

思想的两种尺度：神经动力学与计算分解主义

——Continuous Thought Machines与MAKER的对比研究

摘要： 本文对2025年NeurIPS收录的《Continuous Thought Machines》与arXiv预印本《Solving a Million-Step LLM Task with Zero Errors》（MAKER）进行系统性对比研究。前者提出了一种以神经元级时间处理与神经同步化为核心表征的新型神经网络架构，后者则基于极致任务分解与多智能体投票纠错机制，首次实现了百万步LLM推理的零错误执行。表面上看，两者分属不同的技术路线——一个向内探索神经动力学的表征潜能，一个向外构建多智能体系统的工程可靠性。然而，本文论证，两者实际上构成了当代人工智能研究中两条互补的思想轴线：**时间性与组合性**。CTM通过将时间重新嵌入计算单元内部来逼近认知的的时间性本质，MAKER则通过将任务拆解为原子步骤并用冗余投票来逼近执行的确定性。两者的交汇之处，恰恰指向了通用人工智能的核心张力：表征的动态丰富性与执行的可靠可扩展性之间如何取得平衡。

关键词： 神经动力学；极致分解；多智能体系统；时间表征；错误纠正；神经同步化

1.1 研究背景与问题意识

2025年，人工智能研究呈现出一种深刻的二元性：一方面，以Transformer架构为基础的大语言模型在越来越多标准化基准测试上逼近或超越人类水平；另一方面，研究者们日益清醒地认识到，这些模型在**长程推理、持续执行和鲁棒泛化**方面存在根本性的瓶颈。

正是在这一背景下，两篇论文从截然不同的方向回应了同一个核心问题：**当代AI系统的根本局限在哪里，突破的方向是什么？**

《Continuous Thought Machines》（以下简称CTM）给出的答案是：问题出在**表征的层次**上。现代神经网络为了计算效率而抽象掉了神经元活动的时间精细结构，这种“去时间化”的表征方式使得网络无法捕捉认知过程的内在时间性。CTM的回应是重新将时间嵌入网络的核心运算——通过神经元级模型（Neuron-Level Models, NLMs）和神经同步化表征（Neural Synchronization Representation）让网络的内部动力学本身成为计算的媒介。

《Solving a Million-Step LLM Task with Zero Errors》（以下简称MAKER）给出的答案则截然不同：问题出在**执行的规模**上。LLM的基本推理步骤并不可靠——即使错误率低至千分之一，在万步以上的任务中失败也是必然事件。MAKER的解决方案不是改进底层模型，而是通过极致任务分解将不可靠的基本单元组织成可靠的系统——这一思路与经典计算理论中的纠错码思想一脉相承。

1.2 对比的基本框架

本文的对比围绕以下三个维度展开：

维度	CTM	MAKER
核心隐喻	大脑即时间	计算即分工

主要方法论	生物启发、自下而上的动力学建模	工程驱动、自上而下的系统分解
表征哲学	连续性、涌现性、不可还原性	离散性、可组合性、可纠错性
时间观	时间是内在的、构成的	时间是外在的、序列化的
对"智能"的界定	智能即神经动力学的适应性展开	智能即可靠执行长程任务的能力
基本单元	神经元级模型 (NLM)	微观智能体 (Microagent)
主要瓶颈	计算开销与参数增长	分解的完备性与错误相关性
可扩展策略	增加内部时间步 (ticks)	增加并行投票智能体
误差处理方式	通过确定性监测实现自适应计算	通过冗余投票实现显式纠错

第二章 核心思想的对立与互补

2.1 表征的本体论：同步化 vs 分解

CTM 最核心的创新在于将神经同步化 (neural synchronization) 直接作为潜在表征。这一选择的哲学含义极为深刻。在传统神经网络中，表征是一个静态的、离散的向量——在某个时间点上，网络的隐藏状态向量是其所"知道"的全部内容。CTM打破了这一范式：表征不再是快照式的向量，而是神经元活动历史之间的内积关系：

$$S'_t = Z''_t \cdot Z''_t \in \mathbb{R}^{D \times D}$$

这一公式意味着：在CTM中，"知道什么"等价于"神经元群体以什么方式同步活动"。表征不是一个存在于某个时间点的状

态，而是一个跨越时间窗口的关系结构。这是一种**关系性表征本体论**——知识的载体不是单个神经元的活动水平，而是神经元之间的时间相关性模式。

从哲学层面来看，CTM的表征观接近于柏格森的“绵延”（durée）概念：时间不是匀质的外在容器，而是事物内在构成的要素。CTM的神经元并不在一个外部的时间序列中被动地接收输入——它们通过内部的神经元级模型展开自己的时间轨迹，并通过与其他神经元的同步关系来“表达”信息。这意味着CTM的表征是**不可还原为瞬时的**——任何快照式的观察都会丢失其本质。

MAKER的表征哲学则是彻底的**分析性和原子主义的**。在MAKER框架中，一个百万步的任务被分解为一百万个单步子任务，每个子任务由一个独立的LLM智能体负责执行。每一步的表征就是一个三元组——当前状态，上一步动作，下一步动作。这种表征方式在根本上依赖于**将复杂对象分解为简单组元的可能性**。

MAKER的出发点暗合了休谟的经验主义认识论：复杂观念可以还原为简单观念的集合，而知识就是观念之间的恒常联结。在MAKER中，一个“复杂推理”仅仅是多个“简单步骤”的线性序列，每个步骤的正确性可以通过多数投票独立验证。

这两种表征哲学形成了尖锐的对立：CTM认为复杂认知涌现于不可还原的时间动力学，MAKER则认为复杂任务可以完全分解为独立可验证的原子步骤。这一对立在哲学史上可以追溯到柏格森与实证主义、怀特海的过程哲学与罗素的分析哲学之间的争论。

2.2 计算的时间性：内在时间与外在时间

CTM与MAKER最深刻的差异之一在于它们对“时间”的处理方式。

CTM构建了一个**内在时间维度**：内部时间步（internal ticks） $\ell \in \{1, \dots, \tau'\}$ 是完全与数据时间维度解耦的。即使输入是静态图像，CTM也可以在自身生成的时间线上进行多步处理，就像一个人凝视一幅画时，目光在画面上移动，思维在处理过程中演变。CTM的每个内部tick都产生一个新的预测 y'_t ，整个网络在时间中展开其认知过程。

更精妙的是，CTM通过可学习的指数衰减因子 r_{ij} 来调节不同时间尺度上的同步化影响：

$$R'_{ij} = [\exp(-r_{ij}(\ell-1)), \dots, \exp(-r_{ij}(\ell-t))] \in \mathbb{R}^t$$

不同的神经元对可以关注不同的时间尺度——有些关注最近的几个tick，有些则整合更长的历史。这创造了一个**多层次的时间表征架构**，类似于人类认知中工作记忆与长期记忆的分化，但又远比这个二元区分更为精密和连续。

MAKER的时间观则是**纯粹序列化的**。任务的执行顺序是固定的（从步骤1到步骤s），每个步骤由不同的智能体独立执行。智能体之间没有时间上的重叠——一个步骤完成后，其输出状态传递给下一个步骤。时间是外在的、均匀的、离散的“步数”，是任务执行的载体而非构成要素。

有趣的是，MAKER论文中提到了一个与CTM隐含相关的观察：当单智能体系统执行长序列任务时，“随着context的增长，其输出变得越来越不可靠”[§ 3.1, 引文52]。这正是CTM所诊断的问题——传统神经网络（无论是Transformer还是LSTM）在处理时间序列时，时间信息仅通过位置编码或隐状态传递，缺乏真正的时间动力学结构。CTM的回应是让每个神经元都有自己的时间处理模型，而MAKER的回应则是彻底避免让任何单个智能体处理长序列。

2.3 生物隐喻与工程隐喻

CTM的灵感源是**生物神经系统**。论文明确表示其目标是“在神经抽象与生物真实性之间取得平衡”（"strike a balance between neural abstraction and biological realism"）。
具体而言：

- **神经元级模型（NLM）**是对生物神经元复杂性的抽象模拟。每个神经元拥有自己的参数化模型来处理输入历史——这对应于生物神经元多样化的时间整合特性。
- **神经同步化**直接类比生物大脑中神经元群体的同步放电现象。在生物神经科学中， γ 频段同步被认为与特征绑定、注意力和意识等高级认知功能密切相关。
- **涌现的动态行为**（如低频行波、注意力图的类人扫视模式）被视为模型“像大脑一样工作”的证据。

但CTM对生物隐喻的运用是**有选择性的**。它没有试图模拟动作电位、突触可塑性、神经递质调节等微观生物机制，而是在一个适合梯度下降学习的抽象层次上捕捉“时间处理”与“同步化”这两个计算原则。这是一种**功能主义**的生物启发：保留生物神经系统在计算层面的关键特性，但抛弃其物理实现细节。

MAKER的灵感源则是**工程系统和计算理论**：

- **任务分解**借鉴了软件开发中的微服务架构（microservices），论文明确列出了两者之间的八条对应关系。
- **纠错机制**直接类比通信工程中的信道编码（Shannon, Hamming码）、量子纠错和生物系统中的DNA修复机制。
- **First-to-ahead-by-k投票**源于序贯概率比检验（SPRT）和赌徒破产问题，是经典的统计决策过程。
- **红标机制（Red-flagging）**则类似于工业质量控制中的异常检测。

MAKER的核心隐喻可以概括为：**LLM智能体是一种新型的“计算基板”**，正如经典计算中我们将不可靠的晶体管组织成可靠的逻辑门，再将逻辑门组织成可靠的CPU，MAKER将不可靠的LLM调

用组织成可靠的推理过程。论文中使用的"linguistic computing"（语言计算）和"language-based algorithms"（基于语言的算法）等术语明确指向了这一类比。

这两种隐喻的选择不仅仅是修辞上的偏好，它们深刻地塑造了两篇论文的问题框架、方法论选择和成功标准。CTM以"是否展现出更丰富的神经动力学"作为验证其方法有效性的一个重要维度，MAKER则以"能否在百万步级别实现零错误"作为衡量标准。前者关心的是内部过程的**丰富性**，后者关心的是外部结果的**可靠性**。

第三章 技术架构的深度比较

3.1 核心运算机制

CTM: 神经动力学作为运算

CTM的运算流程可以概括为以下循环：

1. **突触模型** f_{syn} 在共享的D维潜在空间中对当前状态 z'_t 和注意力输出 o'_t 进行处理，产生预激活 a'_t
2. 最近的M个预激活构成历史矩阵 $A'_t \in \mathbb{R}^{D \times M}$
3. 每个神经元的**私有NLM** g_d 处理其对应的M维预激活历史，产生后激活 z''_{dt}
4. 所有后激活构成 $Z''_t \in \mathbb{R}^{D \times t}$
5. **神经同步化** $S'_t = Z''_t \cdot Z''_t \in \mathbb{R}^{D \times D}$ 作为表征被投影为输出和注意力查询

这一架构的关键在于：**每个神经元都有自己私有的、不可共享的时间处理模型**。在标准神经网络中，同一层的神经元共享激活函数；在CTM中，每个神经元的NLM参数是独立的（深度

为1的MLP，宽度为 D_{hidden} ）。这意味着D个神经元有D个不同的时间处理"视角"。

隐藏在这种设计背后的哲学假设是：**时间整合的个体差异是信息处理的关键资源**。生物大脑中的神经元确实表现出极大的多样性——不同神经元有不同的离子通道组成、不同的时间常数、不同的突触可塑性规则。CTM通过私有NLM参数捕捉了这一多样性，而神经同步化则将这种多样性整合为统一的表征。

从计算复杂度的角度看，NLM的参数数量与 $D \times D_{\text{hidden}}$ 成正比（每个神经元一个单独的MLP），这意味着CTM的参数规模随宽度线性增长。这在实践上是可行的，但也构成了一个显著的开销——正如论文承认的那样，"NLMs increase parameter counts compared to standard activation functions" [§ 8]。

MAKER: 极致分解与投票纠错

MAKER的运算流程基于三个核心组件：

1. **极致智能体分解 (MAD)**：将一个s步任务分解为s个子任务，每个子任务由一个独立的LLM智能体执行。每个智能体的context被严格限制在"执行其单一步骤所需的足够信息"范围内。
2. **First-to-ahead-by-k投票**：对于每个子任务，独立采样多个候选回答，当某个候选的得票超过其他所有候选k票时，该候选被选中。数学上，如果正确候选的单次采样概率为 p (>0.5)，则经过该投票过程后，正确候选被选中的概率为：

$$p_{\text{full}} = (1 + ((1-p)/p)^k)^{-s/m}$$

其中 m 是每个子任务包含的步骤数（在MAD中 $m=1$ ）， s 是总步数。

3. **红标机制 (Red-flagging)**: 检测LLM输出中的异常信号 (过长响应、格式错误), 将这些响应直接丢弃而不参与投票。其理论依据是: 不良行为在LLM中具有相关性——一个格式错误的回答很可能内部推理也是错误的。

MAKER的核心洞察在数学上表达为: 当 $m=1$ 时, 系统的期望成本为 $E[\text{cost}] = \Theta(s \ln s)$, 呈**接近线性的缩放**, 而当 m 增加时, 成本呈指数级增长。这意味着将任务分解到极致粒度并采用投票纠错, 其效率远高于让单个智能体处理多个步骤。

3.2 信息流动方式

在CTM中, 信息流动是**循环的、并行的、整体性的**。每个内部tick中, 所有神经元同时更新自己的活动, 信息通过突触模型 (全连接变换) 和注意力机制在神经元之间传播。同步化矩阵 S'_t 编码了所有神经元对之间的时间相关性信息, 是一种全连接的二阶统计量。信息流动的拓扑结构是**完全图**——每个神经元的活动最终会影响所有其他神经元的表征。

在MAKER中, 信息流动是**前馈的、局部的、线性链式的**。每个智能体只接收前一步的输出状态, 生成当前的行动和下一状态。虽然不同智能体在投票层面上是并行的 (同一个步骤的多个采样可以并行生成), 但步骤之间的依赖关系是严格线性的。信息流动的拓扑结构是**有向链**——每个节点只与其直接前驱和后继交换信息。

这两种信息流动方式的差异体现了两种根本不同的“计算观”:

- CTM的计算观是**网络中心式**的: 计算发生在网络的整体动力学中, 表征是分布式和涌现性的。
- MAKER的计算观是**节点中心式**的: 计算发生在离散的节点 (智能体) 内部, 系统层面的可靠性来自节点间的关系组织。

3.3 自适应计算与灵活性

两篇论文都涉及"自适应计算"的概念，但实现方式截然不同。

CTM的自适应计算是其内部神经动力学的**天然副产品**。网络在每个内部tick产生一个预测 y'_t ，并计算该预测的确定性 C_t （1减去归一化熵）。最终的损失函数同时考虑了两个动态选择的tick：最小损失的tick t_1 和最大确定性的tick t_2 ：

$$\mathcal{L} = (\mathcal{L}_1 + \mathcal{L}_2) / 2$$

这一损失函数设计使CTM能够为不同的数据点分配不同数量的内部计算资源——对于简单的输入，确定性可能较早达到阈值，网络可以"提前停止"；对于复杂输入，网络继续处理直到形成更确定的判断。论文的实验表明，CTM在ImageNet-1K上确实表现出这种自适应行为：它在图像上"徘徊"（hover）以收集信息，确定性达到0.8以上时提前输出。

MAKER的自适应计算则体现为**投票过程的动态终止**。理论上，每个步骤需要至少 $k+1$ 个样本才能做出决定，但实际的采样次数取决于候选答案之间的竞争情况。论文中的实验显示（Figure 8），大部分步骤在3-4轮投票后即可决定，但少数"困难"步骤需要多达18轮投票。这种自适应性的来源不是内部动力学，而是外部统计过程（投票的竞争性收敛）。

这一对比揭示了一个深层次的差异：CTM的自适应性是**认知性的**——网络"知道"自己知道什么以及知道得有多清楚；MAKER的自适应性是**统计性的**——系统通过群体决策的收敛速率来判断单个步骤的难度，但没有任何单个智能体具有元认知能力。

第四章 哲学维度的展开

4.1 整体论与还原论的AI版本

CTM与MAKER的对比可以理解为**整体论与还原论**之争在AI领域的当代版本。

CTM采取的是强整体论立场：认知过程不可还原为独立步骤的线性组合，因为时间动力学本身就是认知的构成性要素。一个"步"的意义取决于它在整个动力系统演化中的位置——正如一个音符在旋律中的意义取决于其前后关系。CTM中的单个tick没有独立的意义，其价值只能在其与整个时间序列的关系中才能理解。

这一立场与经典的整体论认识论一脉相承——从奎因的"信念之网"到库恩的"范式"，整体论者始终认为，理解一个事物的前提是理解它所在的整个系统。在CTM中，同步化矩阵 S'_t 就是一个整体论表征：它编码的不是单个神经元的活动，而是所有神经元两两之间的时间相关性——这是一个二阶的整体性质。

MAKER采取的是激进的还原论立场：只要分解足够细致、纠错足够有效，复杂任务可以完全被还原为原子步骤的总和。论文中"a solution to a problem that is not solvable by a monolithic single-agent system"这一表述揭示了其核心信念：系统层面的能力可以通过分解与组合从不可靠的组件中涌现。

MAKER的还原论可以追溯到经典人工智能的"物理符号系统假说"——智能可以被实现为符号的操纵，而复杂的符号操纵可以被分解为简单符号操纵的组合。在MAKER中，LLM充当的正是基本符号操纵器的角色，而多智能体投票系统则是将这些基本操作组合为可靠长程推理的"语法"。

4.2 涌现与设计：两种"智能"概念的对话

CTM和MAKER对"智能是什么"有着不同的隐含假设。

在CTM中，智能是**涌现的**——它不是被显式编程或设计出来的，而是从神经动力学的复杂相互作用中自发产生的。论文多次使用"emergent"一词来描述CTM的行为：注意力图中类似人类

扫视的"注视"行为、低频行波、自适应计算时间——这些都是未经过显式设计却从动力学中自然涌现的现象。智能在CTM中是一个**过程**，而不是一个**产物**。

这与复杂系统科学的基本信念一致：在适当的条件下，简单规则可以产生复杂行为，而这种复杂行为可能具有适应性、创造性和智能性。CTM的架构提供了一个"适当的条件"——神经元级时间处理、同步化表征和注意力机制的组合——然后让智能作为动力学现象从中涌现。

在MAKER中，智能是**设计的**——它体现为精心设计的分解策略、投票机制和红标规则的组合。MAKER不是一个让其组件自主演化出智能行为的系统，而是一个通过明确设计将不可靠组件组合为可靠系统的工程方案。论文中对错误的分类和处理方式（预测格式错误、过长响应）体现了典型的工程设计思维：**识别故障模式，设计针对性的缓解措施**。

有趣的是，MAKER中也存在着涌现的成分：当投票过程收敛时，整体系统的可靠性远高于任何单个组件的可靠性——这是典型的涌现性质。但MAKER在哲学上更接近于**设计认识论**：智能系统的可靠性来自深思熟虑的架构决策，而不是自发的动力学演化。

4.3 确定性与概率性：错误观的根本差异

两篇论文对"错误"的理解和态度形成了鲜明的对比——这或许是最能体现它们哲学差异的维度。

CTM对错误的态度是**认知性的**。在CTM中，"错误"不是需要被消除的外部噪声，而是内部不确定性的信息性指标。当CTM对一个预测的确定性较低时，它选择进行更多的内部tick来收集信息。确定性监测（certainty monitoring）是CTM认知过程的内在组成部分——错误概率被内化为认知状态的一部分。这意味着CTM本质上是一个**贝叶斯认知架构**：它维护着对自己认知状态的置信度分布，并利用这一分布指导计算资源的分配。

CTM论文中特别强调了校准性能 (calibration) ——“置信度为0.75的预测恰好有75%的正确率”。这种完美校准表明CTM不仅在做预测，还在**知道自己在做什么以及做得如何**。简言之，CTM拥有一种初级的元认知能力。

MAKER对错误的态度是**工程性的**。错误是一个需要被最小化的外部风险，其来源是LLM生成过程中的固有随机性。MAKER的策略不是让系统更好地感知自己的不确定性，而是通过统计冗余（多次独立采样）来压制不确定性。在MAKER的世界观中，错误是**伯努利试验**——每次LLM调用都有概率 p 成功、 $1-p$ 失败，而系统设计的目标是通过投票使有效错误率趋近于零。

这一态度的技术表现是：MAKER的参数选择 (k_{\min}) 完全基于概率计算。给定目标成功率 t 和单步成功率 p ，通过不等式

$$t \geq (1 + ((1-p)/p)^k)^{-s/m}$$

求解最小 k 值。错误在这里是一个纯统计量，没有认知维度。

这两种错误观的哲学根源可以追溯到两种科学传统：

- CTM的认知性错误观植根于**控制论与认知科学**传统——Wiener、Ashby、Conant等人在控制系统中将不确定性视为信息，而有效行为要求系统具有关于自身状态的模型。
- MAKER的工程性错误观植根于**信息论与通信工程**传统——Shannon将信道噪声视为需要被编码克服的随机扰动，而纠错码就是通过冗余来对抗不确定性的最优策略。

有意思的是，这两种传统在维纳和香农的时代曾是同一枚硬币的两面——维纳的控制论既关心通信也关心控制，而香农的信息论则为这两种视角提供了统一的数学语言。CTM和MAKER的分歧在某种程度上折射了这两个领域在20世纪后期的分流。

5.1 CTM的局限：计算复杂性与可扩展性

CTM面临的首要挑战是计算开销。论文坦诚地指出了两点局限：

第一，内部时间维度（internal ticks）的使用意味着训练时间大幅延长。对于每个输入，CTM需要执行 τ' 个内部tick的前向传播，每一步都涉及完整的突触模型、NLM处理和注意力计算。相比于前馈网络的一次前向传播，CTM的计算开销大约是 τ' 倍。

第二，NLM的参数数量与网络宽度呈线性关系（ $D \times D_{\text{hidden}}$ ），当网络规模增大时参数数量的增长不容忽视。虽然这一增长是线性的（在可接受范围内），但考虑到每个NLM还需要处理M维的预激活历史（ $M \approx 10 - 100$ ），实际的计算和存储开销是显著的。

从更根本的层面看，CTM面临的挑战是：**在保持神经动力学丰富性的同时实现规模扩展**。生物大脑通过模块化、层次化和稀疏连接来管理大规模神经动力学的计算复杂度，而CTM目前采用的是全连接式的设计。如何在不牺牲动力学丰富性的前提下引入结构化的稀疏性，是CTM未来发展的关键问题。

此外，CTM的实验目的明确是“分享创新方法而非追求新的SOTA结果”——这意味着在标准基准测试上（ImageNet-1K 72.47% top-1），其性能尚未达到当前最优水平。这一方面是因为研究者优先关注的是理解和展示动力学特性而非性能调优，另一方面也反映了神经动力学方法在纯准确率指标上尚未展现出压倒性优势。

5.2 MAKER的局限：分解假设与错误相关性

MAKER面临的核心挑战是**分解的前提条件是否总能被满足**。论文明确指出，MAKER有效的前提是：每个步骤足够小，以至于LLM有很大概率采样到正确答案，且没有任何错误答案比正确答

案更可能被采样到。对于哪些类型的任务可以满足这一条件，目前没有通用性的理论指导。

更根本的问题是：**不是所有任务都可以被分解为独立步骤的线性序列**。有些任务需要全局推理，即每个步骤的正确性取决于对未来步骤的理解或对整个问题空间的全局把握。经典的框架问题（frame problem）就与此相关：在复杂环境中，如何确定哪些信息与当前步骤相关？

第二个重要局限是**错误相关性**。MAKER的理论分析假设各步骤之间的错误是独立同分布的，但实验表明这一假设并不总是成立。论文中报告了一个“病理性的步骤”（step 10241）需要18轮投票——某些步骤由于输入状态的特定组合而导致LLM系统性出错。当多个投票代理面对相同的输入时，它们的错误可能是**高度相关的**，这严重削弱了多数投票的有效性。

论文采用的“红标”机制在一定程度上缓解了这一问题，但未能根本解决。当错误来自LLM训练数据中的系统偏差或架构性的盲点时，简单的重新采样和相关信号的丢弃是远远不够的。

5.3 互补性：时间与组合的整合前景

CTM和MAKER各自面对的核心挑战恰好暗示了彼此的优势所在。

CTM的强项——丰富的内部动力学、自适应计算、元认知性的确定性监测——恰恰是MAKER的弱项。MAKER中的LLM智能体缺乏对自己认知状态的感知，无法根据任务难度动态调整计算策略。如果将CTM式的确定性监测机制引入MAKER的智能体中，使其能够“感知”一个步骤的复杂程度并在必要时请求更多投票或更深入的思考，将有效提升系统的自适应性和效率。

反过来，MAKER的强项——极致分解、冗余投票、系统性可靠性保证——提供了CTM所缺乏的可扩展蓝图。CTM的全连接式神经动力学设计在规模扩大时面临严峻的计算挑战，而MAKER式的模块化分解思想，如果应用于CTM架构中——例如将大规模

CTM分解为多个通过同步化接口通信的小规模CTM模块——可能为神经动力学模型的规模扩展提供新的路径。

更一般的理论前景是：**时间性与组合性可能不是对立的，而是同一认知过程的两个维度**。认知既需要在时间中展开的动力学过程来捕捉世界的变化性和情境的连续性，也需要组合式的表征结构来支持逻辑推理和跨情境的泛化。CTM与MAKER的对比表明，AI的未来可能不是在这两个方向中选择其一，而是在它们的交汇处寻找新的架构范式。

第六章 对AI研究范式的启示

6.1 两种"尺度"的智能

CTM和MAKER代表了两种不同"尺度"的智能概念。

CTM关注的是**微观时间尺度**上的智能——在毫秒到秒的时间窗口内，神经群体如何通过同步化动力学实现灵活和高效的认知操作。这种智能是连续的、流动的、多模态整合的——它对应于人如何在瞬间理解一个场景、在对话中理解一个句子的含义、在运动中协调自己的身体。

MAKER关注的是**宏观任务尺度**上的智能——在分钟到小时的量级上，一个智能系统如何可靠地执行包含成千上万步骤的复杂任务。这种智能是离散的、序列化的、目标导向的——它对应于人如何执行一个复杂的项目、按照一份详细的计划工作、管理一个持续数小时或数天的流程。

值得注意的是，人类认知在这两个尺度上同时运作，但AI研究往往只偏重其中之一。端到端的深度学习模型擅长微观尺度的模式识别和即时反应，但在长程推理和持续执行上表现不佳；符号规划系统擅长宏观尺度的任务分解和计划生成，但在感知和灵活适应上存在困难。

CTM和MAKER各自在一个尺度上取得了显著进展，但真正的通用智能可能需要两个尺度的整合——这就是为什么理解它们

之间的互补性具有超出两篇论文本身的重要意义。

6.2 对"可解释AI"的重新思考

两篇论文都为AI可解释性提供了独特的视角，但它们的路径截然不同。

CTM的可解释性来自其内部动力学的**天然可观测性**。由于CTM将神经同步化作为核心表征，理解CTM的"想法"就意味着观察神经同步化模式的演化。论文中展示的UMAP投影行波、注意力图的"注视"行为、确定性随时间的变化曲线——这些都是不需要额外工具即可从CTM内部提取的可解释信号。

CTM的可解释性是一种**过程可解释性**：我们不仅能知道CTM输出了什么，还能观察到它是如何在时间中形成这一输出的。这对于建立对AI系统的信任至关重要——信任不仅来自结果的准确性，还来自对过程的理解。

MAKER的可解释性则源于其架构的**透明分解**。在MAKER系统中，每个决策步骤都有一个明确的"决策路径"：哪个智能体做了什么决定，基于什么上下文，投票结果是什么，是否有红标被触发。这一路径是完全可审计的。论文中对步骤10241的"病理性"案例的分析就展示了这种可追溯性的价值——研究者可以精确定位哪个步骤出现了异常，并分析异常的原因。

MAKER的可解释性是一种**架构可解释性**：系统的透明性不是通过解释黑箱的内部机制实现的，而是通过将任务分解为透明的原子步骤实现的。这与软件工程中"干净代码"的理念一致——代码不需要注释来解释，如果它本身写得足够清晰。

这两种可解释性路径的互补性很明显：CTM提供的是底层认知过程的可解释性，MAKER提供的是高层推理链路的可解释性。一个理想的可解释AI系统应该同时具备这两种可解释性——既能解释"它在想什么"（认知动力学），也能解释"为什么做出这个决定"（推理步骤）。

6.3 迈向统一的视角

综合以上分析，我认为CTM和MAKER共同指向了AI研究的一个可能的前沿：**时间-组合架构**（Temporal-Compositional Architecture）。这一架构的基本假设是：

1. **认知是时间性的**：任何智能系统都需要一个在时间中展开的动力学核心，这个核心提供了连续性、情境敏感性和自适应能力。时间的引入不是计算的一个选项，而是认知的必要条件。
2. **认知是组合性的**：长程、复杂的智能行为需要将基本操作组合为可靠的过程。没有组合性，智能就被限制在动力学的窗口内，无法真正扩展到人类水平的复杂推理。
3. **时间性与组合性不是层次化的**：并非先有"低级"的动力学再有"高级"的组合推理。相反，两者在所有认知层次上共存和相互作用。感知中有组合性（如视觉特征的组合），推理中有时间性（如思维随时间展开的过程）。

实现这一架构需要解决的问题包括但不限于：

- 如何设计既能展开丰富动力学又支持模块化解构的神经网络构件？
- 如何在保持神经同步化表征的丰富性的同时，使其具有可分解和可组合的特性？
- 如何在一个统一的框架中实现微观时间尺度上的自适应性和宏观任务尺度上的可靠性？

第七章 结语：思想的两条河流

CTM和MAKER的关系，让我想起博尔赫斯小说中那张精确到与帝国完全重合的地图——一个用尽全部细节去描摹现实，一个通过对现实的极度拉伸使之变得可管理。两篇论文面对的是同一个挑战：如何让机器完成人类水平的复杂认知任务。一个选择从时间、同步和动力学出发，试图构建一个能够像生物神

经系统那样灵活和适应性强的认知架构；另一个选择从分解、冗余和纠错出发，试图将不可靠的AI组件组织成可靠的系统。

这两条思想河流在当代AI的地形上各自流淌，它们的源头可以追溯到科学思想的古老传统：一条来自生物学的观察——大脑如何通过神经时间实现认知；一条来自工程学的实践——复杂系统如何通过分解和冗余实现可靠性。两条河流可能最终汇入同一个大海——一个既能展开丰富的内部动力学、又能可靠执行长程复杂任务的通用智能系统。

让我用一段话总结CTM和MAKER的根本差异与互补性：

CTM告诉我们：智能需要时间。 没有时间，就没有真正的认知过程——只有输入到输出的静态映射。时间将计算从瞬间的点扩展到持续的流，使自适应、元认知和涌现成为可能。但CTM还没有告诉我们的是：时间性的智能如何可靠地扩展到百万步、千万步的复杂任务。

MAKER告诉我们：智能需要分解。 没有分解，就没有可靠的扩展——任何脆弱的单系统在面对复杂任务时终究会失败。分解将不可靠的组件转化为可靠的系统，通过冗余和纠错实现了前所未有的规模化执行。但MAKER还没有告诉我们的是：可分解的智能如何保持时间性的丰富性和创造性。

两者各自照亮了通往通用人工智能的一条道路，而完整的路径或许就在它们的交汇之处。

参考文献

- [1] Regan, C., Risi, S., Seely, J., Jones, L., & Darlow, L. (2025). Continuous Thought Machines. *39th Conference on Neural Information Processing Systems (NeurIPS 2025)*. Sakana AI, Tokyo, Japan.
- [2] Meyerson, E., Paolo, G., Dailey, R., Shahrzad, H., Francon, O., Hayes, C. F., Qiu, X., Hodjat, B., & Miikkulainen, R. (2025). Solving a Million-Step LLM

- Task with Zero Errors. *arXiv preprint arXiv:2511.09030v1*. Cognizant AI Lab & UT Austin.
- [3] Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S., & Farajtabar, M. (2025). The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. *arXiv preprint arXiv:2506.06941*.
- [4] Meyerson, E. & Qiu, X. (2025). Position: Scaling LLM Agents Requires Asymptotic Analysis with LLM Primitives. In *Forty-second International Conference on Machine Learning Position Paper Track*.
- [5] Sinha, A., Arun, A., Goel, S., Staab, S., & Geiping, J. (2025). The Illusion of Diminishing Returns: Measuring Long Horizon Execution in LLMs. *arXiv preprint arXiv:2110.09624*.
- [6] Dziri, N. et al. (2023). Faith and Fate: Limits of Transformers on Compositionality. In *Advances in Neural Information Processing Systems*, 36, 70293 – 70332.
- [7] Hinz, A. M., Klavžar, S., Milutinović, U., & Petr, C. (2013). *The Tower of Hanoi - Myths and Maths*. Springer.
- [8] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3,4), 379 – 423, 623 – 656.
- [9] Wald, A. (2004). *Sequential Analysis*. Courier Corporation.
- [10] Ross, S. M. (2025). First Ahead by at Least k Multinomial Game. *Annals of Operations Research*, 1 – 6.
-

本文完成于2026年5月13日。作为一篇学术性对比研究，本文力求客观呈现两篇论文的思想、技术与哲学内涵，并将它们置于更广阔的AI研究语境中进行审视。