南方科技大学
Southern University of
Science and Technology

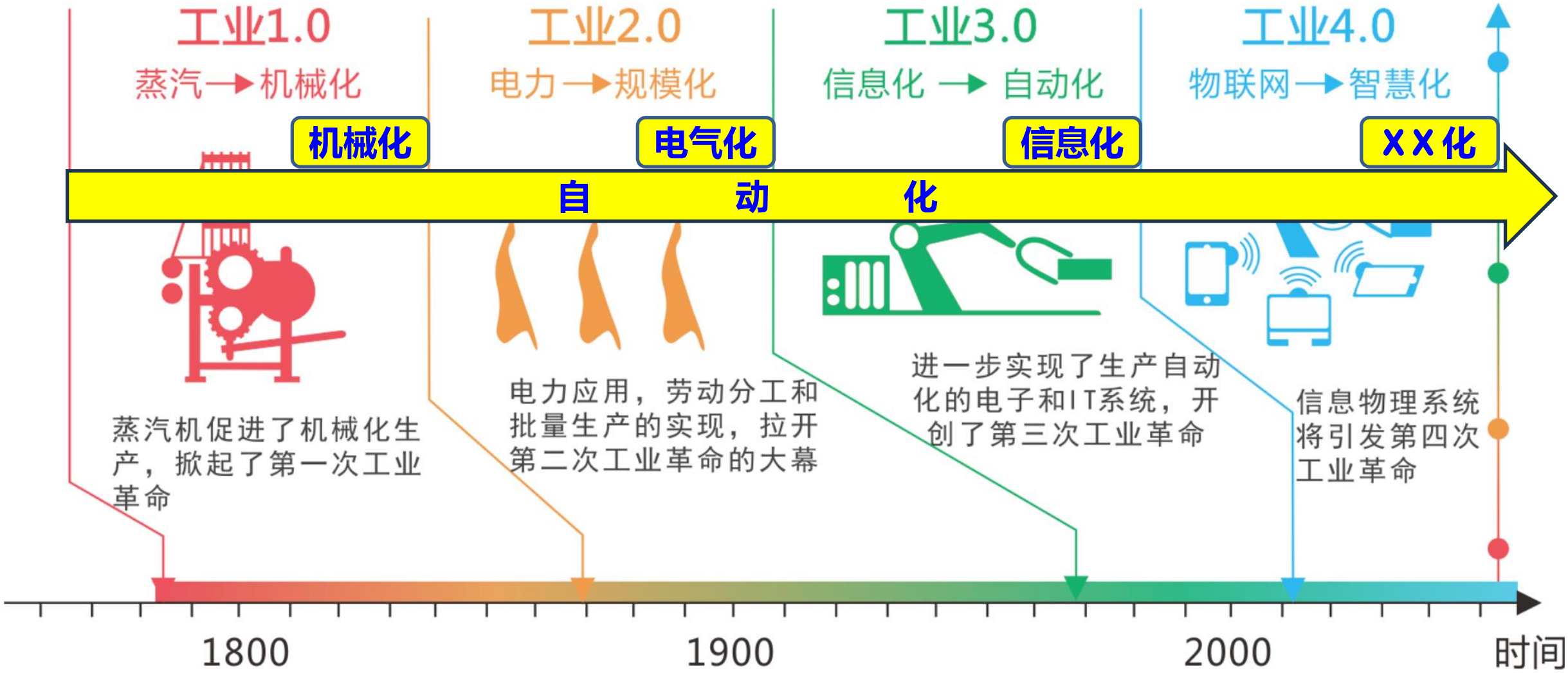
系统设计与智能制造学院
School of System Design and
Intelligent Manufacturing

基于强化学习的控制方法与工业互联网深度融合

刘德荣教授

2023年11月4日

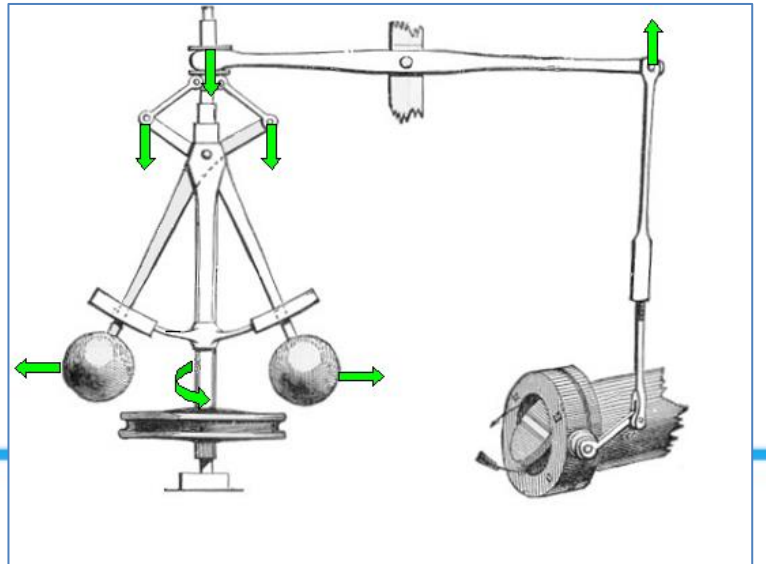
四次工业革命的核心 – 自动化



第一次工业革命 – 机械化（1760–1840）

- 蒸汽机 → 离心自动调速器
- 机械的、模拟的、非数字的
- 机器代替手工劳动

James Watt
1769



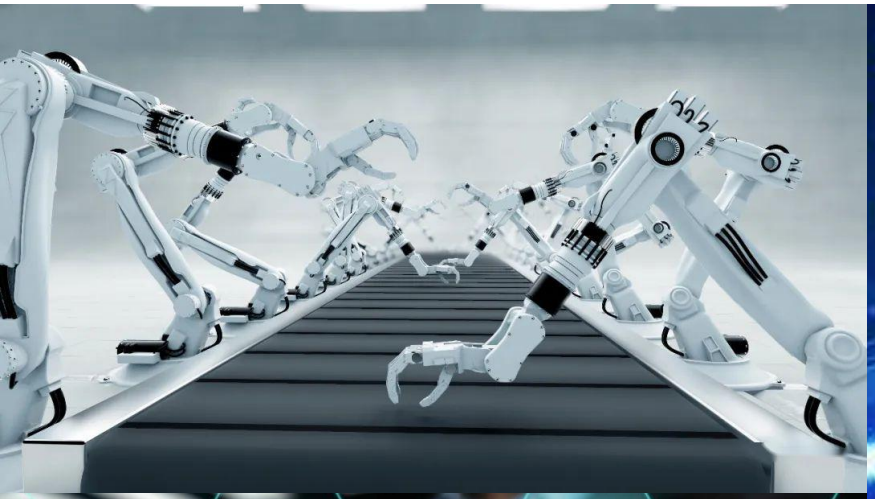
第二次工业革命 – 电气化（1860–1914）

- 电力的大规模应用、非数字的
- 大规模流水线生产 – 自动化或半自动化

Ford Motors
1913

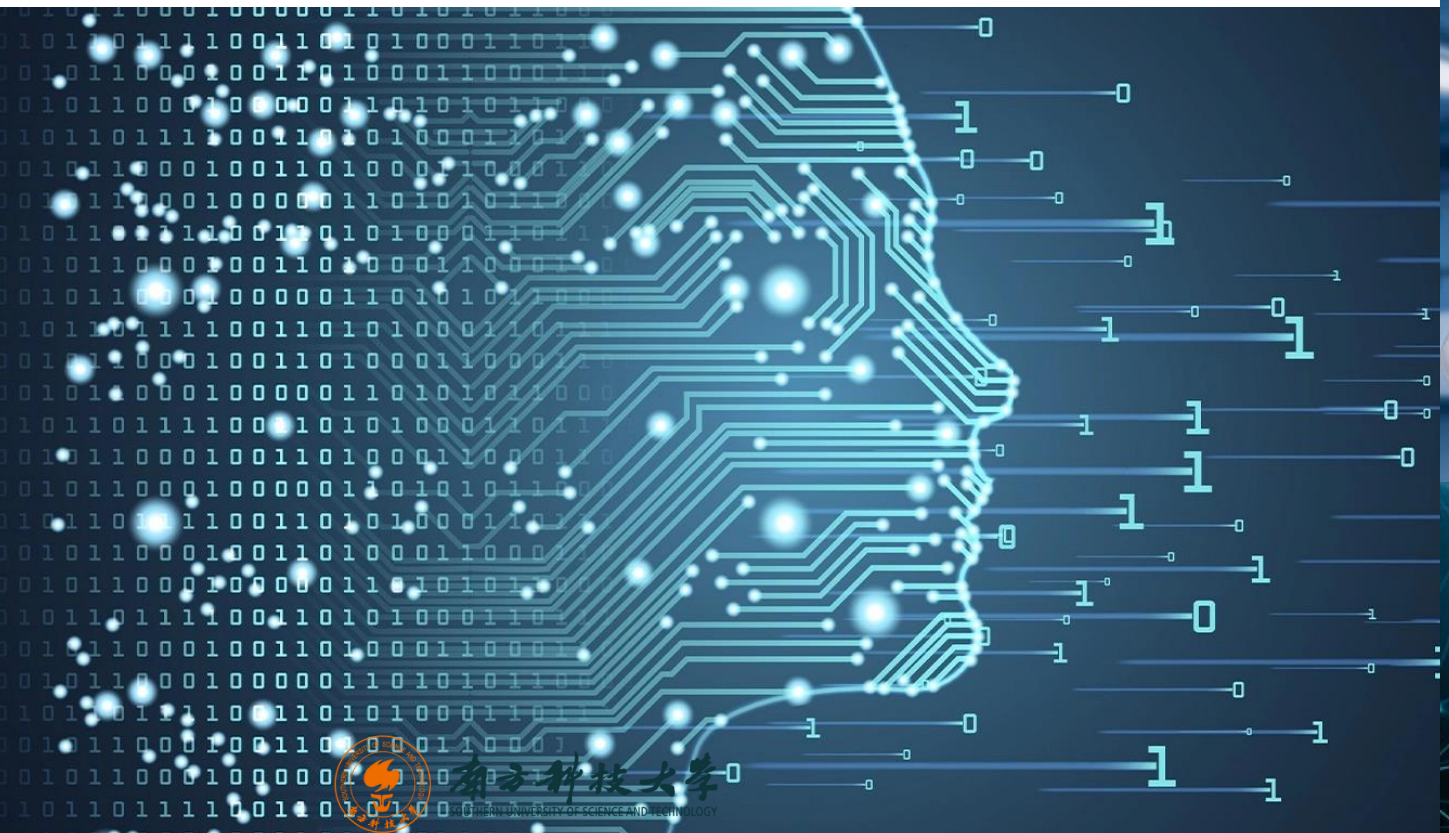
第三次工业革命 – 信息化（20世纪）

- 电子计算机→ 数字化的基石
- 信息化、数字化，自动化得到了空前的发展！
  工业自动化、农业自动化、办公自动化 ……



第四次工业革命 – 智能化（现在）

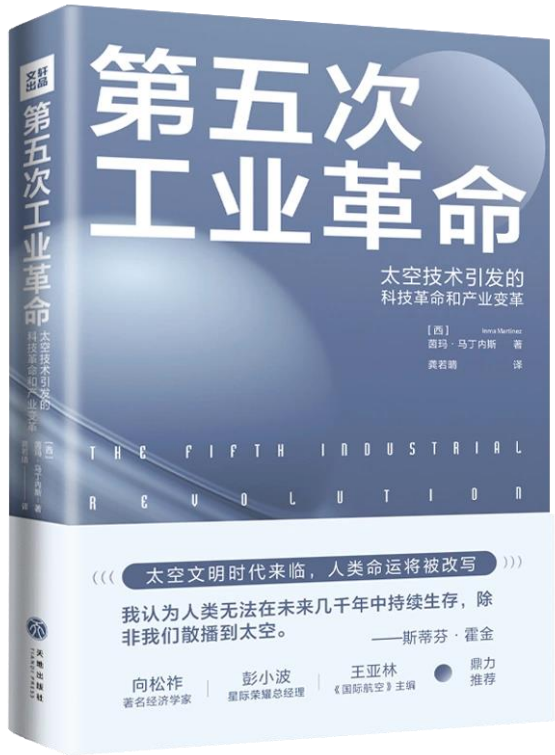
- 互联网时代、人工智能、大数据、机器人等
- 智能化、无人化，离不开自动化!



第五次工业革命？

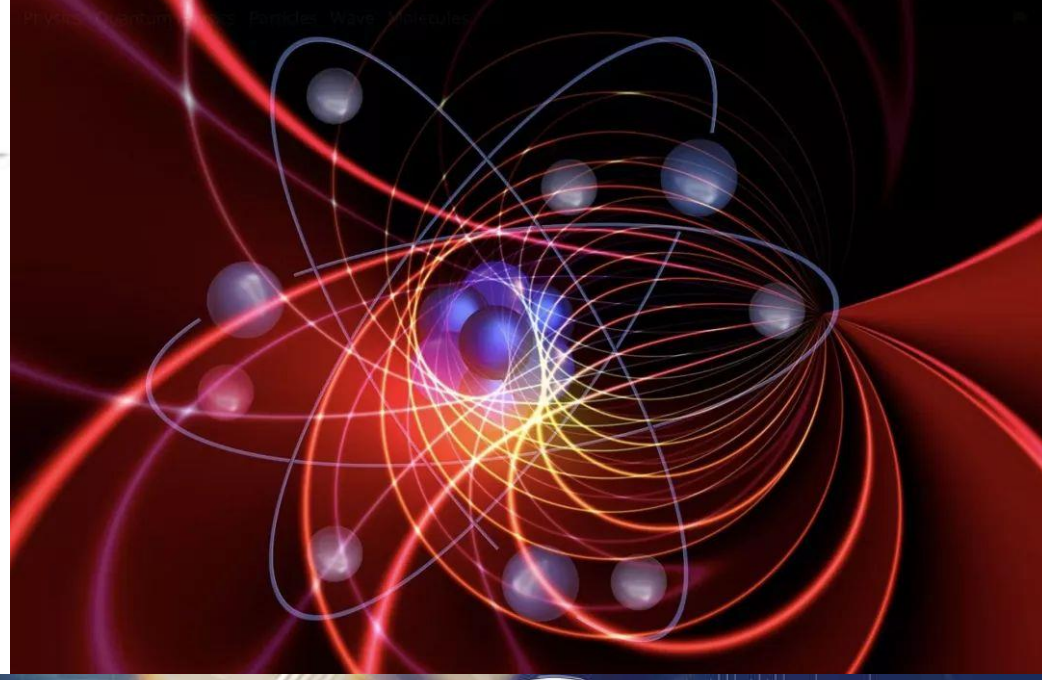
- 碳中和？
- 超级人工智能？
- 机械化 – 电气化 – 信息化 – 智能化 – ？

① ② ③ ④



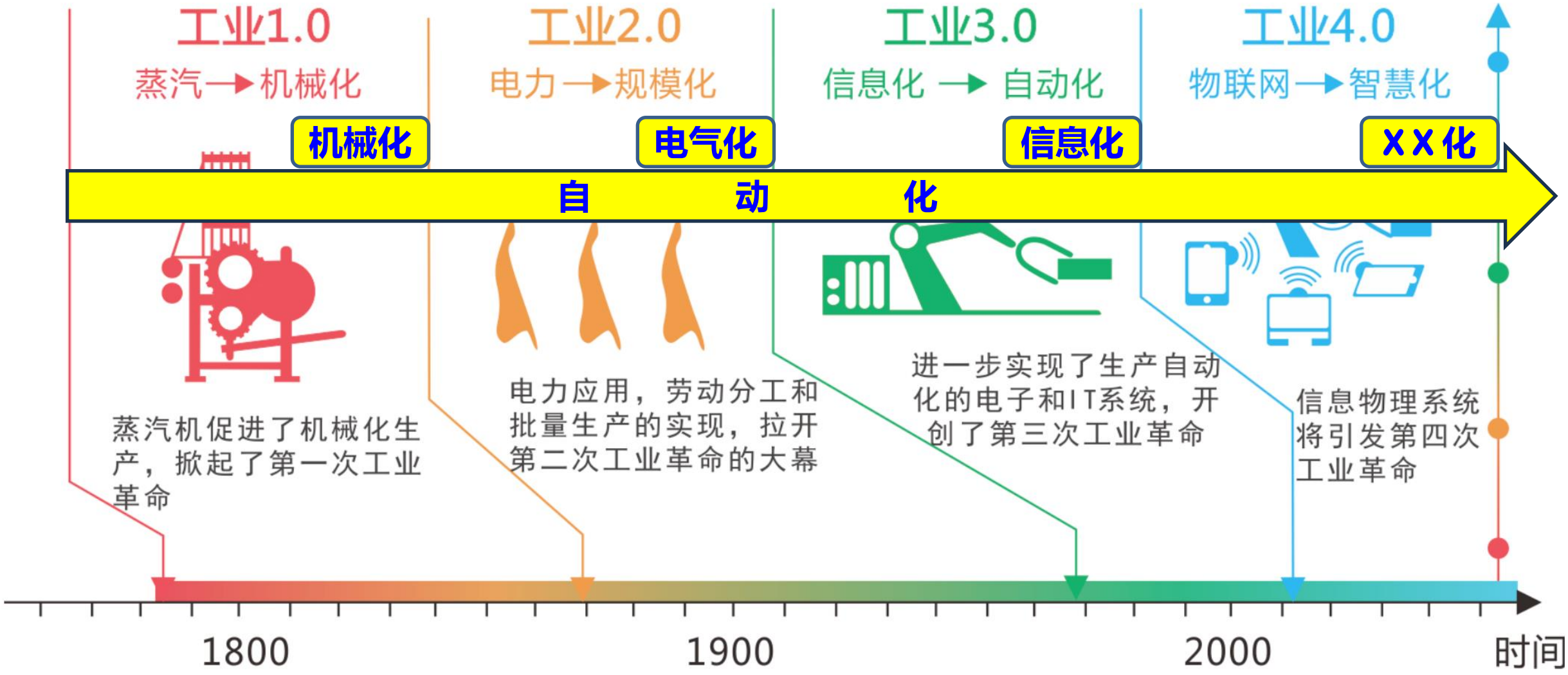
第四次工业革命 – 智能化

- 互联网
- 人工智能
- 大数据
- 机器人
- 虚拟现实
- 量子技术
- 可控核聚变
- 清洁能源



- **智能制造**
  - 自动化
  - 智能化

四次工业革命的核心 – 自动化



智能制造

- 国务院2015年5月8日印发《中国制造2025》：智能制造、绿色制造

- 智能制造：面向产品全生命周期，是传感技术、网络技术、自动化技术、人工智能等的深度融合与集成！

智能制造自动化：几种新的控制方法介绍

- 平行控制方法
- 云控制方法
- 基于代理的控制方法
- 网络化控制方法
- 自适应动态规划与强化学习

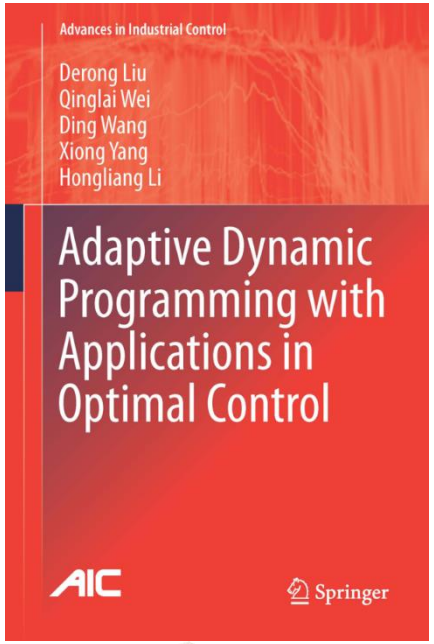
智能制造与自动化技术

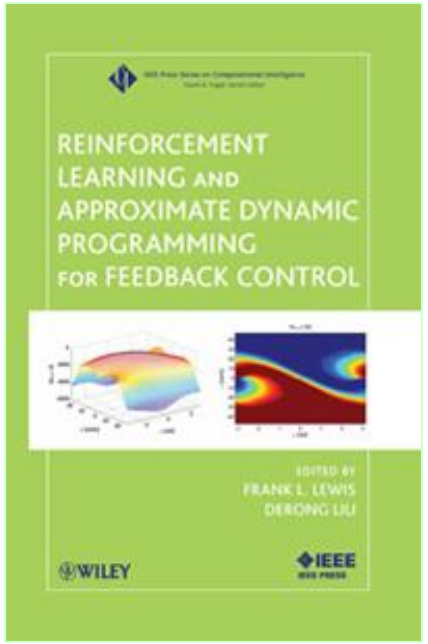
- 平行控制
- 云控制
- 基于代理的控制
- 网络化控制

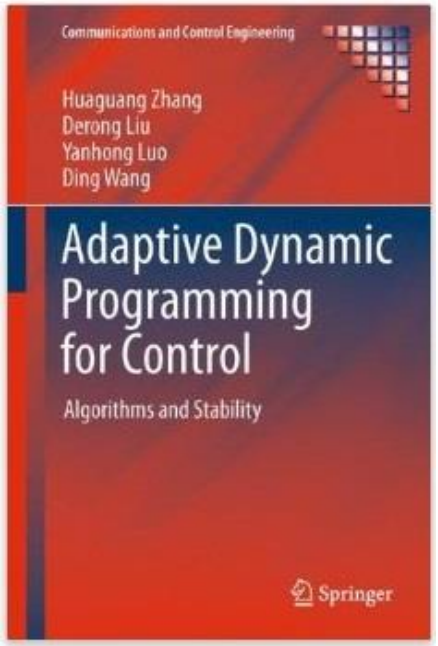
≈≈自适应动态规划与强化学习控制方法

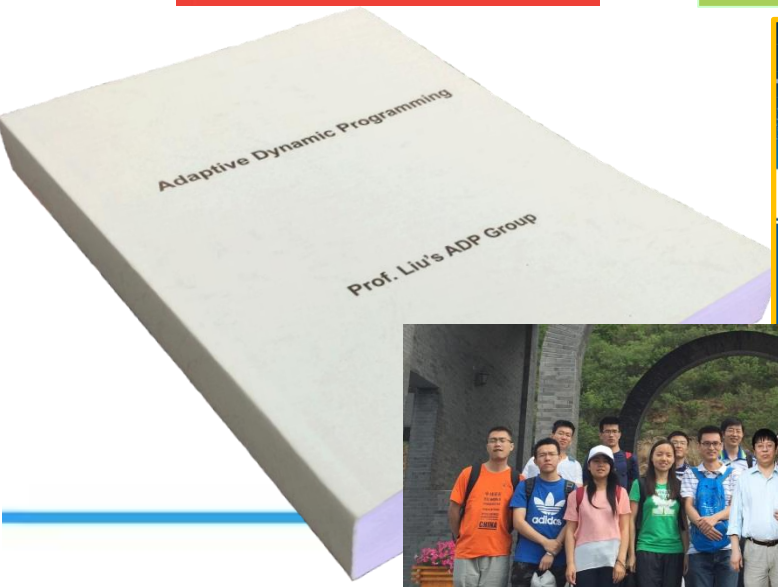
- ④传感技术、③网络技术、①自动化技术、②人工智能等的深度融合与集成！
- 用一个统一的框架来描述！
- 自动化与工业互联网深度融合

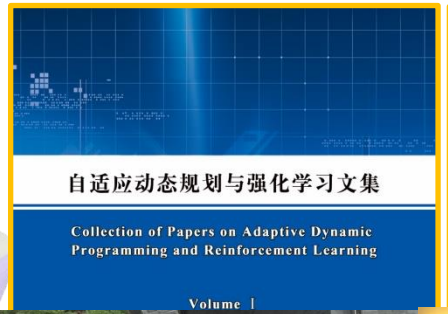
自适应动态规划与强化学习

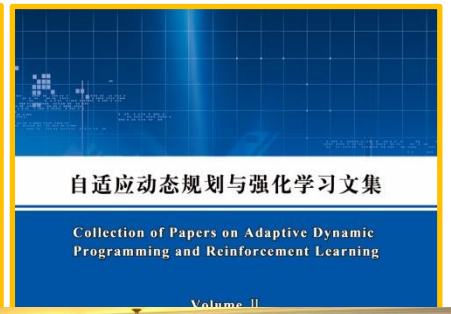






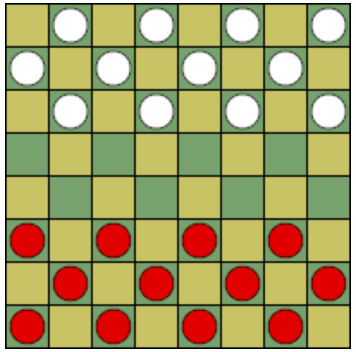
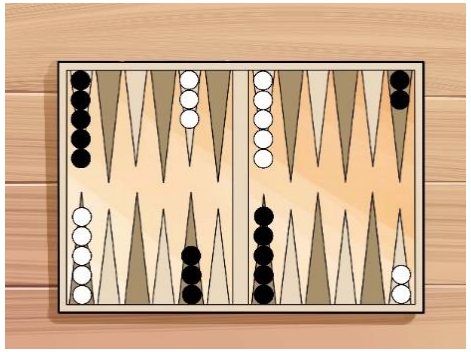


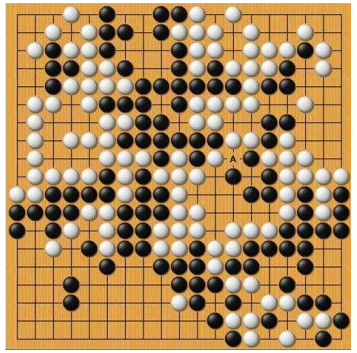


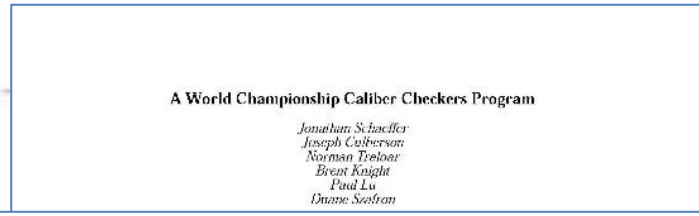


Google's AlphaGO and AlphaZero

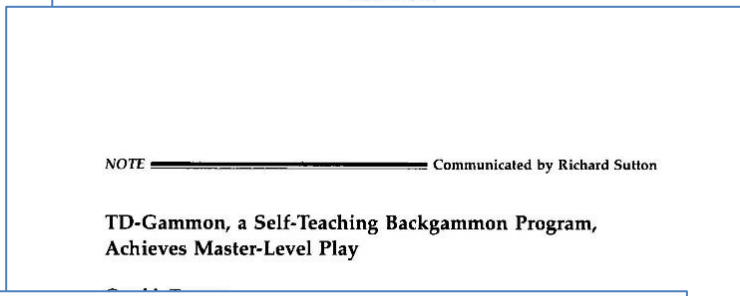
- 计算机刚出现就有人研究计算机下围棋（1964年）
- 计算机下棋成功案例：
  - 西洋跳棋：1992年达到人类世界冠军水平
  - 双　陆　棋：1994年达到人类大师级水平
  - 国际象棋：1997年Deep blue击败人类选手
- 计算机围棋：进展一直缓慢！

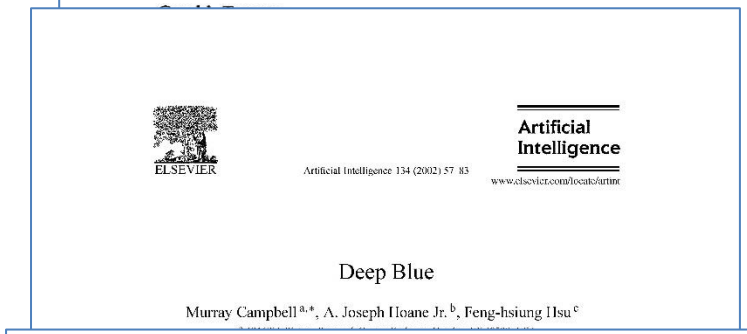



A World Championship Caliber Checkers Program

Jonathan Schaeffer
Joseph Culberson
Norman Treloar
Brent Knight
Paul Lu
Duane Szafron


NOTE — Communicated by Richard Sutton

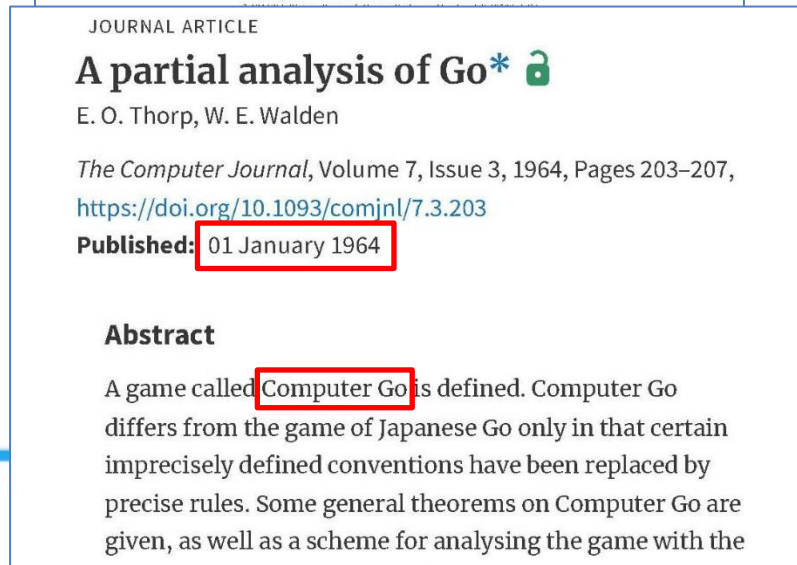
TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play


ELSEVIER

Artificial Intelligence

Deep Blue

Murray Campbell [a,*], A. Joseph Hoane Jr. [b], Feng-hsiung Hsu [c]


JOURNAL ARTICLE

A partial analysis of Go*

E. O. Thorp, W. E. Walden

The Computer Journal, Volume 7, Issue 3, 1964, Pages 203–207, https://doi.org/10.1093/comjnl/7.3.203

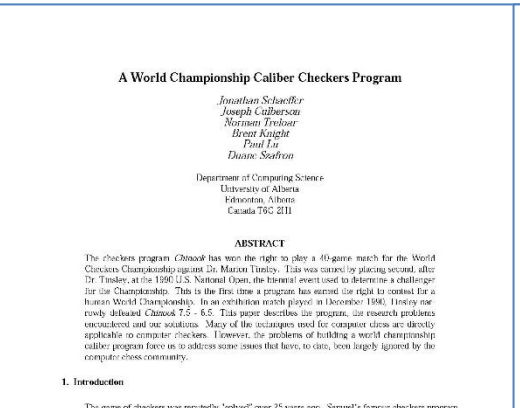
Published: 01 January 1964

Abstract

A game called Computer Go is defined. Computer Go differs from the game of Japanese Go only in that certain imprecisely defined conventions have been replaced by precise rules. Some general theorems on Computer Go are given, as well as a scheme for analysing the game with the

Google's AlphaGO and AlphaZero

- 西洋跳棋：1992年达到人类世界冠军水平
  - 使用评价函数指导每一步棋
- 双陆棋：1994年达到人类大师级水平
  - 成功的原因：使用了强化学习算法TD(λ)
- 国际象棋：1997年Deep blue击败人类选手
  - 使用非常复杂的评价函数和计算机的强大算力（搜索）
- 围棋：进展一直很慢！
  - 复杂评价函数 + 强化学习也不行！



A World Championship Caliber Checkers Program

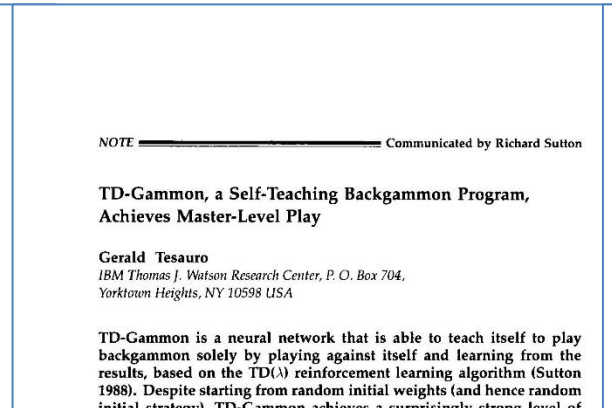
Jonathan Schaeffer
Joseph Culberson
Norman Treloar
Brent Knight
Paul Lu
Duane Szafron

Department of Computing Science
University of Alberta
Edmonton, Alberta
Canada T6G 2H1

ABSTRACT

The checkers program *Chinook* has won the right to play a 40-game match for the World Checkers Championship against Dr. Marion Tinsley. This was earned by placing second, after Dr. Tinsley, at the 1990 U.S. National Open, the biennial event used to determine a challenger for the Championship. This is the first time a program has earned the right to contest for a human World Championship. In an exhibition match played in December 1990, Tinsley narrowly defeated *Chinook* 7.5 - 6.5. This paper describes the program, the research problems encountered and our solutions. Many of the techniques used for computer chess are directly applicable to computer checkers. However, the problems of building a world championship caliber program force us to address some issues that have, to date, been largely ignored by the computer chess community.

1. Introduction

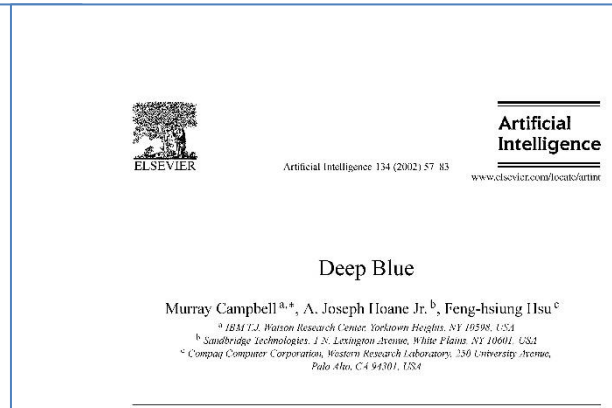


NOTE — Communicated by Richard Sutton

TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play

Gerald Tesauro
IBM Thomas J. Watson Research Center, P. O. Box 704, Yorktown Heights, NY 10598 USA

TD-Gammon is a neural network that is able to teach itself to play backgammon solely by playing against itself and learning from the results, based on the TD(λ) reinforcement learning algorithm (Sutton 1988). Despite starting from random initial weights (and hence random



ELSEVIER

Artificial Intelligence 134 (2002) 57–83

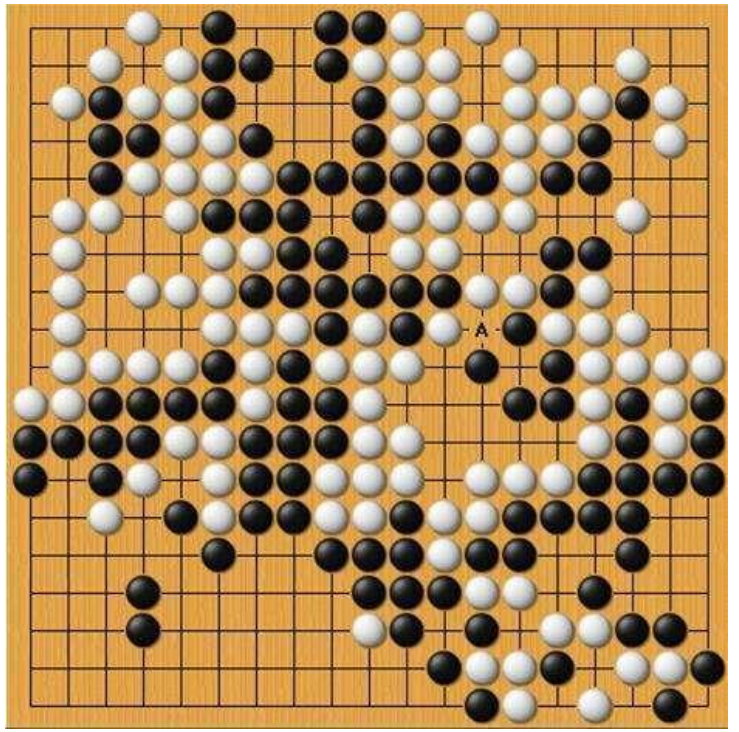
Artificial Intelligence

www.elsevier.com/locate/artint

Deep Blue

Murray Campbell [a,*], A. Joseph Hoane Jr. [b], Feng-hsiung Hsu [c]

[a] *IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA*
[b] *Sandbridge Technologies, 1 N. Lexington Avenue, White Plains, NY 10601, USA*
[c] *Compaq Computer Corporation, Western Research Laboratory, 250 University Avenue, Palo Alto, CA 94301, USA*



Google's AlphaGO and AlphaZero

自适应动态规划与强化学习

- 计算机围棋一直到2015年才获得成功！
  - 2015年击败欧洲围棋冠军樊麾
  - 2016年击败世界冠军李世石
  - 2017年击败世界冠军、世界排名第一的棋手柯洁
  - 超级复杂评价函数 + 强化学习 + 超强算力

ARTICLE

Mastering the game of Go with de
neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dor John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuogl Thore Graepel[1] & Demis Hassabis[1]

The game of Go has long been viewed as the most challenging of classic games for artificial intel enormous search space and the difficulty of evaluating board positions and moves. Here we introd

hanced by policies that These policies are used ability actions, and to

learning[13,21–24]. The SL policy network $p_\sigma(a|$ volutional layers with weights σ, and rectifier max layer outputs a probability distribution

Google's AlphaGO and AlphaZero

计算机围棋:

- **超级复杂评价网络**：深度神经网络，13层之多
- **强化学习**：自动生成评价网络、自己学习下棋
- **超强算力**：上千台服务器!
- 计算机围棋的成功在全世界、尤其在中国掀起了人工智能新热潮!

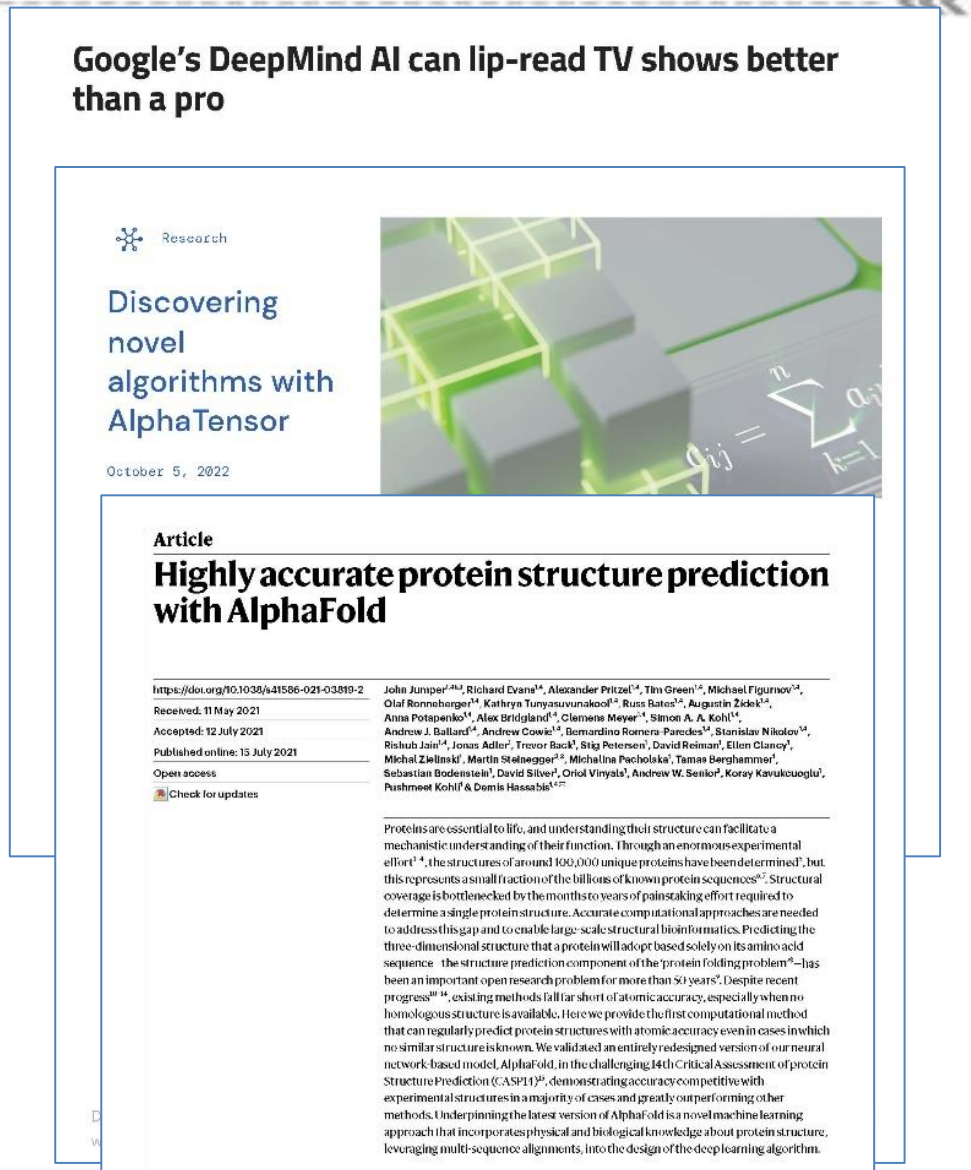
Other Successes of Google's DeepMind

计算机围棋之后:

- 比专业人士更准确地**唇读电视节目**
- 解数学难题，发现新算法：**AlphaTensor**
- 高度准确的蛋白质结构预测：**AlphaFold**
- ……

同样方法一定可以用在控制上

- 我们团队工作：用ADPRL自适应动态规划和强化学习方法做最优控制
- **复杂非线性系统的自学习最优控制**



Google's DeepMind AI can lip-read TV shows better than a pro

Research

Discovering novel algorithms with AlphaTensor

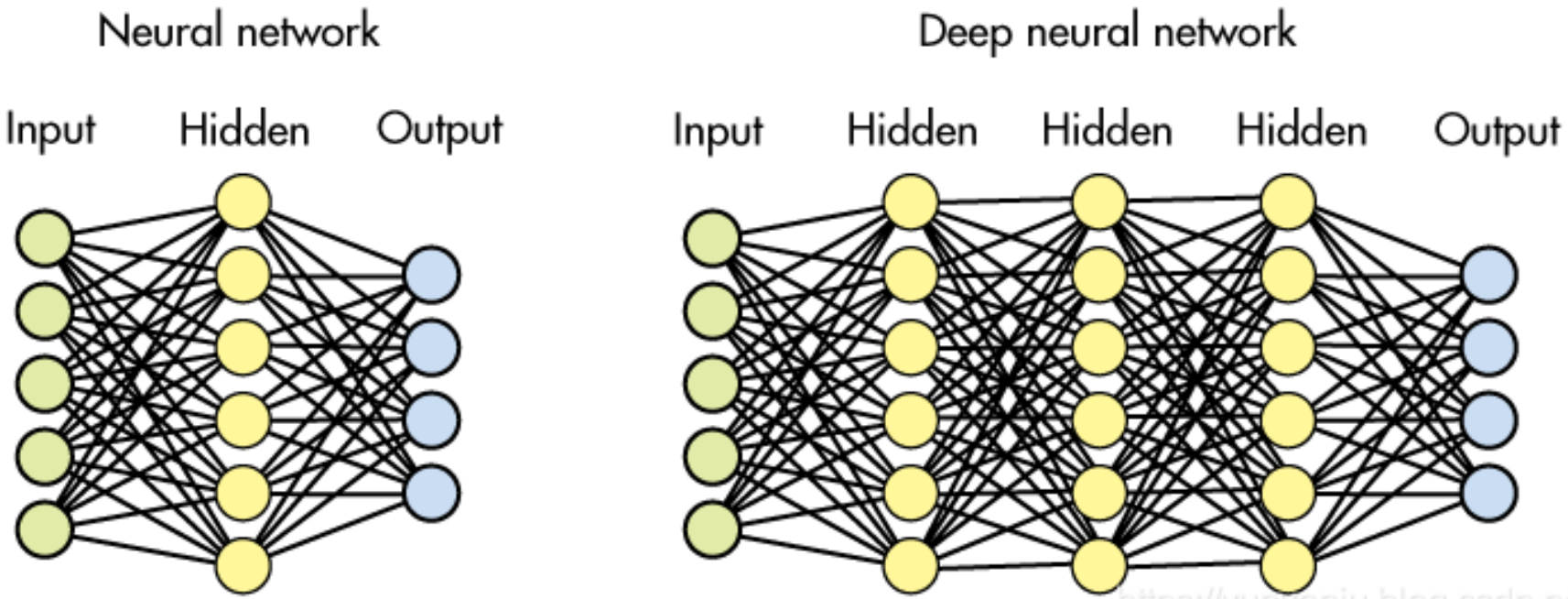
October 5, 2022

Article

Highly accurate protein structure prediction with AlphaFold

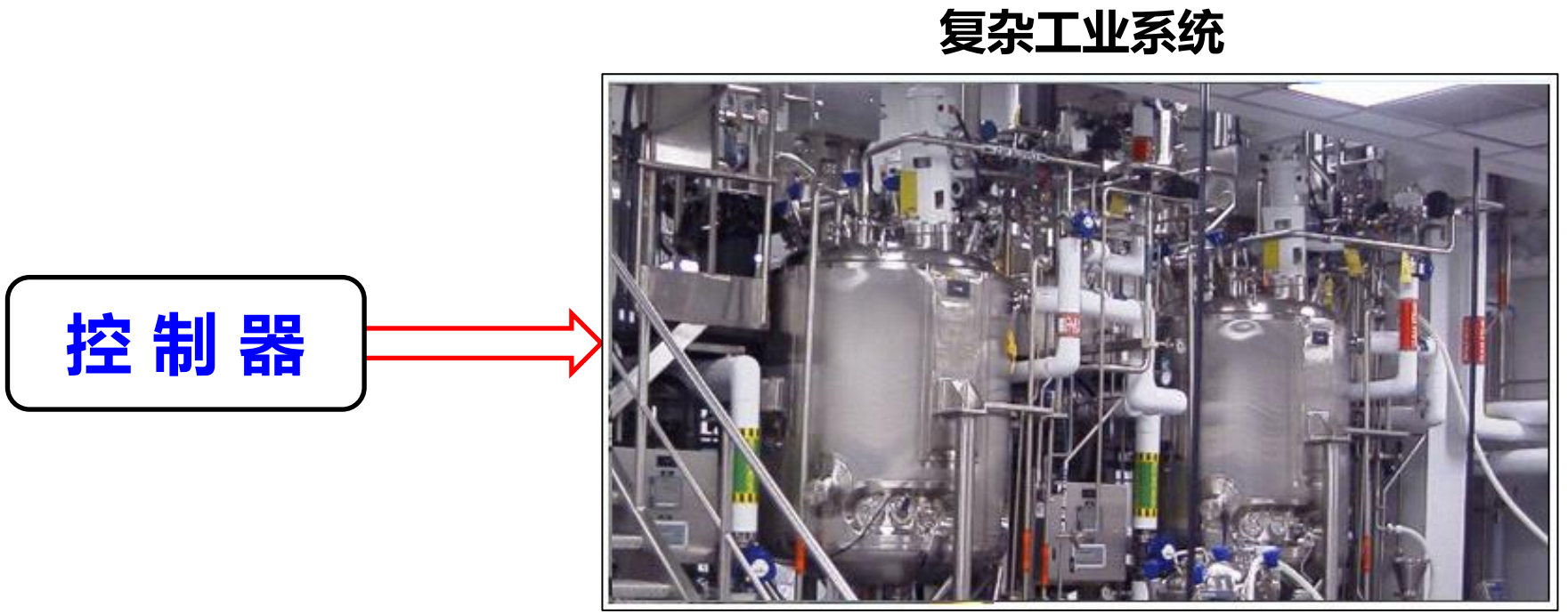
神经网络

- 具有一组可以被学习算法调节的权重或参数
- 可以准确表达输入、输出数据间的复杂非线性关系



复杂非线性系统的自学习最优控制

- **目标：通过神经网络训练来完成强化学习算法自己获取最优控制器**

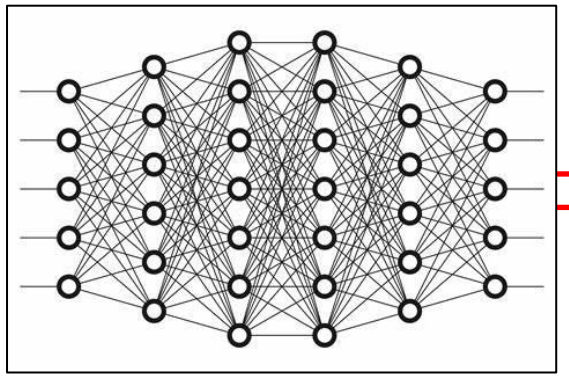


- 无模型、无方程表达式
- 无有效设计方法

复杂非线性系统的自学习最优控制

- **目标：通过神经网络训练来完成强化学习算法自己获取最优控制器**

神经网络强化学习控制器



复杂工业系统

- 无模型、无方程表达式
- 无有效设计方法

复杂非线性系统的自学习最优控制

光有神经网络还不够:

- **神经网络：有许多权重参数可以调整、有训练方法。**
- **还需要强化学习手段！强化学习算法是自学习控制的核心！**
- **两者结合可以发挥巨大威力！**
- **用神经网络作为工具来实现强化学习算法！**
- **进而实现复杂非线性系统的自学习最优控制！**

- 贝尔曼最优化原理：

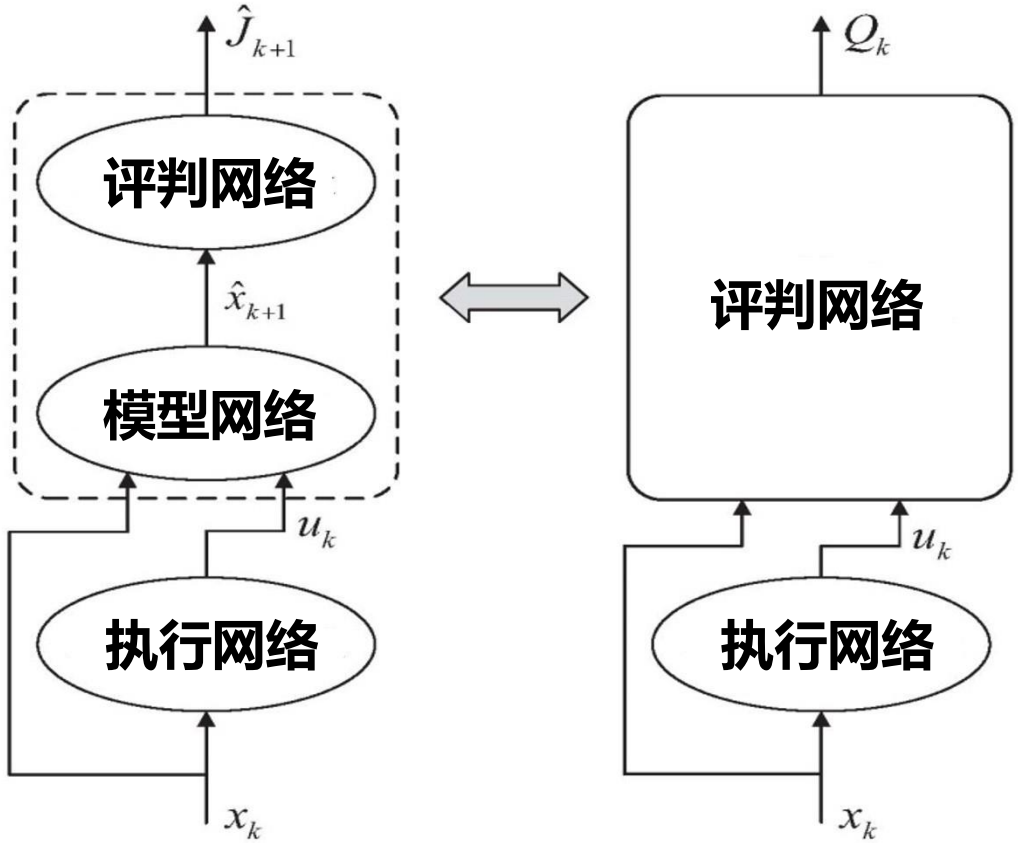
$$J^*(x_k) = \min_{u_k}\{U(x_k, u_k) + J^*(x_{k+1})\}. \qquad (1)$$

评判网络　执行网络

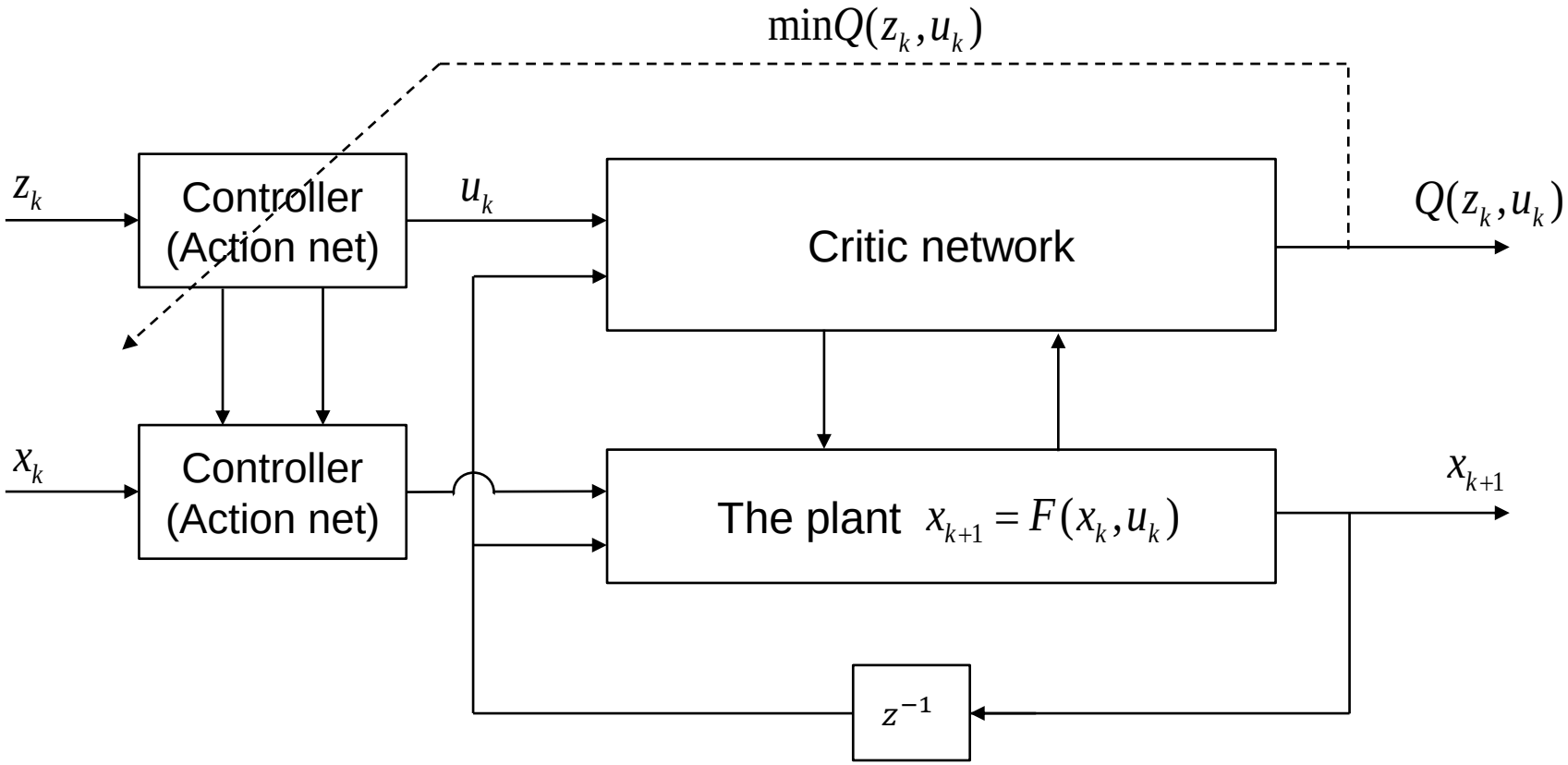
模型网络 $(x_k, u_k) \rightarrow x_{k+1}$

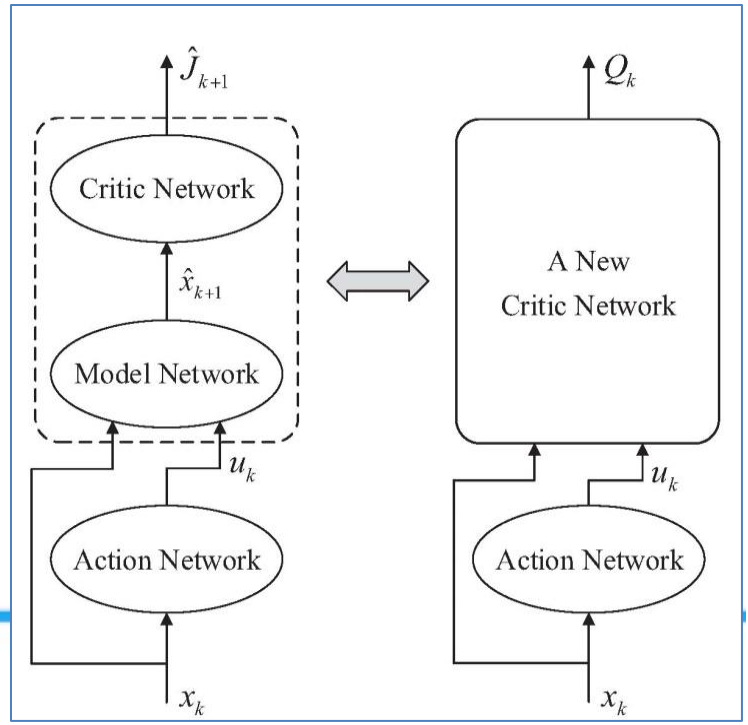
$$E_k = \frac{1}{2}(Q_k - U_k - Q_{k+1})^2$$

Paul Werbos



Self-learning controller

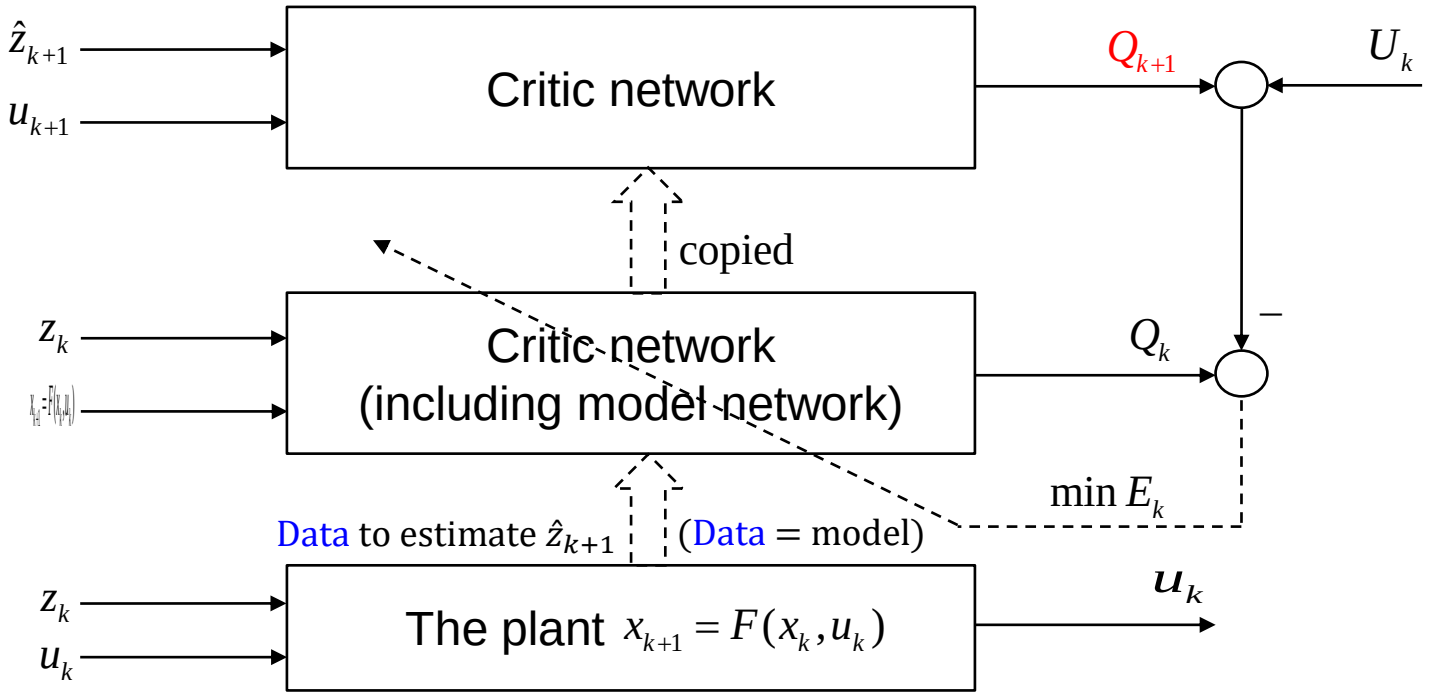




Step 1/3 (critic network)

$$E_k = \frac{1}{2}(Q_k - U_k - Q_{k+1})^2$$

Paul Werbos, 1977



The following data are collected:

$x_0, x_1, x_2, \ldots, x_N$

$u_0, u_1, u_2, \ldots, u_N$

$U_0, U_1, U_2, \ldots, U_N$

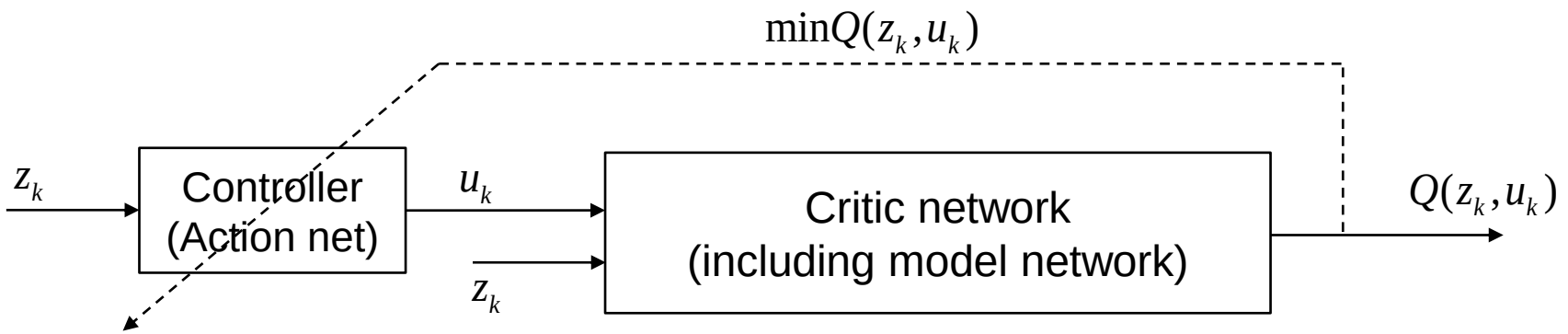
$Q_0, Q_1, Q_2, \ldots, Q_N$

In order to achieve $Q_k - U_k - Q_{k+1} = 0$, the Critic Network learns the following mapping:

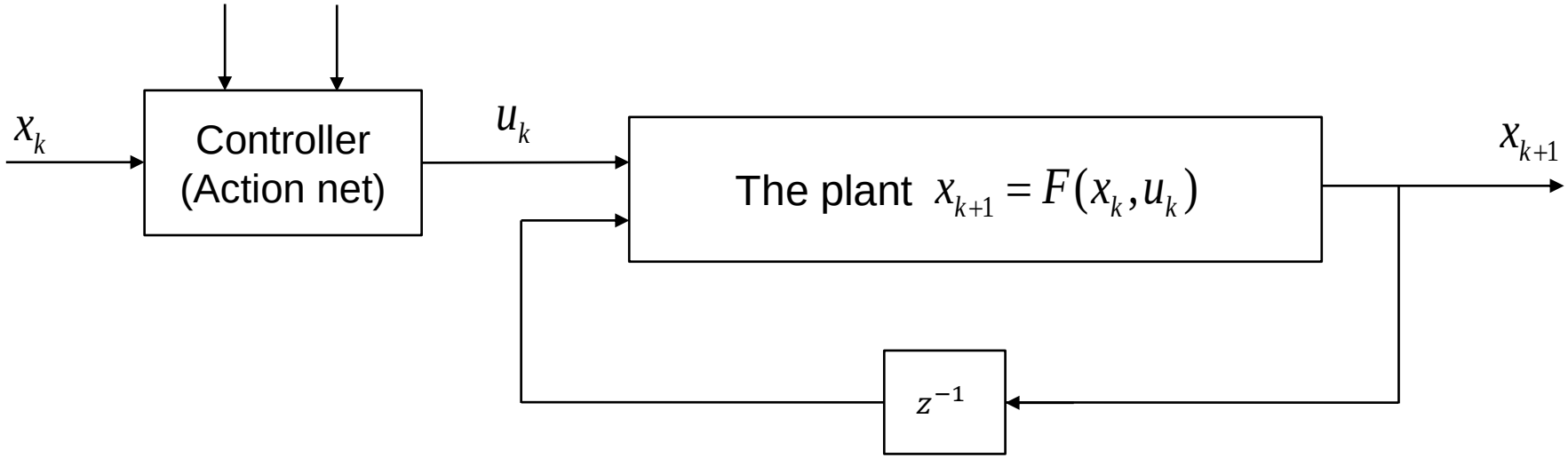
$$NN: \begin{Bmatrix} x_k \\ u_k \end{Bmatrix} \rightarrow \{U_k + Q_{k+1}\}.$$

ADPRL is a data-based method

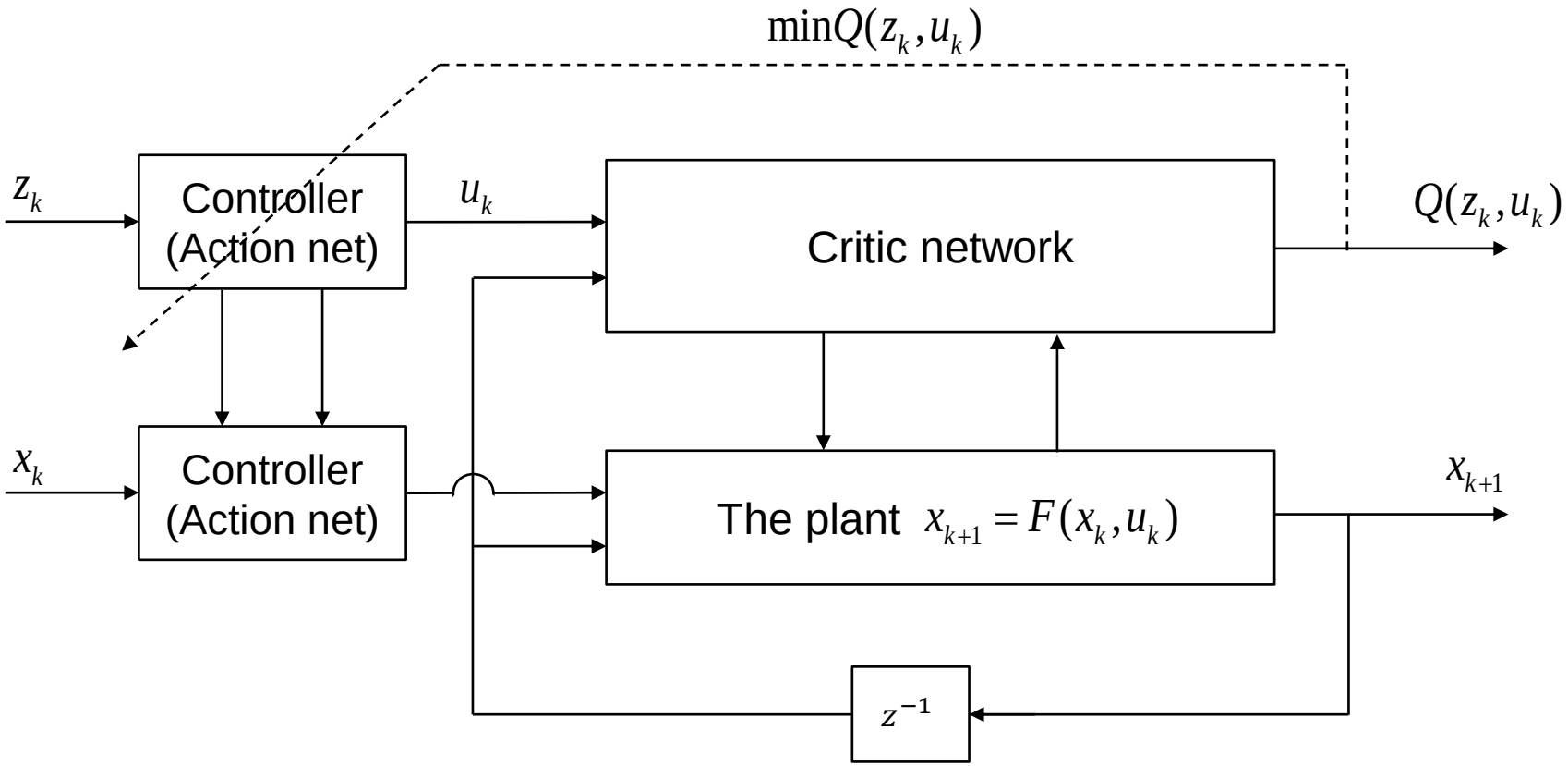
Step 2/3 (controller)



Step 3/3 (implementation)



We actually need Steps 1,2,3,1,2,3, …… repeating

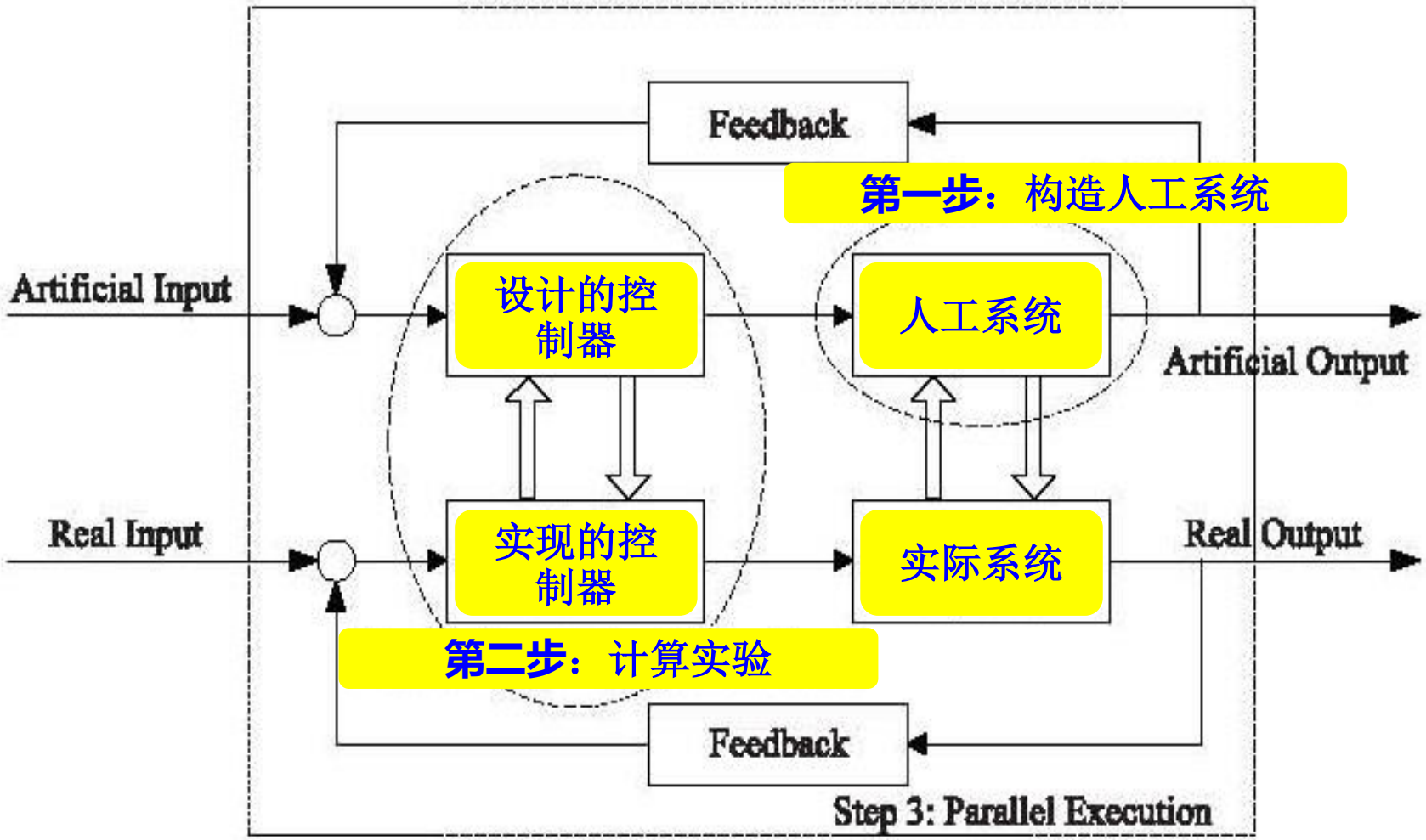


- **Intelligence comes from the critic network – A better critic implies better intelligence and better performance.**
- **It is fulfilled by the action network/controller.**

ADPRL: like a coach, like an athlete

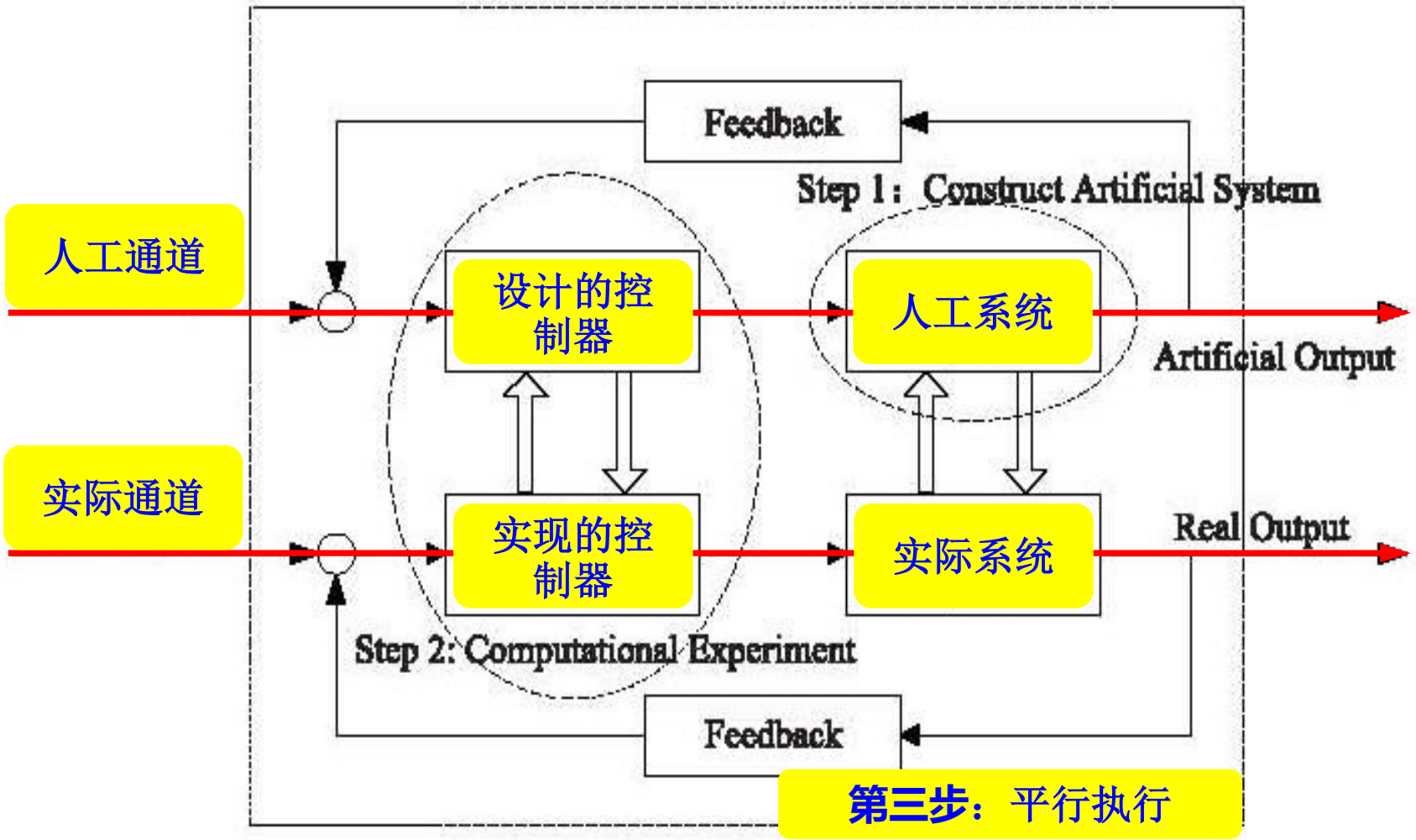
Sherm Chavoor was a coach for the USA's Olympic Swimming teams in 1968 and 1972. Under his coaching, America earned 31 Olympic medals with 21 gold. And he can't swim!!!

平行控制

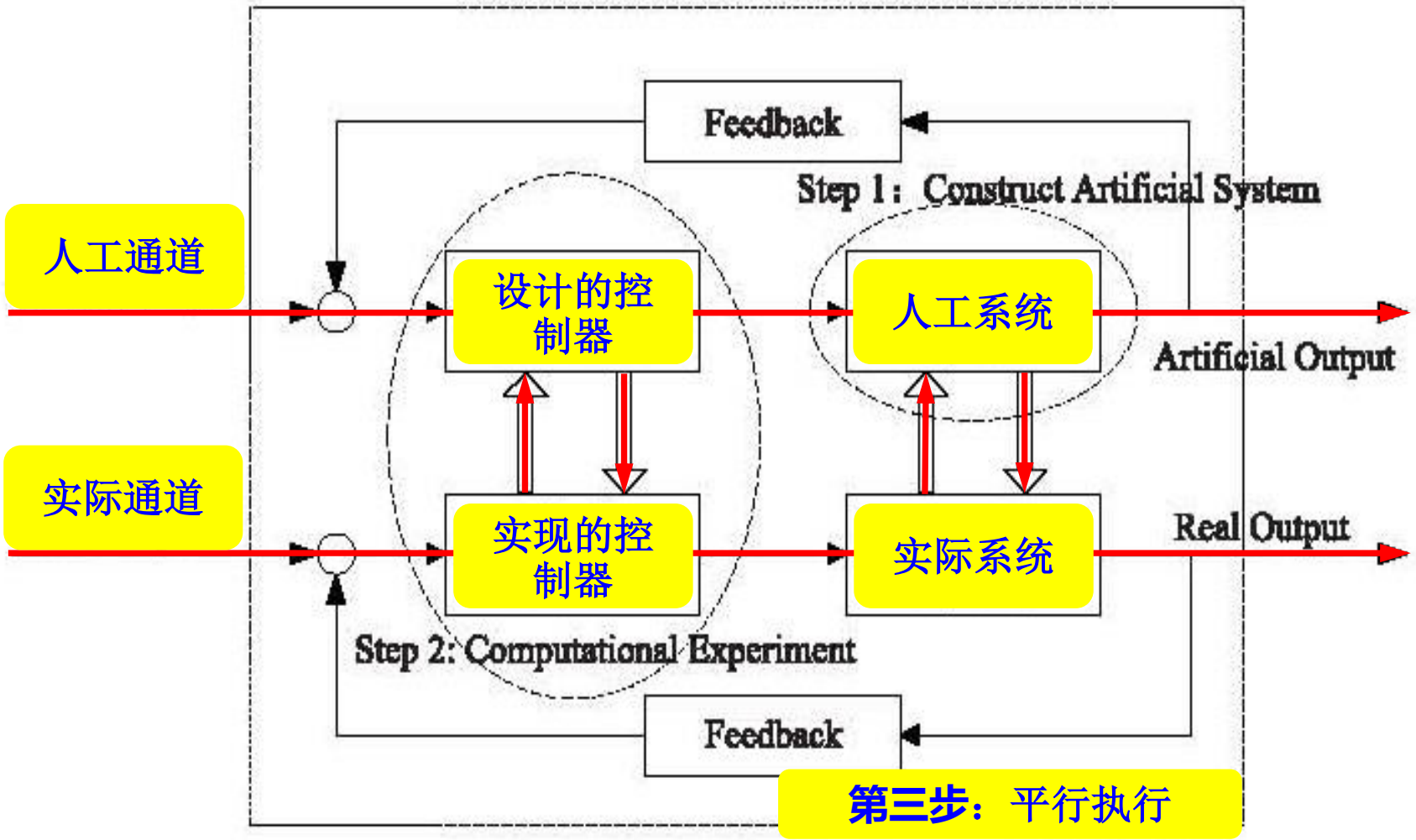


Fei-Yue Wang, 2004

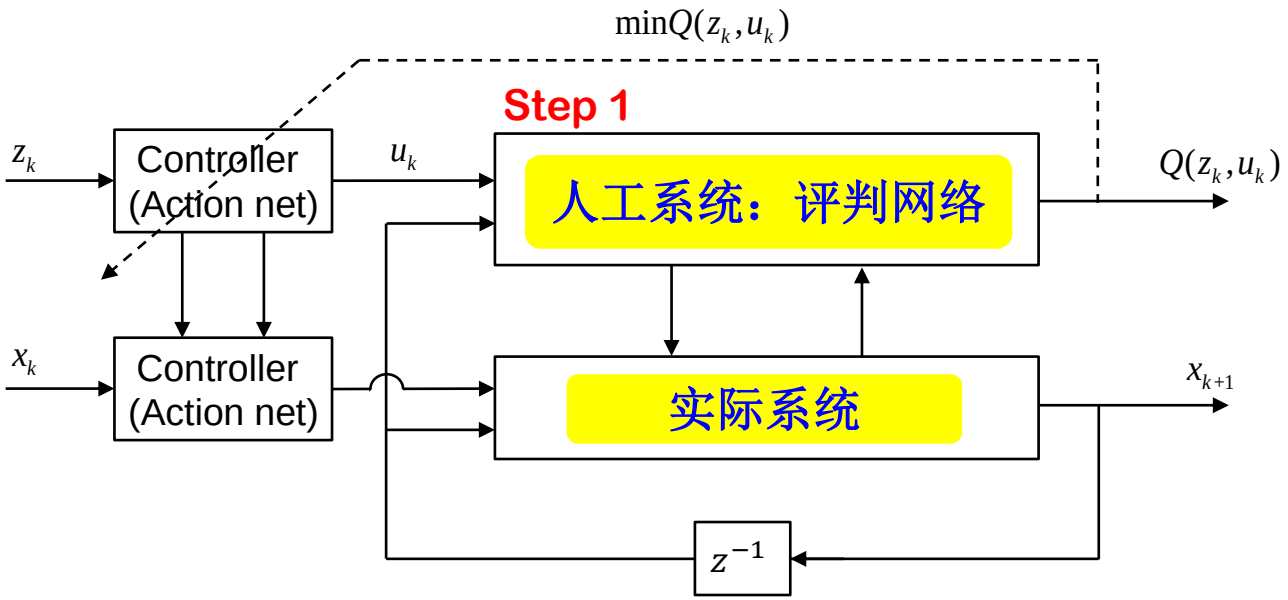
平行控制


Feedback
Step 1：Construct Artificial System
人工通道
设计的控制器
人工系统
Artificial Output
实际通道
实现的控制器
实际系统
Real Output
Step 2: Computational Experiment
Feedback
第三步：平行执行

平行控制

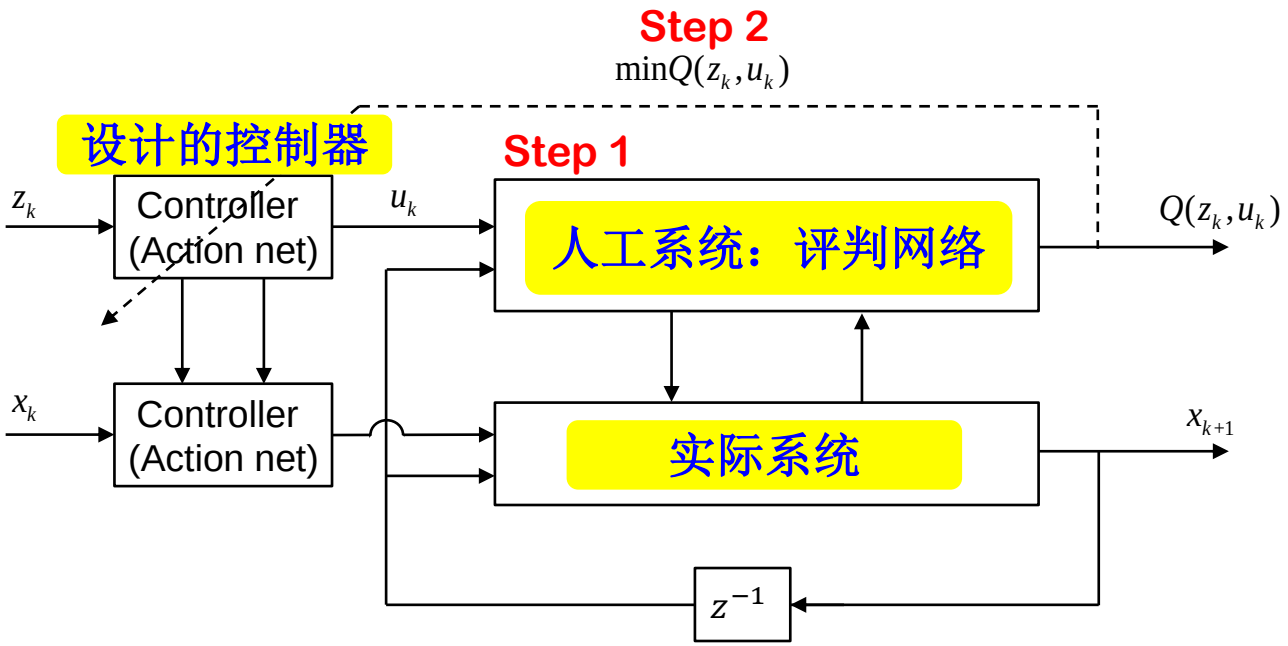

Feedback
Step 1：Construct Artificial System
人工通道
设计的控制器
人工系统
Artificial Output
实际通道
实现的控制器
实际系统
Real Output
Step 2: Computational Experiment
Feedback
第三步：平行执行

ADPRL



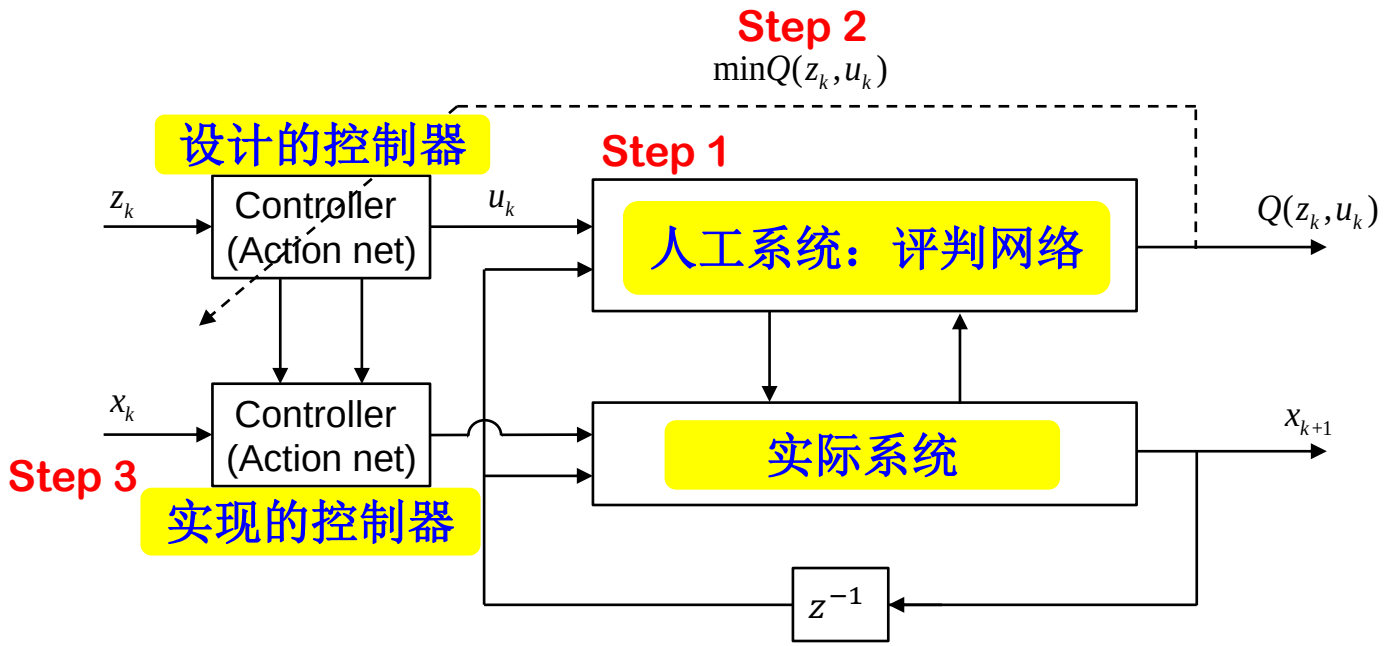
ADPRL	平行控制
评判网络	人工系统
$\min Q$	计算实验
实现	平行执行

ADPRL



ADPRL	平行控制
评判网络	人工系统
$\min Q$	计算实验
实现	平行执行

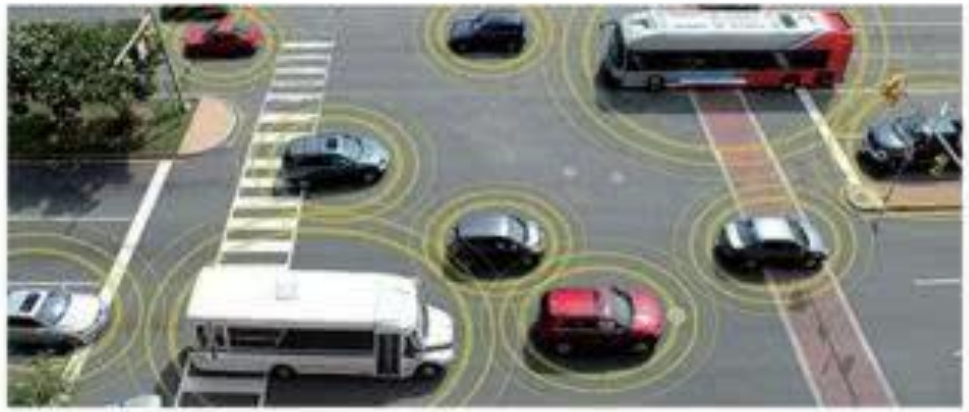
- **ADPRL是平行控制的一个实现**



ADPRL	平行控制
评判网络	人工系统
$\min Q$	计算实验
实现	平行执行

云控制

- **分立系统：具有一定的自主控制能力**
  - **本地控制器：简单具备基本功能.**
- **云端：复杂计算、统筹协调、系统决策**
  - **远程控制器：更复杂的智能控制器、优化、学习、更新等放在云里.**

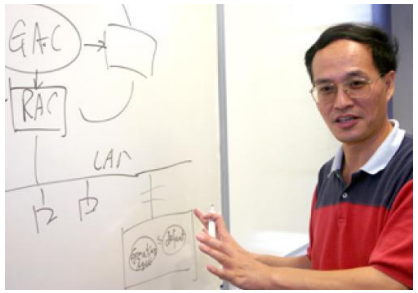


Yuanqing Xia, 2012

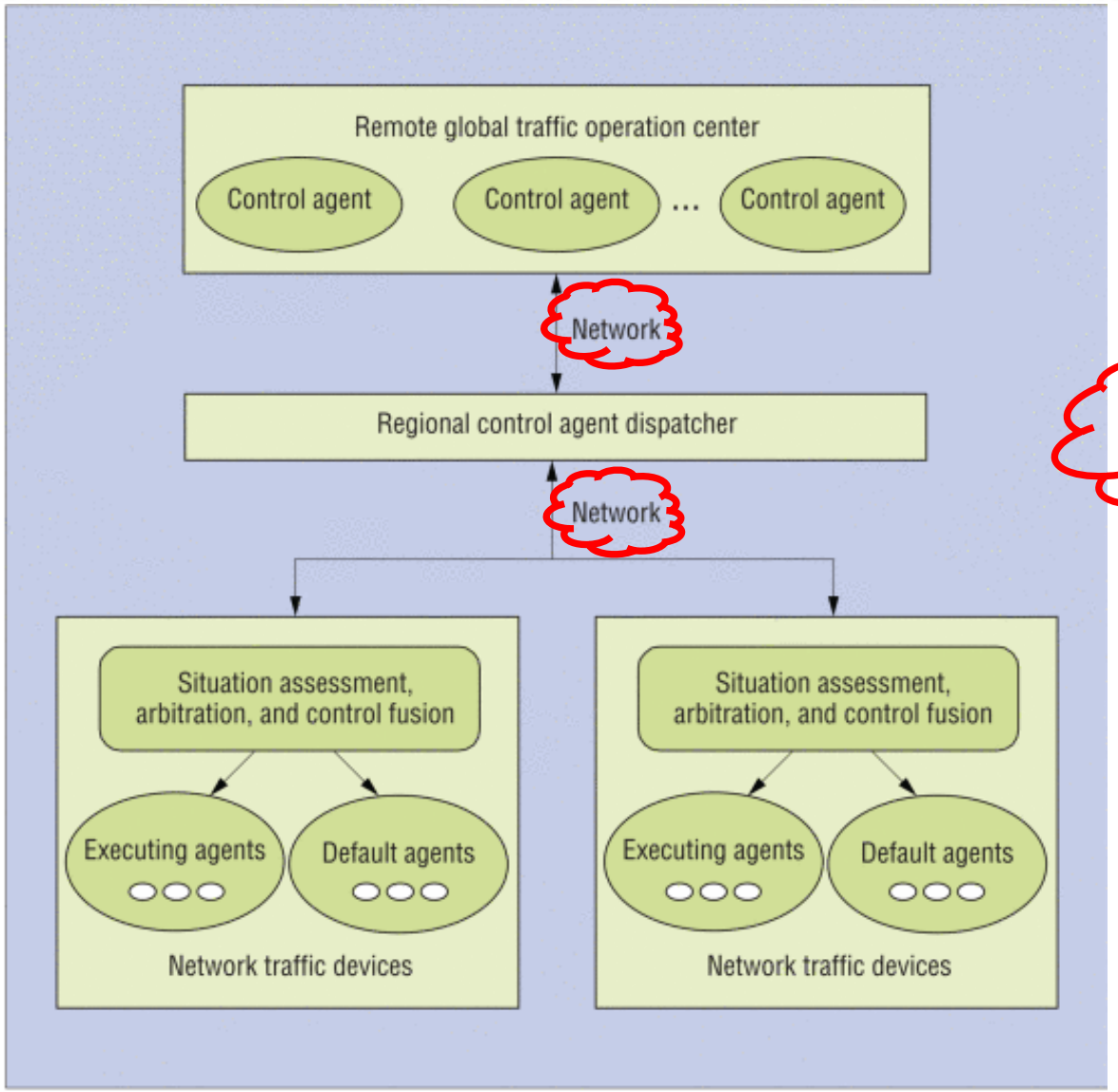
基于代理的控制 – Agent-based control

- **Agent分类的基本思想:**

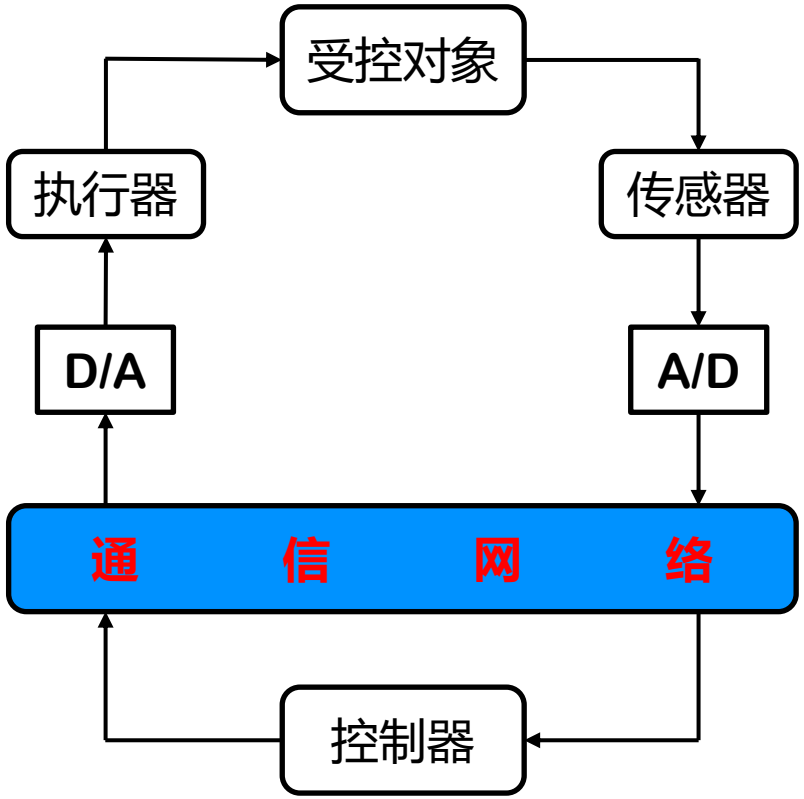
本地简单、远程复杂

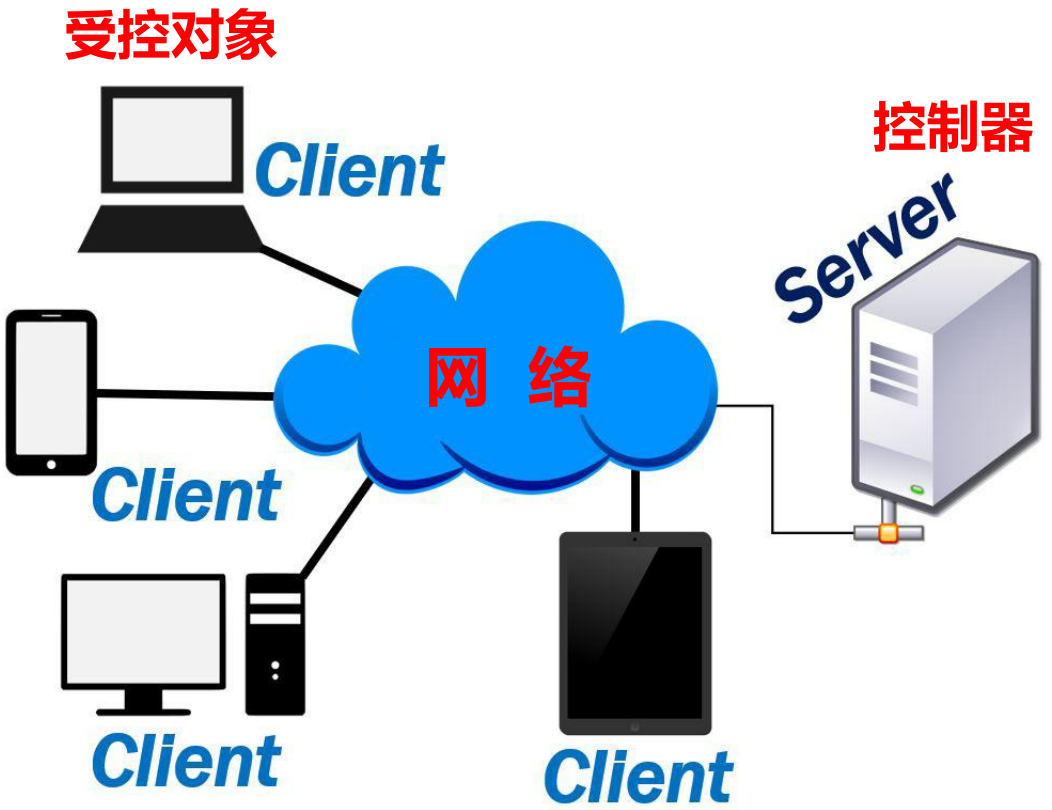


Fei-Yue Wang, 2005

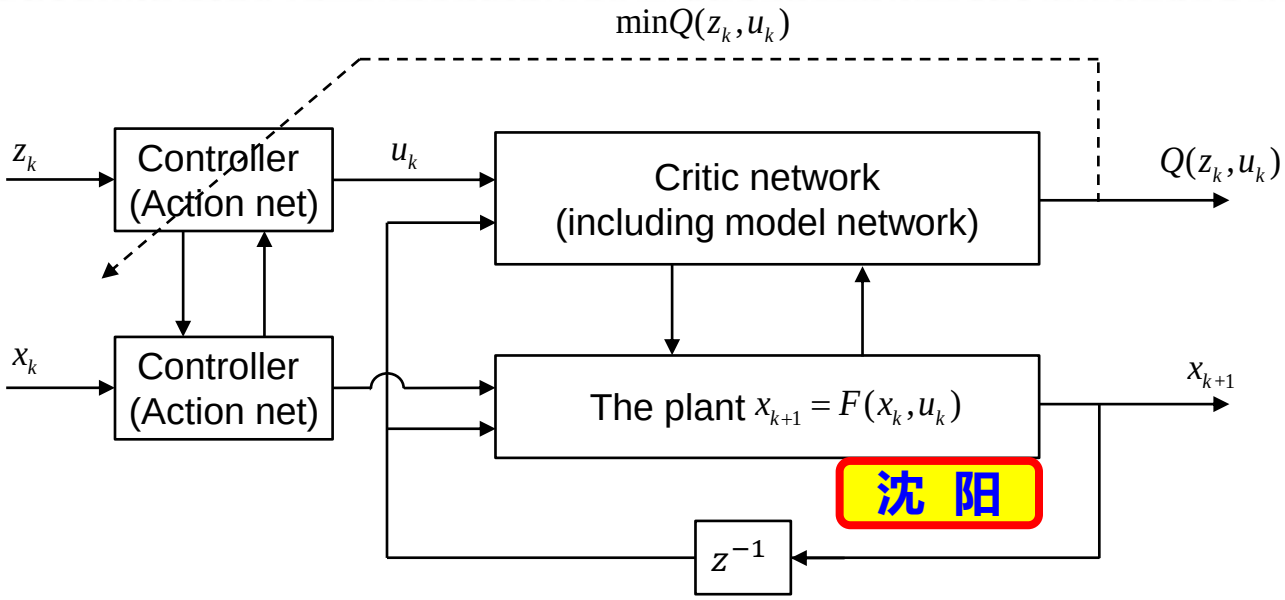


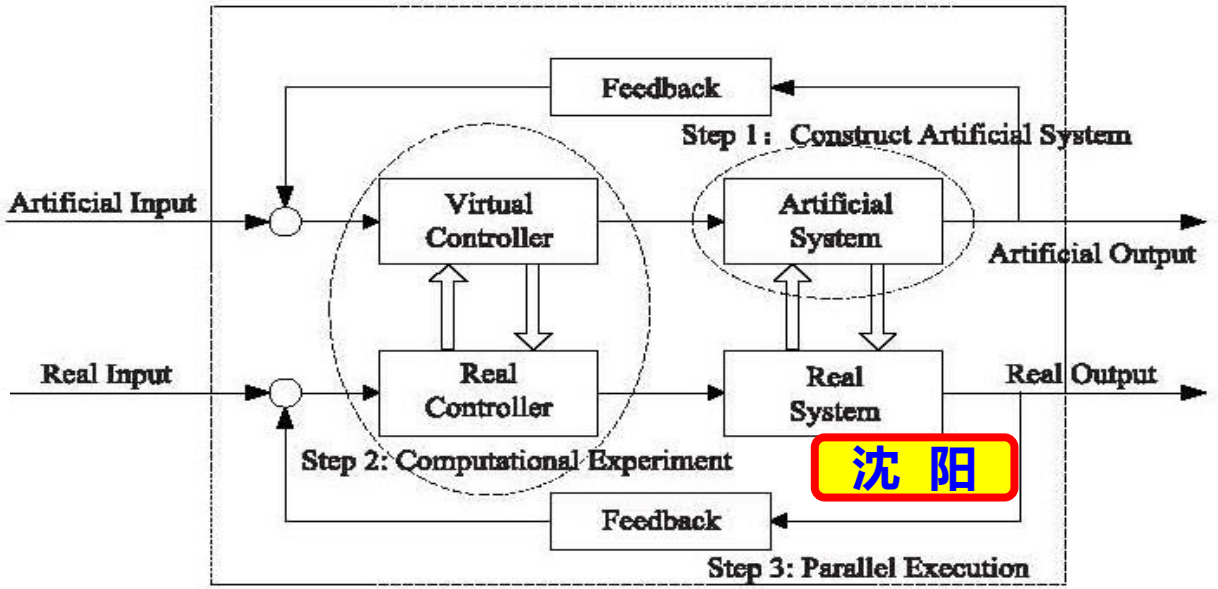
网络化控制


受控对象
执行器
传感器
D/A
A/D
通 信 网 络
控制器

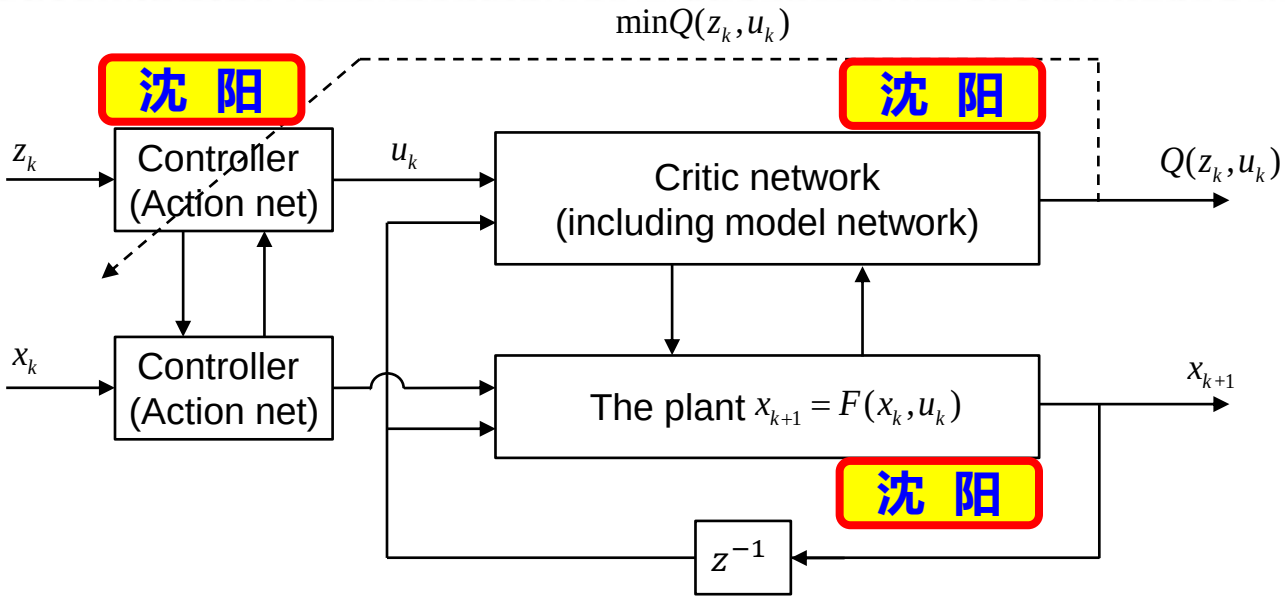

受控对象
控制器
Client
Server
网 络
Client
Client
Client

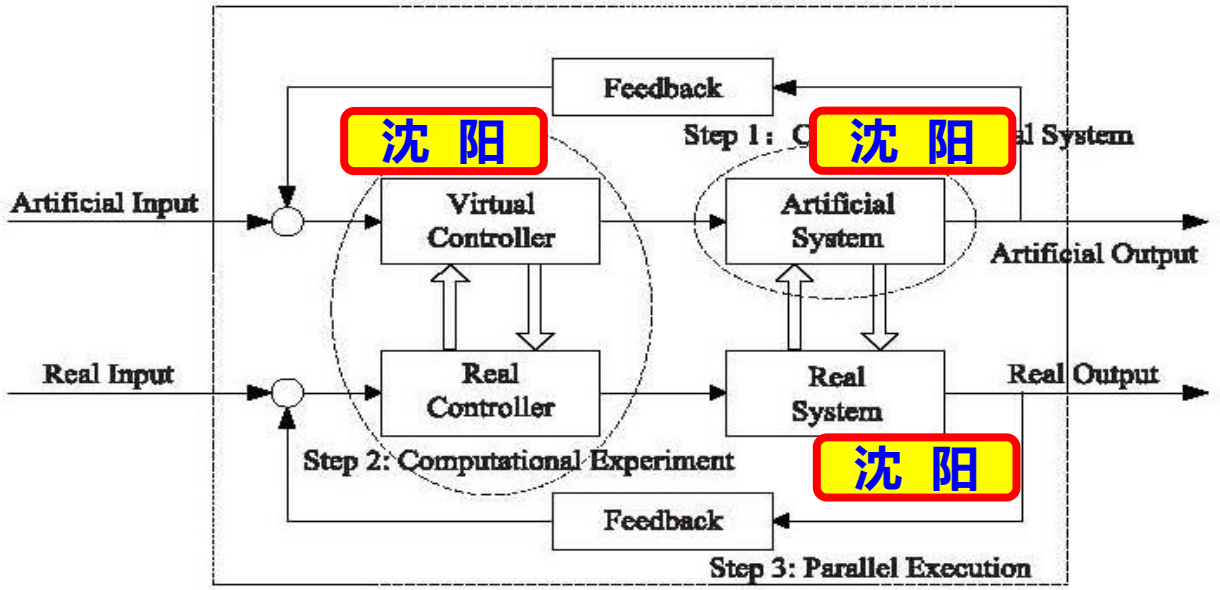
平行控制 / ADPRL / 网络化控制 / 云控制 / ABC



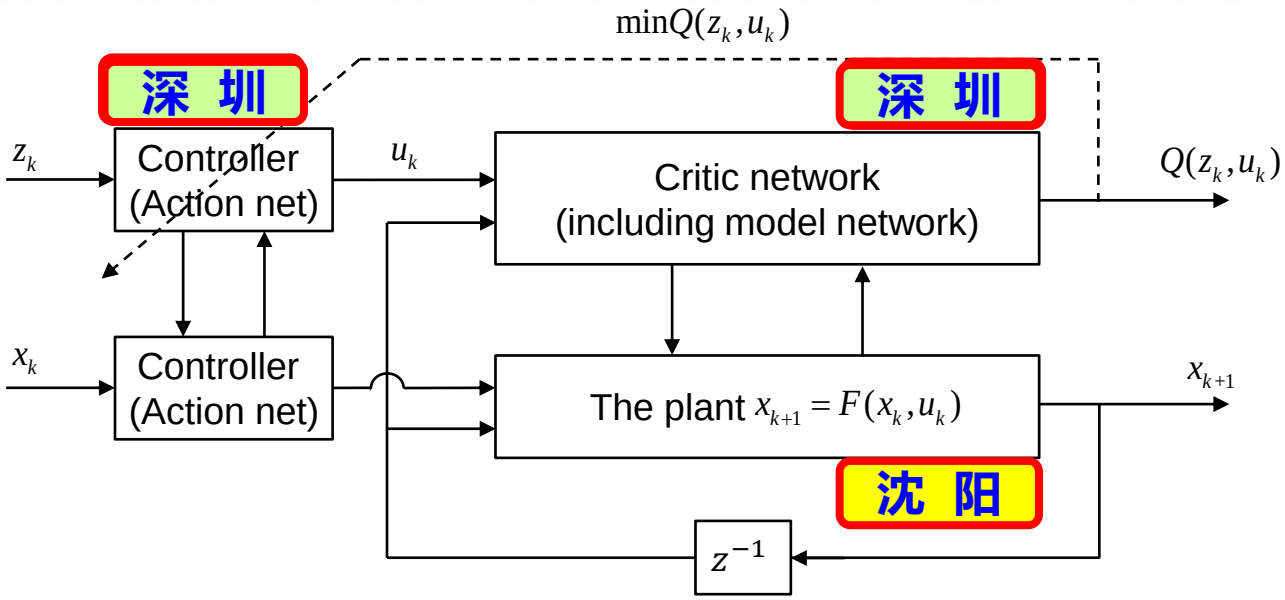


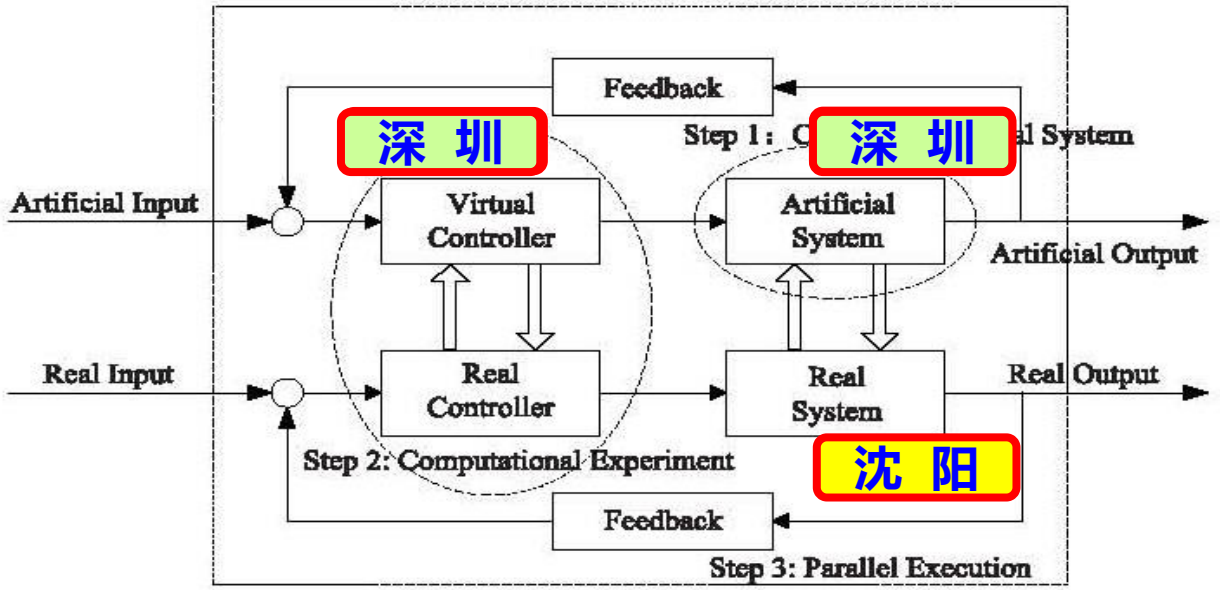
平行控制 / ADPRL / 网络化控制 / 云控制 / ABC



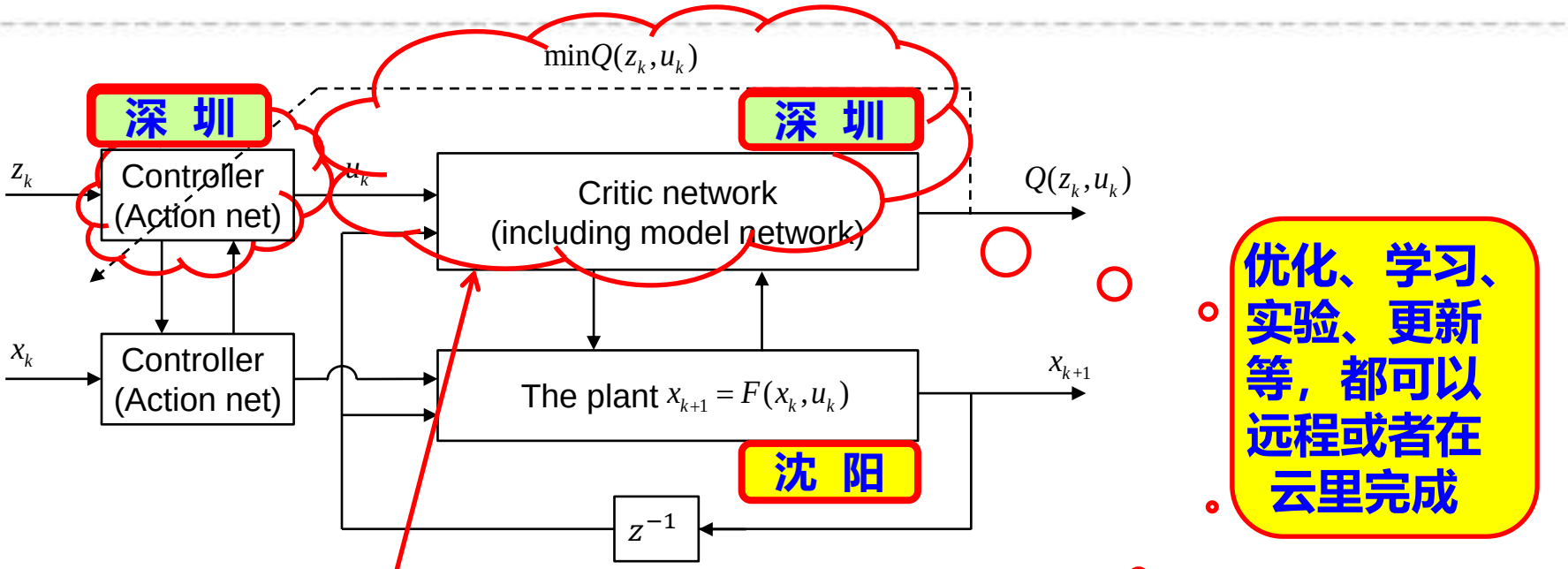


平行控制 / ADPRL / 网络化控制 / 云控制 / ABC



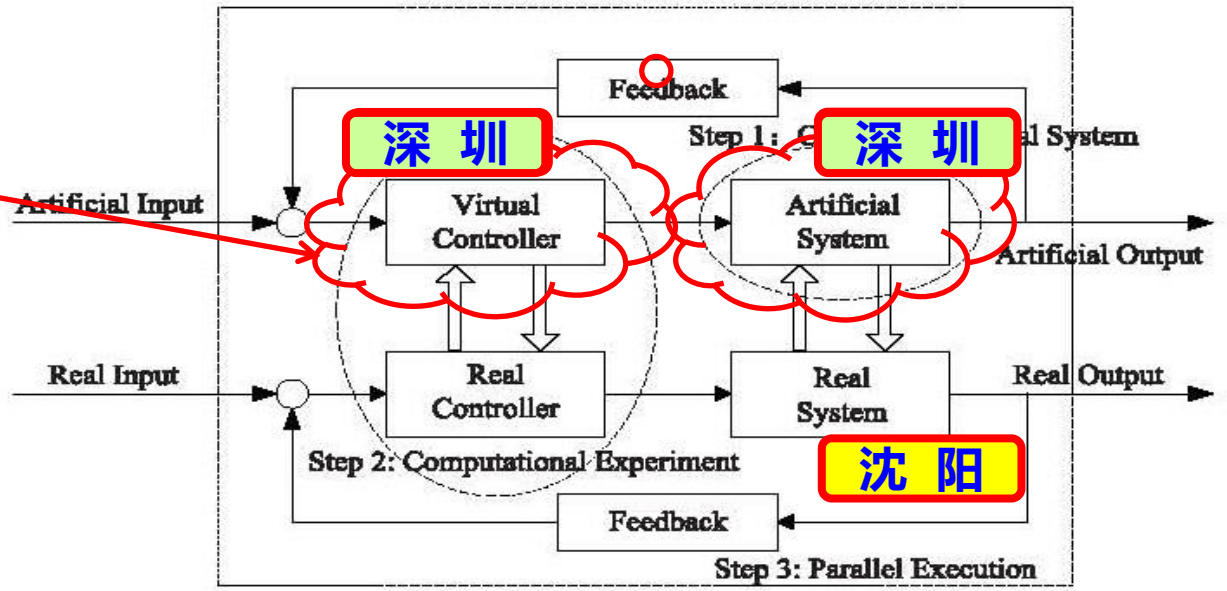


平行控制 / ADPRL / 网络化控制 / 云控制 / ABC



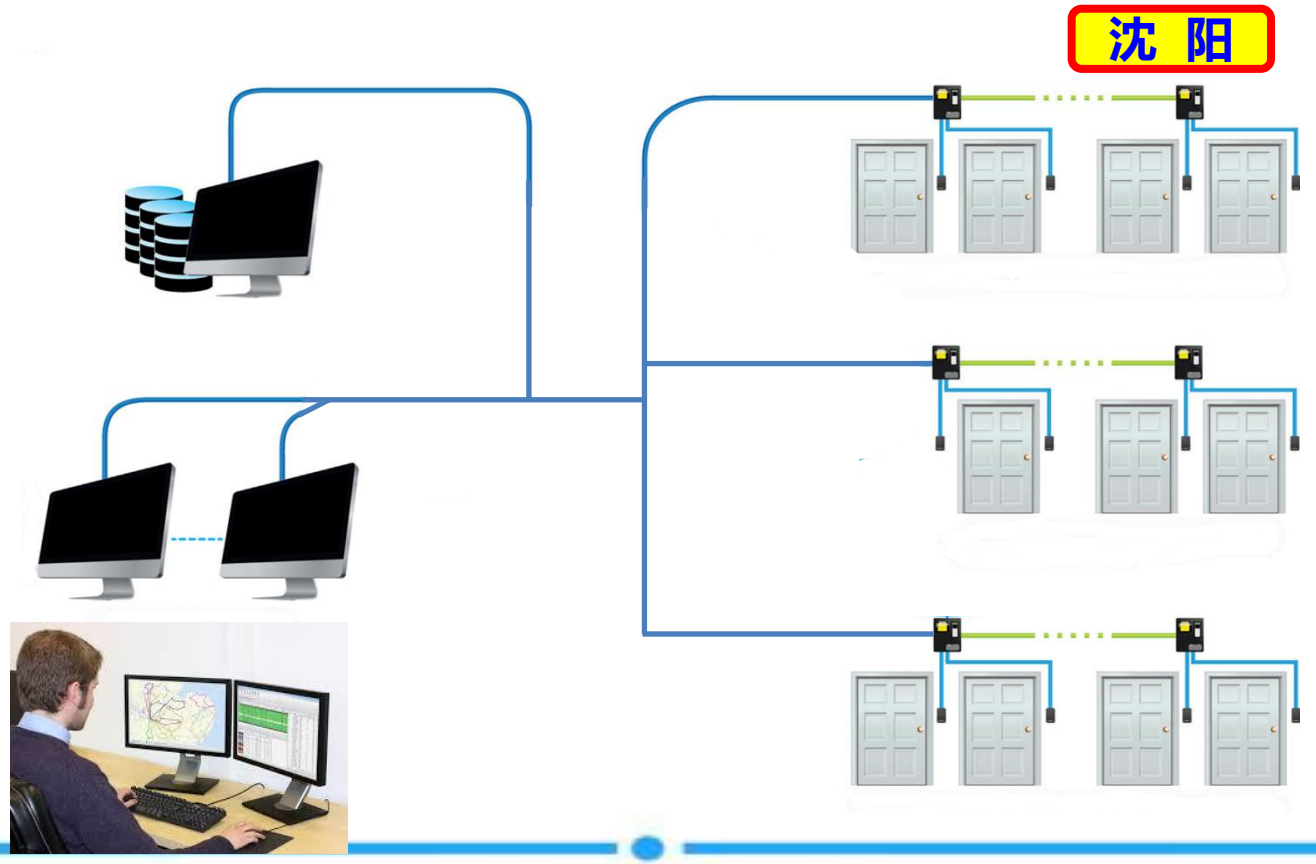
优化、学习、实验、更新等，都可以远程或者在云里完成

远程或者在云里



平行控制 / ADPRL / 网络化控制 / 云控制 / ABC

- 传统的实现方式：控制器需要在现场



平行控制 / ADPRL / 网络化控制 / 云控制 / ABC

- 传统的实现方式：控制器需要在现场
- 网络融合的实现：控制器和现场可以两地

平行控制 / ADPRL / 网络化控制 / 云控制 / ABC

- 第一次工业革命：机械化（1760－1840）
- 第二次工业革命：电气化（1860－1914）
- 第三次工业革命：信息化（20世纪）
- 第四次工业革命：网络化、智能化（现在）

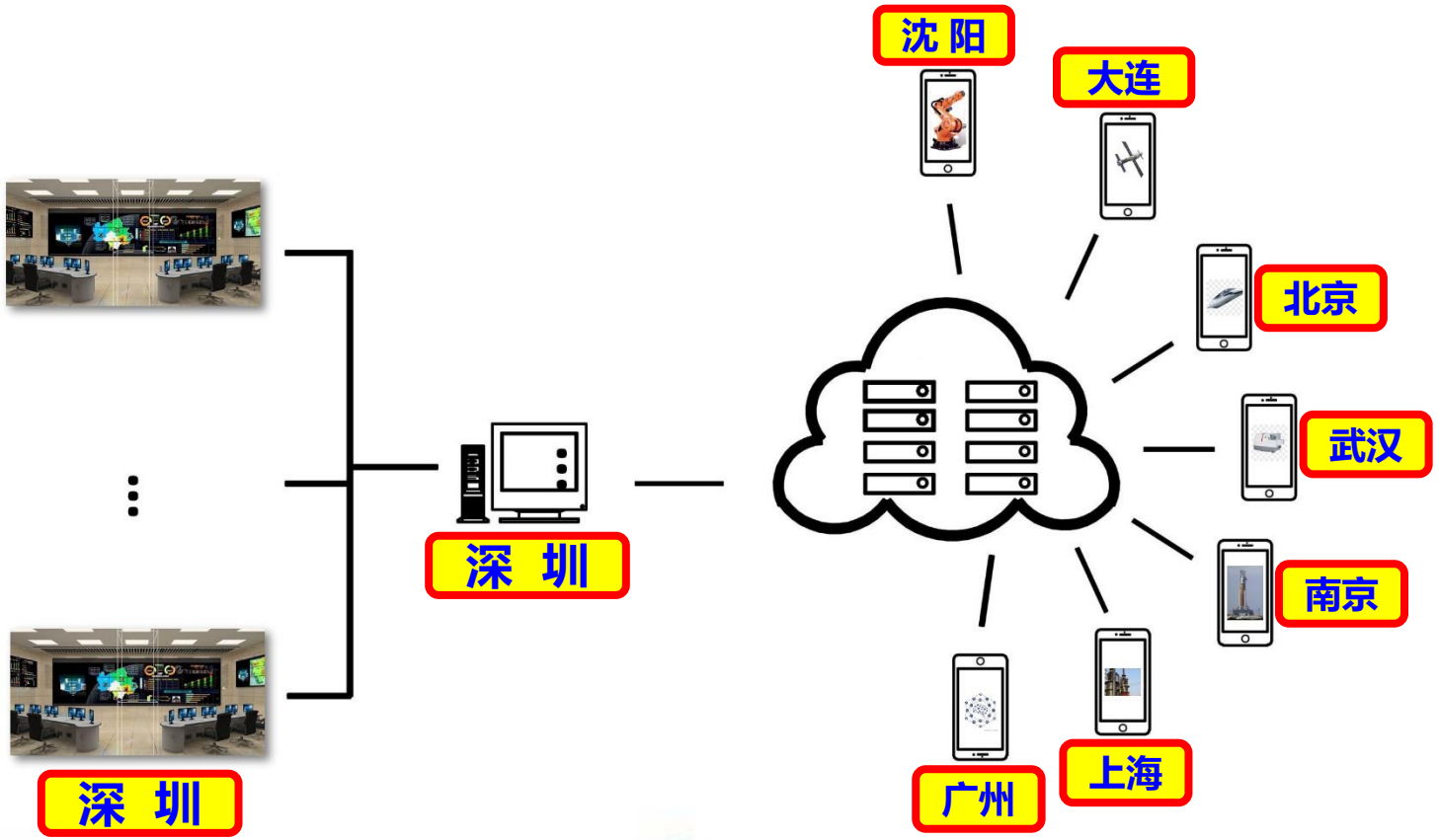
控制方法和网络的深度融合

沈 阳

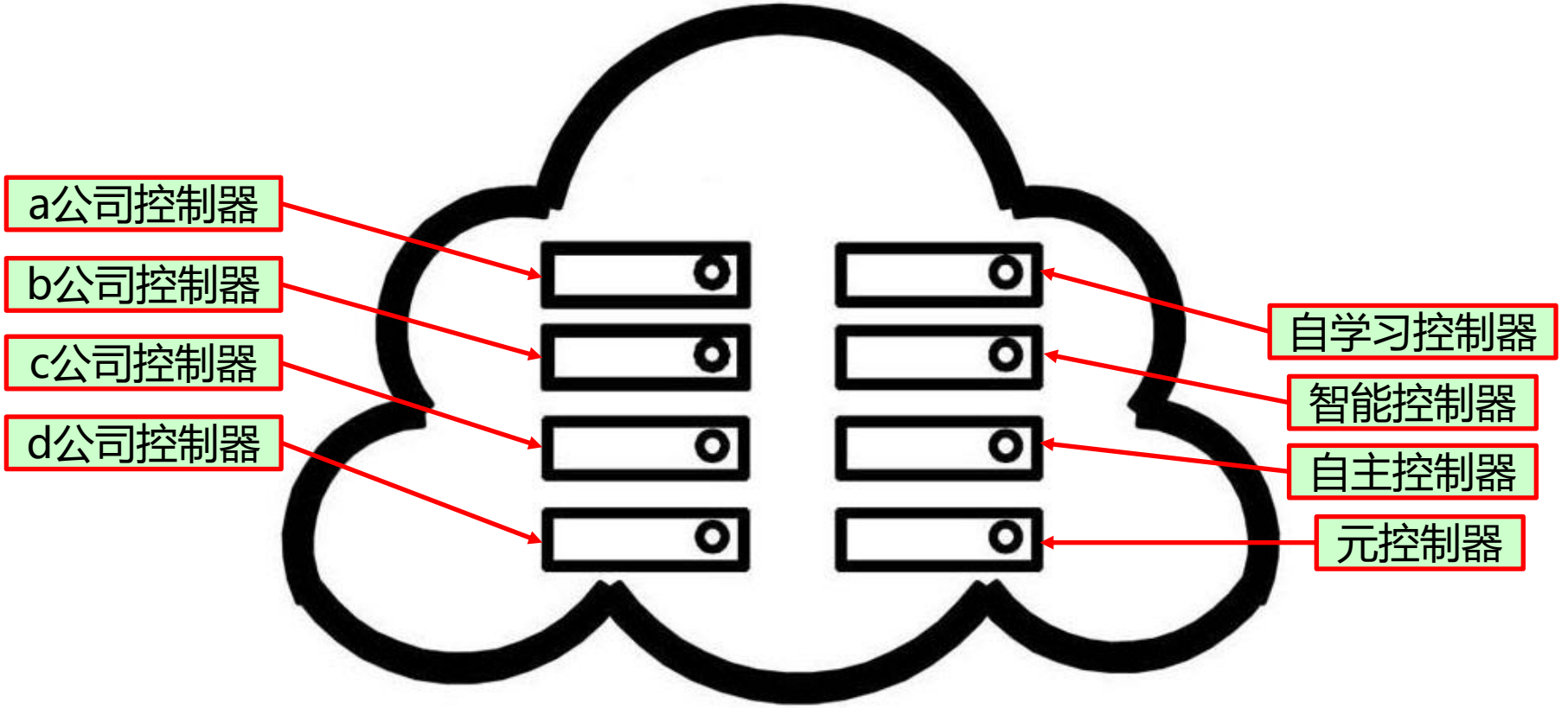
深 圳

总结一下

- 控制方法必须跟互联网做深度融合 = 元控制
- 跟工业互联网深度融合的人工智能是工业智能发展的关键

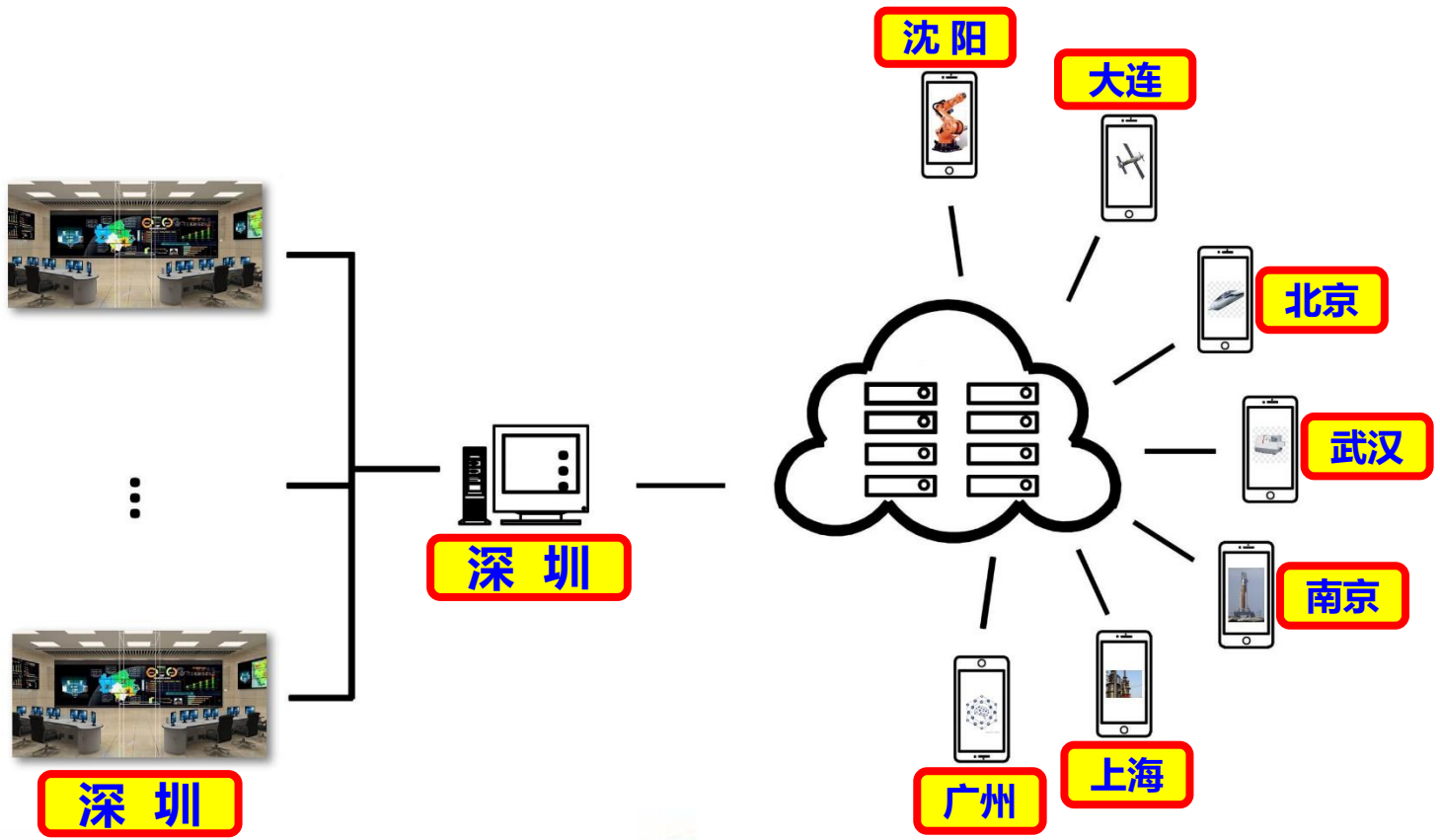


进一步展开



全新的商业模式

- 每个公司研发能力有限
- 客户需求不同、客户按需获取控制器



总结

- 网络化控制方法
- 云控制方法
- 基于代理的控制方法
- 平行控制方法
- 自适应动态规划与强化学习

元控制 通过严格的系统分析满足控制系统性能指标，在**虚实框架下**实现被控系统的预期控制与管理功能。通过检查是否理解、预测结果、评价某个尝试的有效性、计划下一步动作、测查策略，确定当时的时机和努力、修改或变换策略以克服所遇到的困难等。使用计划、监视、调节的办法，获得控制的最符合生产预期的结果。

For more information

- liudr@sustech.edu.cn
- http://www.derongliu.org
- WeChat: L13810670526

衷心感谢!

Thank you!