热湍流环境中以最低能耗实现长距离迁移

徐奥,[1,2,*] 吴华林,[1] 和奚恒东[1,2]

[1]航空学院，西北工业大学，西安 710072，中国[2]极端力学研究所，西北工业大学，西安 710072，中国

（日期：2023年2月9日）

arXiv:2301.04810v2 [physics.flu-dyn] 8 Feb 2023

摘要

我们采用强化学习算法，训练自推进智能体在热湍流环境中进行长距离迁移。我们选择纵横比为 2 的瑞利–贝纳德湍流对流胞作为训练环境（Γ，纵横比定义为胞元长度与单元高度之比）。结果表明，与从原点直线前往目的地的朴素智能体相比，聪明的智能体能够学会利用载流流动来节省推进能量。随后，我们将从 $\Gamma = 2$ 胞元中获得的最优策略用于测试聪明的智能体在 Γ 最高达 32 的对流胞中的迁移表现。在较大的 Γ 胞元中，水平叠置涡胞的主导流动模态较不稳定，高阶流动模态所含的能量也有所增加。我们发现，优化后的策略能够成功推广到纵横比更大的对流胞中。此外，聪明智能体消耗的推进能量与朴素智能体消耗的推进能量之比会随着 Γ的增大而降低，这表明在更大的 Γ 胞元中，聪明的智能体能够节省更多推进能量。我们还评估了智能体从随机选定的原点出发时优化策略的表现，以检验学习框架的鲁棒性，并提出了提高成功率的可能方案。这项工作对长距离迁移问题具有启示意义，例如无人飞行器在湍流对流环境中巡逻；规划节能轨迹有助于提升其续航能力。[a]

I. 引言

人类长期以来一直为飞行着迷，而鸟类能教会我们如何高效飞行。一些翱翔鸟类在旅途中无需扇动翅膀便能飞行很长距离，而且只有在起飞或降落阶段才会耗费最多力气。例如，Weimerskirch 等人 [1] 表明，军舰鸟在跨洋飞行期间可连续翱翔长达48天。Williams 等人 [2] 记录到，一只安第斯神鹫曾连续5小时以上不扇动翅膀，飞行距离达172公里。Croxall 等人 [3]揭示，速度最快的灰头信天翁只需46天便可完成全球环航。直到1885年，Lancaster发表了开创性的观察与推论 [4], 时[a]本文仅供个人使用。如需作其他用途，须事先征得作者和 APS 的许可。本文发表于 Xu 等人所著的《物理评论：流体》8 卷，023502（2023），并可

[a] 见 https://doi.org/10.1103/PhysRevFluids.8.023502。

* 通讯作者：axu@nwpu.edu.cn

鸟类飞行时不扇动翅膀之谜才逐渐得到解答。鸟类的秘诀在于，它们能够利用温暖上升气流（也称为热气流）来减少能量消耗。热气流是大气对流层（即对流层）中形成的对流流动的一部分。晴天时，太阳的热量使地面升温，地面又加热其上方的空气。暖空气膨胀后变得较轻并上升，这些上升气流形成的气柱称为热气流。

鸟类、滑翔机和无人飞行器（UAVs）都能利用热气流中的上升气流来延长续航能力、节省能量。例如，MacCready [5]确定了在热气流之间飞行时的最佳滑翔速度，以最大限度地提高速度和能量增益。Allen 等人[6] 估算，一架名义续航时间为 2 小时的无人飞行器，夏季可将续航时间延长 12 小时，冬季可延长 6 小时。为研究对流升力的利用，人们开发了多种热气流模型，例如烟囱模型和气泡模型 [7]。烟囱热气流是连续上升的空气柱，从地表延伸至对流层的最高处 [8]。气泡热气流是封闭的上升气团，在近地面处形成上升的涡环；涡环核心处的上升气流由空气的浮力提供 [9]。空气离开气泡核心后会冷却并失去浮力，因而沿涡环外侧向下运动，完成一个循环。烟囱热气流模型和气泡热气流模型都描述了简化情形，其中不存在湍流运动，或者将其波动建模为高斯白噪声。然而，在对流层中，风场表现出强烈的湍流波动。Akos 等人 [10] 发现，环境中的湍流波动给识别有效的热气流翱翔策略带来了挑战。Laurent 等人[11] 进一步指出，湍流会在所有飞行模式中留下印记，并揭示了金雕的飞行轨迹与湍流携带粒子的轨迹之间的类比关系。他们还强调，要理解飞行物体的运动和行为，就必须充分考虑湍流。

为模拟强对流天气中的风流动模式，一种称为瑞利–贝纳德（RB）对流的典型湍流对流系统，可用于描述由浮力驱动的湍流流动 [12–14]。RB 系统的控制参数主要包括普朗特数（Pr，本文稍后定义）、瑞利数（Ra）和单元纵横比（Γ）。Pr 描述流体的热物理性质；空气的 $\mathrm{Pr} = 0.71$。Ra 描述温差所产生的浮力与黏性力之比，并且大气中的 $10^{18} \leq \mathrm{Ra} \leq 10^{22}$

[13]。 Γ 表征对流系统的几何信息，而 $\Gamma \approx 100$适用于中尺度对流系统 [15]。在瑞利–贝纳德湍流对流中，重要的相干结构包括小尺度热羽流、大尺度环流涡卷以及超大尺度超结构。热羽流从热边界层或冷边界层脱离，随后相互碰撞并合并，进而自组织形成大尺度环流涡卷。如果对流系统的水平延伸距离是垂直距离的数倍，那么热羽流就会形成由冷下沉流和热上升流体构成的相互连通的脊状结构网，也称为热湍流的超结构 [16, 17]。

采用瑞利–贝纳德湍流环境，Reddy等人 [18] 利用强化学习算法对滑翔机进行了数值训练，使其能够借助热气流上升。训练后的滑翔机可以螺旋上升，从低空升至高空，其飞行模式与自然界中的翱翔鸟类相似 [19]。他们分析了湍流强度变化时滑翔机飞行策略的变化。随后，他们为一架翼展为 2 米的滑翔机配备相关设备，并在野外进行训练，使其能够自主穿行于大气热气流中 [20]。在 Reddy等人的研究中 [18, 20], ，主要目标是训练滑翔机飞得更高；而对于无人机的实际应用，更常见的场景是从一个位置飞往另一个位置。为尽量减少在热湍流环境中点对点迁移时的能量消耗，Xu 等人 [21] 优化了瑞利–贝纳德湍流对流中自推进智能体的轨迹，使其能够尽可能利用热湍流的动能。与直线推进轨迹相比，优化后的轨迹可使智能体节省约三分之二的能量消耗。

在先前关于瑞利–贝纳德湍流环境中翱翔的研究中， [18, 21], 模拟的瑞利–贝纳德对流胞的纵横比为 $\Gamma \leq 2$。然而，迁移通常发生在纵横比较大的对流系统中，且需要跨越较长距离。本研究旨在训练自推进智能体，使其能够在具有多个环流涡卷的、大纵横比瑞利–贝纳德单元中迁移。本文其余部分安排如下：第二节介绍湍流环境模拟的数值细节，包括瑞利–贝纳德对流的数学模型和自编数值求解器。第三节介绍自推进智能体的动力学，以及用于训练智能体寻找节能轨迹的强化学习算法。第四节首先介绍瑞利–贝纳德对流中的一般流动模式，随后展示智能体在 $\Gamma = 2$ 胞元中迁移的训练结果，并测试智能体在更大的 Γ 胞元中迁移的表现。

第五节总结本研究的主要发现。

II. 湍流环境的模拟

II.1. 瑞利–贝纳德湍流热对流的数学模型

我们基于布辛涅斯克近似，对瑞利-贝纳德对流胞中的湍流环境进行模拟。我们假设流体流动不可压缩，并将温度视为主动标量，它通过浮力影响速度场。忽略黏性耗散和压缩功，并假设所有输运系数均为常数。于是，瑞利–贝纳德热对流的控制方程可写为

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho_0}\nabla P + \nu \nabla^2 \mathbf{u} + g\beta_T (T - T_0)\hat{\mathbf{y}}, \tag{2}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \alpha_T \nabla^2 T, \tag{3}$$

其中，$\mathbf{u} = (u, v)$、P 和 T 分别表示流体的速度、压力和温度。ρ_0 和 T_0 分别表示参考密度和参考温度。$\hat{}\mathbf{y}$ 是平行于重力方向的单位向量。采用如下尺度变换

$$\begin{aligned} &\mathbf{x}^* = \mathbf{x}/H, \quad t^* = t/\sqrt{H/(\beta_T g \Delta_T)}, \quad \mathbf{u}^* = \mathbf{u}/\sqrt{\beta_T g H \Delta_T}, \\ &P^* = P/(\rho_0 g \beta_T \Delta_T H), \quad T^* = (T - T_0)/\Delta_T. \end{aligned} \tag{4}$$

于是，方程 (1)–(3) 可改写为无量纲形式：

$$\nabla \cdot \mathbf{u}^* = 0, \tag{5}$$

$$\frac{\partial \mathbf{u}^*}{\partial t^*} + \mathbf{u}^* \cdot \nabla \mathbf{u}^* = -\nabla P^* + \sqrt{\frac{Pr}{Ra}}\nabla^2 \mathbf{u}^* + T^* \tilde{\mathbf{y}}, \tag{6}$$

$$\frac{\partial T^*}{\partial t^*} + \mathbf{u}^* \cdot \nabla T^* = \sqrt{\frac{1}{PrRa}}\nabla^2 T^*. \tag{7}$$

其中，H是单元高度，并将其选为特征长度；$t_f = \sqrt{H/}$（$\beta_T g \Delta_T$）是自由落体时间，并将其选为特征时间。Δ_T是加热壁和冷却壁之间的温差。这两个无量纲参数为瑞利数和普朗特数，其定义如下：

$$\mathrm{Ra} = \frac{g\beta_T \Delta_T H^3}{\nu \alpha_T}, \quad \mathrm{Pr} = \frac{\nu}{\alpha_T}. \tag{8}$$

II.2. 热对流的格子玻尔兹曼方法

我们采用格子玻尔兹曼（LB）方法模拟热对流。格子玻尔兹曼方法的优点包括易于实现、便于并行化以及计算效率高 [22]。具体而言，我们选用 D2Q9 模型求解纳维–斯托克斯方程，以模拟流体流动；选用 D2Q5 模型求解能量方程，以模拟热传递。为提高数值稳定性，在密度分布函数和温度分布函数的演化方程中均采用多松弛时间碰撞算子。密度分布函数的演化方程写为：

$$f_i(\mathbf{x}+\mathbf{e}_i\delta_t, t+\delta_t) - f_i(\mathbf{x},t) = -\left(\mathbf{M}^{-1}\mathbf{S}\right)_{ij}\left[\mathbf{m}_j(\mathbf{x},t) - \mathbf{m}_j^{(\mathrm{eq})}(\mathbf{x},t)\right] + \delta_t F_i', \tag{9}$$

其中， f_i 为密度分布函数，x为流体微团的位置， t 为时间， δ_t为时间步长，$\mathbf{e}_i$ 为沿第 i 个方向的离散速度。方程（9）右端的强迫项 F_i'表示为 $\mathbf{F}' = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{S}/2)\,\mathbf{M}\tilde{}\mathbf{F}$，而项 $\mathbf{M}\tilde{}\mathbf{F}$ 表示为 [23] $\mathbf{M}\tilde{}\mathbf{F} = [0,\ 6\mathbf{u}\cdot\mathbf{F},\ -6\mathbf{u}\cdot\mathbf{F},\ F_x,\ -F_x,\ F_y,\ -F_y,\ 2uF_x - 2vF_y,\ uF_x + vF_y]^T$，其中 $\mathbf{F} = \rho g \beta_T (T - T_0)\hat{}\mathbf{y}$。宏观密度 ρ 和速度 $\mathbf{u}$ 由$\rho = \sum_{i=0}^{8} f_i$得到，$\mathbf{u} = \left(\sum_{i=0}^{8}\mathbf{e}_i f_i + \mathbf{F}/2\right)/\rho$。

T温度分布函数的演化方程写为 s

$$g_i(\mathbf{x}+\mathbf{e}_i\delta_t, t+\delta_t) - g_i(\mathbf{x},t) = -\left(\mathbf{N}^{-1}\mathbf{Q}\right)_{ij}\left[\mathbf{n}_j(\mathbf{x},t) - \mathbf{n}_j^{(\mathrm{eq})}(\mathbf{x},t)\right], \tag{10}$$

其中， g_i 为温度分布函数。宏观温度 T 由 $T = \sum_{i=0}^{4} g_i$得到。关于格子玻尔兹曼方法的更多数值细节以及自研求解器的验证，请参阅我们此前的工作 [24–26]。

II.3. 湍流热对流的仿真设置

我们考虑一个长度为 L 、高度为 H 的二维瑞利–贝纳德单元。单元的上下壁分别保持恒定的低温和高温，另外两面竖直壁面则为绝热壁面；四面壁均采用无滑移速度边界条件。我们将单元纵横比（$\Gamma = L/H$）设为 $2 \le \Gamma \le 32$，并将普朗特数设为 $\mathrm{Pr} = 0.71$（对应于空气这一工作流体），瑞利数设为 $\mathrm{Ra} = 10^8$。尽管由于计算资源有限，无法模拟超高瑞利数对流，因此此处的瑞利数远低于大气中的数值，但我们注意到，$\mathrm{Ra} = 10^8$ 时的瑞利–贝纳德对流已经呈现出强烈的湍流涨落，流动处于“硬湍流”状态 [27]。

我们还检查了湍流数据库，并确认系统已达到统计平稳态，模拟的初始瞬态效应也已消退。

III. 动　　力学与自推进智能体的控制

III.1. 自推进智能体的运动学模型

自推进智能体的动力学可描述为

$$\mathbf{u}_{\text{agent}}(t) = \mathbf{u}_{\text{fluid}}(t) + \mathbf{u}_{\text{propel}}(t) \tag{11}$$

$$\mathbf{x}_{\text{agent}}(t + dt) = \mathbf{x}_{\text{agent}}(t) + \mathbf{u}_{\text{agent}}(t) \cdot dt \tag{12}$$

其中，dt 为时间步长，$\mathbf{u}^{\text{agent}}$ 和 $\mathbf{x}^{\text{agent}}$ 分别表示智能体的速度和位置。$\mathbf{u}_{\text{fluid}}$ 表示载流流体的速度，$\mathbf{u}^{\text{propel}}$ 表示智能体产生的速度。我们假设，在没有控制时，智能体的速度等于载流流动的速度；而在控制下，智能体的速度是载流速度与智能体推进速度的叠加。Krishna 等人[28]此前也采用了类似的智能体动力学模型。为模拟现实场景中智能体有限的推进能力，我们将智能体的最大推进速度 $\|\mathbf{u}_{\text{agent}}\|$限制为低于最大载流速度的三分之一。另一方面，未来的工作可以考虑更复杂的自推进智能体运动学模型，例如包含惯性和旋转动力学 [29–33], 、推进智能体的振颤和翻滚 [34], 以及多模态运动[35]的模型。

III.2. 通过强化学习实现最优控制

我们采用强化学习算法来优化智能体的控制，使其沿着能量消耗较低的轨迹迁移。强化学习算法的优点包括：无需针对控制和优化任务进行专门设计；易于复用，可加快相似系统配置中的优化；并且能够抵抗混沌系统中的扰动 [36, 37]。在强化学习算法中，智能体观察环境的状态，并决定采取某个动作与环境交互。如果智能体随后获得奖励（或受到惩罚），那么它今后就更有可能重复（或放弃）该动作。总的来说，智能体通过试错进行学习，长期目标是最大化累积期望回报，并最终

提升其性能。应用强化学习算法，我们可以得到最优策略，为智能体提供有利动作的建议 [38–41]。无模型强化学习算法通常可分为基于策略的方法和基于价值的方法。在基于策略的方法中，例如策略梯度法，会优化策略网络 θ 的参数，以最大化性能目标 $J(\pi_\theta)$。其中， π_θ 表示参数化随机策略。基于策略的方法采样效率低，因此学习速度较慢，不适用于复杂流动问题 [42]。在基于价值的方法中，例如 Q-学习方法，智能体采取动作 a ，以尝试最大化最优动作价值函数，即 $a(s) = \arg\max_a Q_\theta(s,a)$。其中， s 表示环境的状态， $Q_\theta(s,a)$用于近似最优动作价值函数$Q^*(s,a)$。使用 Q-学习方法，Colabrese等人 [29] 表明，趋重性游动体能够在稳定的泰勒–格林涡流中到达高处；Mui˜nos-Landin等人[43] 证明，人工自热泳微型游动体能够在布朗运动的影响下导航；Monderkamp等人 [44] 训练主动布朗粒子穿越复杂运动性景观；Gazzola等人 [45] 和Verma等人 [46] 则发现了能够最大限度降低鱼群阻力和能量消耗的最优游动策略。以上五个例子采用了离策略学习技术，这意味着每次更新都会随机抽取训练期间任意时刻收集的数据，即$Q(s_t,a_t) = Q(s_t,a_t) + \alpha\left[r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right]$。此前，Reddy等人[18]采用了状态-动作-奖励-状态-动作（SARSA）方法，该方法可视为 Q-学习方法的一种变体。主要区别在于，SARSA方法采用同策略学习技术，利用当前策略执行的动作来学习 Q-值，即 $Q(s_t,a_t) = Q(s_t,a_t) + \alpha\left[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\right]$。然而，基于价值的方法只能用于离散状态与动作空间；而在大多数现实场景中，例如训练车辆进行导航，人们更倾向于采用连续状态与动作空间，以发展出更灵活多样的运动方式，从而应对复杂的导航任务。

在本研究中，我们采用软演员-评论家（SAC）算法，该算法融合了基于策略的方法与基于价值的方法。在SAC算法中，智能体通过演员网络决定下一步动作，同时由评论家网络对该动作进行进一步评估，以指导训练过程。此外，演员旨在最大化期望奖励（即成功完成任务），同时也最大化熵（即尽可能随机地采取动作）。在熵正则化的强化学习中，优化

问题可描述为

$$\pi^*(\theta) = \arg\max_{\pi} E_{\tau\sim\pi}\left[\sum_{t=0}^{\infty}\{R(s_t, a_t, s_{t+1}) + \alpha H[\pi(\cdot|s_t)]\}\right] \tag{13}$$

在上述方程中， π^* 是最优策略。奖励函数 r 取决于环境的当前状态 s_t、当前动作 a_t以及环境的下一状态s_{t+1}。 α 是权衡系数。 H 的熵 τ 根据其分布π 计算为 $H[\pi(\cdot|s_t)] = E_{\tau\sim\pi}[-\log\ \pi(\tau)]$。关于 SAC算法的更多详情，可参见 Haarnoja等人 [47]。

强化学习框架的关键要素包括智能体能够观测到的环境线索（即环境的当前状态 s_t ）、智能体采取的动作（即 a_t），以及智能体对其行为的响应（即奖励 r_t）。在本研究中，为了在大长宽比对流胞中迁移，智能体的观测变量包括载流速度 $\mathbf{u}_{\text{fluid}}$、智能体在垂直方向上的空间坐标 y_{agent}以及流体温度 T。在第 IV.2 节中，我们将详细讨论观测变量的选择。动作变量是智能体生成的推进速度 $\mathbf{u}^{\text{propel}}$ 。遵循 Xu 等人 [21], 先前的研究，我们假设智能体获得的奖励同时受到其当前状态、能量消耗和时间消耗的影响

$$r(t) = r_s(t) + r_e(t) + r_h(t). \tag{14}$$

这里， r_s 表示受智能体当前状态影响的奖励。如果智能体从顶部或底部边界迁出流动区域，则会受到 $-\phi$的惩罚；如果智能体向右运动并靠近右侧边界，则会获得基本奖励 e_{basic} ，其经验前因子为 ε。因此， r_s可写为

$$r_s(t) = \begin{cases} -\phi, & \text{agent is out of the flow domain} \\ \varepsilon e_{\text{basic}}, & x^t_{\text{agent}} > x^{t-1}_{\text{agent}} \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

本文采用 $\phi = 10$ 和 $\varepsilon = 10$。关于奖励函数中超参数敏感性的详细讨论见附录A。在公式（14）中， r_e 表示受智能体能量消耗影响的奖励。如果智能体的推进速度 $\mathbf{u}^{\text{propel}}$ 与背景流的速度 $\mathbf{u}_{\text{fluid}}$方向一致，即这两个向量之间的角度（记为 θ）小于 90°，则智能体将获得奖励

$\varepsilon[e_{\text{basic}}+(e_{\max}-e)]$,，其中 $e=0.5||\mathbf{u}_{\text{propel}}||^2$ 和 $e_{\text{basic}}=e_{\max}=0.5(||\mathbf{u}_{\text{propel}}||)^2_{\max}$；如果 $\mathbf{u}^{\text{propel}}$ 与 $\mathbf{u}_{\text{fluid}}$ 之间的角度大于 $90°$，则智能体将受到$-2\varepsilon(e_{\text{basic}}+e)$的惩罚。上述设计还意味着，当智能体沿载流方向迁移时，推进速度越高，获得的奖励越低；当智能体逆载流方向迁移时，推进速度越高，受到的惩罚越大。因此， r_e 可写为

$$r_e(t)=\begin{cases}\varepsilon[e_{\text{basic}}+(e_{\max}-e)], & 0°\le\theta\le 90°\\ -2\varepsilon(e_{\text{basic}}+e), & 90°\le\theta\le 180°\end{cases} \tag{16}$$

在公式（14）中， r_h 表示受智能体时间消耗影响的奖励。如果智能体经由右侧边界离开区域，我们便假定智能体完成了任务，并会获得与所用时间成反比的奖励。这一设计意味着，智能体越早到达目的地，获得的奖励就越高。因此， r_h 写作

$$r_h(t)=\begin{cases}(t_{\max}-t)/\varepsilon, & x^t_{\text{agent}}>L\\ 0, & \text{otherwise}\end{cases} \tag{17}$$

IV. 结果与讨论

IV.1. 大宽高比瑞利-贝纳德对流中的一般流动模式

图 1 展示了 $\text{Ra}=10^8$、$\text{Pr}=0.71$ 和 $2\le\Gamma\le 32$ 条件下温度场的典型瞬时图。我们可以辨识出小尺度热羽流和大尺度环流卷。薄的热边界层出现在底部加热壁和顶部冷却壁附近。从边界层释放的羽流向上（或向下）穿透，朝向温度较低（或较高）的对侧壁面，并在中心区域发生强烈混合。热气流几乎以周期性的方式从相对固定的位置释放出来，而相邻且向相反方向运动的热气流卷吸周围流体，并自组织形成环流涡卷。此前，Wang 等人 [48] 发现，取决于初始条件，在给定纵横比下，流动系统会演化为具有不同涡卷数量的不同最终湍流状态。在我们的研究中，对流涡卷数量 n 分别为 2、4、9、18 和 34，对应于 $\Gamma=2$、4、8、16 和 32 个对流胞；其对应的平均纵横比分别为 $\Gamma_r=\Gamma/n=1$、1、0.889、0.889 和 0.941，这与椭圆不稳定性理论 [48] 的预测一致。

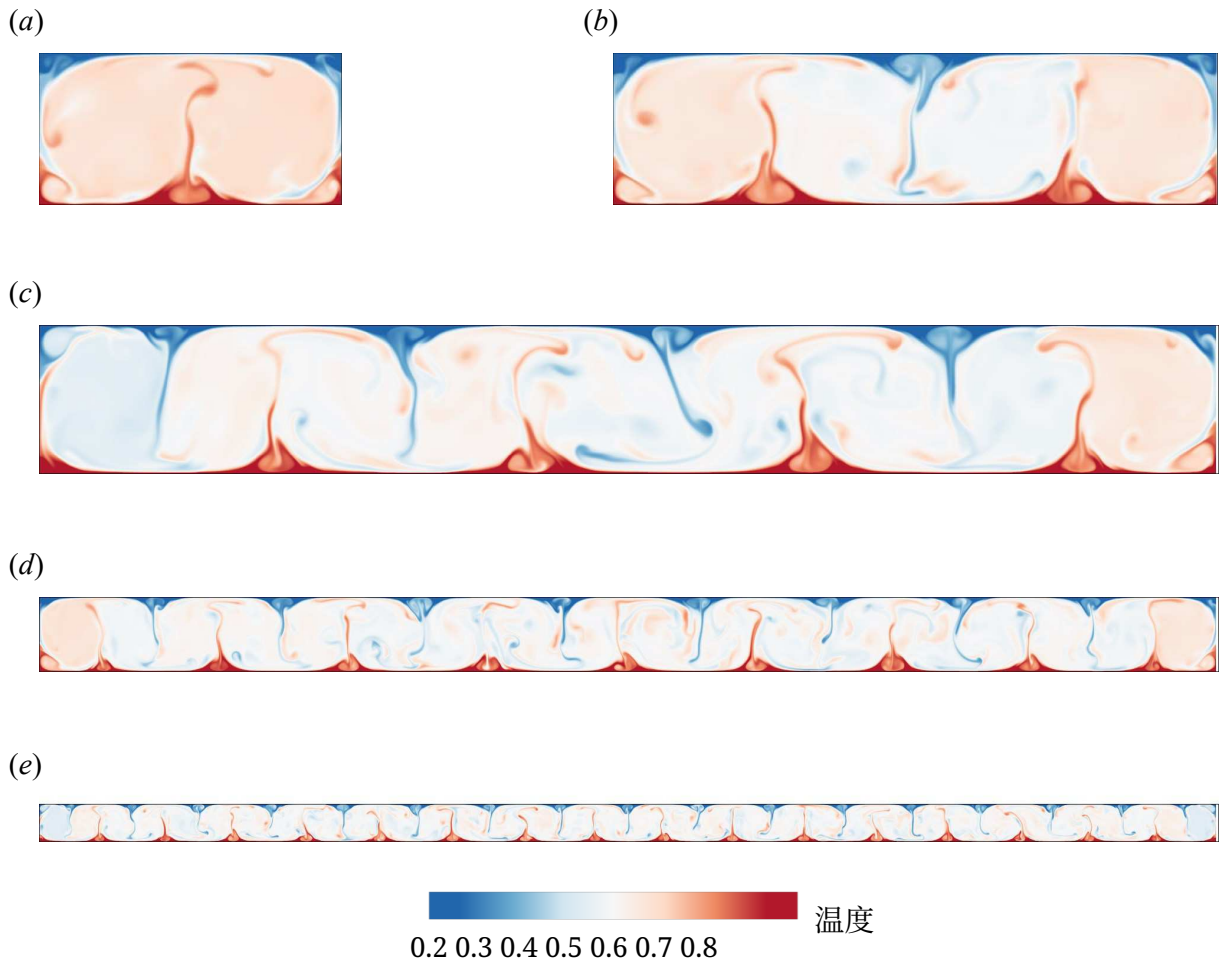

图 1。在 Ra $= 10^8$、Pr $= 0.71$ 条件下的典型瞬时温度场，其中 (a) $\Gamma = 2$、(b) $\Gamma = 4$、(c)$\Gamma = 8$、(d) $\Gamma = 16$ 和 (e) $\Gamma = 32$。

随后，我们考察雷诺数（Re）这一全局响应参数对控制参数 Γ 的响应。这里，全局流动强度计算为 Re $= \sqrt{\langle\ (u^2 + v^2)\ \rangle_{V,\ t}} H/\nu$ 其中$\langle\cdots\rangle_{V,\ t}$表示时空~~平均值~~。测得的 Re 随 Γ 的变化如图 2(a) 所示，我们可以观察到，随着单元纵横比增加，全局流动强度增强。此外，随着 Γ 增加，Re 逐渐趋于渐近值，这与三维（3D）对流腔体 [16] 中的情况类似。此前，Xu 等人 [21] 发现，通过强化学习算法获得的节能策略对全局流动强度的小幅扰动并不敏感，但当全局流动强度增加（或减少）超过一个数量级时，该策略会发生变化。由于自推进智能体是在 $\Gamma = 2$ 单元中训练的，我们进一步验证了，即使在 $\Gamma = 32$ 单元中，与 $\Gamma = 2$ 单元相比，流动强度增加了约 22%，这表明在更大纵横比的单元中，流动强度并不会增加

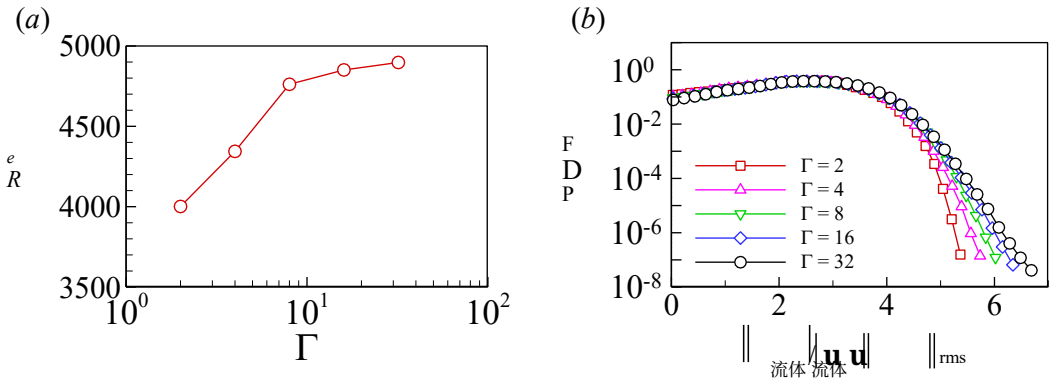

图 2. (a) 雷诺数，以及 (b) 在不同 Γ 条件下，$\mathrm{Ra} = 10^8$、$\mathrm{Pr} = 0.71$ 时归一化速度大小 $\|\mathbf{u}_{\mathrm{fluid}}\|/\|\mathbf{u}_{\mathrm{fluid}}\|_{\mathrm{rms}}$ 的概率密度函数（PDF）。

显著。为量化速度大小的波动，我们绘制了不同Γ条件下归一化速度大小 $\|\mathbf{u}_{\mathrm{fluid}}\|/\|\mathbf{u}_{\mathrm{fluid}}\|_{\mathrm{rms}}$ 的概率密度函数（PDF），如图 2(b) 所示。可以看到，不同 Γ 下的概率密度函数头部重合，而随着 Γ 增大，概率密度函数尾部略微延伸，这表明速度大小 $\|\mathbf{u}_{\mathrm{fluid}}\|$ 的波动程度有所增加。

为从湍流数据库中提取相干流动结构，我们采用本征正交分解（POD）分析。在 POD 中，时空矢量场 $\mathbf{X}(\mathbf{r}, t)$ 被分解为正交特征函数 $\phi_i(\mathbf{r})$ 及其振幅 $a_i(t)$ 的叠加形式：

$$\mathbf{X}(\mathbf{r}, t) = \sum_{i=1}^{\infty} a_i(t)\phi_i(\mathbf{r}) \tag{18}$$

这里将矢量场 $\mathbf{X}(\mathbf{r}, t)$ 选作流速场 $\mathbf{X} = (u, v)$。实际上，我们可以对数据集 $\mathbf{X}$ 进行奇异值分解（SVD），从而获得流动模态$\phi_i(\mathbf{r})$及其对应的模态振幅 $a_i(t)$ [49]。由于大宽高比单元中的湍流对流数据库规模庞大，其中包含高空间分辨率和长时间演化数据，执行标准 SVD 会消耗大量内存。为此，我们采用随机奇异值分解来降低计算负荷 [50]。图 3 展示了不同 Γ 下第一 POD模态 $\phi_1(\mathbf{r})$ 的形状。能量最高的 POD模态由水平堆叠的环流主涡卷组成，这些主涡卷分别沿顺时针方向或逆时针方向旋转，并呈现周期性模式。涡旋边缘附近出现较大的速度大小，表明智能体跨越涡旋边缘时存在较强的能量势垒。值得注意的是，在高得多的 Ra（即 $\mathrm{Ra} > 10^{10}$）下，此时大尺度环流较弱，且流动由

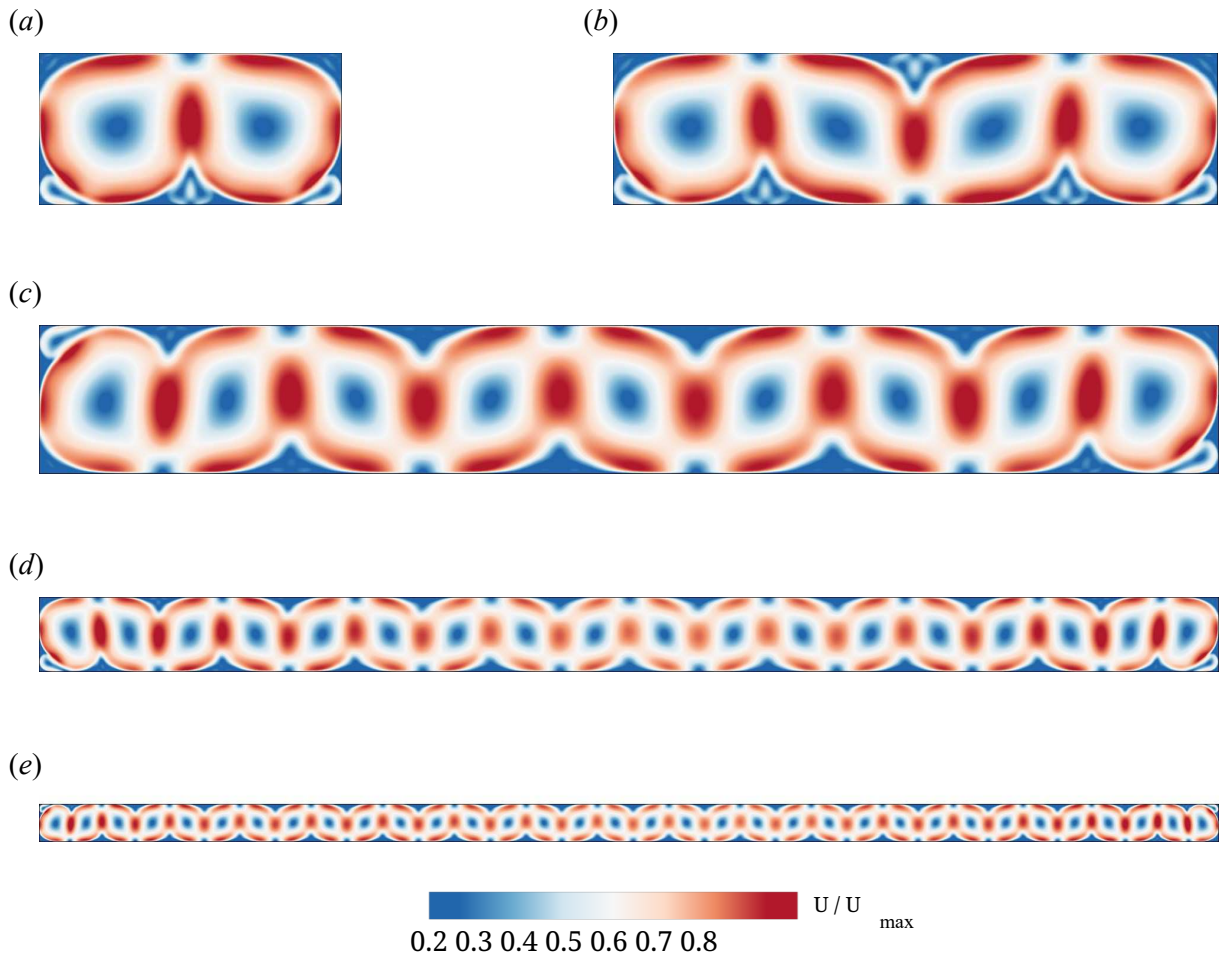

图 3. 瑞利数 $= 10^8$、普朗特数 $= 0.71$ 时第一阶本征正交分解（POD）模态 ϕ_1(r) 的等值线：(a) $\Gamma = 2$、(b) $\Gamma = 4$、(c) $\Gamma = 8$、(d) $\Gamma = 16$ 和 (e) $\Gamma = 32$。其中，$U = \sqrt{u^2 + v^2}$ 表示速度大小，$U_{\max}$ 表示用于归一化的 U 的最大值。

多个可移动且具有轨道运动的小涡旋组成 [51–53], ，当前学习框架是否仍然有效，值得进一步研究。

我们进一步计算了第一 POD 模态的稳定性，记为 $S_1 = \sqrt{\langle a_1(t)\rangle_t / \sigma_{a_1}}$；$S_1$ 的值越大，表示涡胞环流模式越稳定。此前，研究人员曾在湍流热对流的傅里叶模态分解中采用类似方法估算涡胞的稳定性 [49, 54]。从图 4(a) 可以看出，第一 POD 模态的稳定性随着 Γ 的增加而降低。我们还分析了第一 POD 模态所含的能量，并计算能量占比，记为 $\lambda_1 / \sum_{i=1}^{\infty} \lambda_i$。这里，$\lambda_i \delta_{ij} = \langle a_i(t) a_j(t) \rangle_t$，其中 λ_i 表示第 i 个 POD 模态的能量，δ_{ij} 为克罗内克符号，$\langle \cdots \rangle_t$ 表示时间平均。从图 4(b) 可以看出，第一 POD 模态的能量占比也随着 Γ的增加而降低。尽管第一

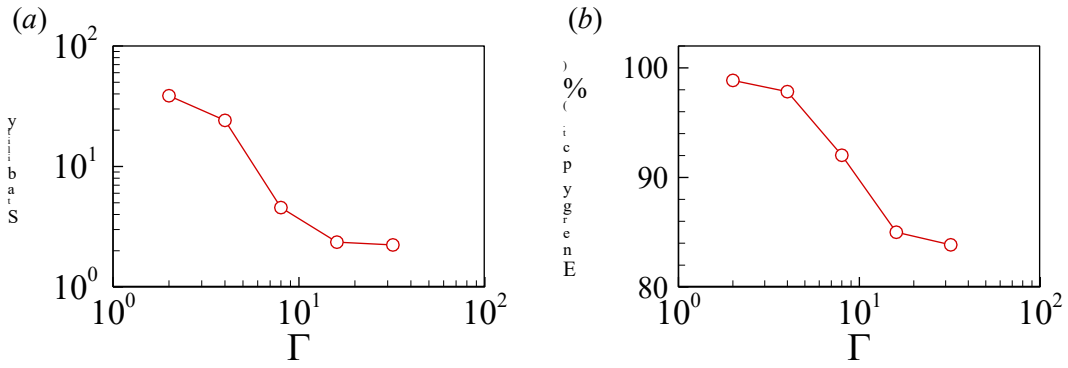

图4。（a）第一POD模态的稳定性，以及（b）在不同 Γ下第一POD模态所包含的能量。

模态仍然是占主导地位的流动结构（例如，在 $\Gamma = 32$ 胞元中，第一 POD 模态所含能量占总能量的 83.8% 以上），但随着 Γ的增加，高阶流动模态会变得更强。因此，在大宽高比单元中，水平堆叠、形成周期性模式的环流涡卷稳定性较低，而这些高阶模态会导致流动模式更加不规则。由于智能体是在 $\Gamma = 2$ 胞元中训练的，上述复杂的流动特征会使智能体难以在更大的 Γ 胞元中识别节能轨迹。

IV.2. T训练智能体在纵横比为2的瑞利–贝纳德单元中迁移

我们首先训练智能体，使其以 $\Gamma = 2$ 穿越湍流瑞利–贝纳德单元。智能体从单元左下角的点 (0, 0) 出发，目标是到达单元的右侧边界（即 $x = 2$ 所在的垂直线）。这里，简单的 $\Gamma = 2$胞元由典型的大尺度相干结构构成（即一个顺时针旋转的主涡和一个逆时针旋转的主涡），因此可作为学习智能体的典范环境。在图 5 中，我们展示了训练后聪明的智能体的瞬时轨迹，对应的视频可在补充视频 [55] 中观看。起初，智能体在左下角顺时针旋转的角涡驱动下向上运动 [见图 5(a)]。到达逆时针方向旋转的主涡边缘后，它便随水平流运动[见图 5(b)]，，直到遇到上升的热气流。随后，智能体沿热气流上升，继续向上攀升 [见图 5(c)]。到达单元顶层后，由于推进速度指向右侧，智能体向右运动并利用水平流

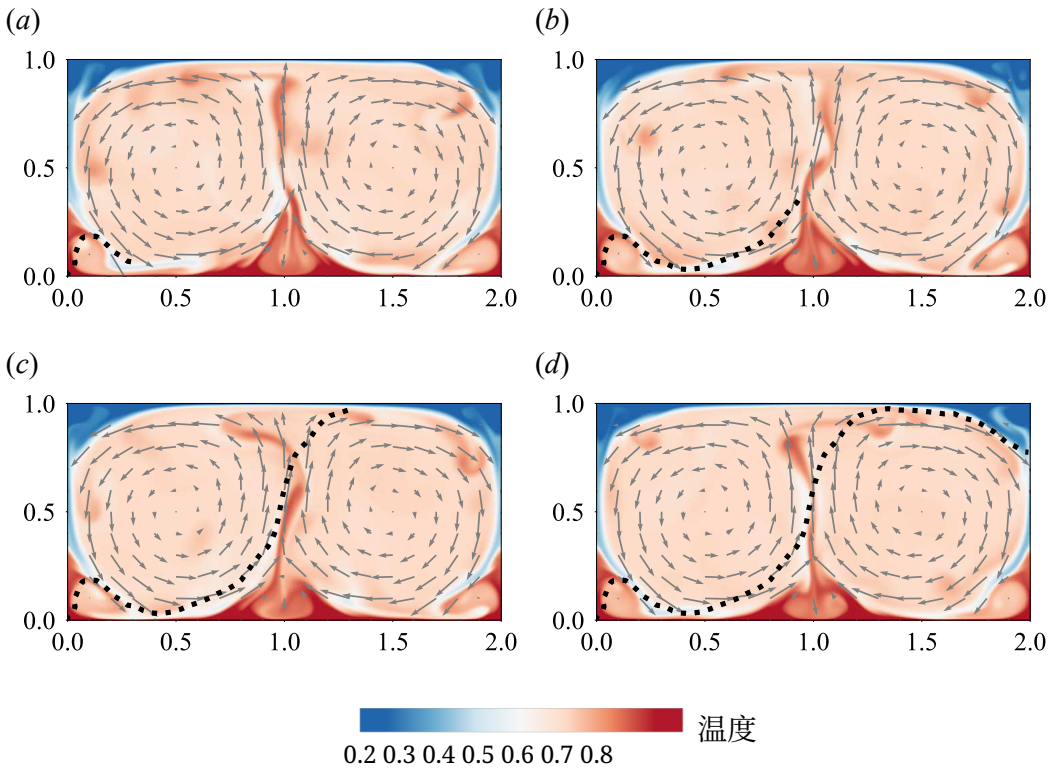

图 5. 瑞利-贝纳德对流中聪明的智能体的轨迹（黑色虚线）：(a) $t = 18$、(b)$t = 36$、(c) $t = 54$ 和 (d) $t = 72$。等值线表示典型的瞬时温度场，向量表示对流的速度场。

[见图 5(d)]。当智能体到达单元的右侧边界时，迁移任务即告完成。总体而言，聪明的智能体会尽可能顺着载流运动，并找到了沿涡胞边缘移动的有效策略。

我们比较聪明智能体与朴素智能体，以突出二者在推进行为和能量消耗方面的差异。这里，朴素智能体指的是从原点直线移动到目的地的智能体。我们设定朴素智能体完成迁移任务所用的总时间与聪明智能体相同，即 t_{total} ，其目的地也与聪明智能体离开右侧边界的位置相同。因此，朴素智能体的速度方向始终保持不变，指向从原点到目的地的方向，速度大小也始终保持为 $\|\mathbf{u}_{\text{agent}}\| = \|\mathbf{x}^{\text{goal}} - \mathbf{x}_{\text{start}}\|/t_{\text{total}}$。图 6(a) 和图 6(b) 分别展示了聪明智能体和朴素智能体的轨迹。从按颜色编码的轨迹所示的瞬时速度大小可以看出，朴素智能体通常比聪明智能体迁移得慢，因为朴素智能体行进的距离较短

，而两者所用的时间相同 t_{total}。尽管聪明智能体迁移得更快，但这并不意味着它会消耗更多能量，因为它可以利用载流流来节省能量。除了轨迹之外，我们还绘制了推进速度矢量（以红色箭头表示）和流体速度矢量（以蓝色箭头表示）。我们计算推进速度矢量的方向（即 θ_{propel}）与流体速度矢量的方向（即 θ_{fluid}）之间的相关系数，结果为

$$C = \langle [\theta_{\text{propel}}(t) - \langle \theta_{\text{propel}} \rangle] \, [\theta_{\text{fluid}}(t) - \langle \theta_{\text{fluid}} \rangle] \rangle / \left(\sigma_{\theta_{\text{propel}}} \sigma_{\theta_{\text{fluid}}} \right). \tag{19}$$

此处方向的范围为 -180° 至 180°。正方向定义为绕 x 轴逆时针旋转，负方向定义为绕 x 轴顺时针旋转。对于智能智能体，所得相关系数 0.56 表明两者之间存在正向统计相关性，说明智能智能体会根据载流的变化调整迁移方向，从而实现节能迁移；对于朴素智能体，所得相关系数 -0.74则意味着存在负向统计相关性。此外，为了定量描述推进速度矢量与流体速度矢量之间的夹角，我们在图 6(c) 和图 6(d) 中分别绘制了智能智能体和朴素智能体的这些夹角直方图。可以看到，对于智能智能体，夹角通常小于 90°，且频率在 20° 附近达到峰值，这进一步证明智能体试图顺应载流。对于朴素智能体，这些夹角的频率在 120°附近达到峰值，表明朴素智能体必须生成与载流流动相反的推进速度，以保持最短迁移路径。

我们假设只有推进速度 $\mathbf{u}_{\text{propel}}$ 会影响智能体的推进能量消耗。因此，累积能耗计算如下：

$$E_{\text{propel}}(t) = \int_0^t \frac{1}{2} \|\mathbf{u}_{\text{propel}}(\tau)\|^2 d\tau \tag{20}$$

在图7(a)中，我们绘制了智能体累积能耗的时间序列。完成迁移任务后，聪明的智能体消耗的推进能量约为朴素智能体的 38%，这意味着沿最短路径迁移并不总能节省能量。我们还将瞬时能耗计算如下：

$$e_{\text{propel}}(t) = \frac{1}{2} \|\mathbf{u}_{\text{propel}}(t)\|^2 \tag{21}$$

从图7(b)可以看出，聪明智能体的 e_{propel} 通常低于朴素智能体的相应数值，并且在整个迁移

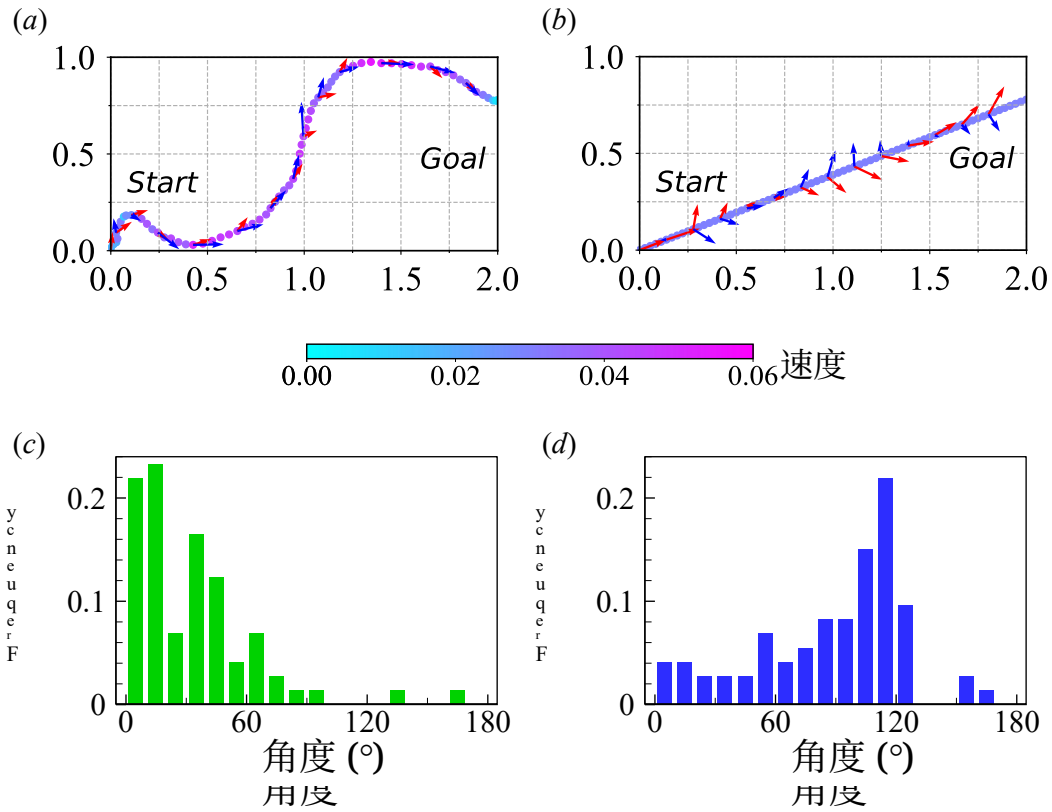

图6. [(a) 和 (b)] 分别为实现节能迁移的智能智能体和直线移动的天真智能体的轨迹。红色箭头表示推进速度，蓝色箭头表示流体速度。轨迹根据智能体的瞬时速度大小进行着色。 [(c) 和 (d)] 分别为智能智能体和天真智能体的推进速度矢量与流体速度矢量之间夹角的直方图。

过程中，聪明智能体的 $t \approx 3$始终较小。对于朴素智能体， e_{propel} 在约 $t \approx 3$时出现第一个峰值，此时它正穿过左下角涡的边缘，而这需要克服较高的能量势垒。在约 $t \approx 23$时，e_{propel} 降至最低，因为朴素智能体到达了流动方向与迁移方向一致的位置。 e_{propel}的第二个峰值出现在约 $t \approx 38$时，此时朴素智能体穿过主涡边缘，并漂移至顺时针旋转的主涡。e_{propel} 的第三个峰值出现在约$t \approx 67$时，此时朴素智能体接近右侧边界。我们还比较了智能体的累积总动能 [见图7(c)], ，其计算方式为

$$E_{\text{total}}(t) = \int_0^t \frac{1}{2} \|\mathbf{u}_{\text{agent}}\|^2 d\tau \tag{22}$$

完成迁移任务后，聪明智能体的总动能几乎是朴素智能体的两倍，其中大部分来自载流的动能。

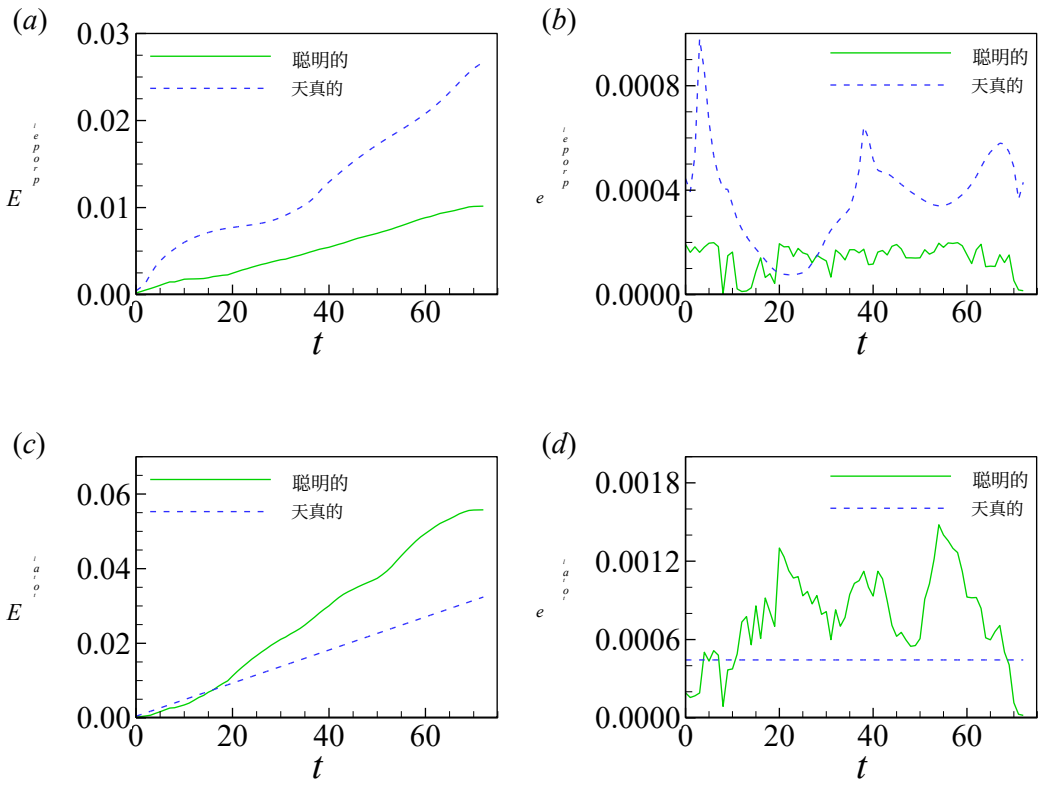

图7. 聪明智能体与朴素智能体的（a）累积能耗 E_{propel}、（b）瞬时能耗 e_{propel}、（c）累积总动能 E_{total}和（d）瞬时总动能 e_{total} 的比较。

同样，我们计算瞬时总动能 e_{total} 如下：

$$e_{\text{total}}(t) = \frac{1}{2}\|\mathbf{u}_{\text{agent}}(t)\|^2 \tag{23}$$

从图7(d)可以看出，聪明智能体的 e_{total} 通常高于朴素智能体，并且在整个迁移过程中始终保持较大的数值。当聪明智能体沿流动方向迁移时，会出现三个峰值，从而利用载流的更多动能。另一方面，由于仿真设置中 $\|\mathbf{u}_{\text{agent}}(t)\|$ 的值恒定，朴素智能体的 e_{total} 保持不变。

选择智能体可以观测到的合适环境线索，对于强化学习训练至关重要。本文通过比较训练过程中累积奖励的变化，数值确定观测变量集合。我们使用不同的随机种子训练每种观测变量集合的五个不同实例，并让每个实例每经过 1000 个环境步数进行一次评估回合。实线表示均值，阴影区域表示五次试验中回报的最小值和最大值；它

表示以 20 个时间步为窗口的移动平均。我们首先考虑智能体无法感知周围流动和温度信息的情形。智能体可以获取自身的位置信息，但仅限于垂直分量，即 $s = \{y\}$。这里不考虑水平位置信息（即 $x \notin \{s\}$ ）；否则，在 $\Gamma = 2$ 胞元中训练的智能体一旦处于 $x > 2$, 便无法在更大的 Γ 胞元中找到最优轨迹。如图8(a)所示，智能体在无法感知流动的情况下表现不佳，其行为与穿越非定常圆柱绕流时的表现类似 [56]。随后，我们考虑智能体还可以感知载流速度，即 $s = \{y, u, v\}$，并将结果绘于图8(b)。可以看到，智能体的表现明显更好，累积奖励也高于无法感知流动的智能体。此外，我们还考虑智能体能够感知额外的涡量信息（即 $s = \{y, u, v, \omega\}$） [29, 57]，并将结果绘于图8(c)。考虑额外的流场信息后，与仅使用速度信息（即 $t \approx 1.0 \times 10^5$，对应 $s = \{y, u, v\}$）相比，累积奖励更早收敛（即 $t \approx 0.5 \times 10^5$，对应 $s = \{y, u, v, \omega\}$）。另一方面，在湍流瑞利–贝纳德对流中，温度作为影响速度的主动标量。接下来，我们考虑智能体能够感知额外的温度信息（即 $s = \{y, u, v, T\}$） [21]。如图8(d)所示，该智能体表现优于前述智能体，累积奖励也更早收敛并几乎保持稳定，这表明温度是智能体在热对流中迁移时的重要环境线索。在附录 B 中，我们评估了更多组合，并据此选择 $s = \{y, u, v, T\}$ 作为观测变量。另一方面，Kubo 和 Shimizu [58] 提出了一种能够基于部分可观测量进行流体流动控制的框架。我们预计，将该框架拓展到湍流中，将有助于简化观测量的选择。

IV.3. 测试智能体在更大长宽比的瑞利–贝纳德单元中迁移 tio

现在，我们将所得策略应用于更大的 Γ 对流胞，以检验聪明的智能体能否以节能方式迁移。第 IV.1 节中的流动模态分析表明，在更大的 Γ 胞元中，水平叠置涡卷的主导流动模态稳定性较差，高阶流动模态所含的能量也会增加。尽管复杂流动特征带来了这些挑战，我们仍能获得优化轨迹，如图 9 所示。从原点 (0, 0) 出发，聪明的智能体首先从左下方胞元角落的角涡中脱离，随后向上攀升；如果第一个主涡卷顺时针旋转，它便沿着热流上升，

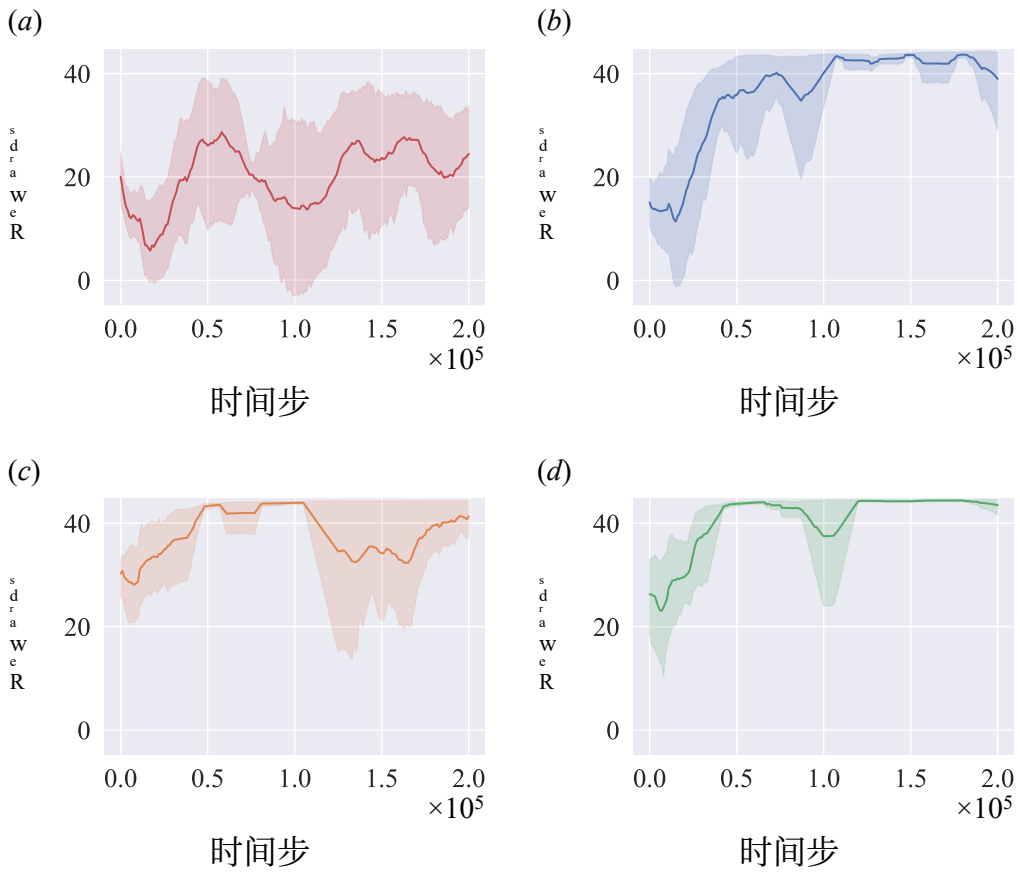

图 8. 不同观测变量组合下训练期间累积奖励的变化：（a） $s = \{y\}$，（b） $s = \{y, u, v\}$，（c） $s = \{y, u, v, \omega\}$ 和（d） $s = \{y, u, v, T\}$。

如果第一个主涡卷逆时针旋转，它便顺着水平流前进。之后，聪明的智能体总是尝试沿着主涡卷的边缘迁移；在这些位置，载流流体流动迅速，并且可以利用流动所提供的大量动能。

随后，我们比较聪明的智能体和朴素智能体在具有不同 Γ 的对流胞中消耗的推进能量。在图 10(a)中，我们绘制了智能体完成迁移任务时的累积推进能量。总体而言，对于两种智能体，随着 Γ增加，能量消耗也会增加，这是因为迁移距离变长。聪明的智能体通过沿着水平堆叠的多个主涡卷的边缘迁移，实现了节能迁移策略，因此我们得到 $E_{\text{propel}} \propto \Gamma$。为了定量描述推进能量的节省幅度，我们绘制了聪明的智能体与朴素智能体的能量消耗比值，如图 10(b)所示。可以看出，在更大的 Γ 胞中， $E_{\text{smart}}/E_{\text{naive}}$ 的比值更小，这意味着可以节省更多推进能量

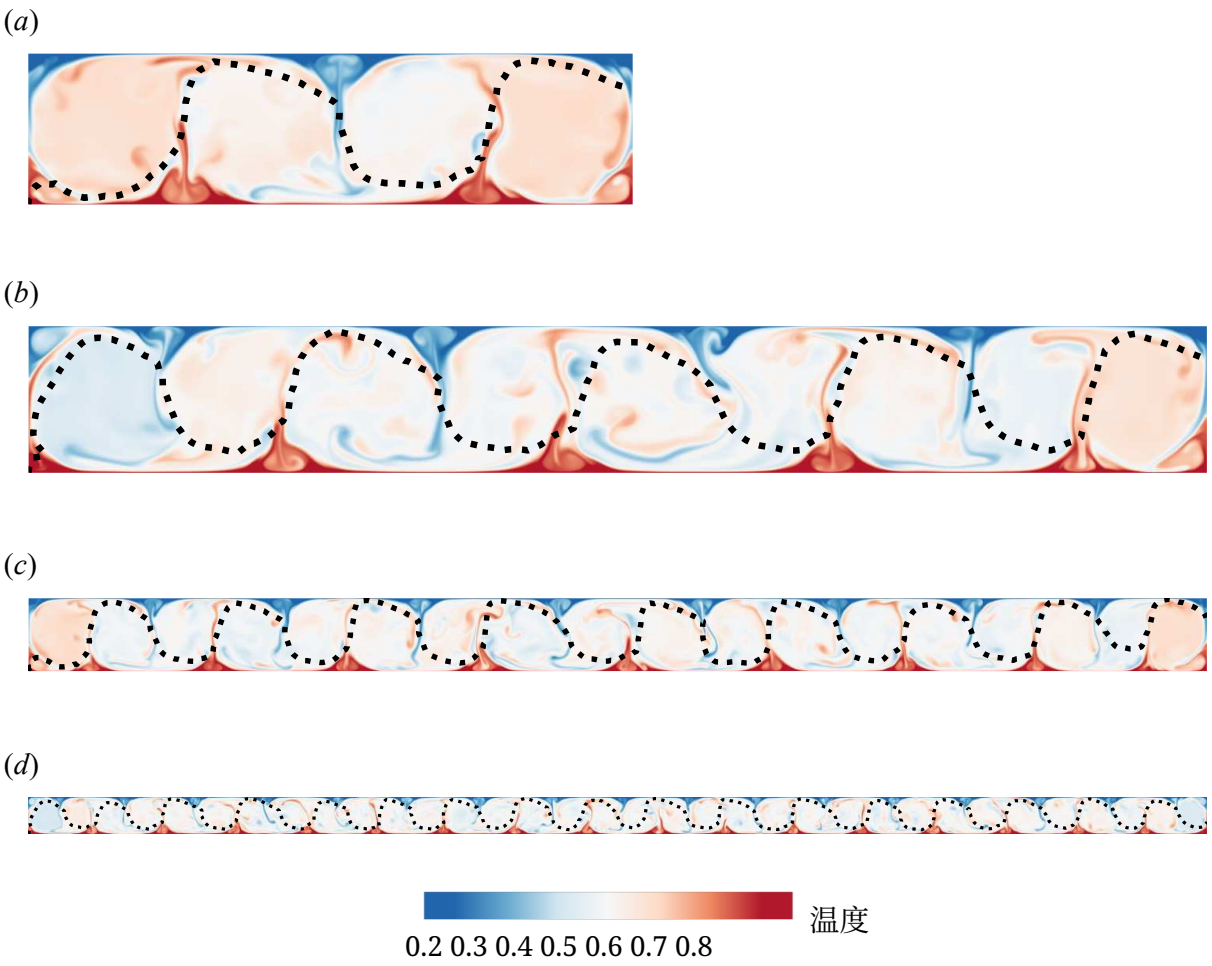

图 9. 对流胞中聪明的智能体的轨迹（黑色虚线）：(a) $\Gamma = 4$、(b) $\Gamma = 8$、(c) $\Gamma = 16$ 和 (d) $\Gamma = 32$。等值线表示典型的瞬时温度场。

。在更大的 Γ胞中，朴素智能体必须穿越更多环流涡卷的边缘并克服更高的能量势垒，而聪明的智能体则以节能方式顺应载流。

上述结果是在原点位置预先设定且固定的情况下得到的，即位于 (0, 0) 点的位置。我们进一步测试节能策略对随机原点位置的鲁棒性。首先，我们将智能体释放在局部流速较弱的位置，即 $\|\mathbf{u}_{\text{fluid}}\| < 0.03$，并在图 11 中绘制了 100 条示例轨迹。可以看到，无论原点位置如何随机变化，成功的尝试都会逐渐汇聚到相似的路径线上，智能体沿着主涡卷的边缘迁移。需要指出的是，我们将这些“随机”位置限制为 $x < \Gamma/2$，以防止智能体被释放在离出口过近的位置。平均成功

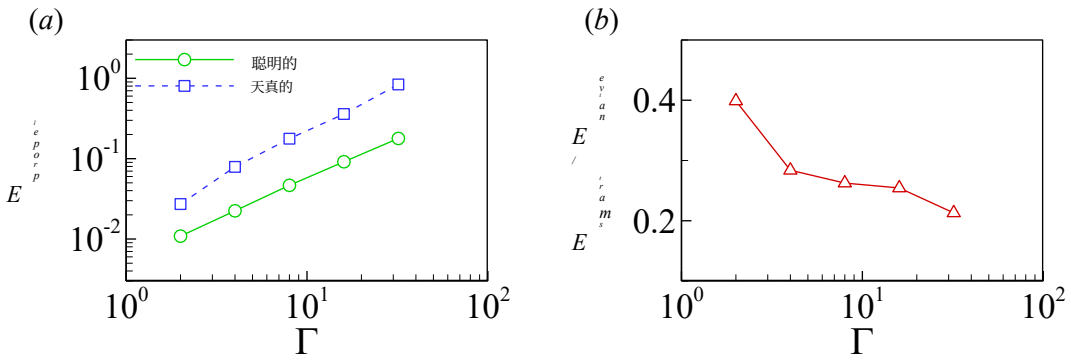

图 10. (a) 聪明智能体和朴素智能体消耗的推进能量；(b) 聪明智能体消耗的能量与朴素智能体消耗的能量之比随 Γ 的变化。

率方面，在 $\Gamma = 4$、8、16 和 32 个胞元中完成迁移任务的成功率分别为 74%、87%、92% 和 97%。随后，我们将智能体释放在局部流速较强的位置，即 $\|\mathbf{u}_{\text{fluid}}\| > 0.03$。平均成功率高得多，在 $\Gamma = 4$、8、16 和 32 个胞元中分别为 100%、100%、100% 和 99%。为简洁起见，我们不再重复绘制示例轨迹。

上述结果表明，完成迁移任务的成功率会随着 Γ的增加而提高。另一方面，在更大的 Γ 胞元中，流动结构更为复杂，我们预计智能体完成迁移任务并获得更高成功率将更具挑战性。为了解这种行为，我们进一步考察以下情形：（i）在载流速度为 $\|\mathbf{u}_{\text{fluid}}\| < 0.01$的位置释放智能体；（ii）在载流速度为 $0.01 < \|\mathbf{u}_{\text{fluid}}\| < 0.02$的位置释放智能体；以及（iii）在载流速度为 $0.02 < \|\mathbf{u}_{\text{fluid}}\| < 0.03$的位置释放智能体。从图12可以看出，在 Γ相同的对流胞中，原点处载流速度越大，平均成功率越高。由于较大的 Γ 胞元具有更强的全局流动强度（见第IV.1节的讨论），智能体更有可能在载流较强的位置被释放。借助更强的载流流，智能体能够在更短时间内完成迁移任务并获得更高奖励，因此智能体未来更有可能重复该动作。在流动强度较强的区域，智能体的采样频率更高，这使得其在较大的 Γ 胞元中成功率更高。需要指出的是，只有当智能体原点处的载流速度较弱时，这一趋势才明显；载流速度较强时，成功率始终接近100%。

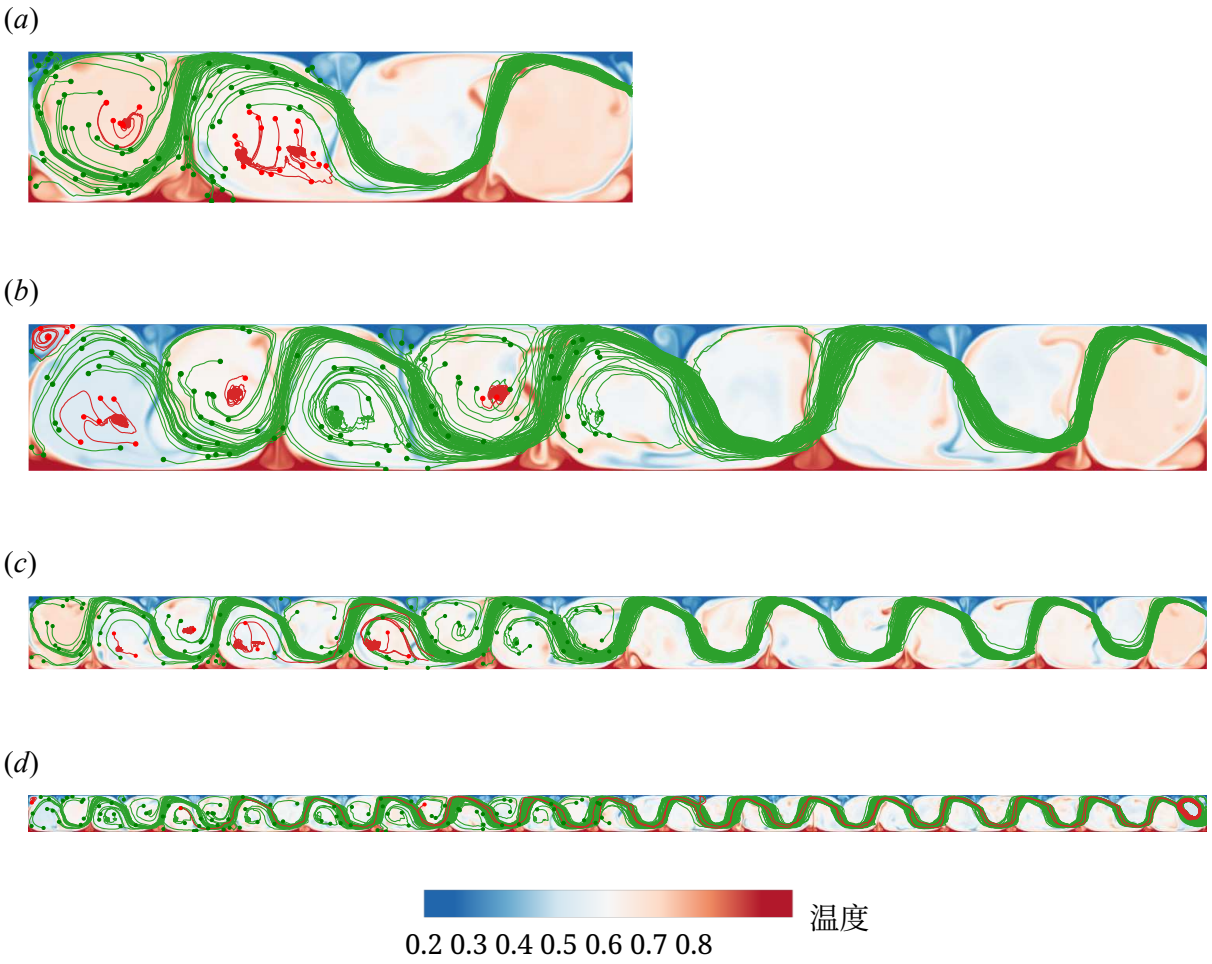

图 11. (a) $\Gamma = 4$、(b) $\Gamma = 8$、(c) $\Gamma = 16$ 和 (d) $\Gamma = 32$ 胞元中的示例轨迹（智能体在 $\|\mathbf{u}_{\text{fluid}}\| < 0.03$ 处释放）。绿色线条表示成功完成迁移的尝试，红色线条表示未成功的尝试。等值线显示了典型的瞬时温度场。

V. 结论

在本研究中，我们采用强化学习算法，对自推进智能体在热湍流环境中的长距离迁移进行了数值训练。我们选择典型的湍流瑞利–贝纳德对流胞作为流动环境，该环境可包含强烈的速度和温度波动。为构建强化学习框架，我们设计了一种奖励函数，同时考虑智能体的当前状态、能量消耗和时间消耗。我们还比较了不同观测变量组合下累积奖励的演变。我们选取智能体的位置，以及载流的速度和温度

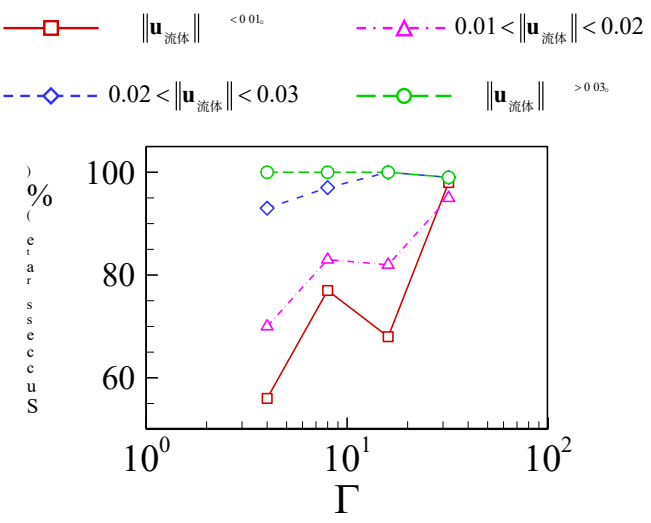

图 12. 当智能体随机释放在局部载流速度不同的位置时，平均成功率随 Γ 的变化。

作为适当的环境线索。在 $\Gamma = 2$ 瑞利–贝纳德单元中的模拟结果表明，与从原点直线移动到目的地的天真智能体相比，聪明的智能体能够学会利用载流来节省推进能量。

随后，我们将在 $\Gamma = 2$ 胞元中获得的最优策略应用于在 Γ更大的对流胞中迁移的聪明智能体，并对其进行测试。通过流动模态分析，我们发现，较大的 Γ瑞利–贝纳德单元中的主导流动模态由稳定性较低的水平叠置涡卷构成，且高阶流动模态所含能量随 Γ增大而增加。尽管这些复杂的流动特征给聪明智能体的轨迹优化带来了挑战，我们仍能利用在 $\Gamma = 2$瑞利–贝纳德单元中训练的策略，获得节能的迁移轨迹。此外，我们发现，聪明智能体消耗的推进能量与天真智能体消耗的推进能量之比随 Γ增大而减小，这意味着聪明的智能体在更大的 Γ 胞元中能够节省更多推进能量。原因在于，在更大的 Γ 胞元中，天真智能体必须穿越更多环流涡卷的边缘并克服更高的能量势垒，而聪明的智能体则始终尽可能沿着载流移动。

我们还评估了智能体从随机选择的原点释放时的优化策略，以测试学习框架的鲁棒性。我们发现，尽管较大的 Γ 胞元中流动结构更加复杂，成功率仍会随着 Γ 的增加而提高。主要原因是，较大的 Γ 胞元中全局流动强度更强（从 Re 与 Γ 之间的关系可以明显看出），而智能体更有可能

在载流速度更强的位置被释放。利用更强的载流速度，智能体可以在更短的时间内完成迁移任务并获得更高的奖励，从而提高成功率。我们的研究对长距离迁移问题具有启示意义，例如在大气对流层中巡逻的无人机。无人机沿着节能轨迹迁移，可以提升续航能力并覆盖更广的范围。

致谢

本研究得到国家自然科学基金委员会（NSFC）项目（编号：12272311、12125204）、国家重点项目（编号：GJXM92579）以及中国111计划（项目编号：B17037）的资助。作者感谢北京北龙超级云计算有限责任公司提供高性能计算资源，为本文报告的研究成果提供了支持。

附录A：奖励函数中超参数的敏感性

在我们设计的奖励函数 [（见方程 (14)-(17)）], 中，有两个超参数：一个是 ϕ，表示智能体位于流动区域之外时受到的惩罚；另一个是 ε，即奖励缩放系数。我们分别调整这两个参数，以确定最优超参数。需要指出的是，我们比较的是归一化奖励值（即介于 0 和 1 之间），而非奖励的绝对值。从图 13(a) 可以看出，较小的惩罚值 ϕ （例如 $\phi = 1$和 5）会导致性能大幅下降；较大的惩罚值（例如 $\phi \geq 10$）则几乎带来相同的性能。至于奖励缩放系数 ε，在所考察的取值范围内，它几乎不影响结果，所有取值都能得到最优策略，如图 13(b) 所示。不过，较大的取值（例如 $\varepsilon = 100$，为清晰起见未在此处展示）会导致训练期间出现 $\sum (r_s + r_e) > \sum r_h$，并使智能体无法在给定的 $t_{\max}$ 时间内探索到成功轨迹。

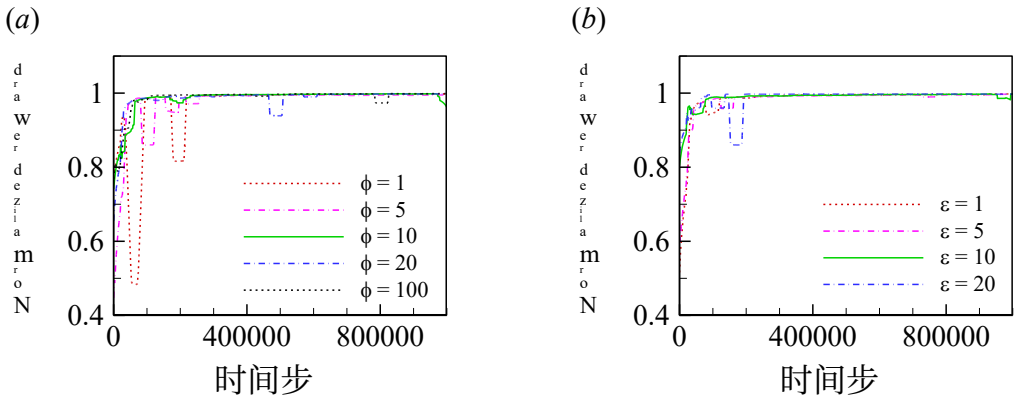

图 13. 奖励函数中超参数的敏感性：(a) 智能体位于流动区域之外时的惩罚 ϕ，以及 (b) 奖励缩放系数 ε。

附录 B：不同观测变量组合的评估

除第 IV.2 节所述的观测变量外，我们还考虑以下不同组合：（i）智能体能够获取位置和速度信息，并能感知应变率 [59, 60],，即 $s = \{y, u, v, s_{xx}\}$ 和 $s = \{y, u, v, s_{xy}\}$。此处 $s_{xx} = \partial_x u$ 和 $s_{xy} = (\partial_y u + \partial_x v)/2$。我们没有考虑应变率张量的 $s_{yy} = \partial_y v$ 分量，因为流动连续性方程在二维流动中给出 $s_{xx} + s_{yy} = 0$，这意味着 s_{xx} 和 s_{yy} 呈负相关。（ii）智能体能够获取位置和速度信息，并能感知温度梯度，即 $s = \{y, u, v, (\nabla T)_x\}$ 和$s = \{y, u, v, (\nabla T)_y\}$。此处 $(\nabla T)_x$ 本质上表示二维对流流动 [61]中由浮力产生的涡量。（iii）智能体能够获取位置、速度和温度信息，并能额外感知涡量、应变率或温度梯度，即 $s = \{y, u, v, T, s_{xx}\}$、$s = \{y, u, v, T, s_{xy}\}$、$s = \{y, u, v, T, \omega\}$、$s = \{y, u, v, T, (\nabla T)_x\}$以及 $s = \{y, u, v, T, (\nabla T)_y\}$。图 14 展示了上述九种观测变量组合在训练期间累积奖励的变化情况。可以看出，这些组合只会略微改变训练的收敛速度，而不会影响渐近累积奖励值。其中，图 14(h) 所示的 $s = \{y, u, v, T, (\nabla T)_x\}$ 优于其他组合。在实际应用中，可通过多种方式测量速度或温度，例如使用皮托管、热线等；而涡量、剪切应变分量和温度梯度则需要根据多个速度或温度传感器的测量值计算，这会增加智能体需要感知的信息的复杂性。因此，如第 IV.2 节所述，我们有意简化环境

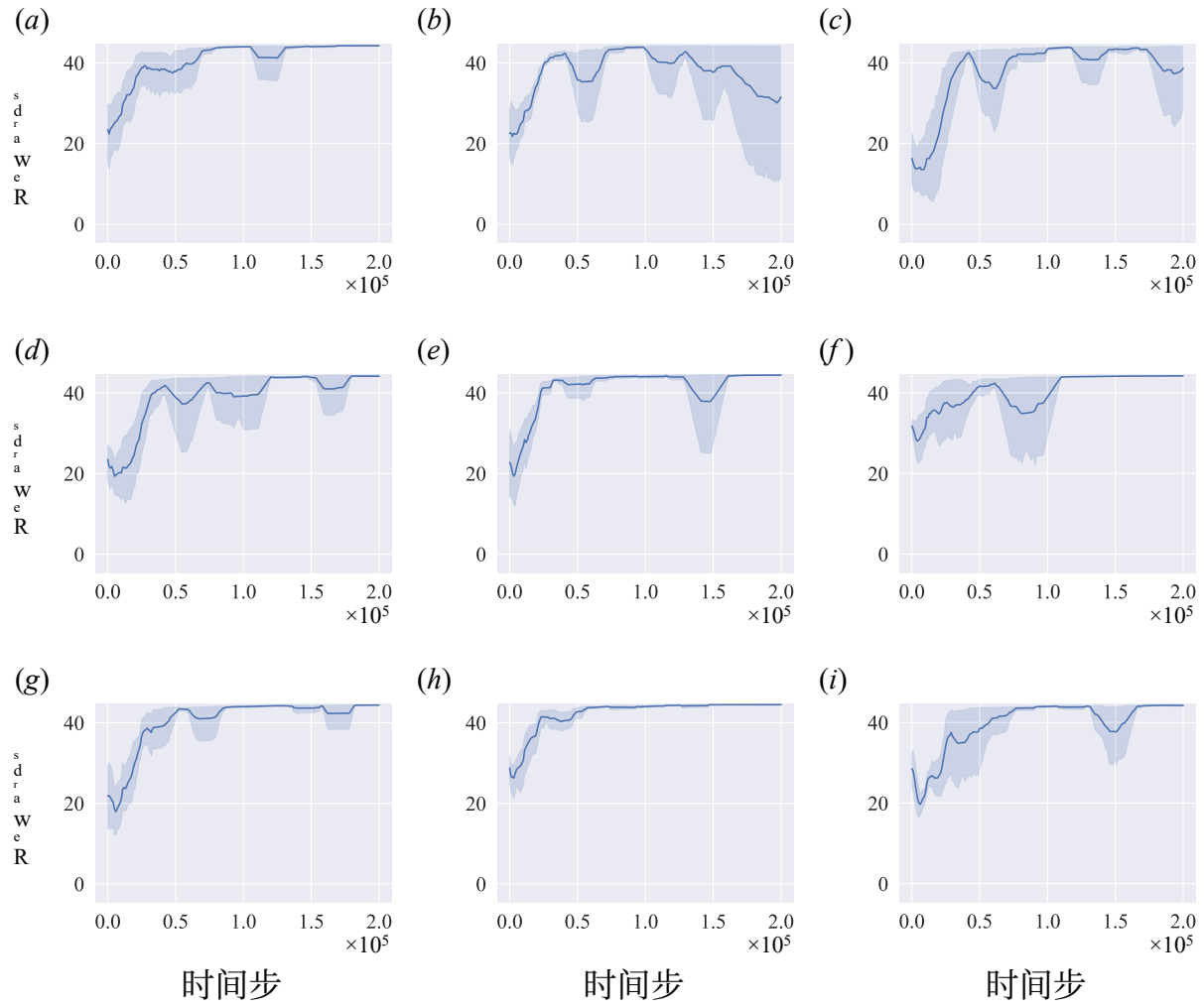

图14. 不同观测变量组合下训练期间累积奖励的变化：

(a) $s = \{y, u, v, s_{xx}\}$、(b) $s = \{y, u, v, s_{xy}\}$、(c) $s = \{y, u, v, (\nabla T)_x\}$、(d)$s = \{y, u, v, (\nabla T)_y\}$、(e) $s = \{y, u, v, T, s_{xx}\}$、(f) $s = \{y, u, v, T, s_{xy}\}$、(g) $s = \{y, u, v, T, \omega\}$、(h) $s = \{y, u, v, T, (\nabla T)_x\}$和(i) $s = \{y, u, v, T, (\nabla T)_y\}$。

智能体可见的局部信息线索 $s = \{y, u, v, T\}$，用于引导其迁移，从而减少实际应用中智能体的数据存储量。

[1] H. Weimerskirch、C. Bishop、T. Jeanniard-du Dot、A. Prudor 和 G. Sachs，《军舰鸟在为期数月的跨洋飞行中追踪大气条件》，《科学》353，74（2016）。

[2] H. J.Williams、E. Shepard、M. D. Holton、P. Alarc´on、R. Wilson 和 S. Lambertucci，《最重滑翔鸟类飞行性能的物理极限》，《美国国家科学院院刊》117，

17884（2020）。

[3] J. P. Croxall, J. R. Silk, R. A. Phillips, V. Afanasyev, 和 D. R. Briggs, 《全球环航：追踪非繁殖期信天翁的全年活动范围》, 《科学》307, 249（2005）。[4] I. Lancaster, 《滑翔鸟类的问题》, 《美国博物学家》19, 1055（1885）。[5] P. B. MacCready, 《最佳空速选择器》, 《翱翔》10, 10（1958）。[6] M. Allen, 《通过自主翱翔提升小型无人飞行器的续航能力》, 载于《第43届AIAA航空航天科学会议暨展览会论文集》（美国航空航天学会, 弗吉尼亚州雷斯顿, 2005年）, 第1025页。[7] R. Bencatel, J. T. de Sousa, 和 A. Girard, 《适用于延长飞机续航时间的大气流场模型》, 《航空航天科学进展》61, 1（2013）。[8] M. Allen, 《用于开发自主翱翔无人飞行器的上升气流模型》, 载于《第44届AIAA航空航天科学会议暨展览会论文集》（美国航空航天学会, 弗吉尼亚州雷斯顿, 2006年）, 第1510页。[9] N. Lawrance 和 S. Sukkarieh, 《面向小型滑翔无人飞行器的基于风能的路径规划》, 载于《AIAA制导、导航与控制会议论文集》（美国航空航天学会, 弗吉尼亚州雷斯顿, 2009年）, 第6112页。[10] Z. ´Akos, M. Nagy, S. Leven, 和 T. Vicsek, 《鸟类和无人飞行器的热气流翱翔飞行》, 《生物启发与仿生》5, 045003（2010）。[11] K. M. Laurent, B. Fogg, T. Ginsburg, C. Halverson, M. J. Lanzone, T. A. Miller, D. W. Winkler, 和 G. P. Bewley, 《湍流解释了鹰在自然飞行中的加速度》, 《美国国家科学院院刊》118, e2102588118（2021）。[12] G. Ahlers, S. Grossmann, 和 D. Lohse, 《湍流瑞利–贝纳德对流中的热传递和大尺度动力学》, 《现代物理评论》81, 503（2009）。´ F. Chill`a 和 J. Schumacher, 《湍流瑞利–贝纳德对流的新视角》, 《欧洲物理学杂志E辑》35, 1（2012）。[14] K.-Q. Xia, 《湍流热对流的当前趋势与未来方向》, 《理论与应用力学快报》3, 052001（2013）。[15] B. Atkinson 和 J. Wu Zhang, 《大气中的中尺度浅对流》, 《地球物理学评论》34, 403（1996）。[16] R. J. Stevens, A. Blass, X. Zhu, R. Verzicco, 和 D. Lohse, 《瑞利–贝纳德对流中的湍流热超结构》, 《物理评论：流体》3, 041501(R)（2018）。

[17] A. Pandey, J. D. Scheel, 和 J. Schumacher, 《瑞利–贝纳德对流中的湍流超结构》, 自然通讯9, 1（2018）。[18] G. Reddy, A. Celani, T. J. Sejnowski, 和 M. Vergassola, 《学习在湍流环境中翱翔》, 《美国国家科学院院刊》113, 4877（2016）。[19] Z. ´Akos, M. Nagy, 和 T. Vicsek, 《比较鸟类和人类的翱翔策略》, 《美国国家科学院院刊》105, 4139（2008）。[20] G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, 和 M. Vergassola, 《通过强化学习实现滑翔机在野外翱翔》, 自然562, 236（2018）。[21] A. Xu, H.-L. Wu, 和 H.-D. Xi, 《湍流环境中自推进智能体的最低能耗迁移》, 《流体物理学》34, 035117（2022）。[22] A. Xu, W. Shyy, 和 T. Zhao, 《燃料电池和液流电池中输运现象的格子玻尔兹曼建模》, 《力学学报》33, 555（2017）。[23] Z. Guo 和 C. Zheng, 《微尺度气体流动的格子玻尔兹曼方程分析：弛豫时间、边界条件和克努森层》, 《国际计算流体动力学杂志》22, 465（2008）。[24] A. Xu, L. Shi, 和 T. Zhao, 《采用图形处理器和OpenACC并结合数据管理加速格子玻尔兹曼模拟》, 《国际传热传质杂志》109, 577（2017）。[25] A. Xu, L. Shi, 和 H.-D. Xi, 《高瑞利数下三维热对流流动的格子玻尔兹曼模拟》, 《国际传热传质杂志》140, 359（2019）。[26] A. Xu 和 B.-T. Li, 《使用OpenACC和消息传递接口进行多图形处理器热格子玻尔兹曼模拟》, 《国际传热传质杂志》201, 123649（2023）。[27] B. Castaing, G. Gunaratne, F. Heslot, L. Kadanoff, A. Libchaber, S. Thomae, X.-Z. Wu, S. Zaleski, 和 G. Zanetti, 《瑞利–贝纳德对流中强热湍流的标度关系》, J. Fluid Mech. 204, 1（1989）。´K. Krishna, Z. Song, 和 S. L. Brunton, 《非定常流动中的有限时域节能轨迹》, 《英国皇家学会会刊A辑：数学、物理与工程科学》478, 20210255（2022）。[28] S. Colabrese, K. Gustavsson, A. Celani, 和 L. Biferale, 《通过强化学习实现智能微型游泳者的流场导航》, 《物理评论快报》118, 158004（2017）。[29] S. Colabrese, K. Gustavsson, A. Celani, 和 L. Biferale, 《智能惯性粒子》, 《物理评论：流体》3, 084301（2018）。[30] E. Schneider 和 H. Stark, 《智能活性粒子的最优操控》, 《欧洲物理快报》127, 64003（2019）。

[32] J. K. Alageshan, A. K. Verma, J. Bec, 和 R. Pandit, 《用于湍流中路径规划的微型游动体的机器学习策略》, Phys. Rev. E 101, 043110 (2020).[33] F. Borra, L. Biferale, M. Cencini, 和 A. Celani, 《低雷诺数下微型游动体追逐与逃避的强化学习》, 《物理评论：流体》7, 023103 (2022).[34] G. Novati, L. Mahadevan, 和 P. Koumoutsakos, 《通过深度强化学习实现受控滑翔与停栖》, 《物理评论：流体》4, 093902 (2019).[35] Z. Zou, Y. Liu, Y.-N. Young, O. S. Pak, 和 A. C. Tsang, 《通过深度强化学习实现微型游动体的步态切换与目标导航》, Commun. Phys. 5, 158 (2022).[36] L. Biferale, F. Bonaccorso, M. Buzzicotti, P. Clark Di Leoni, 和 K. Gustavsson, 《泽梅洛问题：利用强化学习在二维湍流中实现最优点到点导航》, Chaos 29, 103138 (2019).[37] P. Garnier, J. Viquerat, J. Rabault, A. Larcher, A. Kuhnle, 和 E. Hachem, 《流体力学深度强化学习综述》, Comput. Fluids 225, 104973 (2021).[38] R. S. Sutton 和 A. G. Barto, 《强化学习：导论》（麻省理工学院出版社，马萨诸塞州剑桥，2018）。[39] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, 和 D. J. Schwab, 《面向物理学家的机器学习导论：高偏差、低方差》, Phys. Rep. 810, 1 (2019).[40] S. L. Brunton, B. R. Noack, 和 P. Koumoutsakos, 《流体力学中的机器学习》, Annu. Rev. Fluid Mech. 52, 477 (2020).[41] F. Cichos, K. Gustavsson, B. Mehlig, 和 G. Volpe, 《活性物质中的机器学习》, Nat. Mach. Intell. 2, 94 (2020).[42] A. C. H. Tsang, P. W. Tong, S. Nallan, 和 O. S. Pak, 《自学习如何在低雷诺数下游泳》, 《物理评论：流体》5, 074101 (2020).[43] S. Mui˜ños-Landin, A. Fischer, V. Holubec, 和 F. Cichos, 《人工微型游动体的强化学习》, Sci. Robot. 6, eabd9285 (2021).[44] P. A. Monderkamp, F. J. Schwarzendahl, M. A. Klatt, 和 H. L¨owen, 《利用强化学习在复杂运动能力景观中导航的活性粒子》, Mach. Learn.-Sci. Technol. 3, 045024 (2022).[45] M. Gazzola, A. A. Tchieu, D. Alexeev, A. de Brauer, 和 P. Koumoutsakos, 《在流体动力学相互作用下学习群游》, J. Fluid Mech. 789, 726 (2016).

[46] S. Verma, G. Novati, 和 P. Koumoutsakos, 《通过深度强化学习利用涡旋实现高效集体游动》, 《美国国家科学院院刊》115, 5849 (2018).[47] T. Haarnoja, A. Zhou, P. Abbeel, 和 S. Levine, 《软演员-评论家：采用随机演员的离策略最大熵深度强化学习》, 载于《国际机器学习会议论文集》（PMLR, 2018）, 第1861–1870页。[48] Q. Wang, R. Verzicco, D. Lohse, 和 O. S hishkina, 《湍流大宽高比热对流中的多种状态：对流滚的数量由什么决定？》, 《物理评论快报》125, 074501 (2020).[49] A. Xu, X. Chen, F. Wang, 和 H.-D. Xi, 《湍流瑞利–贝纳德对流中内部流动结构与传热效率的相关性》, 《流体物理学》32, 105112 (2020).[50] N. Halko, P.-G. Martinsson, 和 J. A. Tropp, 《利用随机性发现结构：构造近似矩阵分解的概率算法》, SIAM Rev. 53, 217 (2011).[51] Y. Zhang, Q. Zhou, 和 C. Sun, 《二维湍流瑞利–贝纳德对流中动能与热能耗散率的统计特性》, J. Fluid Mech. 814, 165 (2017).[52] X. Zhu, V. Mathai, R. J. Stevens, R. Verzicco, 和 D. Lohse, 《二维瑞利–贝纳德对流向终极态的转变》, 《物理评论快报》120, 144502 (2018).[53] B.-F. Wang, Q. Zhou, 和 C. Sun, 《振动诱导的边界层失稳实现大幅热输运增强》, Sci. Adv. 6, eaaz8239 (2020).[54] X. Chen, S.-D. Huang, K.-Q. Xia, 和 H.-D. Xi, 《大尺度环流内部子结构的出现诱发湍流热对流中流动反转的转变》, J. Fluid Mech. 877, R1 (2019).[55] 有关智能智能体与朴素智能体在湍流瑞利–贝纳德对流中迁移的比较，请参见补充材料：http://link.aps.org/supplemental/10.1103/physrevfluids.8.023502, .[56] P. Gunnarson, I. Mandralis, G. Novati, P. Koumoutsakos, 和 J. O. Dabiri, 《学习在涡旋流场中高效导航》, 自然通讯 12, 7143 (2021).[57] K. Gustavsson, L. Biferale, A. Celani, 和 S. Colabrese, 《通过强化学习在三维混沌流中寻找高效游动策略》, 《欧洲物理学杂志E辑》40, 110 (2017).[58] A. Kubo 和 M. Shimizu, 《利用部分可观测量进行流体流动控制的高效强化学习》, Phys. Rev. E 105, 065101 (2022).

[59] J. Qiu、N. Mousavi、K. Gustavsson、C. Xu、B. Mehlig 和 L. Zhao，《稳态流动中微型游动体的导航：对称性的重要性》，J. Fluid Mech. 932，A10（2022年）。[60] J. Qiu、N. Mousavi、L. Zhao 和 K. Gustavsson，《利用流体力学信号实现微型游动体的主动趋陀螺稳定性》，《物理评论：流体》7，014311（2022年）。[61] A. Xu、B.-R. Xu、L.-S. Jiang 和 H.-D. Xi，《二维瑞利-贝纳德对流胞中的涡量产生与输运》，《流体物理学》34，013609（2022年）。