[[1]](#footnote-1)

Deterministic Option Actor-Critic Framework

First A. Author, *Fellow, IEEE*, Second B. Author, and Third C. Author, Jr., *Member, IEEE*

***Abstract*—This document provides a guide for preparing articles for IEEE Transactions, Journals, and Letters*.* Use this document as a template if you are using Microsoft *Word*. Otherwise, use this as an instruction set. The electronic file of your article will be formatted further at IEEE. Titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., "Nd–Fe–B"). Do not write “(Invited)” in the title. Full names of authors are preferred in the author field but are not required. Put a space between authors’ initials. ORCIDs can be provided here as well. In the title, all variables should appear lightface italic; numbers and units will remain bold. Abstracts must be a single paragraph. In order for an Abstract to be effective when displayed in IEEE *Xplore* as well as through indexing services such as Compendex, INSPEC, Medline, ProQuest, and Web of Science, it must be an accurate, stand-alone reflection of the contents of the article. They shall not contain displayed mathematical equations, numbered reference citations, nor footnotes. They should include three or four different keywords or phrases, as this will help readers to find it. It is important to avoid over-repetition of such phrases as this can result in a page being rejected by search engines. Ensure that your abstract reads well and is grammatically correct.**

***Index Terms*—Enter keywords or phrases in alphabetical order, separated by commas. For a list of suggested keywords, send a blank e-mail to** [keywords@ieee.org](mailto:keywords@ieee.org) **or visit** <http://www.ieee.org/organizations/pubs/ani_prod/keywrd98.txt>

# I. INTRODUCTION

T

HIS document is a template for Microsoft *Word*. If you would prefer to use LaTeX, download IEEE’s LaTeX style and sample files from [https://template-selector.ieee.org/secure/templateSelector/publicationType.](https://journals.ieeeauthorcenter.ieee.org/create-your-ieee-journal-article/authoring-tools-and-templates/tools-for-ieee-authors/ieee-article-templates/) You can also use the Overleaf editor at [https://www.overleaf.com/blog/278-how-to-use-overleaf-with-ieee-collabratec-your-quick-guide-to-getting-started#.Vp6tpPkrKM9](https://www.overleaf.com/blog/278-how-to-use-overleaf-with-ieee-collabratec-your-quick-guide-to-getting-started%23.Vp6tpPkrKM9).

人工智能的主要目标之一是开发能够在环境中通过规划最佳行动序列执行各种复杂任务的自主代理[80]。强化学习（RL）是一种计算范式，用于学习在任务环境的各种状态下采取最优行动的策略，以使行动的代理人获得最大的累积奖励[91]。为了学习最优策略，代理人通过执行各种状态-行动-下一个状态的转换序列来探索与任务相关的状态和行动空间。这种序列的平均长度1被称为任务水平线。如果水平线很长，而任务涉及大的状态和行动空间，那么探索空间也会变得很大。这导致标准RL算法[56, 68, 100]在没有复杂的探索技术[10, 71, 75]的情况下，在这种长视野的任务上表现不佳。

强化学习（RL）允许代理人通过与环境的互动，以试错的方式解决连续的决策问题。当这些环境非常复杂时，对可能解决方案的纯随机探索往往会失败，或者效率非常低，需要与环境进行不合理的互动。分层强化学习（HRL）利用时间和状态抽象的形式来解决这些挑战，同时为行为重用和增加RL系统的可解释性铺平道路。

分层强化学习（HRL）将一个长跨度的强化学习任务分解为一个层次的子问题或子任务，这样，上层策略通过选择最优子任务作为上层行动来学习执行任务。一个子任务本身可能是一个强化学习问题，由较低级别的策略来学习解决它[39]。这个策略的层次结构共同决定了代理人的行为。任务分解有效地将原始任务的长水平线减少到子任务序列的短水平线中。这是因为每个子任务都是一个较高层次的行动，与较低层次的行动相比，其持续的时间尺度较长，这一特性通常被称为时间抽象[7, 20, 93]。时间抽象性也可以在较长的时间范围内实现有效的信用分配[99]。

同时，一个子任务本身可能更容易学习，而且在HRL代理的训练过程中，学到的子任务会导致更有条理的探索[71]。

HRL的核心机制之一是时间性抽象。解决复杂的问题往往需要在多个时间尺度上进行推理。例如，当你开车去上班时，你通常不会开始计划你应该向左还是向右打方向。

你是在一个更高的层次上开始的，例如计划走什么路，以及是否应该停下来加油。

一个时间上的扩展行动（子行为），由一连串的原始行动和可能的其他时间上的扩展行动组成。这些时间上的抽象可以被学习代理所利用，以便在更高的抽象层次上做出决定。构成时间上的扩展行动的行动序列可以是固定的，也可以由一个政策来支配。

利用时间上扩展的行动的分层组合的学习算法将必须定义这些子行为应该如何被发现，如何被开发，如何学习有效的状态抽象，以及如何组成。

在本节中，我们将简要介绍这些挑战

HRL最重要的问题之一包括：我们如何能够自动发现有意义的子行为？很多实证研究（例如，[140,148-150]）都是关于能够自动学习有意义的子行为的算法，这些算法来自于与环境的交互，没有外部专家提供的任何知识，而且是以一种高效的方式。理论上，已经证明[151]，在有限的步骤中找到一小套子行为是一个NP-hard问题。

Abel等人[152]研究了哪些状态抽象和时间抽象的集合能够保留近似最优策略的表示。然而，这种方法需要访问底层MDP。

在低维环境中，一个小的离散的子行为集可以大大改善学习算法的样本效率。然而，在真正复杂的环境中，将需要一个大的连续的子行为范围。

最常用的自动发现时间抽象的技术，要么利用不同状态的特殊属性[153,154]，要么将轨迹与子行为随机变量关联起来[148-150,155]。

关于HRL代理的最初研究集中在证明在在线RL环境下使用时间抽象的好处。为了证明HRL的能力，人们提出了高度针对问题的模型。它们提供了直观的、通常可以解释的答案，即如何对抽象的层次进行建模

上一节介绍的特定问题模型由于其非通用性和复杂的架构，很难自动发现。选项框架[137]提供了一个替代框架，以更通用的方式对时间上的扩展行为进行建模，这样在多种环境下使用相同的学习算法，自动学习子行为和它们的组成就变得更加可行了。然而，这些针对具体问题的框架所引入的观点在选项框架中也仍然适用。这是由于大多数特定问题的框架都可以被表示为选项的事实

在期权框架中，代理人的行动空间被扩展为时间上的扩展行动，称为期权。SMDP形式主义被用来为利用期权的控制问题建模。如果一个选项的策略不仅取决于MDP的当前状态，而且还取决于自该选项被调用以来观察到的状态和奖励的集合，则该选项被认为是半马尔科夫的。例如，如果一个选项未能在若干步内满足终止条件，这个集合就可以用来终止它。

选项代表了闭环的子行为（图 6），这些行为会在多个时间段内进行，直到触发其终止条件。选项被认为是闭环系统，因为它们根据当前的状态来调整自己的行为。这与开环系统不同，开环系统一旦初始化，在面对新的状态时不会调整其行为。将子行为建模为闭环系统通常比使用开环子行为更现实。例如，在驾驶汽车时，如果我们会致力于一个开环的选项，我们就不会在遇到危险时偏离，一个闭环的选项会根据当前状态改变其行动。

使用一套定义明确的选项将要求代理人在解决问题时做出更少的决定[176-178]。在RL环境中使用选项已被证明可以加快学习速度。例如，[61,179]证明了选项是总结知识的一种方式，是终身学习环境中的一个重要组成部分。Guo等人[180]证明了使用重要性抽样的时间扩展行动的性能提高。

各种算法都利用了选项框架，在本节的剩余部分，我们首先讨论框架的不同组成部分。此外，我们回顾了各种能够通过与环境的交互来自动发现选项的算法，以及如何找到一个超过选项的策略，以便适当地组成选项。

最经常使用的执行模型是调用和返回模型。这种方法也经常被称为分层执行。在这个模型中，一个政策性的选项根据当前状态选择一个选项。代理人跟随这个选项，直到代理人触发了活动选项的终止条件。终止后，代理人会选择一个新的选项来跟随。

另一个模型被称为单步选项模型，有时也被称为非层次执行模型，在每个单一的时间步数上查询政策选项，允许切换选项，即使该选项还没有完全终止。例如，Mankowitz等人[186]建议，当当前执行的期权以外的期权的预期未来总价值变得更高时，可以切换期权。然而，期权大多应该能够运行一定的步数，这样才是有用的。Harb 等人[187]加入了一个终止商议的成本，以防止在每个时间步长上切换期权

选项批评（OC）算法[140]是一个端到端的框架，能够在不使用先验知识的情况下共同发现和开发选项。这种方法受到了 Actor-Critic 框架 [210] 的启发。OC框架中的行动者部分由多个选项内的政策组成。批评者部分能够评估选项和行动的未来折现值。一个选项由一个政策内选项选择，并运行到代理人触发随机选项终止条件。终止条件的梯度使用关于期权未来价值的优势[211]，与其他期权相比。与其他选项相比表现出较高优势的选项会利用这个梯度更新，使其运行时间更长。

# II. Related work

“Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning”: Natasha Jaques et.al. achieved coordination and communication through causal inference, however, they assessed the CI in action level which is infeasible in high-dimensional space.

“Real-World Human-Robot Collaborative Reinforcement Learning”: 作者将隐式合作运用到一个真实世界中的人在回路中的强化学习装置中，来研究人机合作，并证明了隐式合作在实际应用中的潜力。

“Implicit Communication in a Joint Action”: Knepper等人（2017）提出了一个合作中隐性交流的框架,并表明各种问题都可以映射到这个框架中。

“Online Implicit Agent Modelling”: 隐式方式也同样运用在了智能体建模中，其中智能体的目的是估计预先计算的策略的固定投资组合的效用，相比于显式地建模智能体策略的参数，该方法在复杂环境中是可行的。

“Multi-Agent Common Knowledge Reinforcement Learning”: 本文使用一个层次化的策略树，利用适当级别的智能体群体之间的共同知识，端到端地学习复杂的协调策略。尽管其最终可以学到一个协调分散的策略，但仍需要集中式的训练方式，且仅根据共同知识无法学习到非对称信息。

“Modeling Others using Oneself in Multi-Agent Reinforcement Learning”: 从其他玩家的行为中发现他们的隐藏目标，并使用这些估计来选择行动。尽管这项工作与本文都是对其他智能体的意图进行推理，但这项工作将意图直接建模为最终的目标，这样可能导致损失某些有用的信息，而我们则是更抽象的层次。并且这项工作尚未在有两个以上玩家的更复杂环境中评估。

“Bayesian Action Decoder for Deep Multi-Agent Reinforcement Learning”: defines a new Markov process, the public belief Markov decision process (PuB-MDP) and proposed BAD which uses an approximate Bayesian update to obtain a public belief. 然而，由于BAD依赖于一个被视为第三方的公共智能体，因此难以在分散控制中部署。

“Agreeing To Cross: How Drivers and Pedestrians Communicate”: 我们提出了一些关于行人在不同的过马路情况下采取的行动过程和使用的非语言提示的发现。我们表明，横穿马路的行为会受到各种环境因素的影响，如十字路口的结构、司机的行为、与临近车辆的距离等等。

“A Study of Reinforcement Learning Algorithms for Aggregates of Minimalistic Robots”: 研究最低限度的机器人集合体的控制，这些可以感知目标位置和附近的障碍物，但缺乏通过信息传递等方式进行明确的交流。将通讯设置为隐性的，即通过每个机器人对物体施加的聚合推拉来进行调解。

“A Bayesian Approach to Imitation in Reinforcement Learning”: 提出贝叶斯模仿，允许学习者顺利地汇集先前的知识，通过与环境的互动获得的数据，以及从专家智能体行为的观察中推断出的信息，而不需要与这些其他智能体进行明确的沟通或合作。

In interactive POMDPs (I-POMDPs; Gmytrasiewicz & Doshi 2005), agents model each other’s beliefs, beliefs over these beliefs, and so on, but this is often computationally intractable.

“A mathematical theory of cooperative communication”: 我们将合作通信解释为一个最佳运输的问题

“Bi-Level Actor-Critic for Multi-Agent Coordination”: 从博弈论角度将协作问题定义为具有SE均衡的博弈，并定义了寻找Stackelberg均衡的双层强化学习问题。并提出了一种新颖的双级行为体批评学习方法。然而他们假设智能体间的决策具有优先级顺序。

“Coordination Between Individual Agents in Multi-Agent Reinforcement Learning”: introduces three correlation coefficients to analyze the agent’s roles and the correlation between individual agents and proposes correlation-based communication information and reward function. Although we both consider coordination among individual agents, however, this method relies on a specific correlation coefficient and requires modification of the reward function, whereas our approach is end-to-end and does not require any other modifications.

“Learning Efficient Multi-agent Communication: An Information Bottleneck Approach”: proves that when communication capacity is restricted, the limited bandwidth constraint requires low-entropy messages and develops an Informative Multi-Agent Communication (IMAC) method to learn efficient communication protocols as well as scheduling.

“Learning Individually Inferred Communication for Multi-Agent Cooperation”: To reduce information redundancy, proposes Individually Inferred Communication (I2C) to enable agents to learn a prior for agent-agent communication via causal inference.

“Learning Multi-Agent Communication through Structured Attentive Reasoning”: Rangwala and Williams proposes a new communication architecture, the Structured Attentive Reasoning Network (SARNet), where agents extract the relevance of other agents’ information and reason over received communications and past memories before performing an action.(broadcasting)

“Learning to Communicate Implicitly by Actions”: proposes Policy Belief Learning (PBL) to model the other agent’s private information and an auxiliary reward to encourage communication by actions. 然而他们需要以有监督的方式进行训练，且需要对奖励函数进行修改。

“Learning to Ground Multi-Agent Communication with Autoencoders”: present a framework for grounding multi-agent communication through autoencoding allows agents to learn non-differentiable communication in fully decentralized settings. assumes that all agents have the same model architecture.

“Learning to Simulate Self-Driven Particles System with Coordinated Policy Optimization”: develops Coordinated Policy Optimization (CoPO), which incorporates social psychology principle to learn neural controller for Self-Driven Particles (SDP). CoPo is based on neighborhood-level coordination, in which agents from the same neighborhood are integrated into a single virtual agent, and coordination occurs only inside the agent's neighborhood, i.e., agents only communicate their actions and observations with their neighbors.

“Neurosymbolic Transformers for Multi-Agent Communication”: proposes an approach for synthesizing programmatic communication policies for decentralized control of multi-agent systems. significantly reducing the amount of communication

“Succinct and Robust Multi-Agent Communication With Temporal Message Control”: proposes Temporal Message Control (TMC) providing significantly lower communication overhead and better robustness against transmission loss.

“COMMUNICATION IN MULTI-AGENT REINFORCEMENT LEARNING: INTENTION SHARING”: proposed the Intention Sharing scheme, a new communication protocol, based on sharing intention among multiple agents. 我们的方法与前述方法的一个区别是，我们使用推断来协作，而不是通讯。

“Learning to Share and Hide Intentions using Information Regularization”: introduced information-regularizer to share or hide agent’s intention to other agents for a multi-goal MARL setting in which some agents know the goal and other agents do not know the goal. By maximizing (or minimizing) the mutual information between the goal and action, an agent knowing the goal learns to share (or hide) its intention to other agents not knowing the goal in cooperative (or competitive) tasks.

“TRUST REGION POLICY OPTIMISATION IN MULTI-AGENT REINFORCEMENT LEARNING”: apply trust region learning to multi-agent settings by proposing the first MARL algorithm that attains theoretically-justified monotonical improvement property.

When you open the template, select “Page Layout” from the “View” menu in the menu bar (View | Page Layout), (these instructions assume Microsoft *Word*. Some versions may have  **Fig. 1.** This is a sample of a figure caption.

alternate ways to access the same functionalities noted here). Then, type over sections of the template or cut and paste from another document and use markup styles. The pull-down style menu is in the Formatting Toolbar at the top of your *Word* window (e.g., the style at this point in the document is “Text”). Highlight a section that you want to designate with a certain style, and then select the appropriate name on the style menu. The style will adjust your fonts and line spacing. Do not change the font sizes or line spacing to squeeze more text into a limited number of pages.Use *italics* for emphasis; do not underline.

IEEE will do the final formatting of your article. If your article is intended for a conference, please observe the conference page limits.

This is intended as an authoring template, not a final production template. It is not intended to match the final published format. Differences in final formatting are likely in the final IEEE files. Page count in the template is an estimate. Do not adjust line and character spacing to fit your paper to a specific length.

## A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have already been defined in the abstract. Abbreviations such as IEEE, SI, ac, and dc do not have to be defined. Abbreviations that incorporate periods should not have spaces: write “C.N.R.S.,” not “C. N. R. S.” Do not use abbreviations in the title unless they are unavoidable (for example, “IEEE” in the title of this article).

# III. Preliminaries

Use either the Microsoft Equation Editor or the MathType plugin, which can be obtained from <https://store.wiris.com/en/products/mathtype/download>. For help with formatting and placing equations, refer to the *IEEE Editing Math Guide* at <http://journals.ieeeauthorcenter.ieee.org/wp-content/uploads/sites/7/Editing-Mathematics.pdf> and the *IEEE MathType Tutorial for Microsoft Word Users* at <http://journals.ieeeauthorcenter.ieee.org/wp-content/uploads/sites/7/IEEE-Math-Typesetting-Guide-for-MS-Word-Users.pdf>.

TABLE I

This is a Sample of a Table Title



## A. Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin of the column, as in (1). First use the equation editor to create the equation. Then

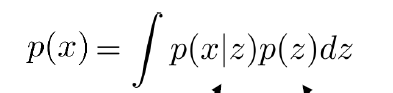
select the “Equation” markup style. Press the tab key and write the equation number in parentheses. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Use parentheses to avoid ambiguities in denominators. Punctuate equations when they are part of a sentence, as in

*Bp* + *H*2 = 40. (1)

Be sure that the symbols in your equation have been defined before the equation appears or immediately following. Italicize symbols (*T* might refer to temperature, but T is the unit tesla). When referring to an equation or formula, use simply “(1),” not “Eq. (1)” or “equation (1),” except at the beginning of a sentence: “Equation (1) is ... .”

# IV. MULTI-AGENT Coordination LEARNING Via inference

## A. Generative Intention Network

在部分可观测环境中，每个智能体只能通过自己的观察进行行动，无法感知全局状态。这里，我们假设在多智能体稳定协作的情况下，将其他智能体的意图与当前观测结合，可以重建或部分重建全局状态。因此，在协作稳定情况下的状态可表示为：

人类合作者具有社会感知方面的能力，被定义为根据基本行为信号解码另一个人的心理状态的能力，由于其反射性和高效的性质，在进化上是适应性的。

当人们理解他人的行为时，还有一种机制在很大程度上依赖于感知者的自我认知：为了理解他人的行动，感知者必须首先观察该行动，然后形成对该行动的心理模仿。在人类中，镜像神经元网络允许在观察和模仿的行动之间进行比较。

因此，根据以上理论，我们提出了一个有条件的深度生成模型来建模人类的社会感知能力，并为每个代理人编码潜伏代码来代表代理的意图。

在每个场景中，我们有N个跨T步的代理。用X 属于 R N\*T表示这个场景中的所有轨迹数据，xti表示第i个代理人在第t步的位置。给定Tobs T为历史帧的数量，我们有观察XP = X1:Tobs，任务是根据观察结果预测XF = X(Tobs+1):T。

我们的动机根据过去的动作轨迹进行意图建模。因此，我们用条件生成模型对数据进行建模，潜变量zI描述代理人意图。这个模型可以通过最大化证据下限（ELBO，[35]）来训练，ELBO通常被用作难以解决的对数似然的替代物。在训练过程中，当收集到一整幕的动作轨迹后，我们将数据分为过去和未来部分，我们使用给定XP和XF的编码器推断潜势，并用SGD优化ELBO。在预测过程中，我们不需要编码器，而是使用给定XP的生成模型来建模意图。我们的模型，GIN，如图1所示。我们在下面的小节中描述细节。

* 1. **Encoder**  
     编码器将基本行为信号解码另一个人的心理状态。

编码器计算出一个近似的后验q(zjXP;XF)，样本z = [zI 以XP和XF为条件。

人类具有时间整合的能力，即构建和整合一段时间内信息的过程和能力，使之成为一个连贯的整体，从而能够理解和预测随着时间推移发生的事件。因此，我们沿时间维度串联XP和XF，并使用递归神经网络（RNN）来计算每个代理的全长轨迹的特征.并使用层归一化进行特征正则化

假设z遵循多变量高斯分布N(; I)，利用重参数化技巧[31, 16]，我们对N(0; I)进行采样，得到z = + ，其中和是由多层感知器（MLP）预测的。

* 1. **Prior**  
     先验p(zjXP)表征了仅给定XP的z的条件分布。在我们的实现中，先验与编码器共享模型权重。在这种情况下，编码器和先验只是在公式1中的RNN的输入数据方面有所不同。通过权重共享，同一个RNN可以看到不同长度的输入，而损失中的KL-分歧项会强制编码器和解码器从输入中提取类似的信息。因此，RNN被鼓励学习从XP和[XP;XF]中编码相同的信息。此外，权重共享减少了其内存成本。
  2. **Decoder**  
     形成对该行动的心理模仿，并在观察和模仿的行动之间进行比较。对互动伙伴的预测和实际行动进行比较，可以帮助个人调整自己的行动，以适应互动伙伴的行动。

解码器p(XF jzA; zG;XP )产生以随机变量z和过去轨迹XP为条件的未来轨迹。按照[18]，我们假设潜伏的z是对距离历史的充分总结，并将似然定义为p(Xt+1jz;X1:t) = p(Xt+1jz;Xt)。因此，给定z和Xt，解码器可以预测^X t+1。对于t > Tobs，解码器的输入是它自己的预测^X t。

在这里，我们采用预测下一个观测值与当前观测值之间差异的动力学函数，即oi t+1 oi t，而不是（Nagabandi等人（2018））提出的下一个观测值oi t+1，以减少学习早期阶段的模型偏差。

## B. Casual Inference

在此有理由提出这样的问题。如果参与者已经在跟踪和预测他们自己和互动伙伴的意图，为什么还需要一个推断他人意图的因果关系，而不是直接将所有意图整合？在我们看来，简单的意图整合会导致信息冗余，甚至可能影响学习过程。而因果推断则将使代理能够捕获高阶关系，甚至形成有共同的目标和意图的动态即时团队。

现在人们普遍认为(Interactive brains, social minds: Neural and physiological mechanisms of interpersonal action coordination)，大脑间的同步化是人际行动协调和社会互动行为的重要和必然的机制。虽然社会知觉涉及感知和解码他人的外部行为信号并推断他们的潜在意图，但知觉者的自我作用相对来说是沉默的。然而，当人们理解他人的行为时，还有一种机制在很大程度上依赖于感知者的自我：感知者可能试图理解其他智能体的意图对感知着智能体的策略的因果效应。

与协作相关的行动理解需要感知者超越单纯的解码，努力将自己的行动与观察到的行动成功匹配。

一个代理人对另一个代理人的影响是由该代理人的意图对另一个人的政策的因果效应来体现的。对于任何能够引起对方政策剧烈变化的代理人，该代理人被认为是相关和有影响力的。

直观地说，一个代理人更有可能与那些可能对其战略施加更多影响的人进行协作，希望获得关于他们倾向于如何行动以及如何作出合作反应的线索。因此，其他代理人的因果效应可以被看作是以其他代理人的意图为条件进行决策的必要性。

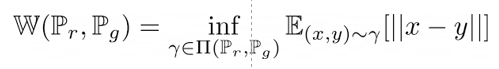
这里，我们通过以不同意图为条件生成的策略分布来测量和量化代理人之间的因果效应来确定代理人-代理人的协：

\documentclass{article}
\usepackage{amsmath}
\pagestyle{empty}
\begin{document}

$ KL\left [ \pi \left ( a\mid o,[z_{intention}^m, z_{intention}^n, \cdots z_{intention}^k] \right )\parallel \pi \left ( a\mid o,[z_{intention}^n, \cdots z_{intention}^k] \right ) \right ]  $


\end{document}

Kullback-Leibler（KL）散度被用来衡量这两个条件概率分布之间的差异。Ij i的大小表明，如果考虑到代理人j的意图，代理人i会对其政策做出多大的调整，也表明代理人j的战略与代理人i的政策有多大的关联。如果Ij i>阈值δ，则将该智能体加入到协作集中，delta是一个超参数。

我们同时也提出了使用Wasserstein Distance直接比较自身意图分布与其他智能体意图分布的距离，以此作为每对代理之间的相关性。基于此，代理人之间的强、弱相关性定义如下：

代理人之间的强关联性。对于任何一对代理人i和j，如果相关系数ci;j大于阈值G，那么这两个代理人就处于彼此的强相关状态。

代理人之间的弱关联性。对于任何一对代理人i和j，如果相关系数ci;j小于阈值G，那么这两个代理人之间就处于弱相关关系。

如果为强相关关系，则加入到协作集中。

阈值G设定为0.5，实验中分析了不同阈值G的性能

每个智能体根据部分其他智能体的意图进行决策，产生的新的决策被智能体感知后，意图也会随之更新，从而进一步更新策略，这样便形成了无限嵌套的信念。

## C. Long Short Term Intention Clip

我们工作中的其中一个主要技巧是使用裁剪的意图重要性比率--这试图限制意图在迭代之间的急剧变化。剪切的强度由超参数控制：大的允许更大的意图变化。

我们认为对意图进行剪裁，在一定程度上控制了由变化的多代理政策引起的非平稳性。如第5.4节所示，我们观察到，虽然较低的数值减缓了学习速度，但它们对应的是更一致的政策改进。另一方面，较高的数值会导致较大的方差和较大的性能波动。



此外，智能体的探索可能会产生不符合宏观意图的动作从而干扰意图的建模，在非合作情形下，对手可能在短时间内做出假动作（与长期意图相反）。因此，为了学习更稳定的意图，我们提出了一个短期意图来消除干扰性动作。

短期意图网络与GIN中**Prior**网络共享参数，和先验只是在公式1中的RNN的输入数据方面有所不同。具体地，先验使用整个历史动作轨迹作为输入，而短期意图网络使用长度为n的动作轨迹作为输入，长度n设定为3，实验中分析了不同n的性能。根据先验和短期意图网络输出的意图，我们比较了他们的KL散度，如果KL散度大于delta，则将n个动作进行masking，否则保留。

## D. Action Mask

在多智能体环境中，每个单元由一个独立的代理控制，该代理只对以该单元为中心的有限视域内的局部观测进行控制(如，smac),智能体不可能在每个时刻都可以观察到其他所有智能体的动作，因此在智能体观察的动作轨迹中会存在缺省值。对这些缺省值进行意图建模，无疑是毫无意义的，甚至会对策略训练造成巨大影响，我们的建议是简单地使用一个特定于代理人的常数向量，即一个带有代理人ID的零向量，作为代理人动作缺省值的输入。我们称这种技术为 "动作屏蔽"。

## E. HAPPO with Intention Ratio

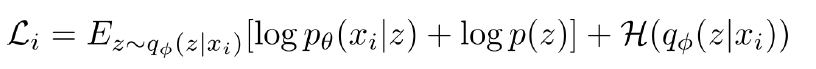
TRUST REGION POLICY OPTIMISATION IN MULTI-AGENT REINFORCEMENT LEARNING开发了异质代理近端政策优化（HAPPO）算法。与许多现有的MARL算法不同，HATRPO/HAPPO不需要代理共享参数，也不需要对联合价值函数的可分解性进行任何限制性假设。最重要的是，他们在理论上证明了HAPPO的单调性改进特性，建立了新的技术水平。

我们在HAPPO的基础上，引入了意图重要性比率，这使得每个代理人的目标都不仅考虑到所有先前代理人的更新，还考虑到意图的更新。

## F. Training

* 1. GIN

对于训练，我们从编码器中取样z = [zI]，q(zjXP ;XF)，然后优化ELBO：



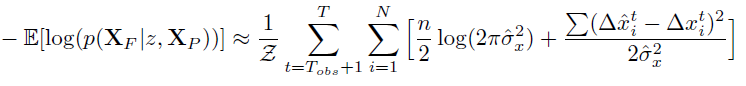
可以从另外一个角度看待ELBO：



最大化对数似然的同时，尽量缩小编码模型和先验分布的KL divergence



由于我们的解码器预测[^xti ;^x]为代理人i在步骤t，并且它遵循高斯分布。我们用样本平均数对期望值进行近似。

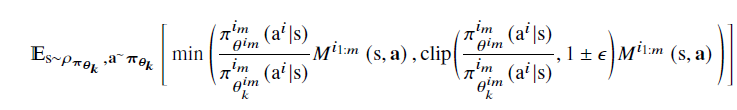


其中Z是归一化项，n是xti的维度。由于先验和编码器的输出是高斯的，所以第二个项可以用分析法计算。



优化的目标为最大化ELBO：

* 1. HAIPPO



## G. Color Space

The term “color space” refers to the entire sum of colors that can be represented within the said medium. For our purposes, the three main color spaces are grayscale, RGB (red/green/blue), and CMYK (cyan/magenta/yellow/black). RGB is generally used with on-screen graphics, whereas CMYK is used for printing purposes.

All color figures should be generated in RGB or CMYK color space. Grayscale images should be submitted in grayscale color space. Line art may be provided in grayscale OR bitmap colorspace. Note that “bitmap colorspace” and “bitmap file format” are not the same thing. When bitmap color space is selected, .TIF/.TIFF/.PNG are the recommended file formats.

## H. Accepted Fonts Within Figures

When preparing your graphics, IEEE suggests that you use one of the following Open Type fonts: Times New Roman, Helvetica, Arial, Cambria, or Symbol. If you are supplying EPS, PS, or PDF files, all fonts must be embedded. Some fonts may only be native to your operating system; without the fonts embedded, parts of the graphic may be distorted or missing.

A safe option when finalizing your figures is to strip out the fonts before you save the files, creating “outline” type. This converts fonts to artwork which will appear uniformly on any screen.

## I. Using Labels Within Figures

1. **Figure Axis Labels**
   1. Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity “Magnetization” or “Magnetization *M*,” not just “*M*.” Put units in parentheses. Do not label axes only with units. For example, write “Magnetization (A/m)” or “Magnetization (Am−1),” not just “A/m.” Do not label axes with a ratio of quantities and units. For example, write “Temperature (K),” not “Temperature/K.”
   2. Multipliers can be especially confusing. Write “Magnetization (kA/m)” or “Magnetization (103 A/m).” Do not write “Magnetization (A/m) × 1000” because the reader would not know whether the top axis label means 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 8- to 10-point type.
2. **Subfigure Labels in Multipart Figures and Tables**

Multipart figures should be combined and labeled before final submission. Labels should appear centered below each subfigure in 8-point Times New Roman font in the format of (a) (b) (c).

## J. Referencing a Figure or Table Within Your Article

When referencing your figures and tables within your article, use the abbreviation “Fig.” even at the beginning of a sentence. Do not abbreviate “Table.” Tables should be numbered with Roman numerals.

## K. Submitting Your Graphics

Because IEEE will do the final formatting of your article, all figures, figure captions, and tables can be placed at the end of your article. However, if you do place your figures within the article, they should be placed at the top of the page, closest to the first mention in the text. Figures should be submitted as individual files, separate from the manuscript in one of the file formats listed above. Place figure captions below the figures; place table headings above the tables. Do not include captions as part of the figures, or put them in “text boxes” linked to the figures. Also, do not place borders around the outside of your figures.

## L. Color Processing / Printing in IEEE Transactions, Journals, and Letters

All IEEE Transactions, Journals, and Letters allow an author to publish color figures on IEEE *Xplore* at no charge, and automatically convert them to grayscale for print versions. In most journals, figures and tables may alternatively be printed in color if an author chooses to do so. Please note that this service comes at an extra expense to the author. If you intend to have print color graphics, you will have the opportunity to indicate this in the Author Gateway and will be contacted by PubOps to confirm the charges.

# V. EXPERIMENTS AND RESULTS

## 我们考虑两个最常见的基准--StarCraftII多Agent挑战赛（SMAC）（Samvelyan等人，2019）和多Agent Mujoco（de Witt等人，2020b）--用于评估MARL算法。所有的超参数设置和实现细节都可以在附录中找到。

星际争霸II多Agent挑战（SMAC）。SMAC包含一组《星际争霸》的地图，其中

一组盟友单位旨在击败对手的团队。IPPO（de Witt等人，2020a）和MAPPO

(Yu et al., 2021)被认为在这个基准上取得了最高的成绩。通过采用参数

共享，这些方法在大多数地图上实现了100%的获胜率，甚至包括有异质代理的地图。

有异质性代理的地图。因此，我们假设，不一定需要非参数共享，共享政策就足以解决

需要，共享政策足以解决SMAC任务。我们在两张硬地图和一张超硬地图上测试了我们的方法。

图2的结果证实，SMAC的难度并不足以展示HATRPO的能力。

与参数共享方法相比，SMAC的难度不足以显示HATRPO/HAPPO的能力。

多Agent MuJoCo。与SMAC相比，我们认为Mujoco环境为我们的方法提供了一个更为

为我们的方法提供了更合适的测试案例。MuJoCo任务要求机器人学习一种最佳的运动方式。

多Agent MuJoCo将机器人的每个部分都建模为一个独立的代理，例如，蜘蛛的腿或手臂。

例如，蜘蛛的腿或游泳者的手臂。随着身体各部分的种类越来越多，建立模型

异质性的政策变得很有必要。图3表明，在所有情况下，HATRPO和

HAPPO享有比参数共享方法更优越的性能。IPPO和MAPPO。

并且在奖励值和方差方面都优于非参数共享的MADDPG（Lowe等人，2017b）。

奖励值和方差。还值得注意的是，HATRPO和它的竞争对手之间的性能差距，随

其竞争对手之间的性能差距随着代理数量的增加而扩大。同时，我们可以看到，HATRPO

在几乎所有的任务中都优于HAPPO；我们认为这是因为HATRPO的硬KL约束。

我们认为这是因为HATRPO中的硬KL约束与HAPPO中的剪切版本相比，更接近于算法1，而算法1可以达到

单调性改进的保证。

# VI. Conclusion

## A conclusion section is not required. Although a conclusion may review the main points of the article, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes, if needed, appear before the acknowledgment.

Acknowledgment

The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank ... .” Instead, write “F. A. Author thanks ... .” In most cases, sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

# References and Footnotes

## A. References

References need not be cited in text. When they are, they appear on the line, in square brackets, inside the punctuation. Multiple references are each numbered with separate brackets. When citing a section in a book, please give the relevant page numbers. In text, refer simply to the reference number. Do not use “Ref.” or “reference” except at the beginning of a sentence: “Reference [3] shows ... .” Please do not use automatic endnotes in *Word*, rather, type the reference list at the end of the paper using the “References” style.

Reference numbers are set flush left and form a column of their own, hanging out beyond the body of the reference. The reference numbers are on the line, enclosed in square brackets. In all references, the given name of the author or editor is abbreviated to the initial only and precedes the last name. Use them all; use *et al*. only if names are not given or if there are more than 6 authors. Use commas around Jr., Sr., and III in names. Abbreviate conference titles. When citing IEEE Transactions, provide the issue number, page range, volume number, month if available, and year. When referencing a patent, provide the day and the month of issue, or application. References may not include all information; please obtain and include relevant information. Do not combine references. There must be only one reference with each number. If there is a URL included with the reference, it can be included at the end of the reference.

Other than books, capitalize only the first word in an article title, except for proper nouns and element symbols. For articles published in translation journals, please give the English citation first, followed by the original foreign-language citation. See the end of this document for formats and examples of common references. For a complete discussion of references and their formats, see the *IEEE Editorial Style Manual* *for Authors* at <https://journals.ieeeauthorcenter.ieee.org/create-your-ieee-journal-article/create-the-text-of-your-article/ieee-editorial-style-manual/>.

## B. Footnotes

Number footnotes separately in superscripts (Insert | Footnote).[[2]](#footnote-2) Place the actual footnote at the bottom of the column in which it is cited; do not put footnotes in the reference list (endnotes). Use letters for table footnotes (see Table I).

# Submitting Your Article for Review

## A. Review Stage Using ScholarOne Manuscripts

Contributions to the Transactions, Journals, and Letters may be submitted electronically on IEEE’s online manuscript submission and peer-review system, ScholarOne Manuscripts. You can get help choosing the correct publication for your manuscript as well as find their corresponding ScholarOne Manuscripts peer review site using the tools listed at <http://www.ieee.org/publications_standards/publications/authors/authors_submission.html> Once you have chosen your publication and navigated to the ScholarOne site, check first to see if you have an existing account. If there is none, please create a new account. After logging in, go to your Author Center and click “Start New Submission.”

Along with other information, you will be asked to select the manuscript type from the journal’s pre-determined list of options. Depending on the journal, there are various steps to the submission process; please make sure to carefully answer all of the submission questions presented to you. At the end of each step you must click “Save and Continue”; just uploading the paper is not sufficient. After the last step, you should see a confirmation that the submission is complete. You should also receive an e-mail confirmation. For inquiries regarding the submission of your paper on ScholarOne Manuscripts, please contact oprs-support@ieee.org or call +1 732 465 5861.

ScholarOne Manuscripts will accept files for review in various formats. There is a “Journal Home” link on the log-in page of each ScholarOne Manuscripts site that will bring you to the journal’s homepage with their detailed requirements; please check these guidelines for your particular journal before you submit.

## B. Final Stage Using ScholarOne Manuscripts

Upon acceptance, you will receive an email with specific instructions regarding the submission of your final files. To avoid any delays in publication, please be sure to follow these instructions. Final submissions should include source files of your accepted manuscript, high quality graphic files (if not embedded in your source file), and a formatted pdf file. The accepted version of your manuscript will also be sent to the IEEE publication teams for a comparison to the final files to ensure no significant or unauthorized changes were made after acceptance. If you have any questions regarding the final submission process, please contact the administrative contact for the journal.

When submitting your final files on a hybrid OA journal you will have the opportunity to designate your article as “open access” if you agree to pay the IEEE open access fee. Please select the appropriate choice. Immediately after you have submitted your final files through ScholarOne Manuscripts you will be automatically redirected to the IEEE electronic copyright form wizard. Please complete the copyright at that time to avoid publication delays.

## C. Copyright Form

Authors must submit an electronic IEEE Copyright Form (eCF) upon submitting their final manuscript files. You can access the eCF system through your manuscript submission system or through the Author Gateway. You are responsible for obtaining any necessary approvals and/or security clearances. For additional information on intellectual property rights, visit the IEEE Intellectual Property Rights department web page at <https://www.ieee.org/publications/rights/index.html>

# IEEE Guidelines and Policies

A full overview of IEEE publishing guidelines and policies can be found at <https://journals.ieeeauthorcenter.ieee.org/become-an-ieee-journal-author/publishing-ethics/guidelines-and-policies/>. They are designed to help authors understand and navigate the publishing process successfully. Learn more about IEEE’s fundamental publishing guidelines and principles, submission and peer review policies, post-publication policies, and guidelines on advertising, accessibility, and data privacy.

References

*Basic format for periodicals:*

J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. x, no. x, pp. xxx-xxx, Abbrev. Month, year, doi: 10.1109.XXX.1234567.

*Periodicals using article numbers:*

J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. x, no. x, Abbrev. Month, year, Art. no. xxxxx, doi: 10.1109.XXX.1234567.

*Examples:*

# J. U. Duncombe, “Infrared navigation—Part I: An assessment of feasibility,” *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959, doi: 10.1109/TED.2016.2628402.

1. E. P. Wigner, “Theory of traveling-wave optical laser,” *Phys. Rev*., vol. 134, pp. A635–A646, Dec. 1965.
2. P. Kopyt *et al., “*Electric properties of graphene-based conductive layers from DC up to terahertz range,” *IEEE THz Sci. Technol.,* to be published, doi: 10.1109/TTHZ.2016.2544142. *(Note: If a paper is still to be published, but is available in early access, please follow ref [5]).)*
3. R. Fardel, M. Nagel, F. Nuesch, T. Lippert, and A. Wokaun, “Fabrication of organic light emitting diode pixels by laser-assisted forward transfer,” *Appl. Phys. Lett.*, vol. 91, no. 6, Aug. 2007, Art. no. 061103.
4. D. Comite and N. Pierdicca, "Decorrelation of the near-specular land scattering in bistatic radar systems," *IEEE Trans. Geosci. Remote Sens.*, early access, doi: 10.1109/TGRS.2021.3072864. (*Note: This format is used for articles in early access. The doi must be included.)*
5. H. V. Habi and H. Messer, "Recurrent neural network for rain estimation using commercial microwave links," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3672-3681, May 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9153027

*Basic format for books:*

J. K. Author, “Title of chapter in the book,” in *Title of Published Book, x*th ed. City of Publisher, (only U.S. State), Country: Abbrev. of Publisher, year, ch. x, sec. *x*, pp. xxx–xxx*.*

*Examples:*

1. G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics,* 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, *Linear Networks and Systems.* Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.
3. Philip B. Kurland and Ralph Lerner, eds., *The Founders’ Constitution.* Chicago, IL, USA: Univ. of Chicago Press, 1987, Accessed on: Feb. 28, 2010, [Online]. Available: http://press-pubs.uchicago.edu/founders/

*Basic format for handbooks:*

*Name of Manual/Handbook, x* ed., Abbrev. Name of Co., City of Co., Abbrev. State, Country, year, pp. xxx-xxx.

*Examples:*

1. *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, USA, 1985, pp. 44–60.
2. *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, USA, 1989.
3. R. J. Hijmans and J. van Etten, “Raster: Geographic analysis and modeling with raster data,” R Package Version 2.0-12, Jan. 12, 2012. [Online]. Available: http://CRAN.R-project.org/package=raster

*Basic format for reports:*

J. K. Author, “Title of report,” Abbrev. Name of Co., City of Co., Abbrev. State, Country, Rep. xxx, year.

*Example:*

1. E. E. Reber, R. L. Michell, and C. J. Carter, “Oxygen absorption in the earth’s atmosphere,” Aerospace Corp., Los Angeles, CA, USA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.

*Basic format for conference proceedings:*

J. K. Author, “Title of paper,” in *Abbreviated Name of Conf.*, City of Conf., Abbrev. State (if given), Country, year, pp. xxxxxx*.*

*Examples:*

1. D. B. Payne and J. R. Stern, “Wavelength-switched passively coupled single-mode optical network,” in *Proc. IOOC-ECOC,* Boston, MA, USA,1985,   
   pp. 585–590.
2. D. Ebehard and E. Voges, “Digital single sideband detection for interferometric sensors,” presented at the 2nd Int. Conf. Optical Fiber Sensors*,* Stuttgart, Germany, Jan. 2-5, 1984.
3. PROCESS Corporation, Boston, MA, USA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at INET96 Annual Meeting. [Online]. Available: http://home.process.com/Intranets/wp2.htp

*Basic format for electronic documents (when available online):*

Issuing Organization. (year, month day). *Title*. [Type of medium]. Available: site/path/file

*Example:*

1. U.S. House. 102nd Congress, 1st Session. (1991, Jan. 11). *H. Con. Res. 1, Sense of the Congress on Approval of Military Action*. [Online]. Available: LEXIS Library: GENFED File: BILLS

*Basic format for patents:*

J. K. Author, “Title of patent,” U.S. Patent *x xxx xxx*, Abbrev. Month, day, year.

*Example:*

1. G. Brandli and M. Dick, “Alternating current fed power supply,” U.S. Patent 4 084 217, Nov. 4, 1978.

*Basic format**for theses (M.S.) and dissertations (Ph.D.):*

J. K. Author, “Title of thesis,” M.S. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

J. K. Author, “Title of dissertation,” Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

*Examples:*

1. J. O. Williams, “Narrow-band analyzer,” Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, USA, 1993.
2. N. Kawasaki, “Parametric study of thermal and chemical nonequilibrium nozzle flow,” M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

*Basic format for the most common types of unpublished references:*

J. K. Author, private communication, Abbrev. Month, year.

J. K. Author, “Title of paper,” unpublished.

J. K. Author, “Title of paper,” to be published.

*Examples:*

1. A. Harrison, private communication, May 1995.
2. B. Smith, “An approach to graphs of linear forms,” 2014, *arXiv:2105.02824*.
3. A. Brahms, “Representation error for real numbers in binary computer arithmetic,” IEEE Computer Group Repository, Paper R-67-85.

*Basic formats for standards:*

a) *Title of Standard*, Standard number, date.

b) *Title of Standard*, Standard number, Corporate author, location, date.

*Examples:*

1. IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.
2. Letter Symbols for Quantities, ANSI Standard Y10.5-1968.

**First A. Author** (Fellow, IEEE) and all authors may include biographies if the publication allows. Biographies are often not included in conference-related papers. Please check the Information for Authors to confirm. Author photos should be current, professional images of the head and shoulders. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author’s educational background is listed. The degrees should be listed with the type of degree in what field, which institution, city, state, and country, and year the degree was earned. The author’s major field of study should be lowercase. 

The second paragraph uses the preferred third person pronoun (he, she, they, etc.) and not the author’s last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. The format for listing publishers of a book within the biography is: *Title of Book* (publisher name, year) similar to a reference. Current and previous research interests end the paragraph.

The third paragraph begins with the author’s preferred title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter, Mx. Riley). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications.

**Second B. Author**, photograph and biography not available at the time of publication.

**Third C. Author, Jr.** (Member, IEEE), photograph and biography not available at the time of publication.

1. This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE. It is IEEE style to display support information, including sponsor and financial support acknowledgment, here and not in an acknowledgment section at the end of the article. For example, “This work was supported in part by the U.S. Department of Commerce under Grant BS123456.” The name of the corresponding author appears after the financial information, e.g. *(Corresponding author: M. Smith).* Here you may also indicate if authors contributed equally or if there are co-first authors.

   The next few paragraphs should contain the authors’ current affiliations, including current address and e-mail. For example, First A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@ boulder.nist.gov).

   Second B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

   Third C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba 305-0047, Japan (e-mail: author@nrim.go.jp).

   Mentions of supplemental materials and animal/human rights statements can be included here.

   Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org [↑](#footnote-ref-1)
2. It is recommended that footnotes be avoided (except for the unnumbered footnote with the receipt date on the first page). Instead, try to integrate the footnote information into the text. [↑](#footnote-ref-2)