一神经处理单元中的代码、以及用于至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态的代码。

[0580] 图18解说了根据本公开的某些方面的用于实现前述操作人工神经元的方法的组件的示例框图1800,其中存储器1802可以经由互连网络1804与计算网络(神经网络)的个体(分布式)处理单元(神经处理器)1806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1802中,并且可从存储器1802经由互连网络1804的连接被加载到每个处理单元(神经处理器)1806中。在本公开的一个方面,处理单元1806可以被配置成:确定人工神经元的第一状态位于第一区域内;至少部分地基于第一组线性方程来确定人工神经元的第二状态,其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数;确定人工神经元的第二状态位于第二区域内;以及至少部分地基于第二组线性方程来确定人工神经元的第三状态,其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。对于某些方面,第一区域或第二区域中的至少一者是由两个或更多个维度来定义的。

[0581] 在本公开的另一个方面,处理单元1806可以被配置成:将从多个神经元模型中选择的第一神经元模型的参数加载到(在神经处理单元系统中的)第一神经处理单元中,并且至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态。

[0582] 图19解说了根据本公开的某些方面的基于分布式权重存储器1902和分布式处理单元(神经处理器)1904来实现前述操作人工神经元的方法的组件的示例框图1900。如图19中所解说的,一个存储器组1902可直接与计算网络(神经网络)的一个处理单元1904对接,其中该存储器组1902可存储与该处理单元(神经处理器)1904相关联的变量(神经信号)、突

触权重和/或系统参数。在本公开的一个方面,处理单元1904可以被配置成:确定人工神经元的第二状态位于第一区域内;至少部分地基于第一组线性方程来确定人工神经元的第二状态,其中第一组线性方程至少部分地基于与第一区域相对应的第一组参数;确定人工神经元的第二状态位于第二区域内;以及至少部分地基于第二组线性方程来确定人工神经元的第三状态,其中第二组线性方程至少部分地基于与第二区域相对应的第二组参数。对于某些方面,第一区域或第二区域中的至少一者是由两个或更多个维度来定义的。

[0583] 在本公开的另一个方面,处理单元1904可以被配置成:将从多个神经元模型中选择的第一神经元模型的参数加载到(在神经处理单元系统中的)第一神经处理单元中,并且至少部分地基于第一神经元模型的参数来确定第一神经处理单元的第一状态。

[0584] 图20解说根据本公开的某些方面的神经网络2000的示例实现。如图20中所解说的,神经网络2000可包括多个局部处理单元2002,它们可执行以上描述的方法的各种操作。每个处理单元2002可包括存储该神经网络的参数的局部状态存储器2004和局部参数存储器2006。另外,处理单元2002可包括具有局部(神经元)模型程序的存储器2008、具有局部学习程序的存储器2010、以及局部连接存储器2012。此外,如图20中所解说的,每个局部处理单元2002可与用于配置处理的单元2014对接并且与路由连接处理元件2016对接,单元2014可提供对局部处理单元的局部存储器的配置,元件2016提供局部处理单元2002之间的路由。

[0585] 根据本公开的某些方面,每一个局部处理单元2002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。

[0586] 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,各个操作可由图17-20中所示的各种处理器中的一个或多个来执行。一般而言,在存在附图中解说的操作的场合,那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14中所解说的操作1400对应于图14A中所解说的装置1400A。

[0587] 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(例如表格、图表或图形)的合适装置。用于处理的装置、用于生成的装置、用于加载的装置、用于重置的装置、用于获取的装置、用于更新的装置、用于计算的装置、用于运算的装置、用于输出的装置、或用于确定的装置可以包括处理系统,该处理系统可以包括一个或多个处理器或处理单元。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)。

[0588] 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。

[0589] 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。

[0590] 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,处理器可以是任何可商业购得的处理器、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、或任何其他此类配置。

[0591] 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。

[0592] 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。

[0593] 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。

[0594] 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,机器可读介质可以包括 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。

[0595] 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。

[0596] 处理系统可以被配置成通用处理系统,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、和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。

[0597] 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输模块和接收模块。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由处理器执行。在以下谈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。

[0598] 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或其它光盘存储、磁盘存储或其它磁存储设备、或能被用来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它介质。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。

[0599] 因此,某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。对于某些方面,计算机程序产品可包括包装材料。

[0600] 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。

[0601] 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设备的布局、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。

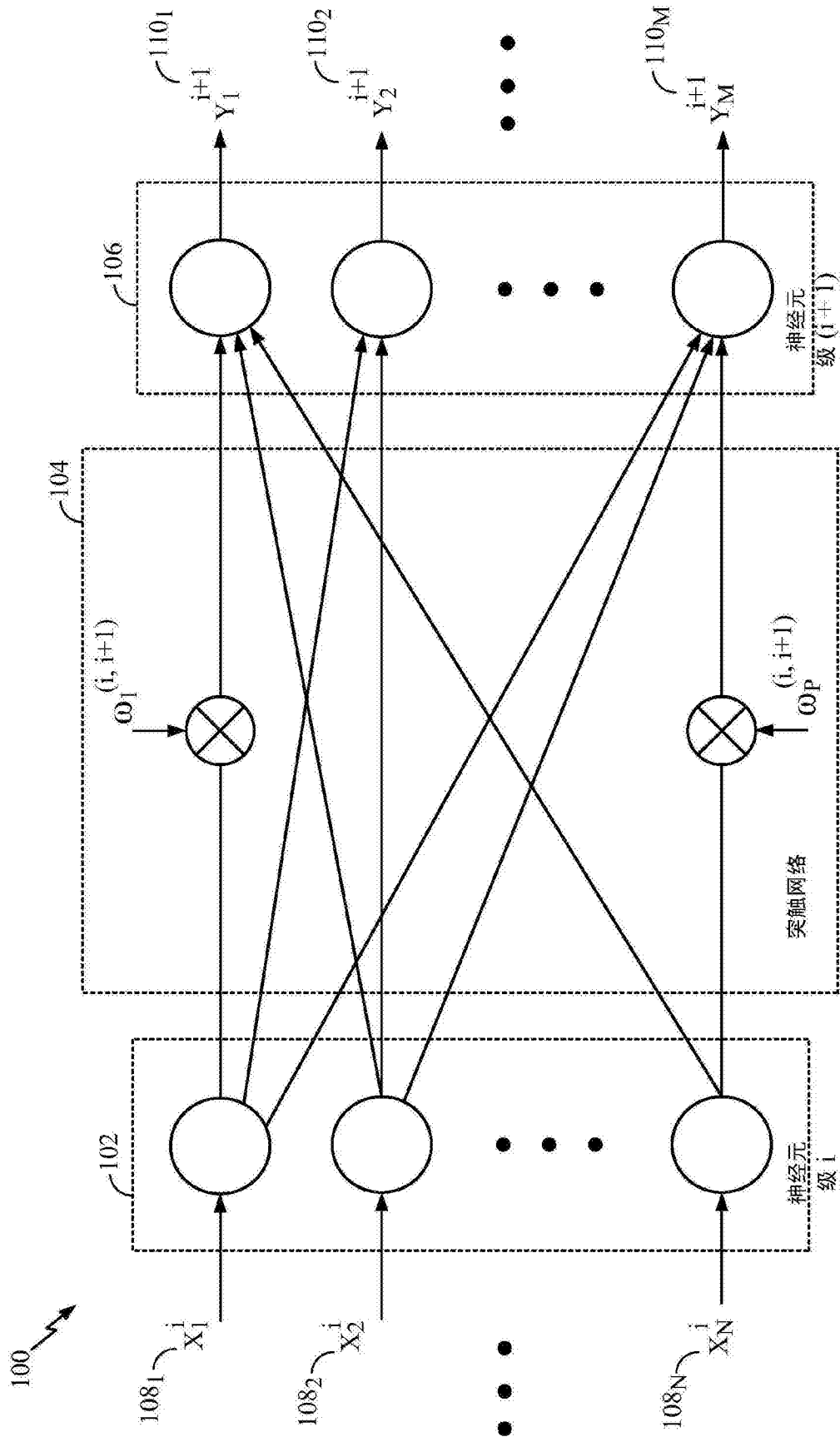


图1

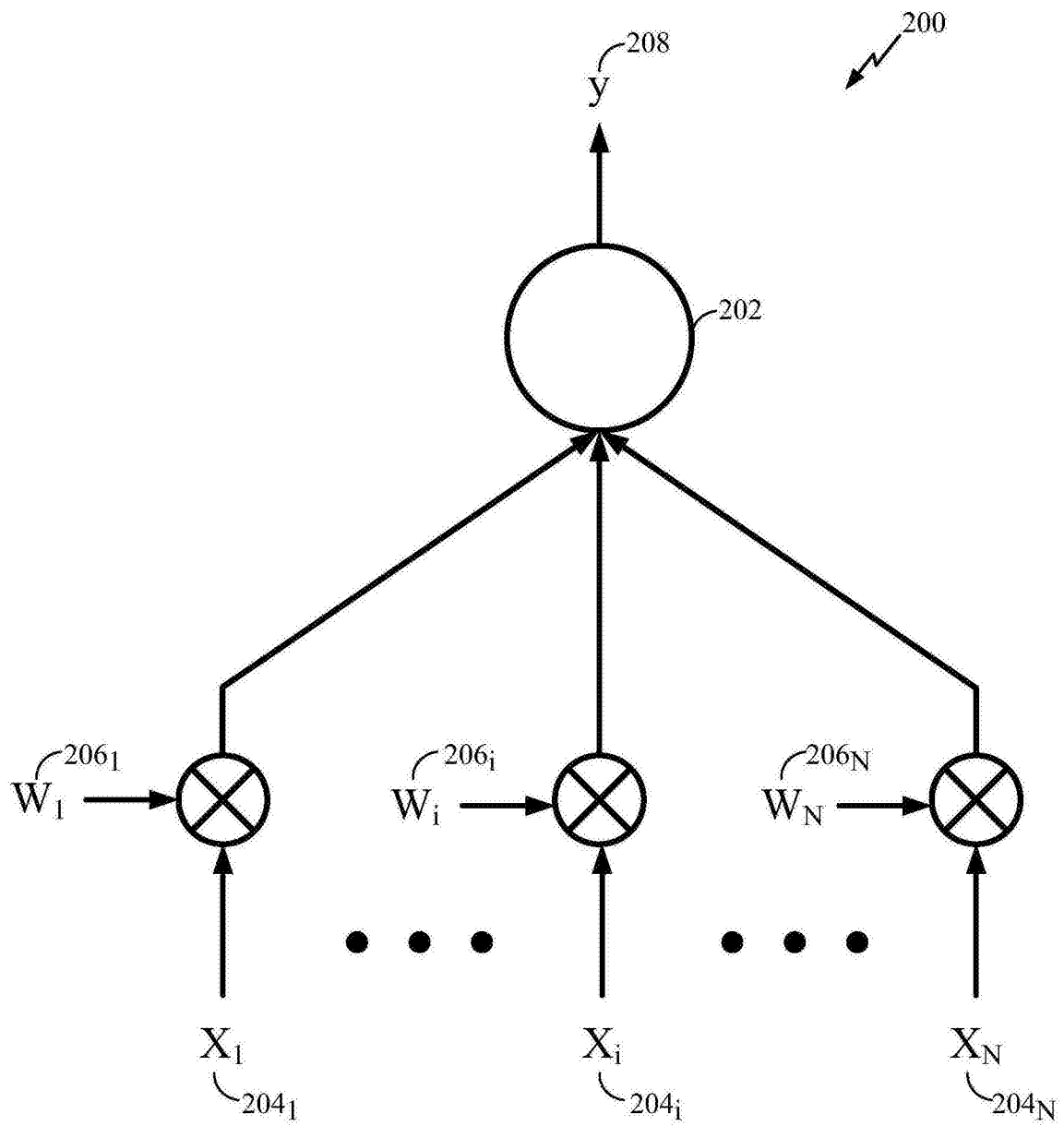


图2

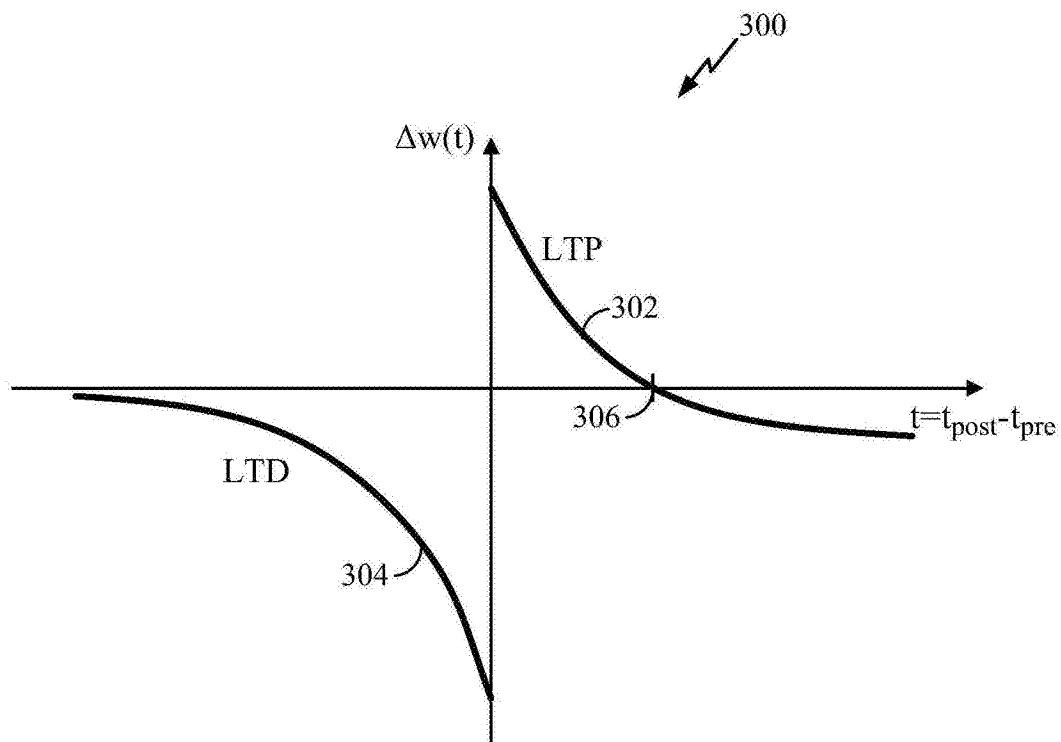


图3

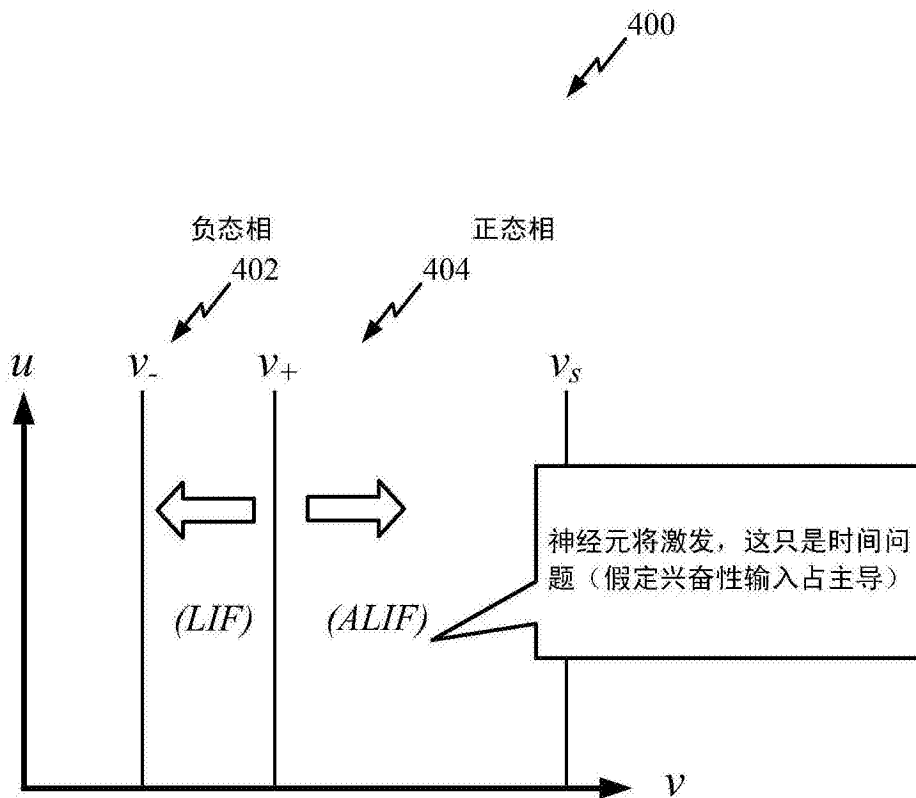


图4

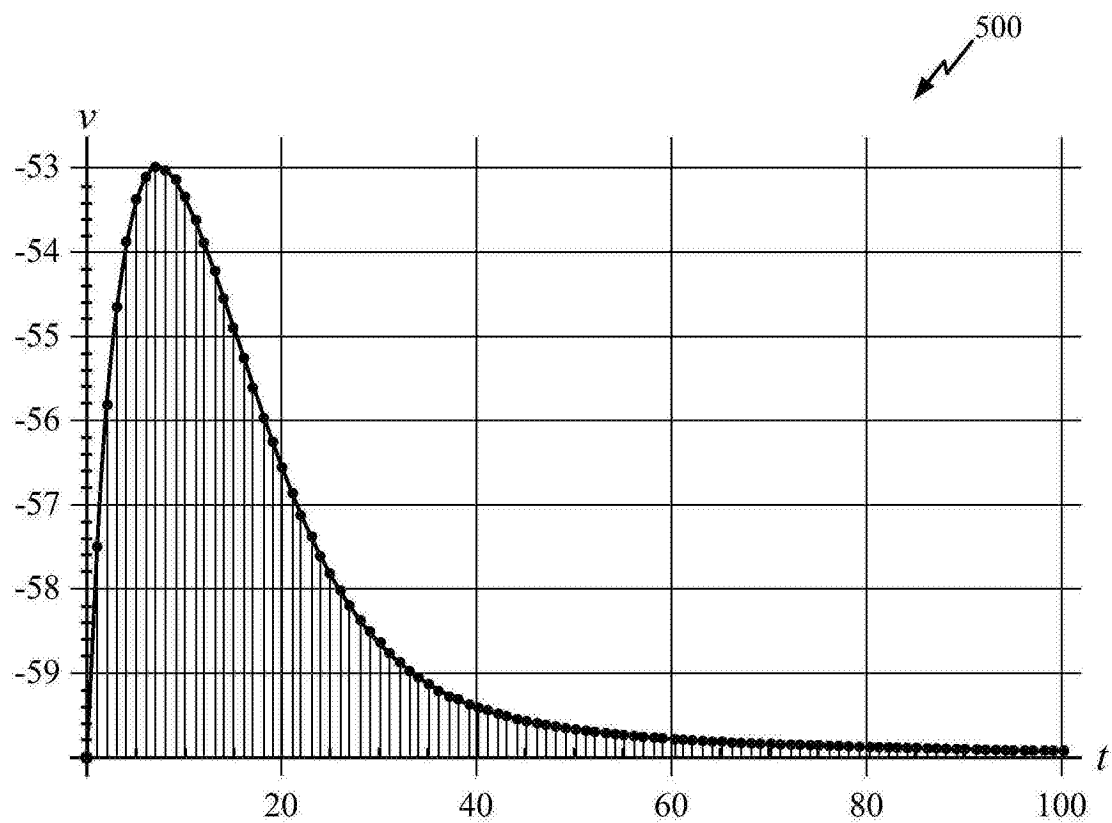


图5A

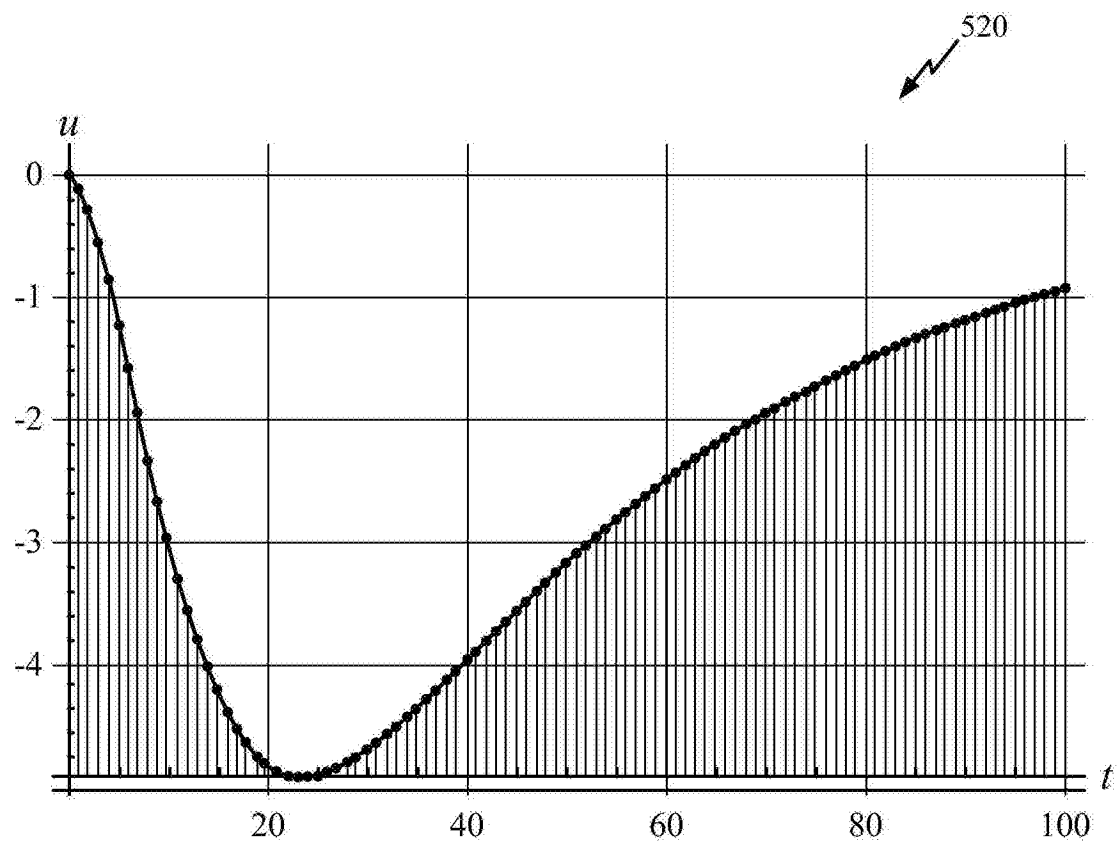


图5B

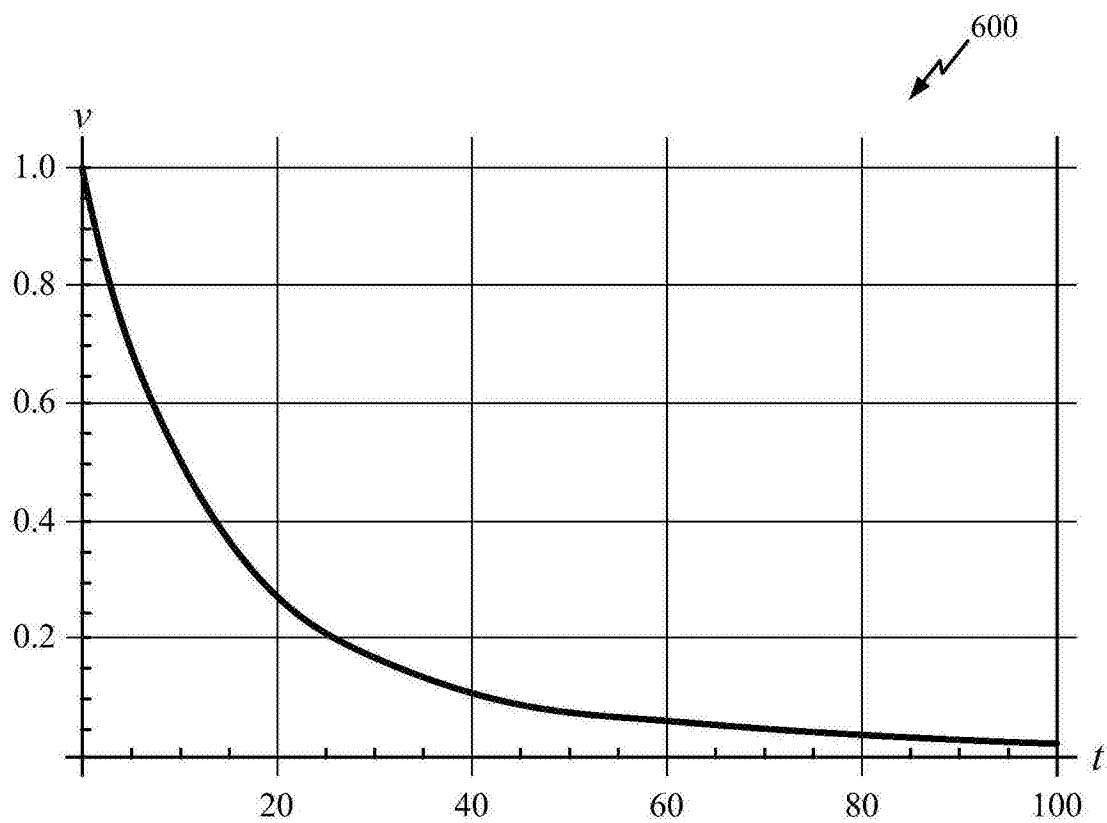


图6A

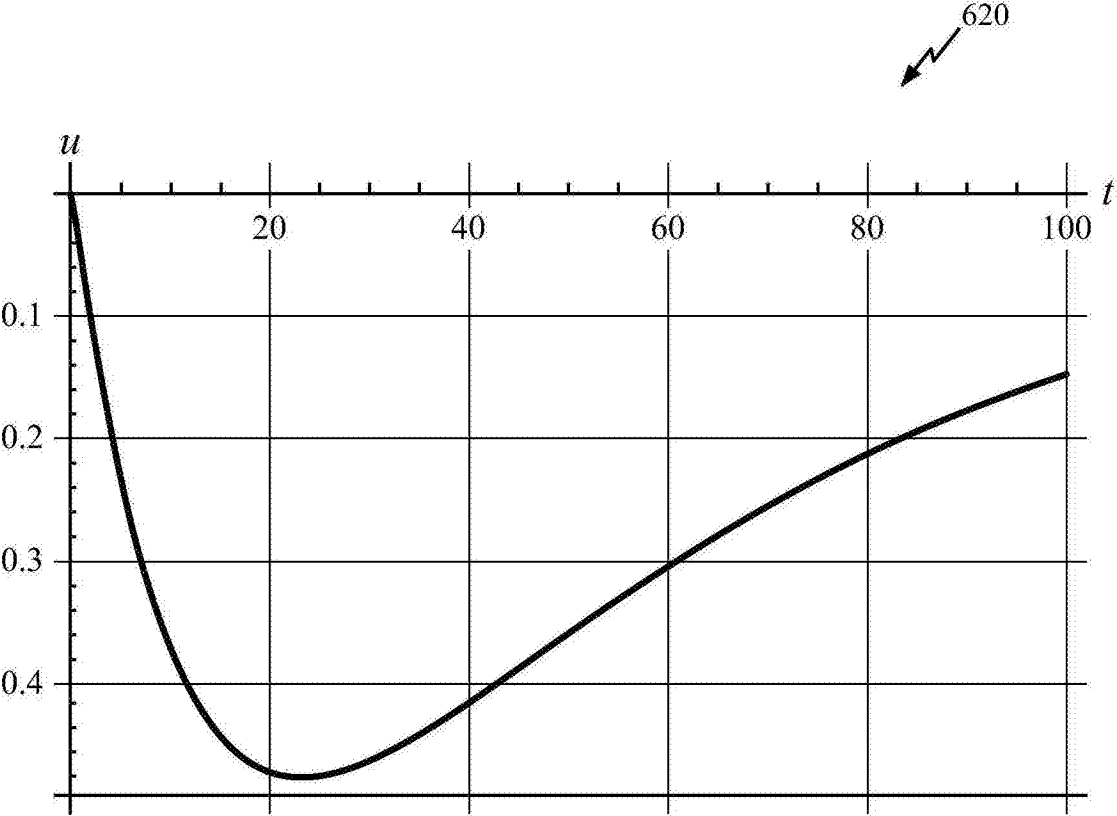


图6B

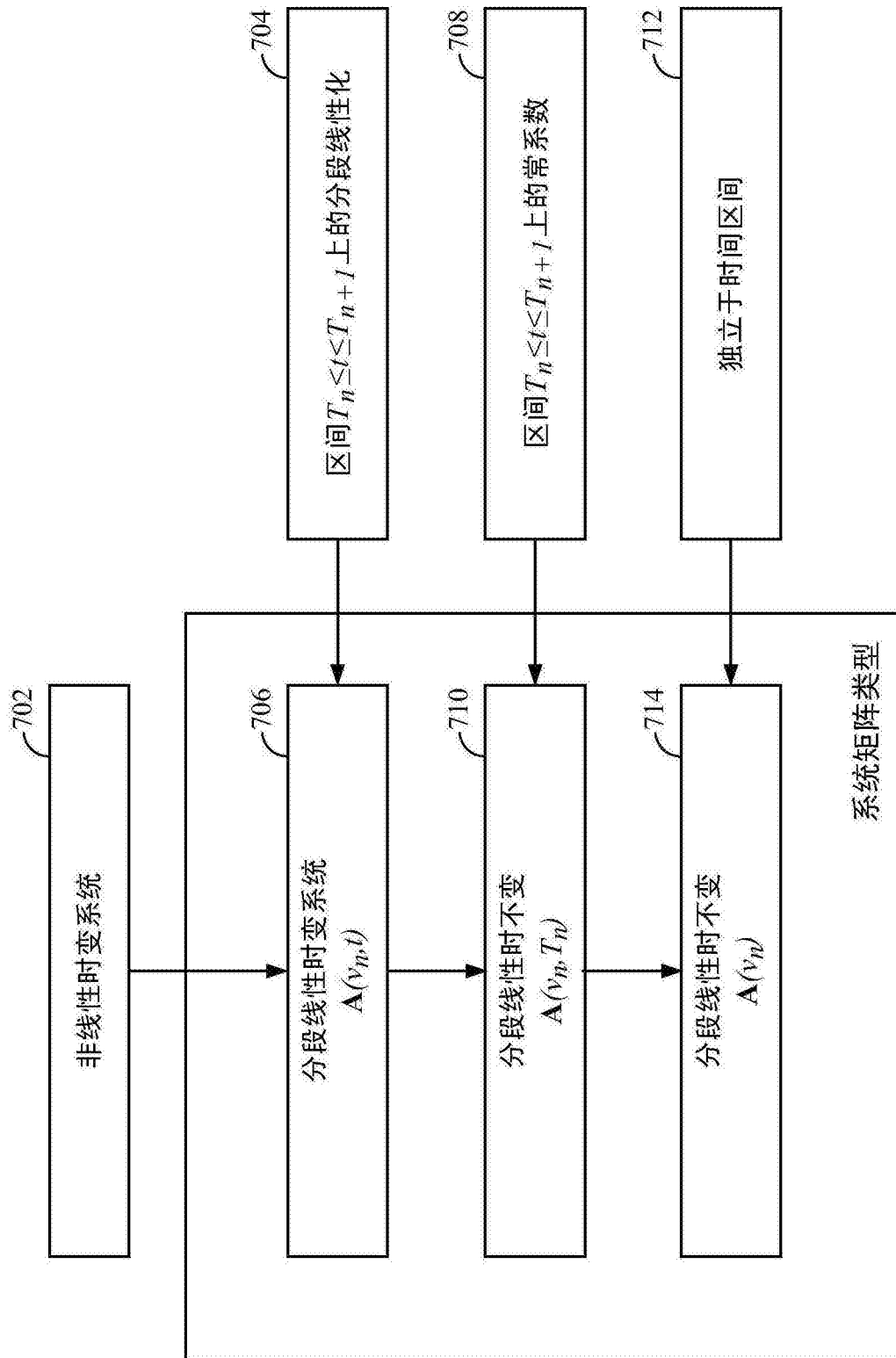


图7

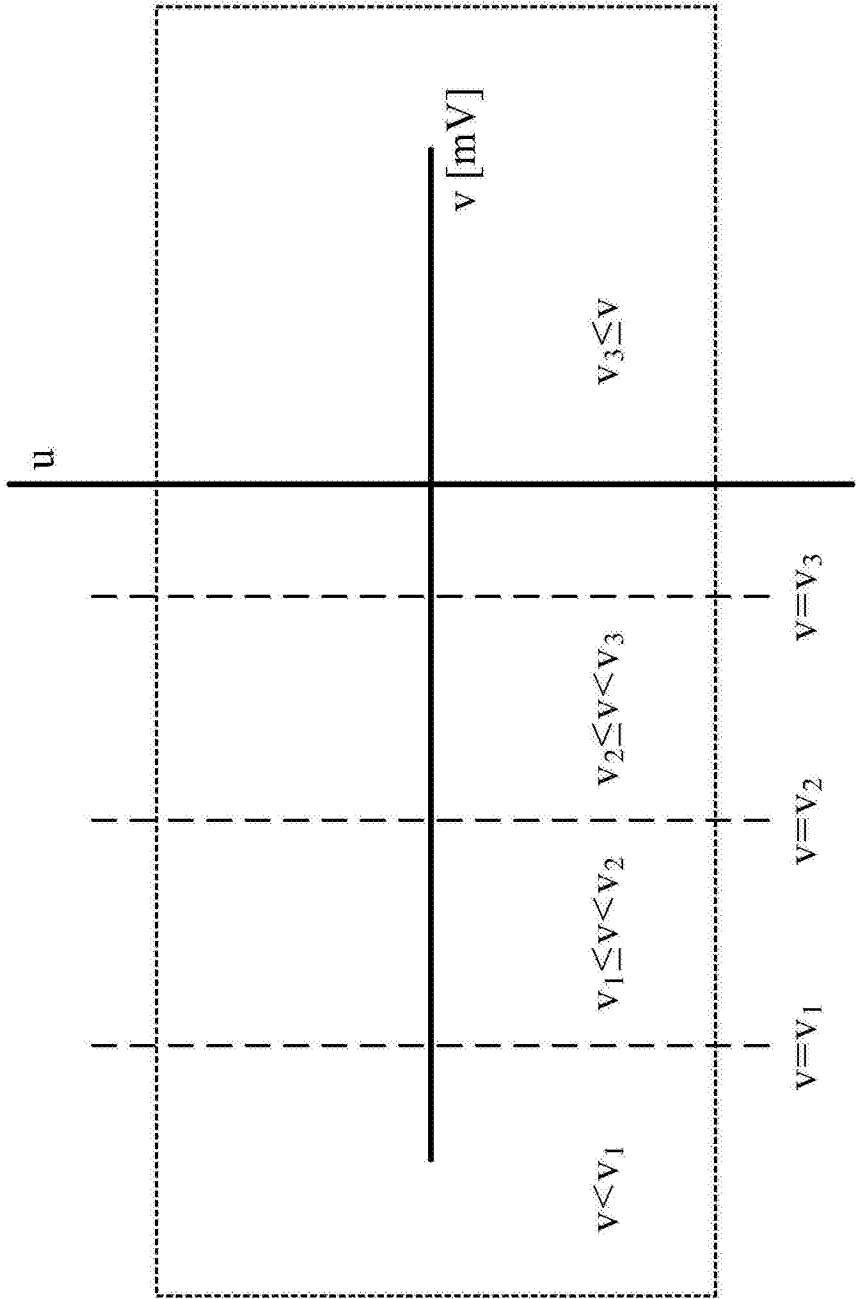


图8

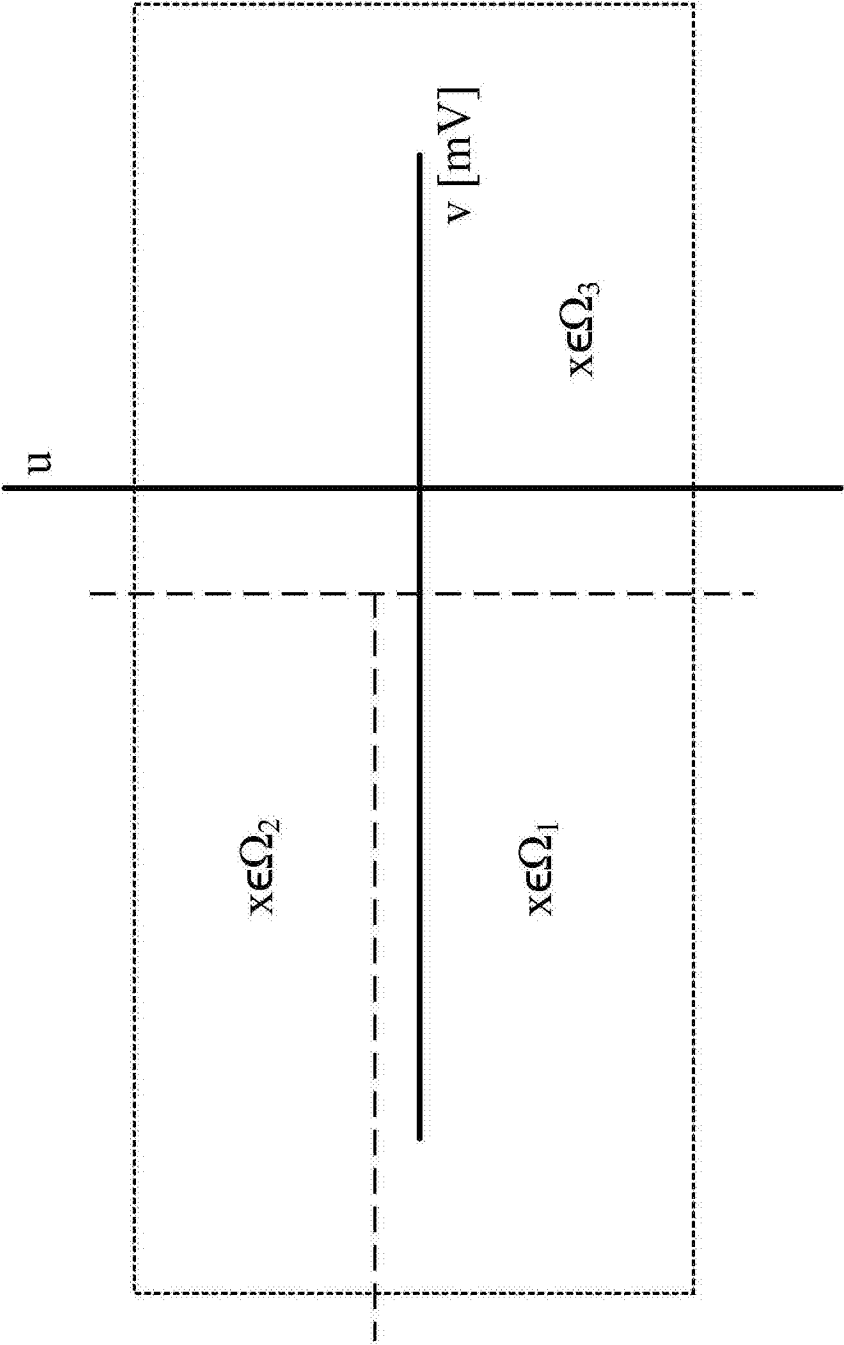


图9

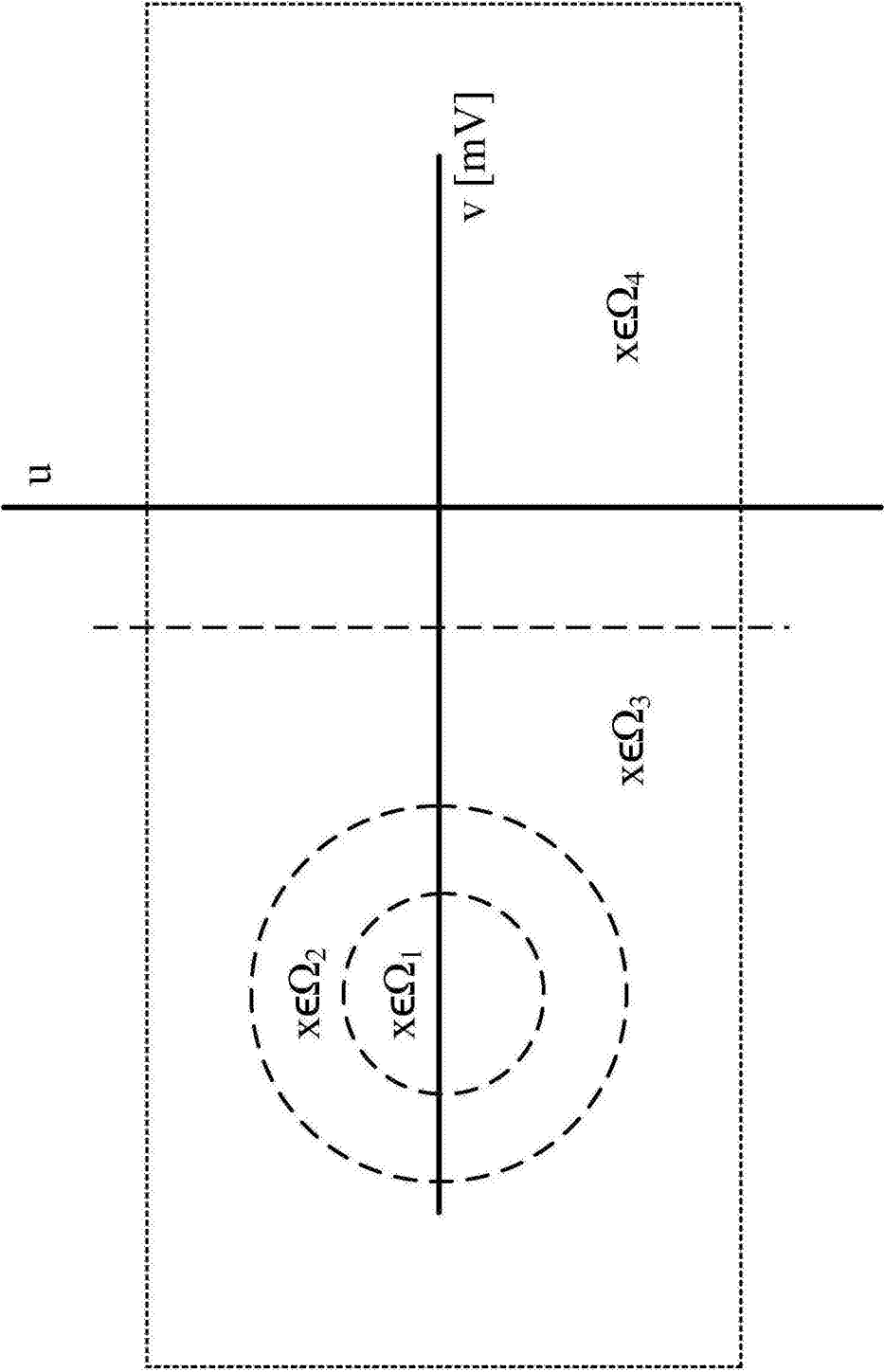


图10

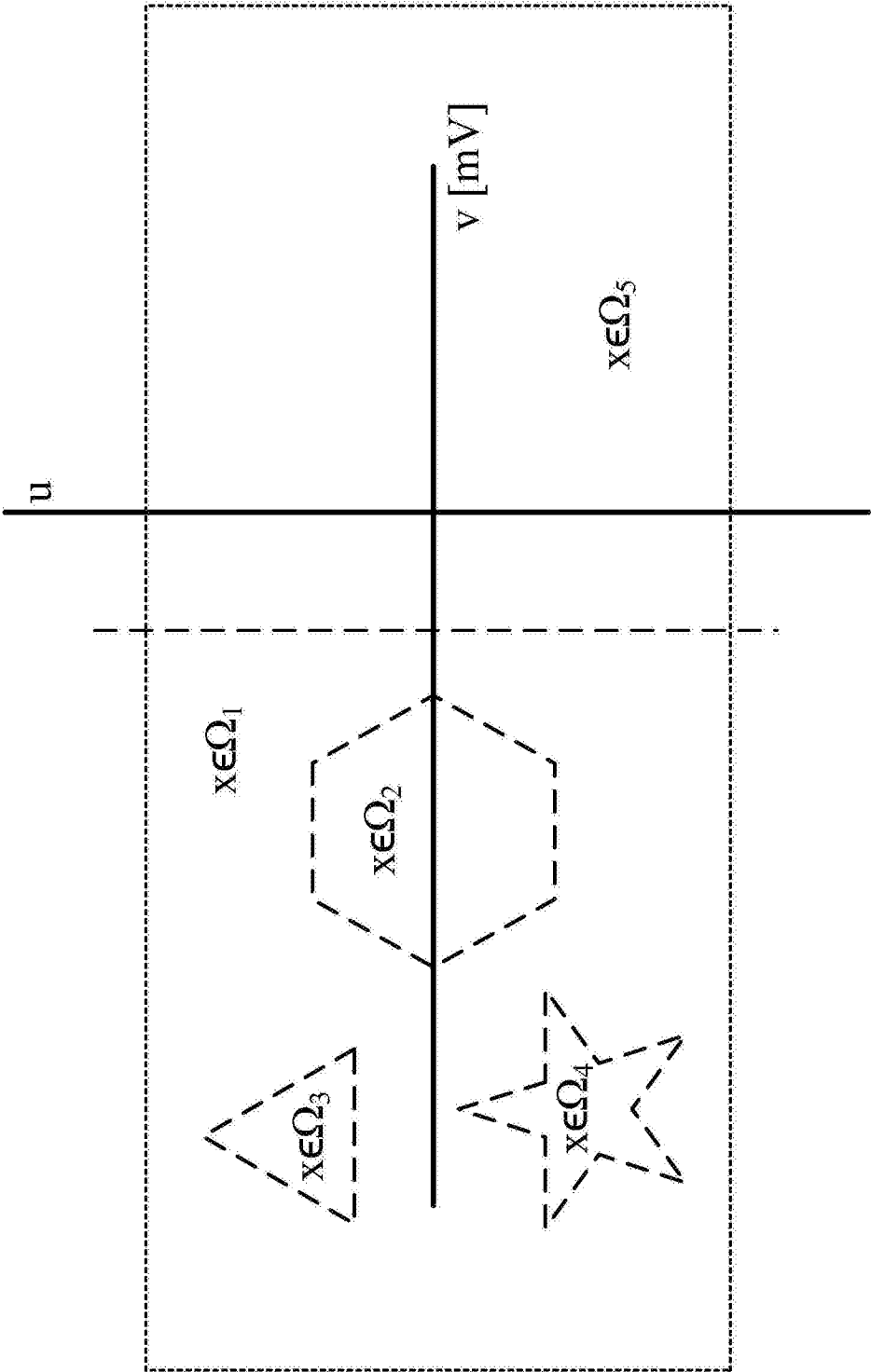


图11

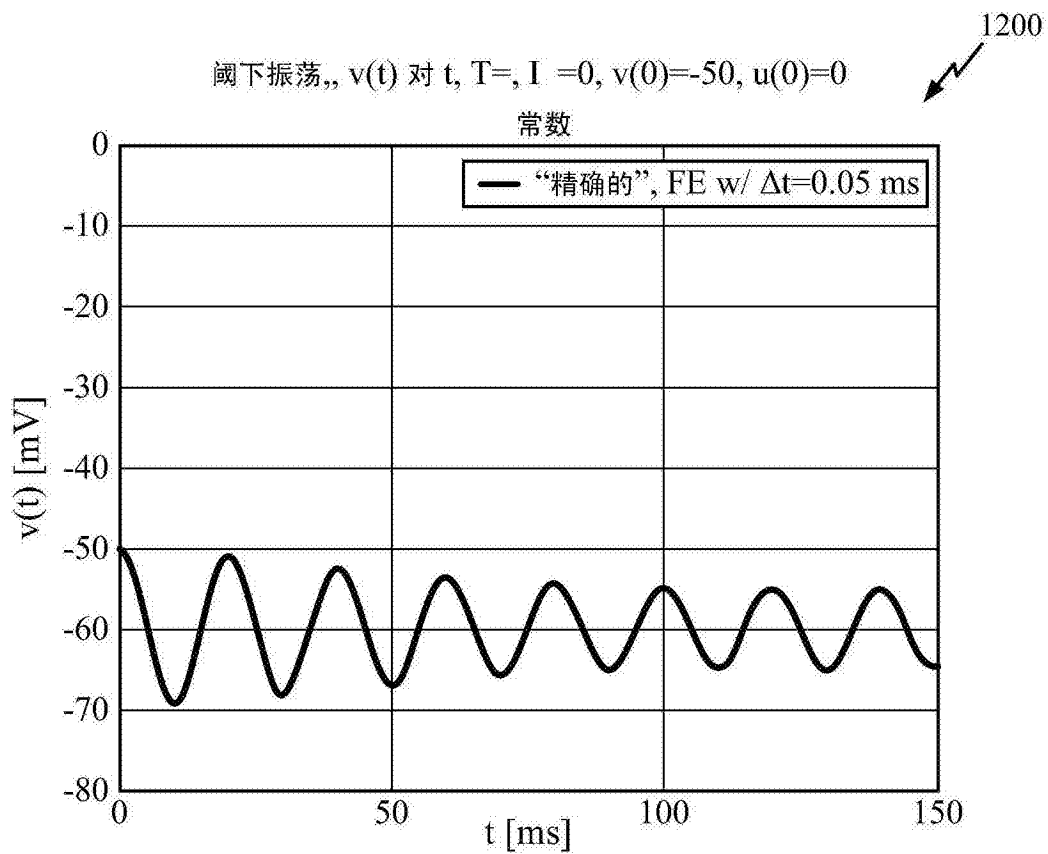


图12A

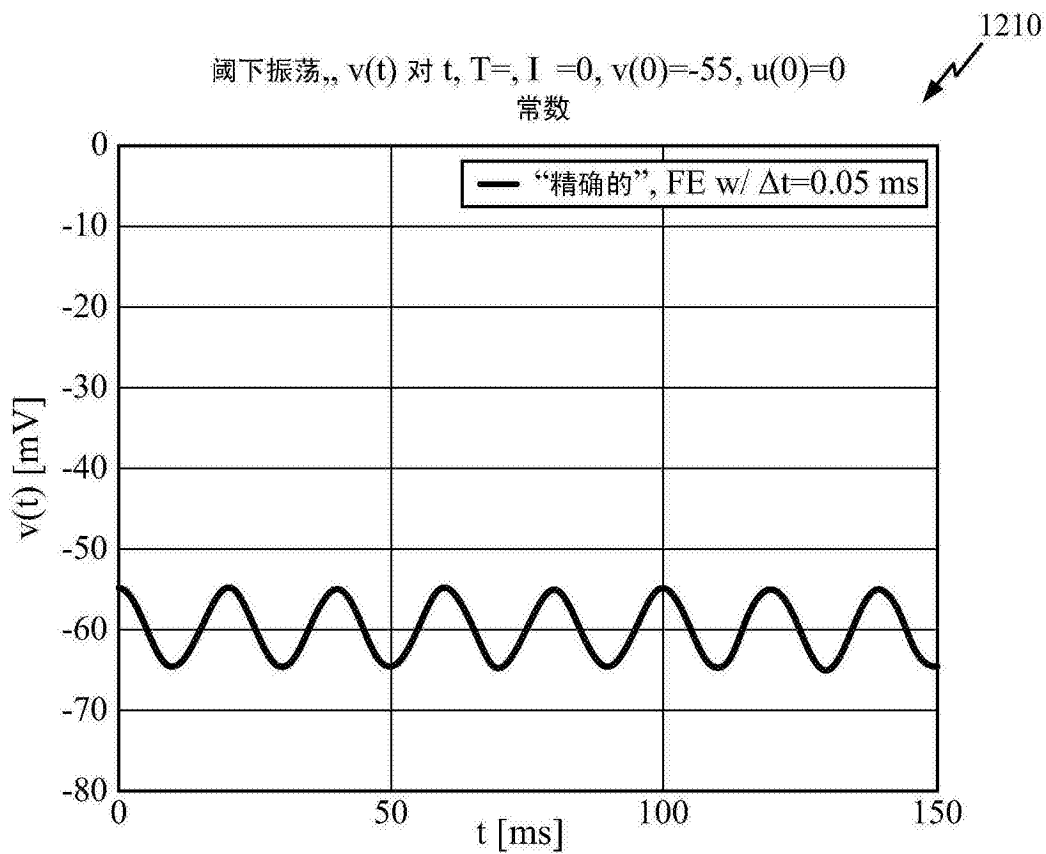


图12B

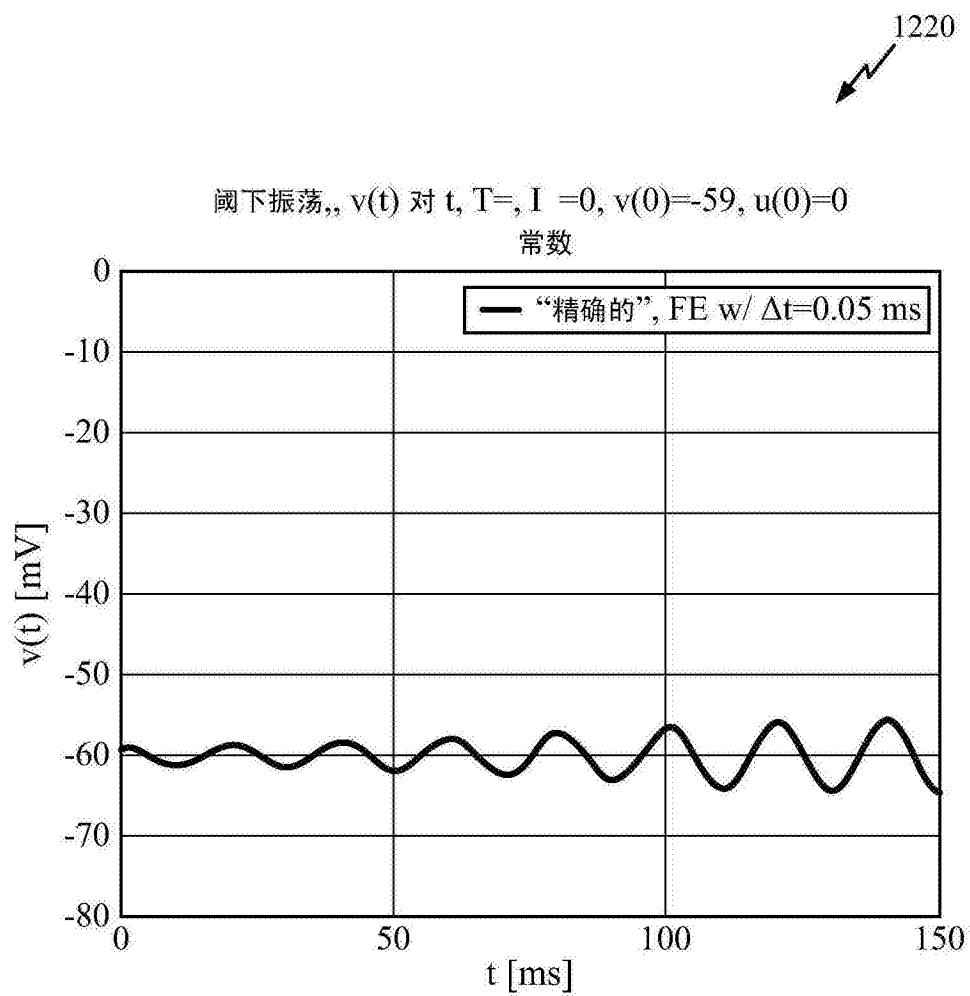


图12C

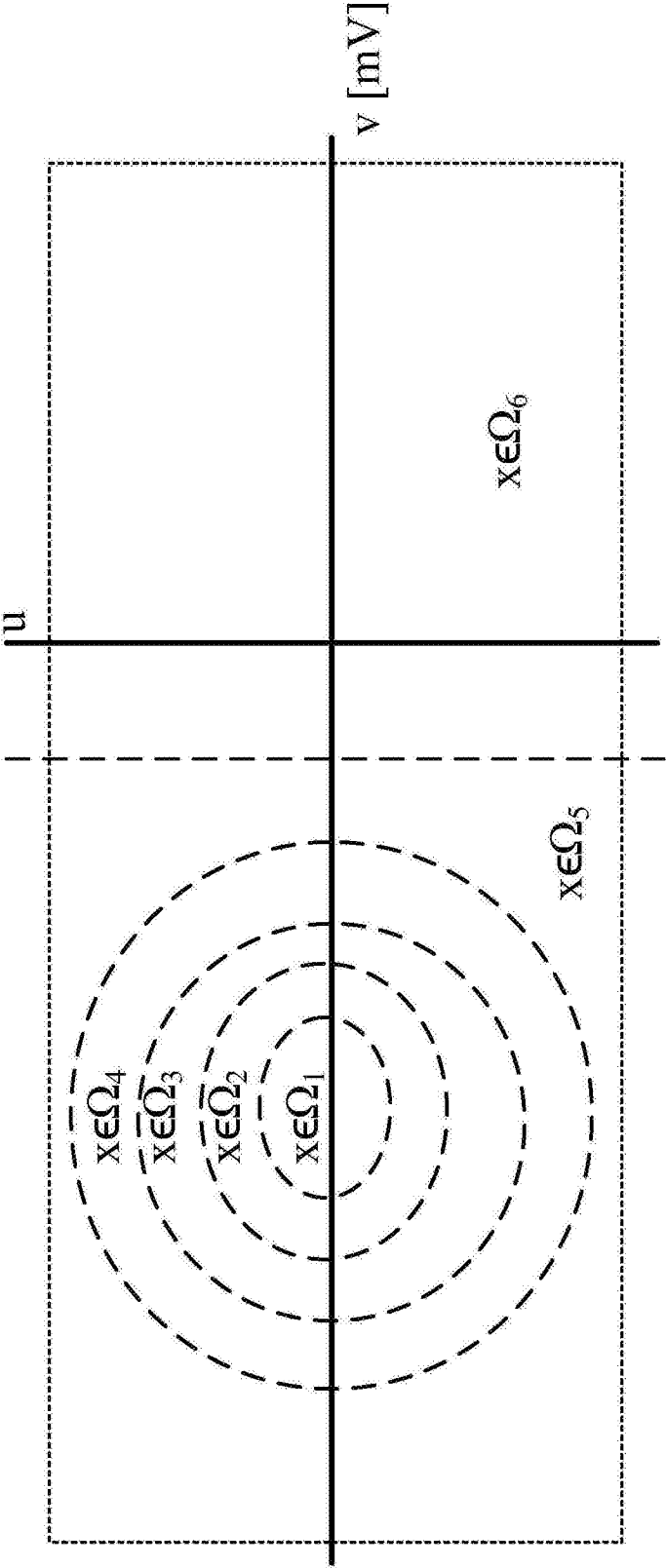


图13

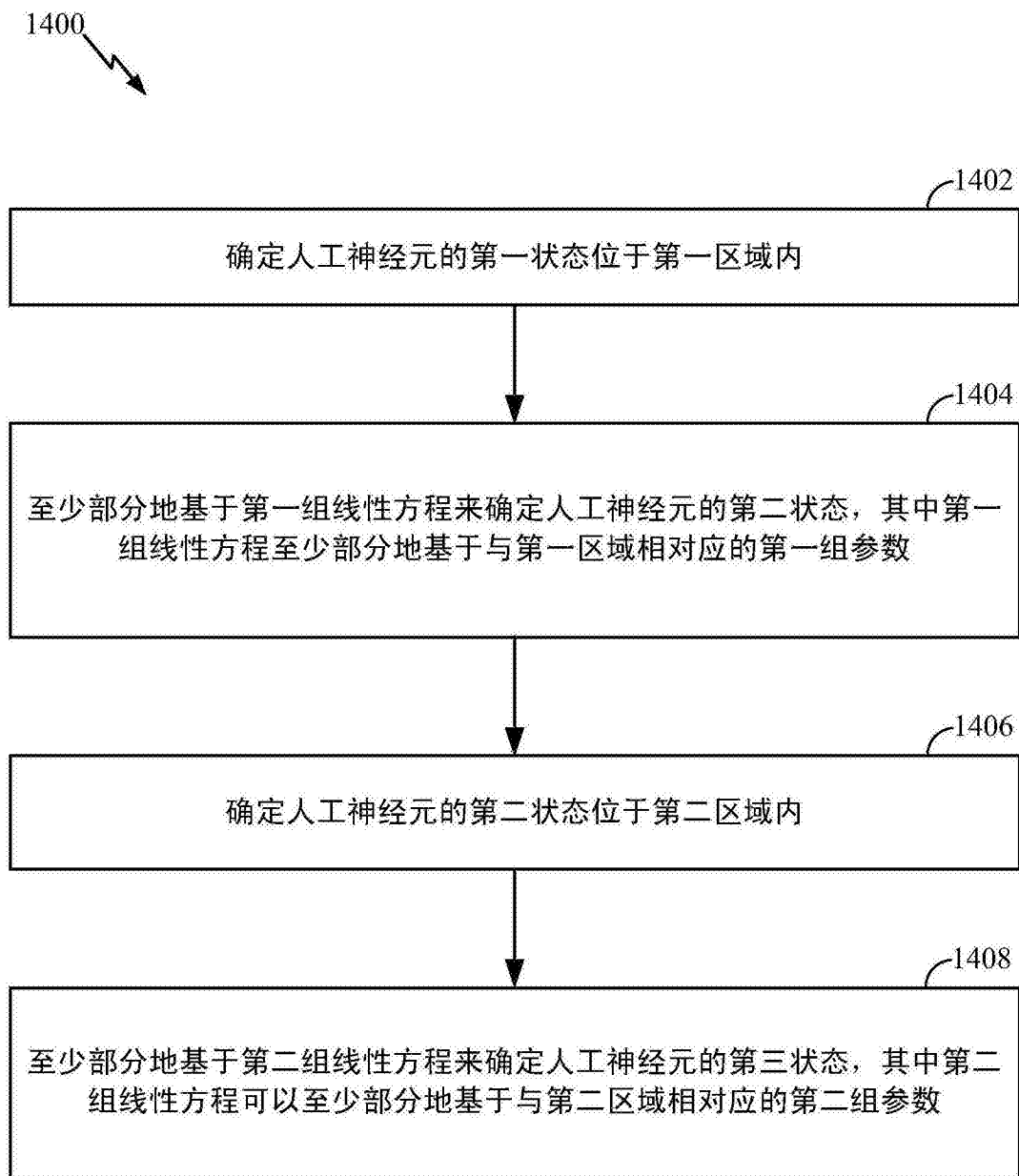


图14

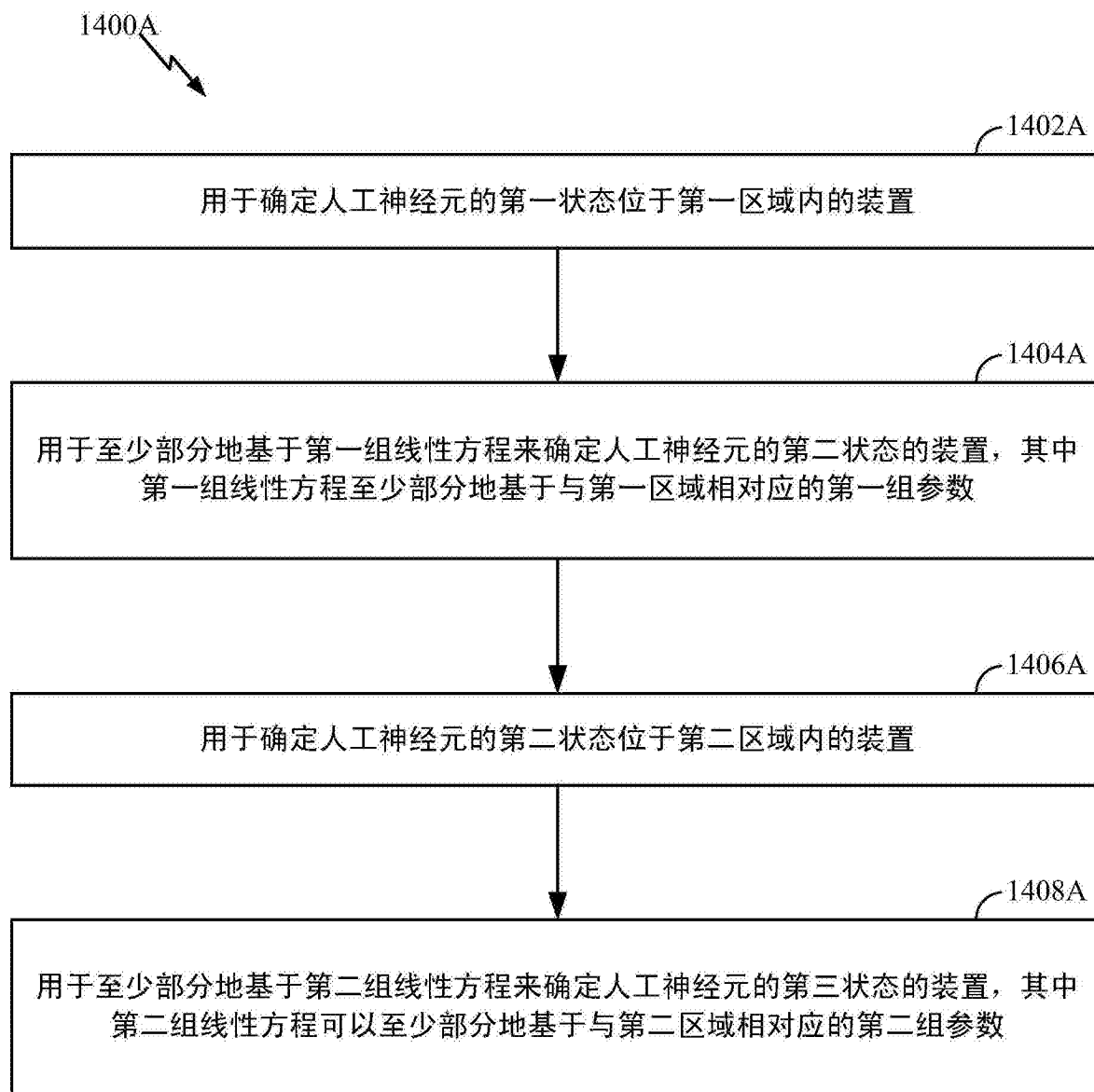


图14A

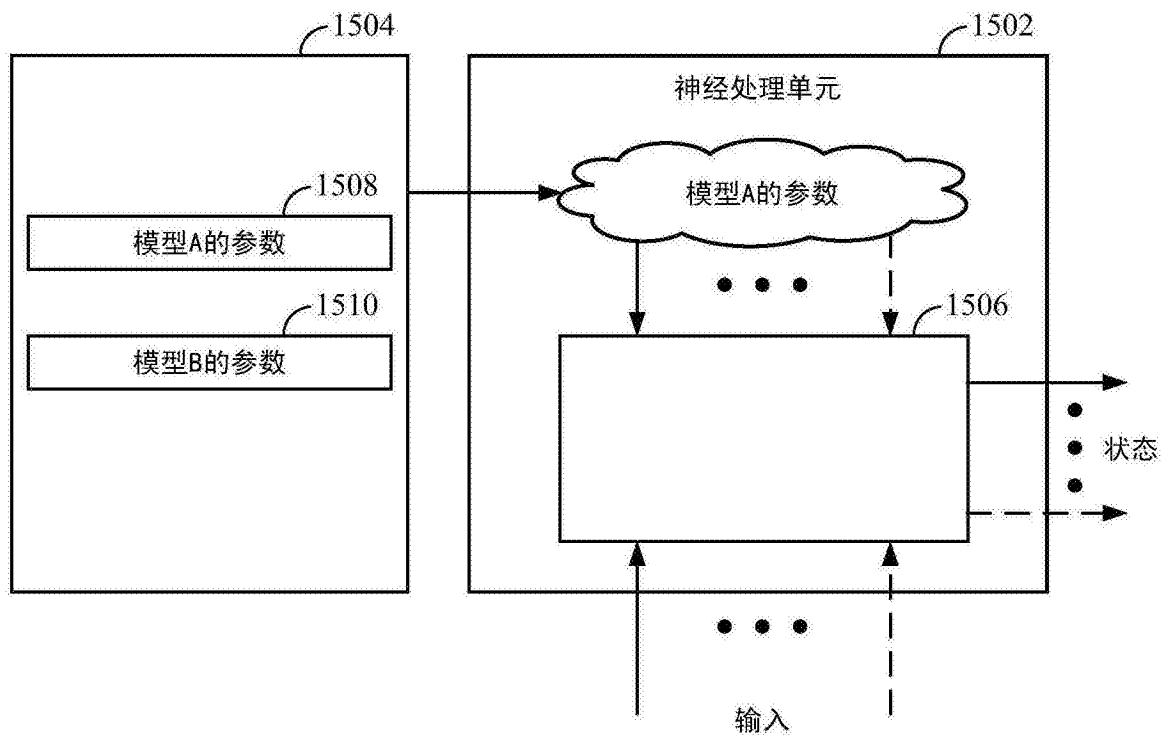


图15A

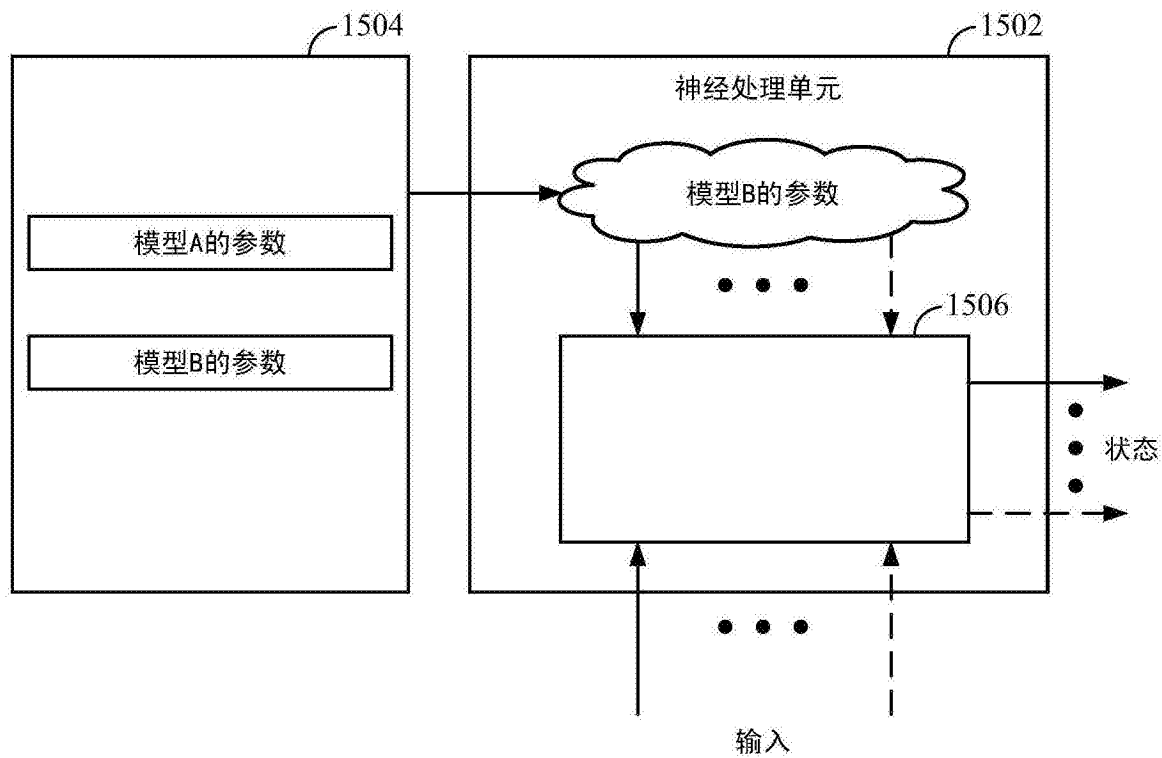


图15B

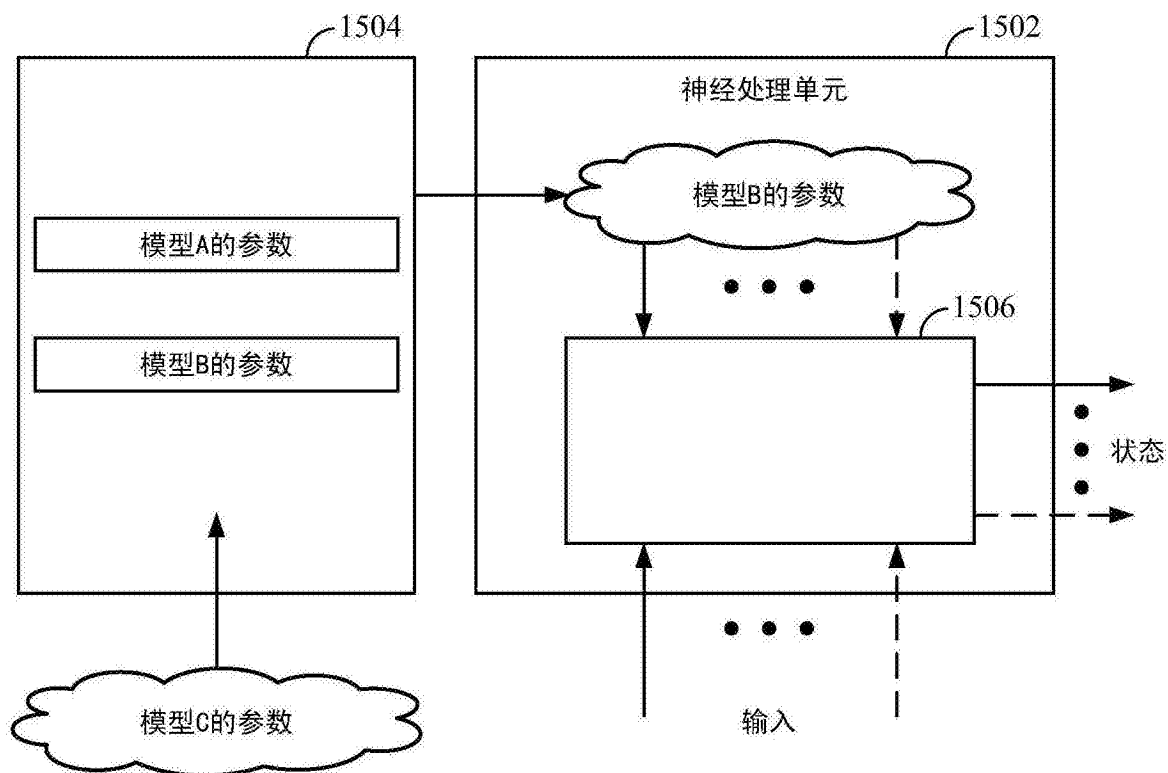


图15C

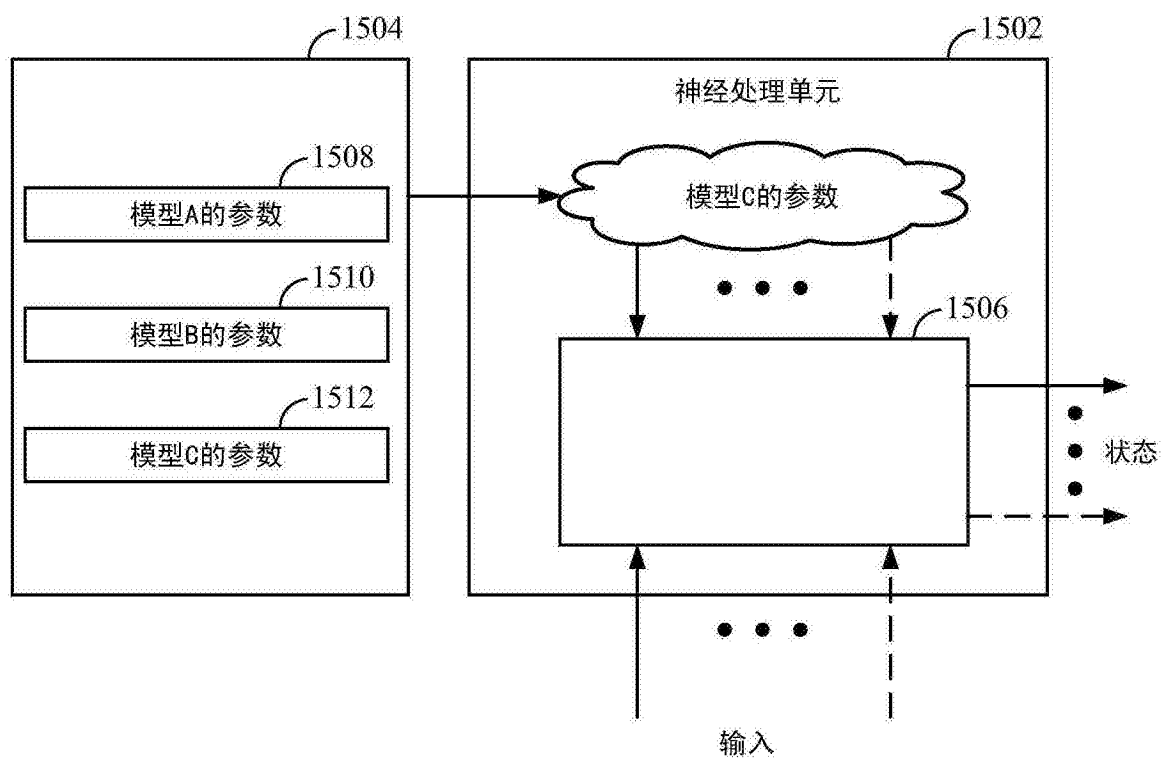


图15D

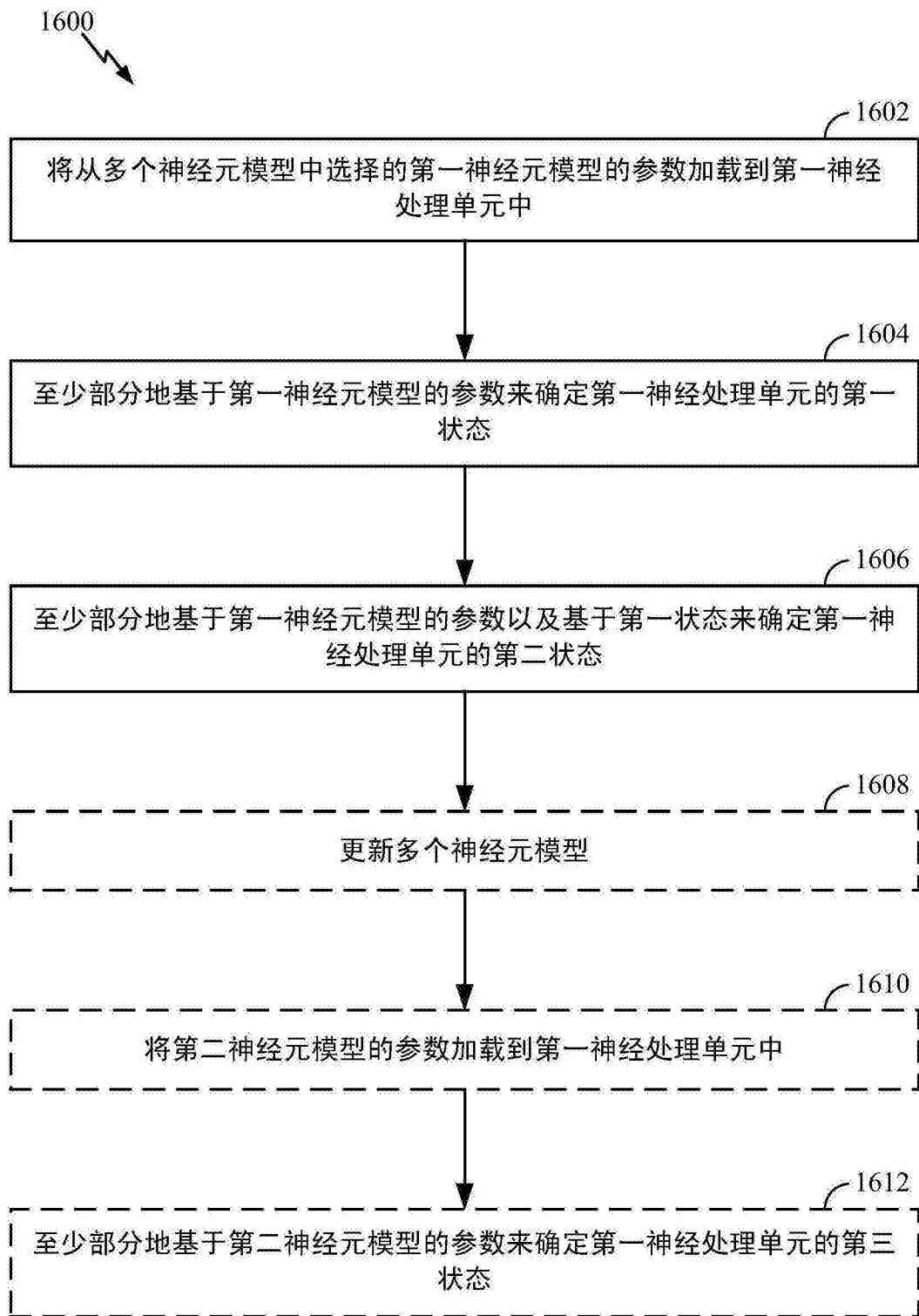


图16

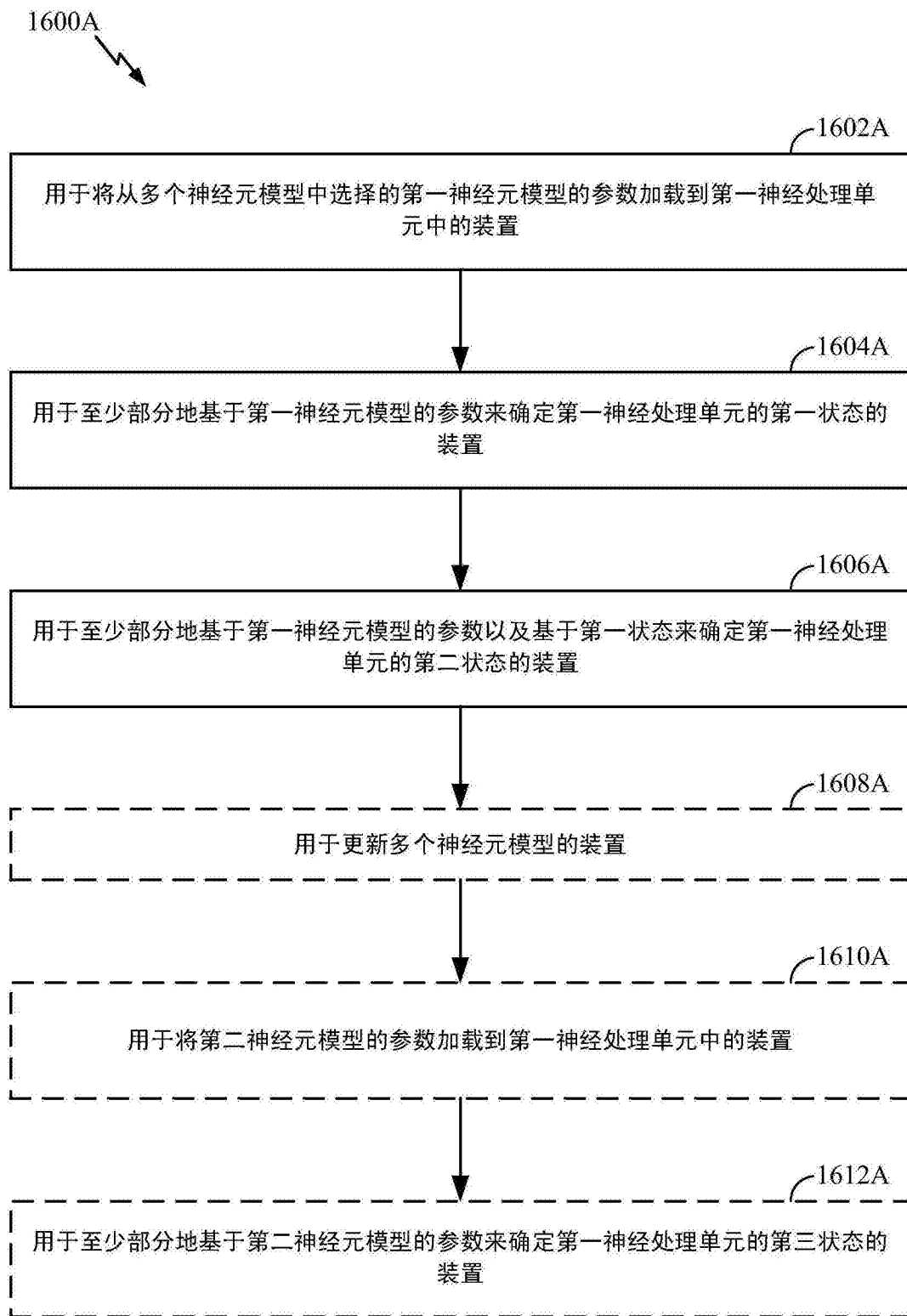


图16A

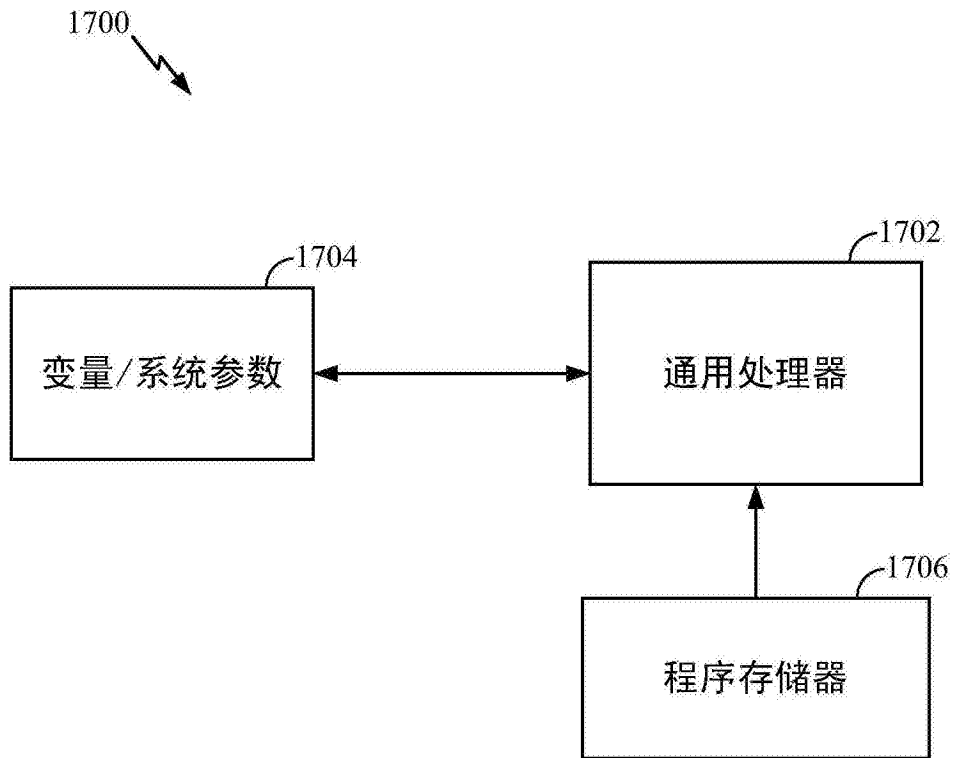


图17

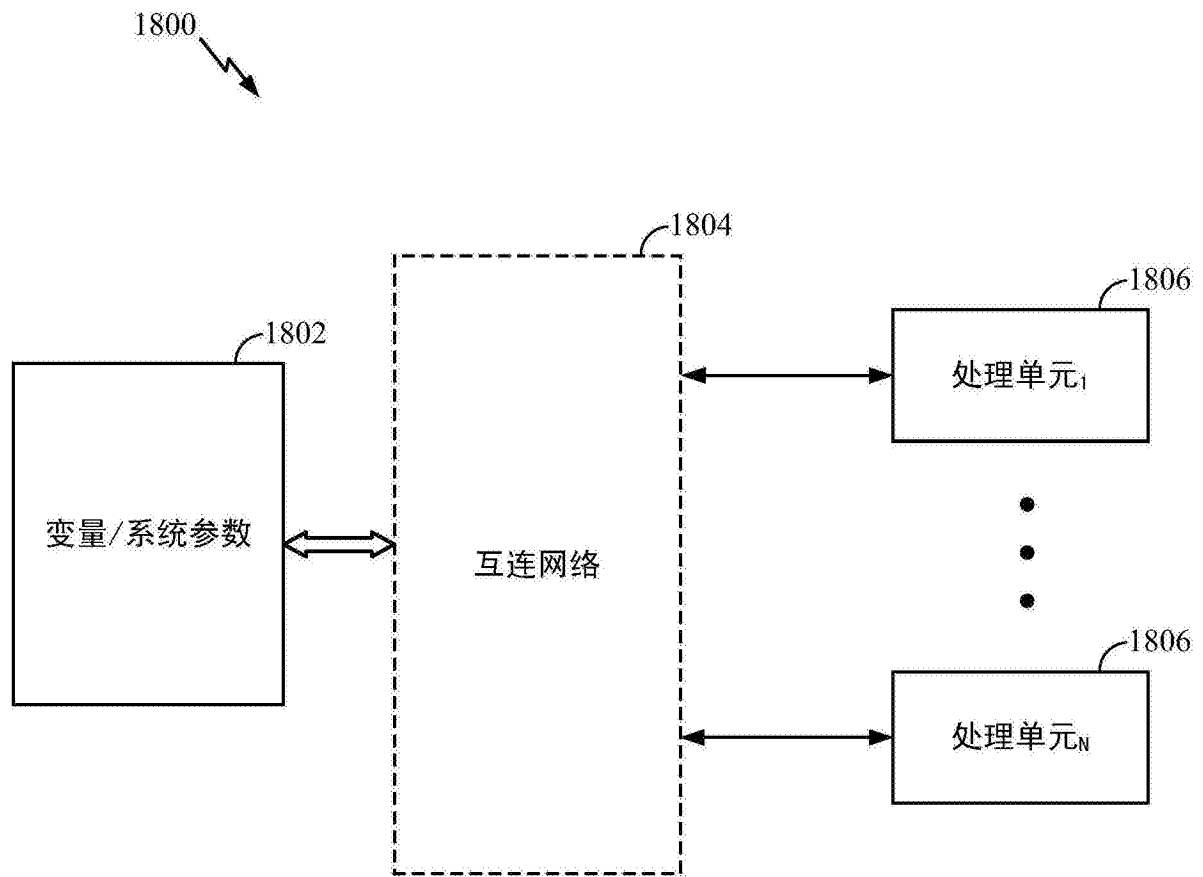


图18

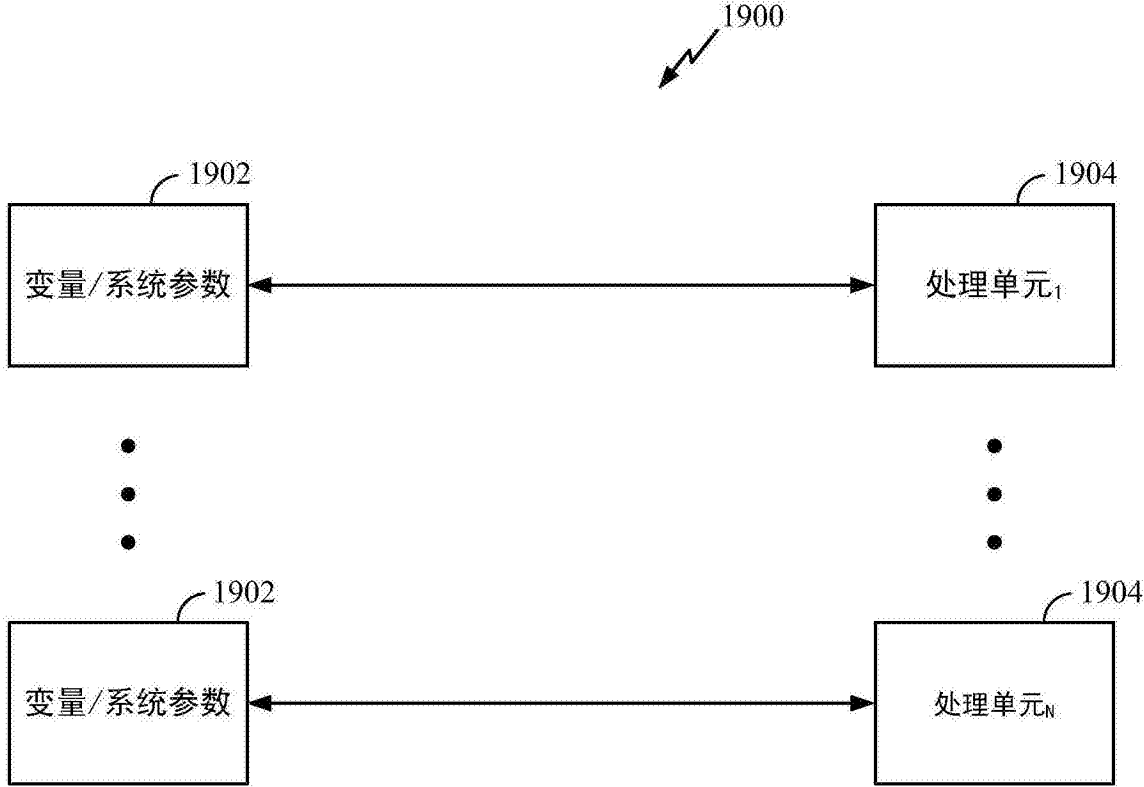


图19

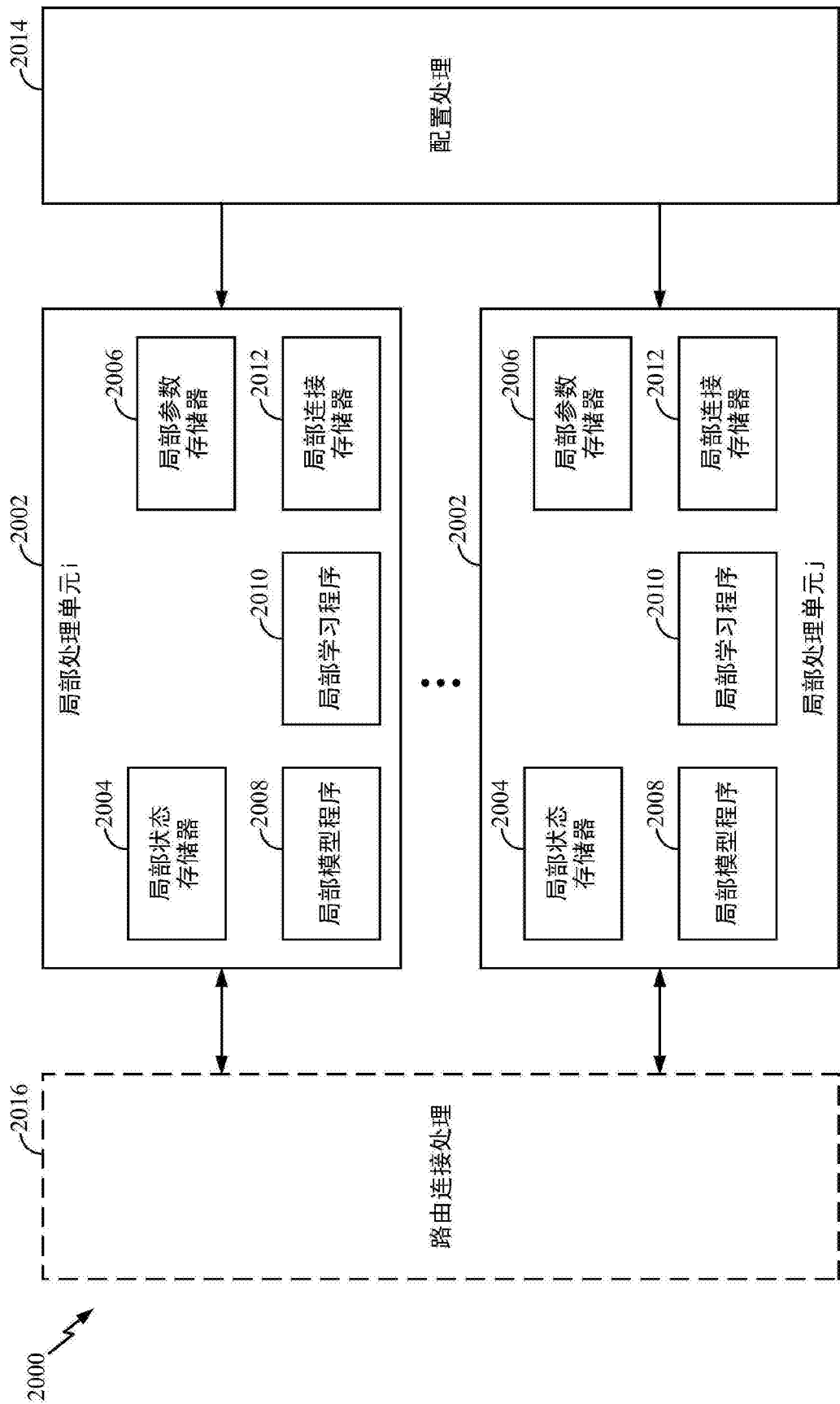


图20

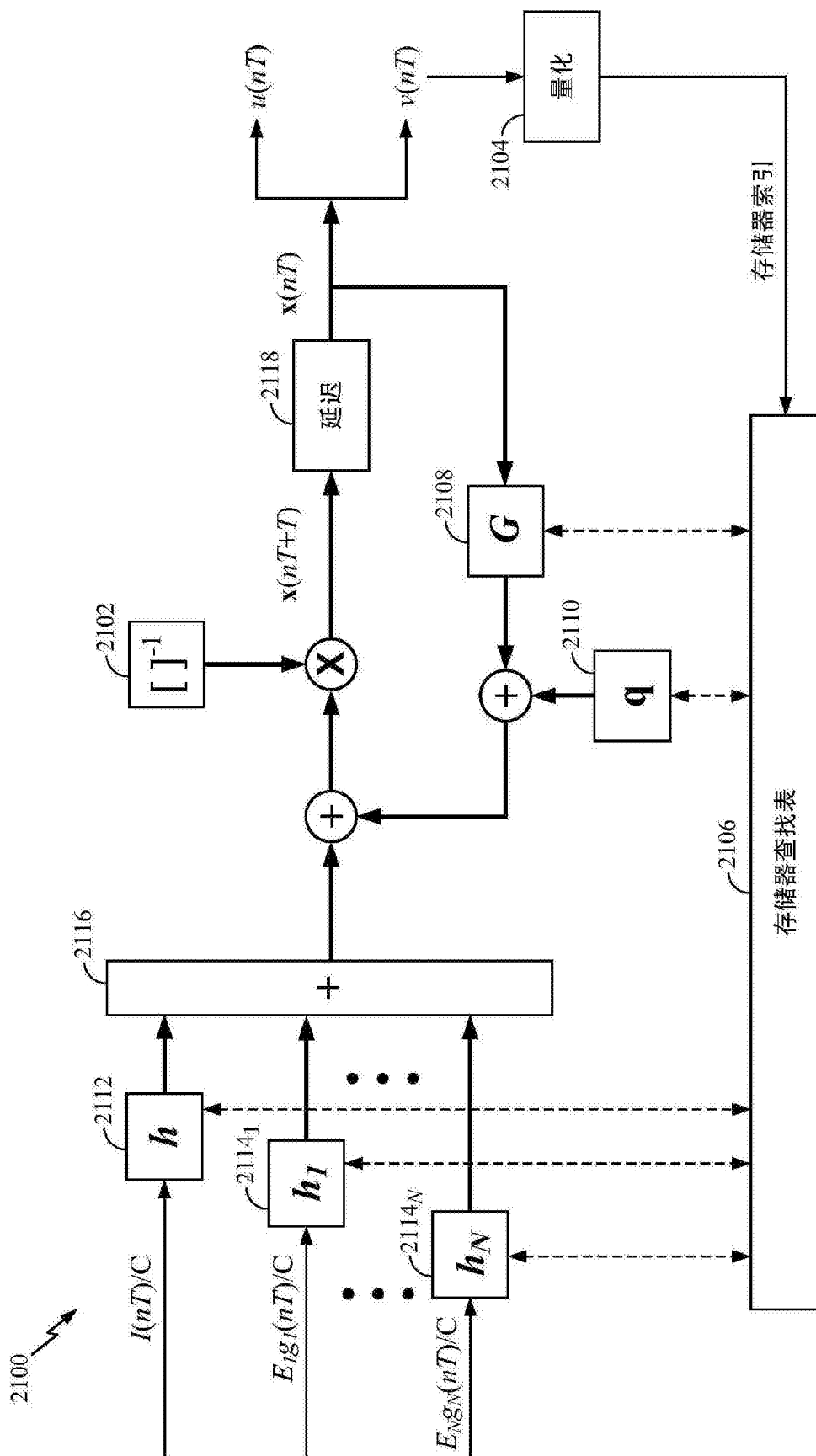


图21