# Reinforcement Learning Fundamentals

## Lecture 2: RL Framework

Dr Sandeep Manjanna
Assistant Professor, Plaksha University
sandeep.manjanna@plaksha.edu.in

# Course Logistics

- Classes on **M, W, and F** at **12:00 to 12:50 pm** in Room **1001**.

- Office hours: **Mondays 3:00 to 4:00 pm** (Office **2407**).

- TFs →      Keshav Sivakumar      Office Hour: Wed 11:00 to 12:00 pm
             Poonam Adhikari      Office Hour: Mon 2:00 to 3:00 pm

# What is RL?

- Learning about stimuli and actions based <u>on rewards and punishments alone</u>.
- No detailed supervision available
- Trial-and-error learning
- Delayed rewards
- Sequence of actions required to obtain reward
- Associative learning required
  - Need to associate actions to states
- Learn about policies not just actions
- Typically in a stochastic world

# Applications of RL

- Stanford's autonomous helicopter flight: extreme aerobatics under computer control (2008)
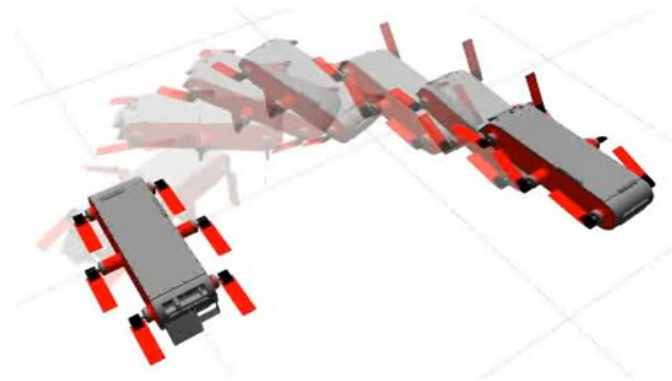
Source: http://heli.stanford.edu/
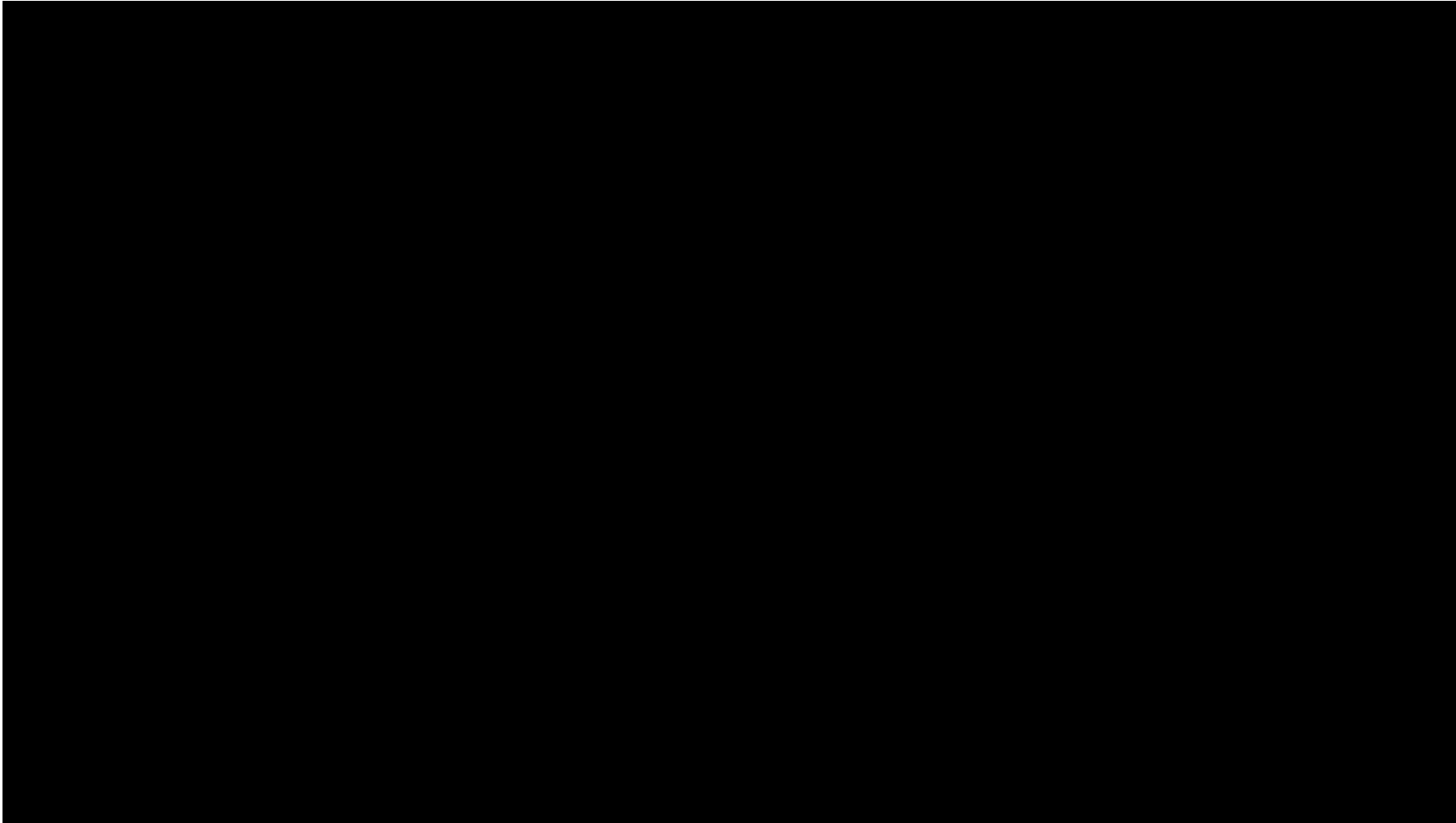
# Applications of RL

- McGill's Adaptive Gait Control for Swimming Robots (2015)

## Learning Legged Swimming Gaits from Experience

ICRA 2015 - Best Paper Award Nominee
http://www.cim.mcgill.ca/~dmeger/ICRA2015_GaitLearning/

David Meger, Juan Camilo Gamboa Higuera,
Anqi Xu, Philippe Giguere and Gregory Dudek

McGill University and Universite Laval

# Applications of RL

- Gerald Tesauro at IBM's Thomas J. Watson Research Center on TD Gammon (Backgammon game) (1992)

# Applications of RL

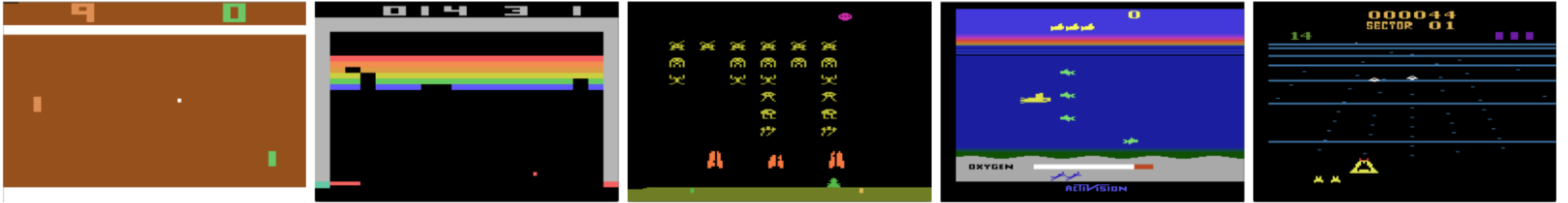- Ref: Playing Atari with Deep Reinforcement Learning (2013)



Figure 1: Screen shots from five Atari 2600 Games: (*Left-to-right*) Pong, Breakout, Space Invaders, Seaquest, Beam Rider

... first deep learning model to successfully learn control policies directly from **high-dimensional sensory input** using reinforcement learning.
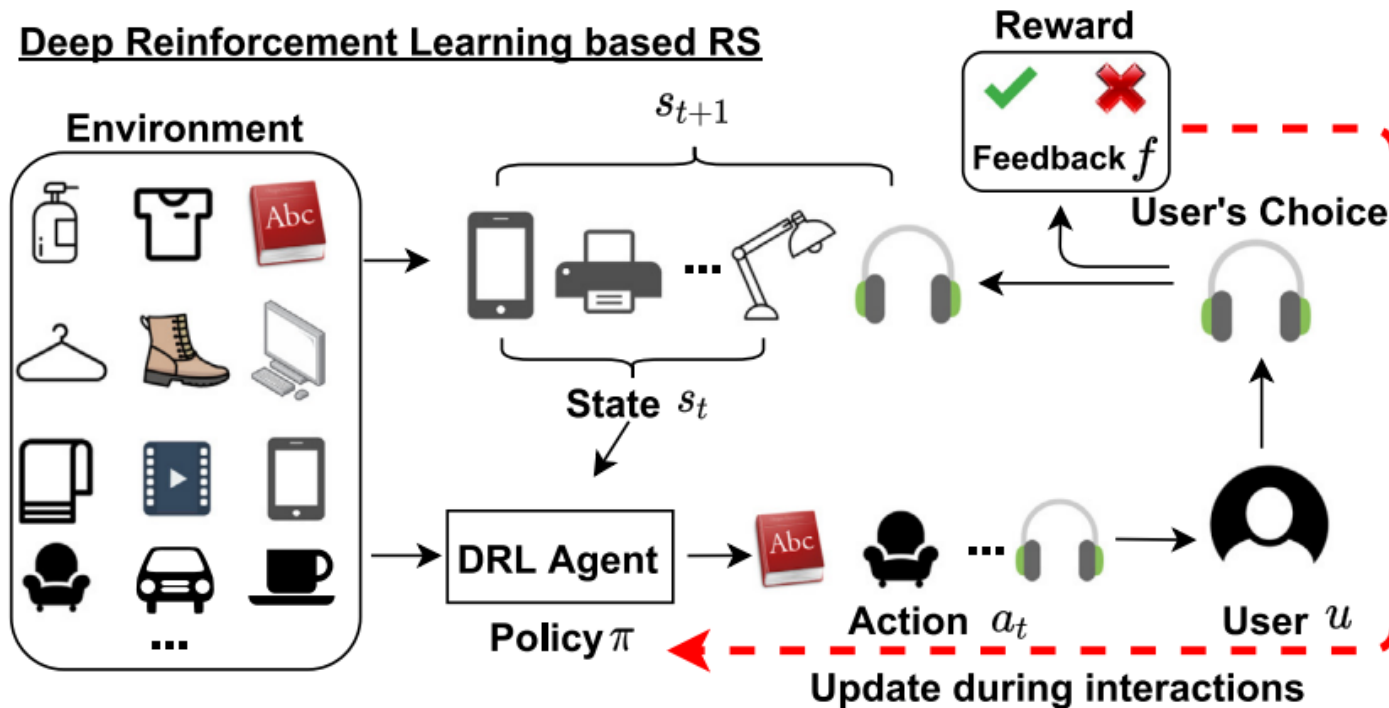
# Applications of RL

- David Silver from DeepMind worked on TD algorithm for Go game (2015)
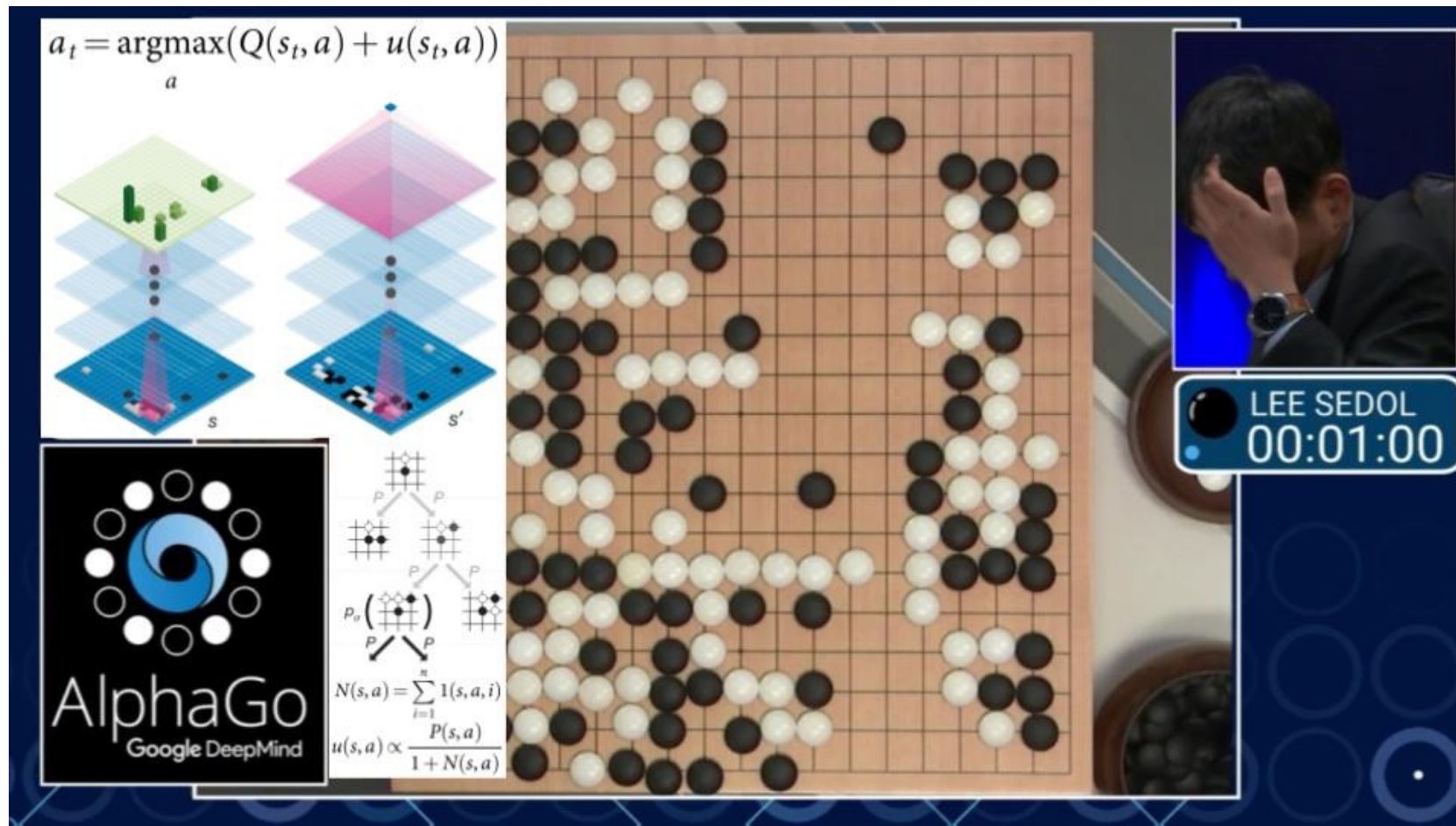
# Applications of RL

- Manage investment portfolio
- RL for online learning ( suggesting what to show on the stories / Ad )
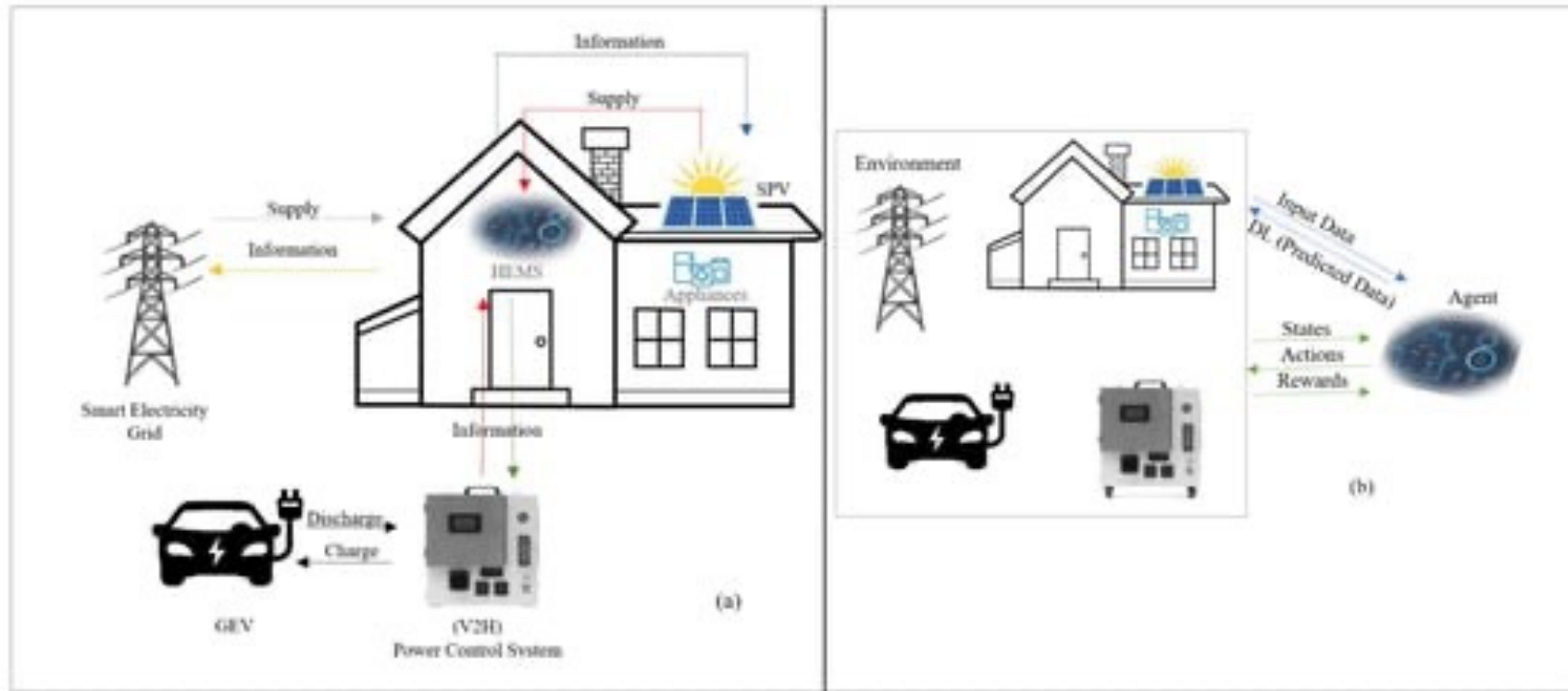


(b) Deep reinforcement learning-based recommender systems

**Source**: *Deep reinforcement learning in recommender systems: A survey and new perspectives* by Xiaocong Chen, et al.

# Applications of RL

- Control a power station



**Source**: *A Reinforcement Learning Approach for Integrating an Intelligent Home Energy Management System with a Vehicle-to-Home Unit* by Ohoud Almughram, et al.

# What are Rewards?

- A reward $R_t$ is a scalar feedback signal

- Indicates how well agent is doing at step $t$

- The agent's goal is to maximize cumulative reward / total future reward

Reward hypothesis in RL:

**All goals can be described by the maximization of expected cumulative reward.**

- What if there are no intermediate reward?
- What if the task is time sensitive?
- Can the rewards be delayed?
- Is it better to sacrifice immediate reward to gain more long-term reward?

A financial investment may take months to mature.

Refueling a helicopter might prevent a crash in several hours.

# What are Rewards?

- Fly stunt maneuvers in a helicopter
  - *+ve feedback for following desired trajectory*
  - *−ve feedback for crashing*

- Defeat the world champion at Backgammon
  - *+/−ve feedback for winning/losing a game*

- Manage an investment portfolio
  - *+ve feedback for each ₹ in bank*

- Control a power station
  - *+ve feedback for producing power*
  - *−ve feedback for exceeding safety thresholds*

- Make a humanoid robot walk
  - *+ve feedback for forward motion*
  - *−ve feedback for falling over*