



湖南大学  
HUNAN UNIVERSITY

# 基于深度强化学习的多区域办公楼 独立新风系统优化控制方法

唐旭东<sup>1</sup> 闫钰凡<sup>2</sup> 方雷<sup>1</sup> 罗治勇<sup>☆3</sup> 张林<sup>3</sup> 尹应德<sup>☆4</sup> 陈江<sup>1</sup> 张泠<sup>1</sup>

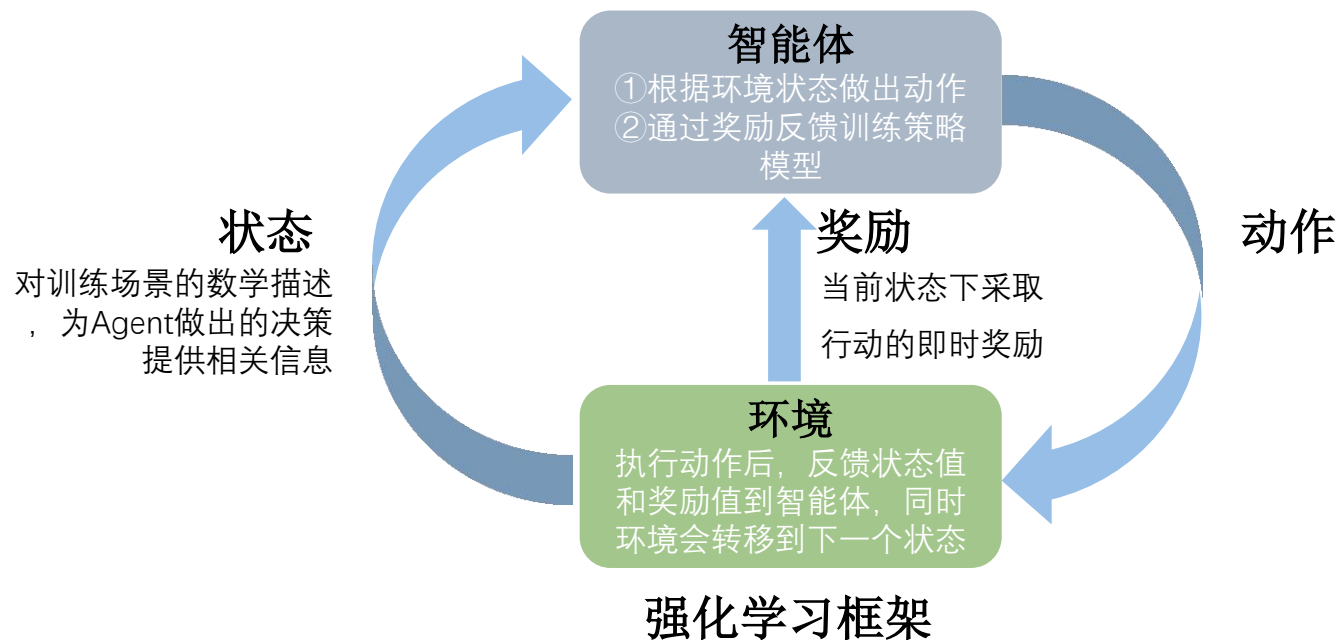
- |                 |       |
|-----------------|-------|
| 1. 湖南大学土木工程学院   | 湖南长沙; |
| 2. 深圳大学南特金融科技学院 | 广东深圳; |
| 3. 湖南省国土资源规划院   | 湖南长沙; |
| 4. 桂林电子科技大学     | 广西桂林。 |

2025.7 湖南 | 长沙

# 强化学习介绍

随着计算能力和传感器技术的提升，**AI技术**逐步渗透到建筑物的全生命周期中，推动了**智慧建筑**的发展，例如利用预测算法降低建筑能耗、优化暖通空调系统的运行策略等。

强化学习是机器学习的一个分支，主要由**智能体**（Agent）、**环境**（Environment）、**状态**（State）、**动作**（Action）、**奖励**（Reward）组成。智能体一般是控制的对象，它在当前状态下依据策略执行一个动作，环境将会根据动作转换到一个新的状态，并给出奖励信号（正奖励或者负奖励）。随后，智能体根据新的状态和环境反馈的奖励，更新控制策略并执行新的动作。



强化学习在于与环境的交互中学习经验，不断试探如何采取最佳的行动，以最大限度地提高累积奖励。这个过程不需要真实数据作为学习素材。

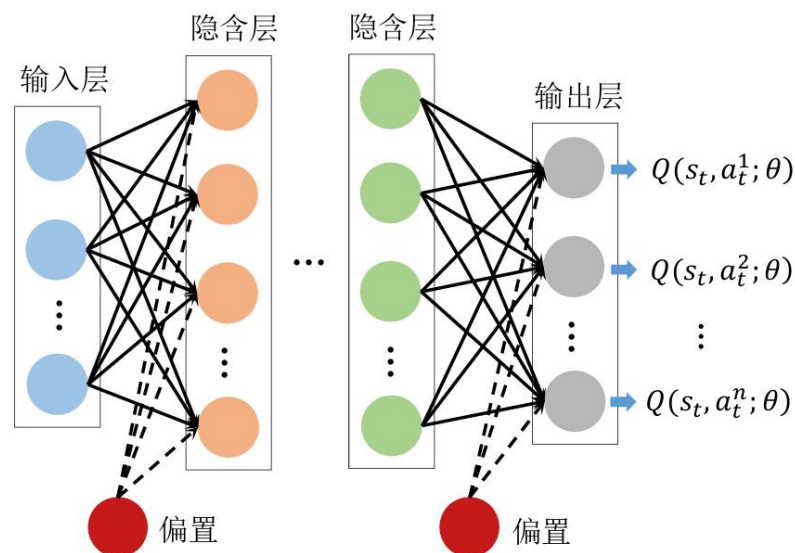
# 强化学习介绍

## 常用的强化学习算法

	a1	a2
s1	0	1
s2	-1	-2
...	...	...
Sn	0.56	0.43

### Q-learning

使用一个 Q 表格记录每个状态-动作对的价值，来衡量状态下采取动作的好坏，并选择最好的动作来执行。不断更新表格从而逼近最优策略



### DQN

使用深度神经网络代替Q表格  
解决状态空间维度高的问题



Actor

Critic

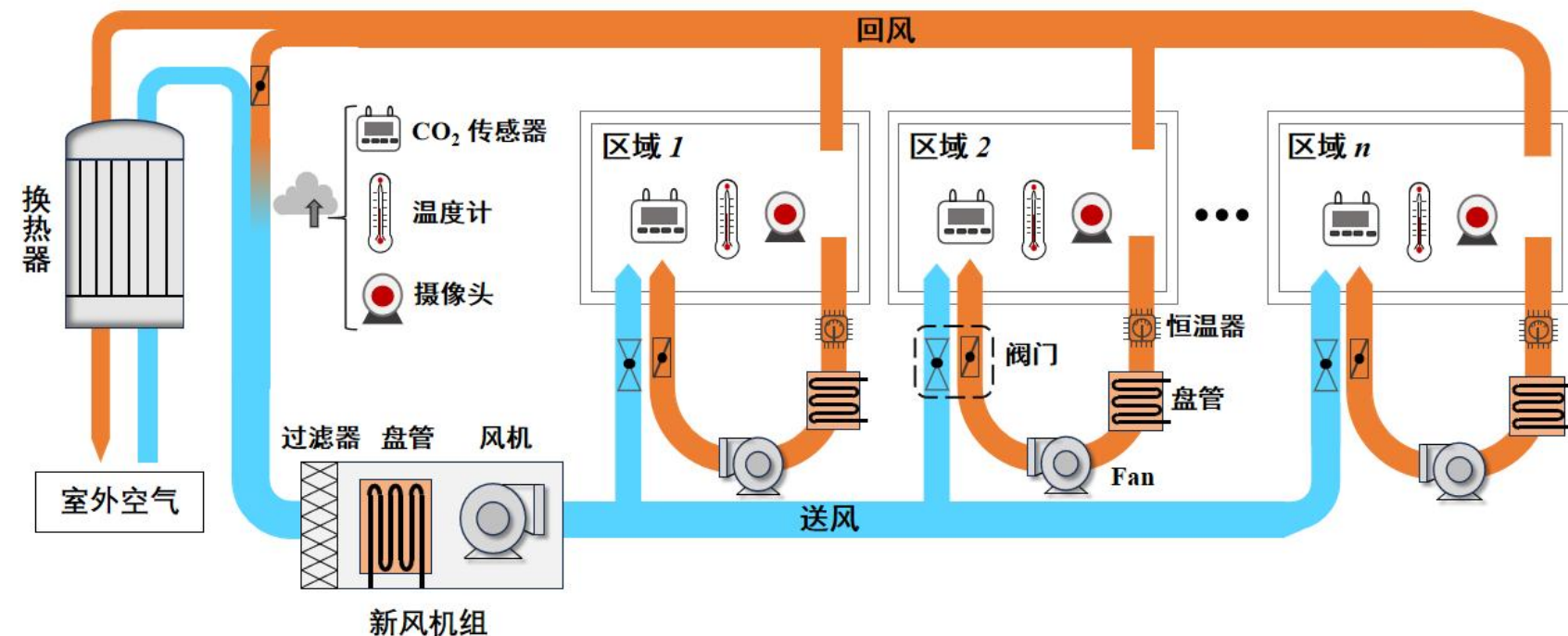
### DDPG

将演员-评论家框架和  
DQN结合，有利于提高  
训练速度和稳定性

# 研究案例

从智慧建筑的发展趋势来说，强化学习实现了从静态控制走向**动态自适应**，以更精细的操作代替了粗狂的管理。然而应用于**大型的HVAC系统**时，面临样本复杂度高的挑战，这可能导致训练时间延长和计算成本增加。

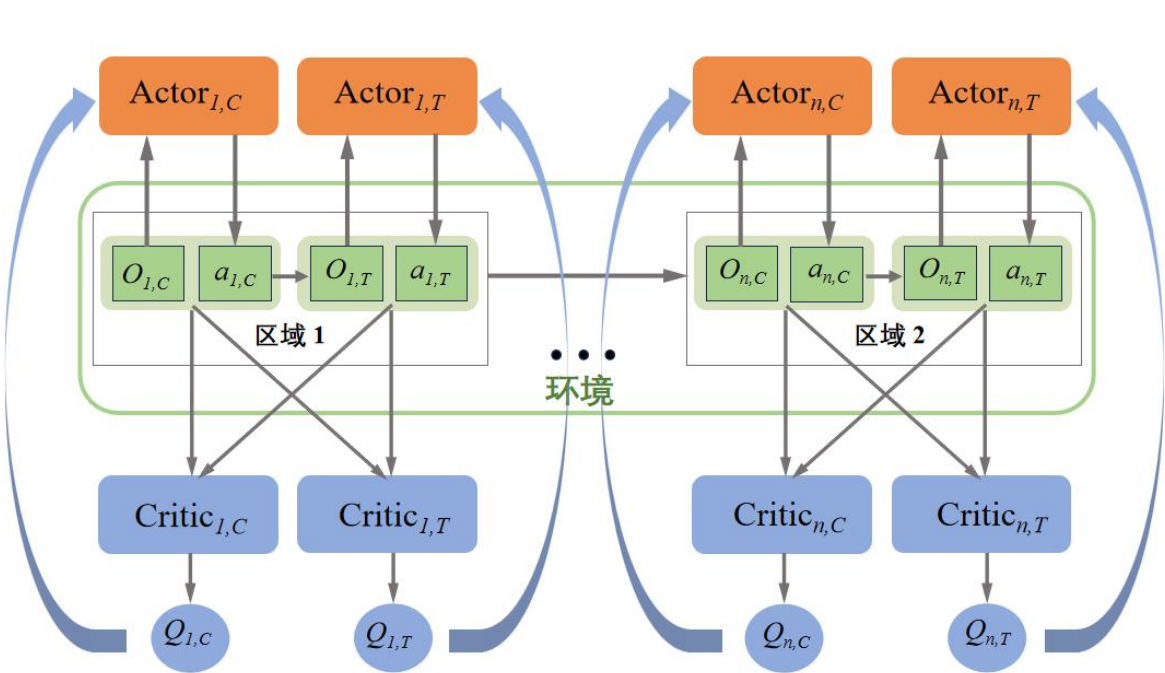
为了解决该问题，提出了一种基于**MADDPG**的最优控制方法，通过控制办公建筑中不同区域的独立新风系统和恒温器，确保每个区域实现**舒适健康的室内环境**，同时降低HVAC系统的**能耗**。



**独立新风系统 (DOAS)** 通过独立的空气处理设备调节进入建筑物的所有外部空气，并输送至每个区域用于通风。由于每个区域都需要对**温度**和**新风量**进行独立调节，在采用强化学习控制时，智能体需要与环境进行大量交互才能学习到有效的策略。

# 研究案例

MADDPG算法是深度确定性策略梯度（DDPG）算法在多智能体系统中的扩展。它除了结合了**演员-评论家**框架，还有**集中训练、分散执行**的理念，使多个智能体能够有效合作解决复杂问题。



应用于独立新风系统的MADDPG框架

**状态:** **恒温器** **通风机**

室内负荷  
太阳辐射  
PMV  
室内温度  
室外温度  
室外渗透  
室内人数  
CO<sub>2</sub>浓度  
新风负荷

**动作:**

**恒温器** { 32, 28, 26, 25, 24 (供冷季)  
13, 18, 21, 22, 23 (采暖季)

**通风机** = 0, 1 (0关1开)

**奖励:**

**温度惩罚:**  $|PMV| > 0.5$  **CO<sub>2</sub>浓度惩罚:**  $CO_2 > 1000ppm$

**能耗惩罚:** - (新风负荷 + 室内负荷)



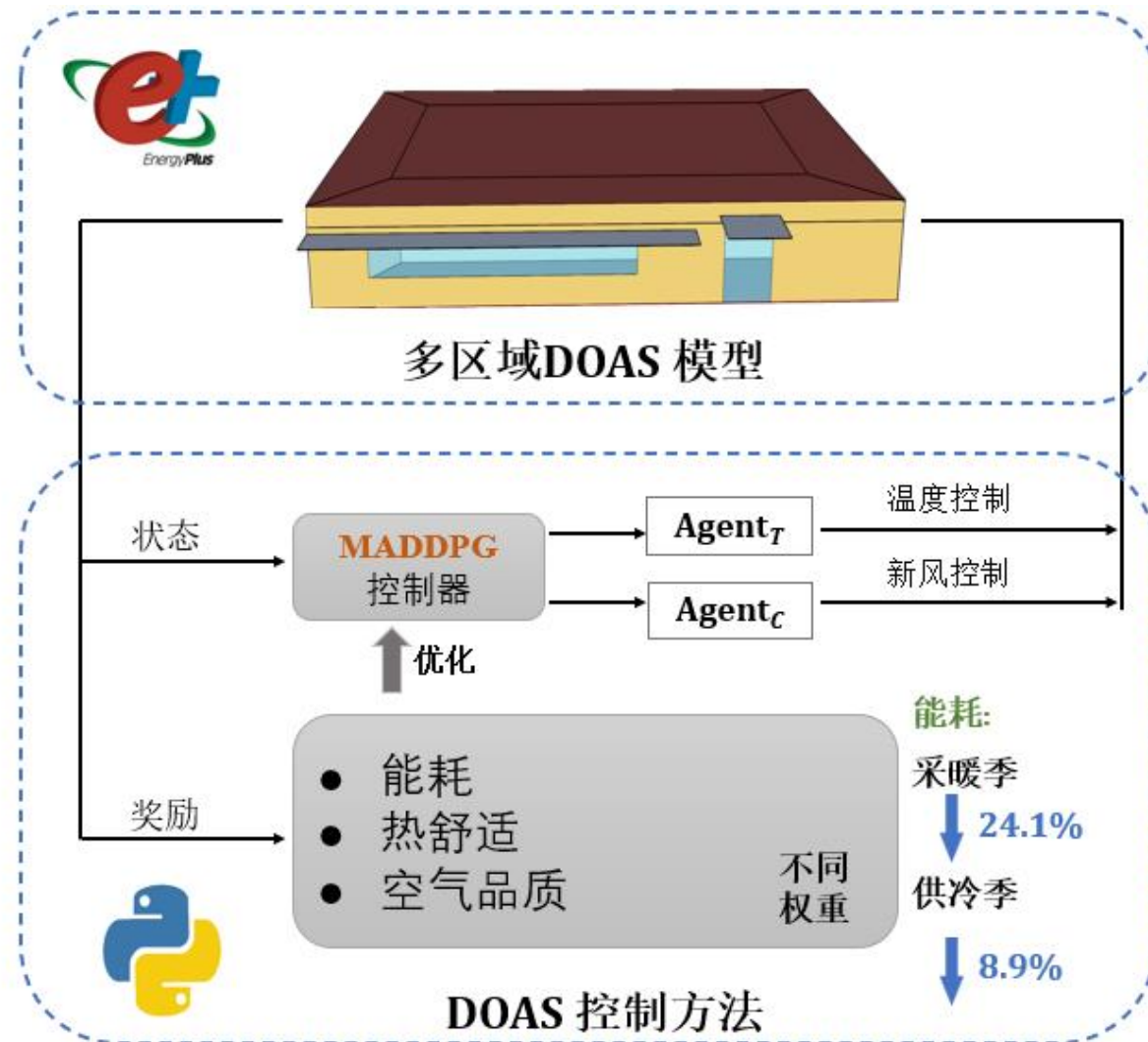
# 研究案例

## 构建联合仿真平台

开发了一个 **EnergyPlus-Python** 联合仿真平台，Energyplus构建了一座五区办公室建筑模型，每个区域均可独立控制温度设定和新风的供应。

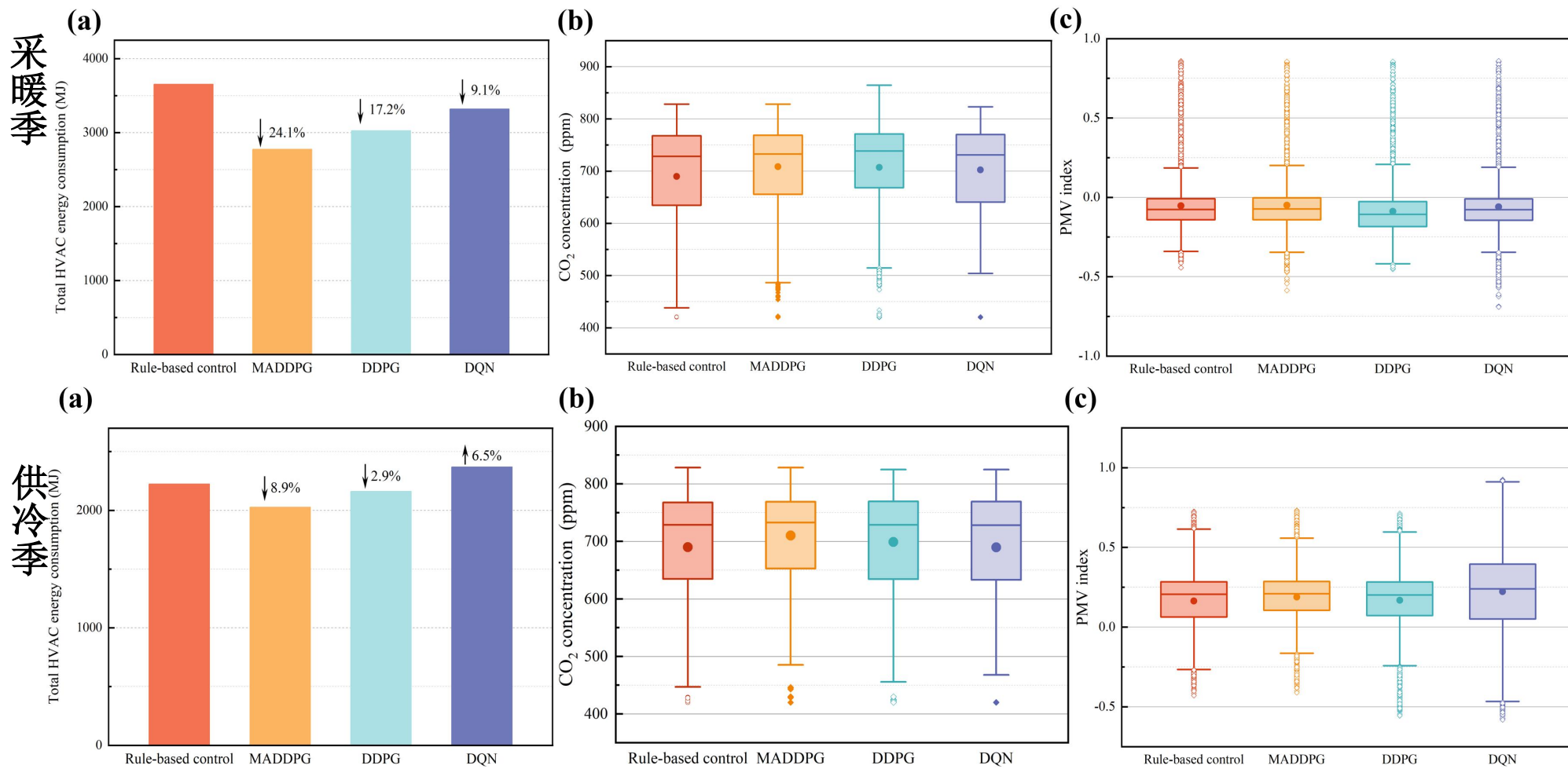
用**基于规则**的控制方法作为对比，该方法在区域被人占用时，持续供应新风，冬季室内温度设定为  $23^{\circ}\text{C}$ ，夏季室内温度设定为  $24^{\circ}\text{C}$ 。未被占用时，不供应新风，冬季工作日提前一小时预热室内空间。

**MADDPG** 相比于基于规则的控制方法，在具有相似的空气质量和热舒适度指数的情况下，采暖季节实现了**24.1%**的节能率，供冷季节实现了**8.9%**的节能率。



# 研究案例

## 仿真测试结果对比

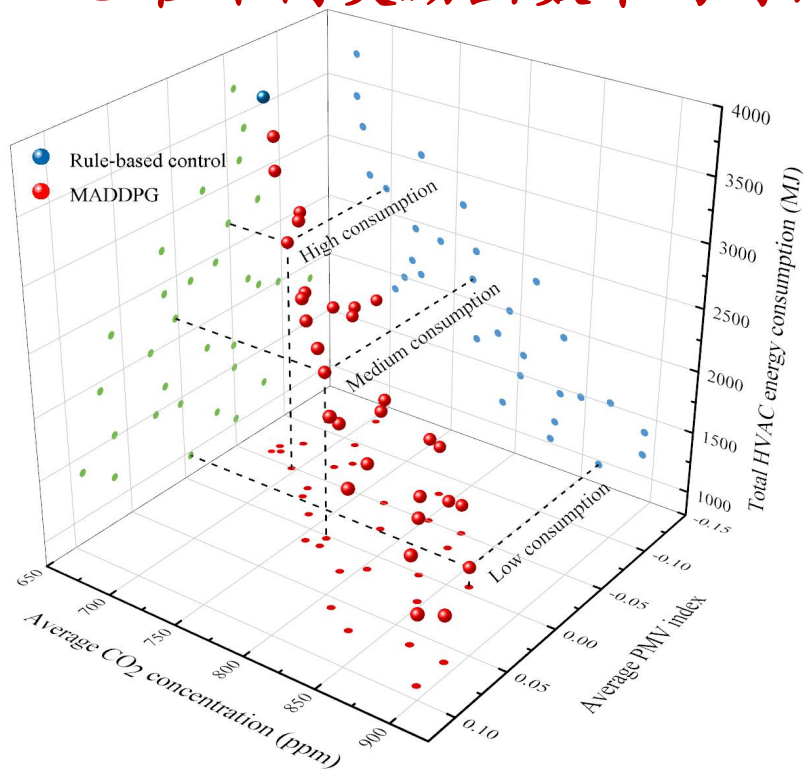


与常用的强化学习方法DQN和DDPG相比，无论是在采暖季还是供冷季，MADDPG都是最节能的控制方法，此外还能够有效维持室内CO<sub>2</sub>浓度和温度，减少过度通风。这表明该方法在热舒适度、空气质量和能耗之间取得了更好的平衡。

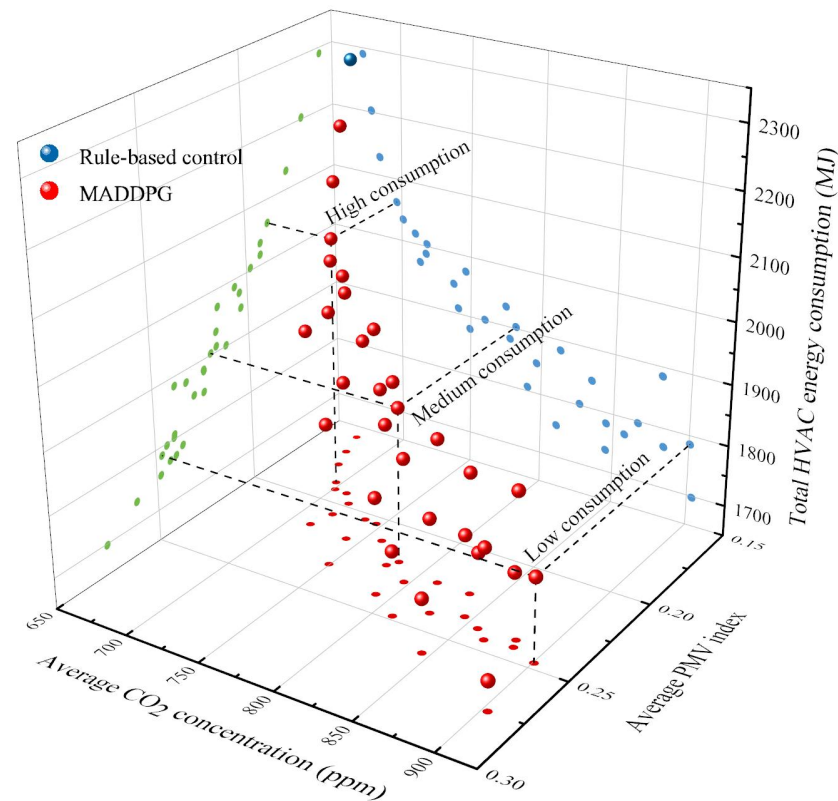
# 研究案例

## MADDPG 在不同奖励函数下的对比

采暖季



供冷季



在采暖季和  
供冷季，处  
理新风都需  
要消耗大量  
能源

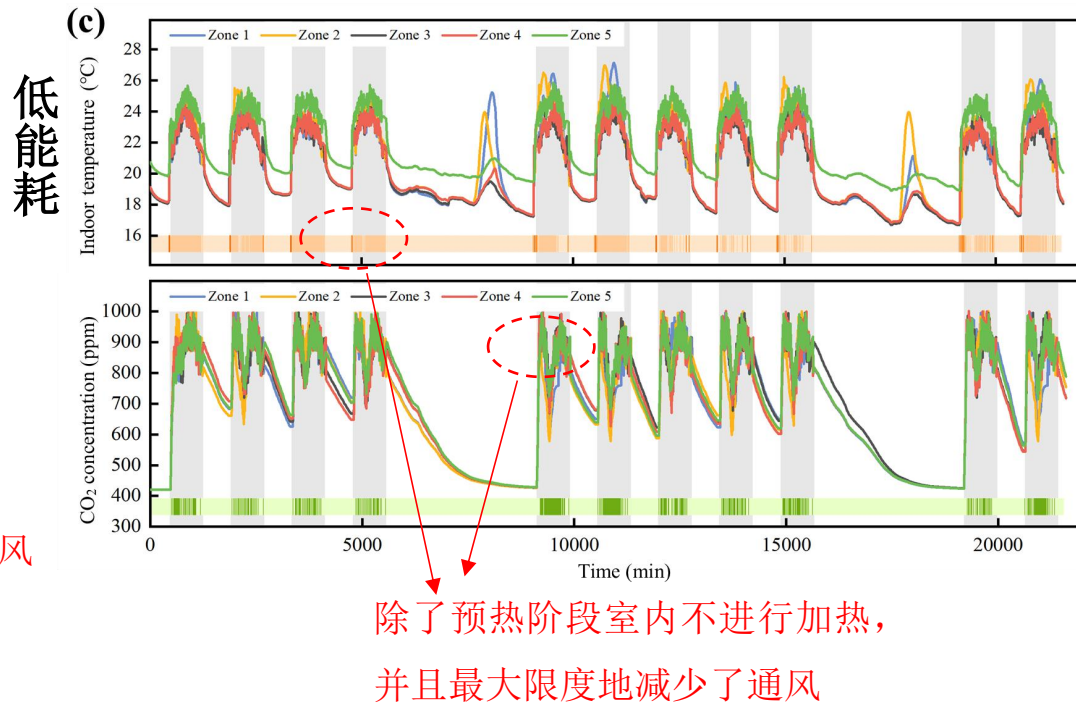
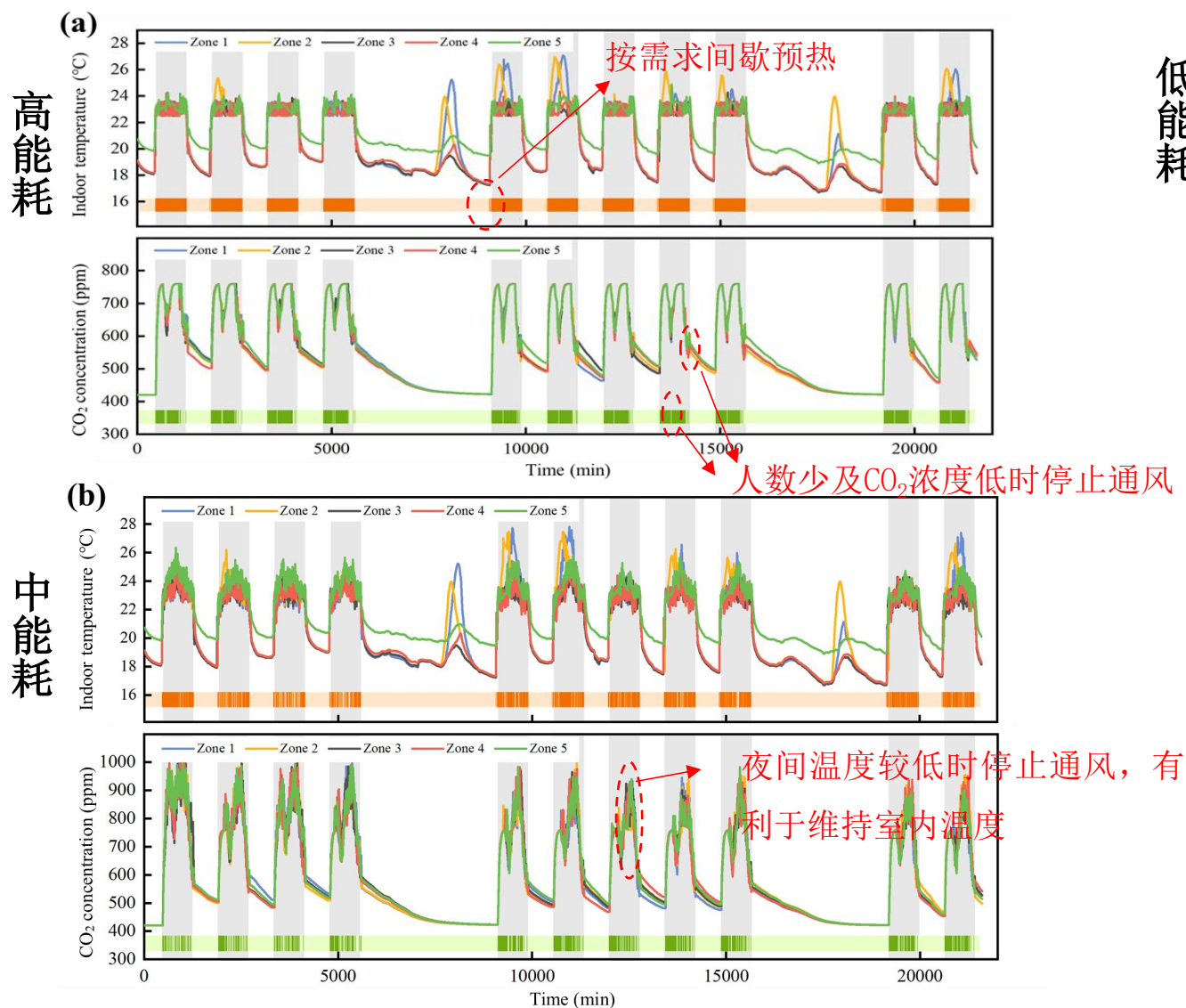
在采暖季，当MADDPG策略偏向于调节更好的室内空气质量时，室内PMV指数也较低。这是因为当新鲜空气被送入室内时，必然会降低室内温度，从而导致PMV指数较低。此冲突表明，控制室内供暖的agent<sub>T</sub>与控制新鲜空气供应的agent<sub>C</sub>存在一定的竞争关系。

供冷季总能耗与室内平均CO2浓度和平均PMV指数的相关性都比较显著。这是由于新风的送风温度低于室温，供应新风有利于维持较低的室内温度，可以认为agent<sub>T</sub>和agent<sub>C</sub>合作将室内温度降低



# 研究案例

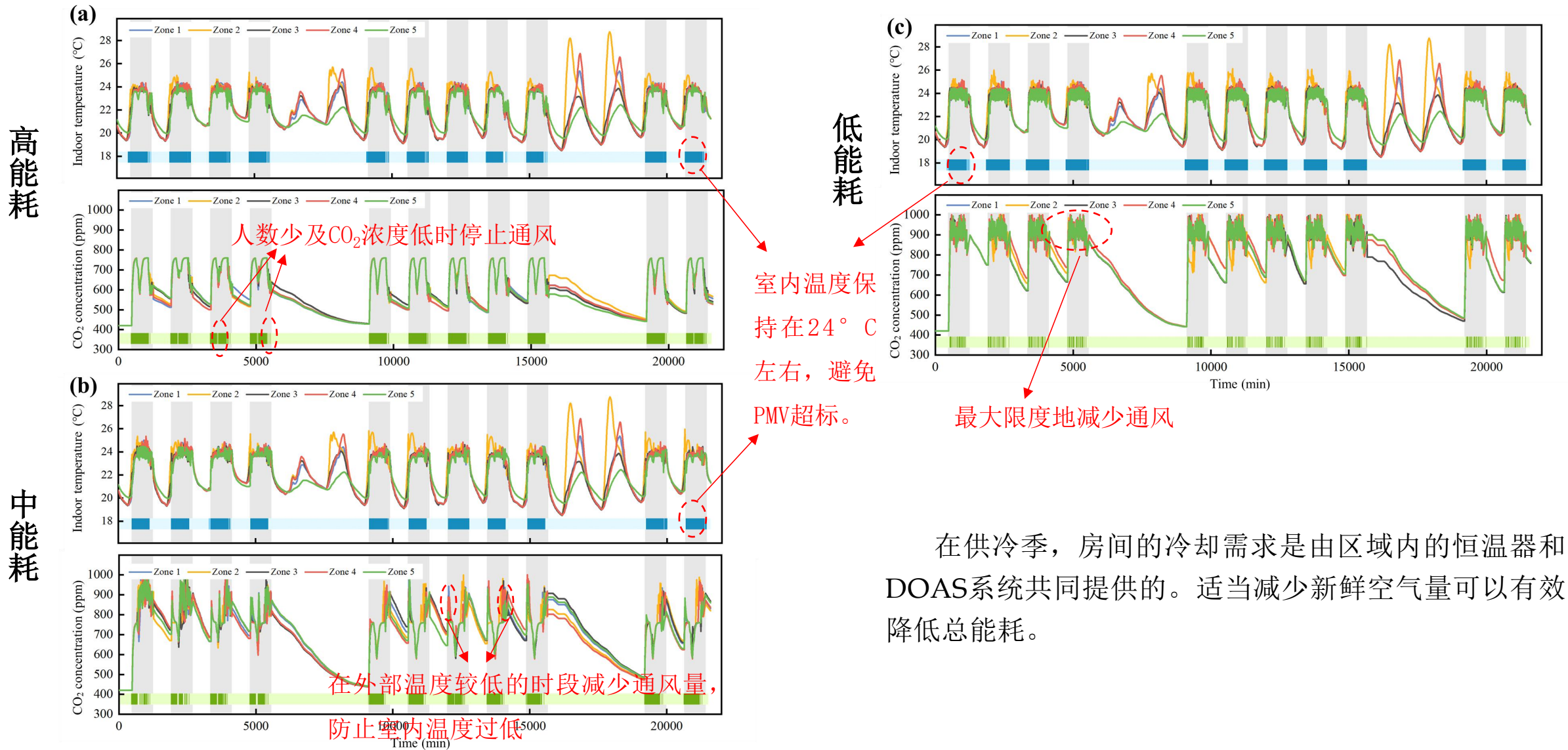
## MADDPG 在不同奖励函数下的对比（采暖季）



在供暖季节，在合理范围内减少新鲜空气供应有助于维持较高的室内温度，并显著减少能耗。然而，这也导致室内空气质量下降，并使室内温度变得不稳定。

# 研究案例

## MADDPG 在不同奖励函数下的对比（供冷季）



# 未来发展

---

强化学习在暖通空调控制中的**节能潜力和自适应能力**日益凸显，但是面临的现实挑战如训练成本、泛化难、可部署性等仍待突破。未来通过仿真训练 + 迁移学习 + 安全保障机制，有望实现强化学习在智能建筑中的可持续落地。

- 构建数字孪生环境，实现低风险的预训练和策略验证；
- 利用跨建筑/跨任务知识迁移，提升泛化能力；
- 使用历史数据训练智能体，降低对真实交互的依赖。

让建筑像人一样“学习”如何节能又舒适！

# THANKS

汇报人

唐旭东  
13893275301