Reinforcement learning in string theory

FABIAN RUEHLE (UNIVERSITY OF OXFORD)

String Pheno 2018 - Warsaw

02/07/2018

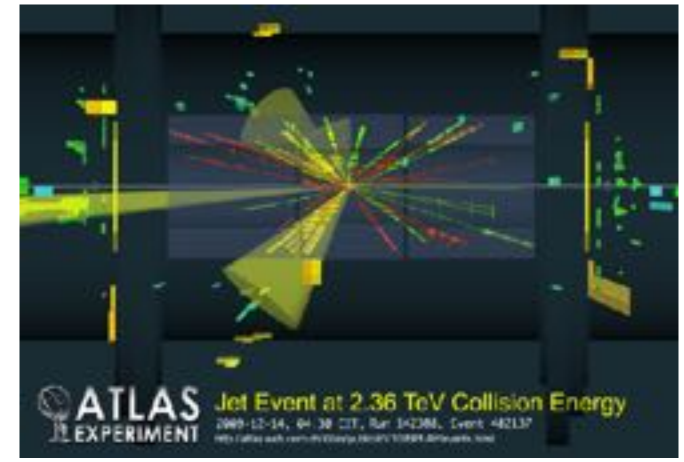
UNIVERSITY OF OXFORD

Based on:

[Brent Nelson, Jim Halverson, Fabian Ruehle]

[Jim Halverson, Hans Peter Nilles, Fabian Ruehle, Patrick Vaudrevange]

Motivation - ML in Science and Society




GOOGLE AI CAN PREDICT WHEN PEOPLE WILL DIE WITH '95 PER CENT ACCURACY'

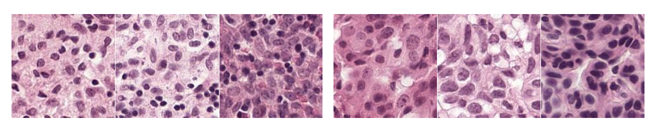


Fig. 1. Left: three tumor patches and **right**: three challenging normal patches.

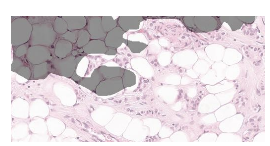


Fig. 2. Difficulty of pixel-accurate annotations for scattered tumor cells. Ground truth annotation is overlaid with a lighter shade. Note that the tumor annotations include both tumor cells and normal cells *e.g.*,white space representing adipose tissue (fat).

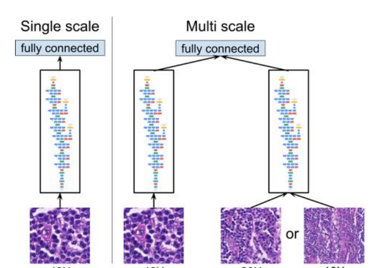
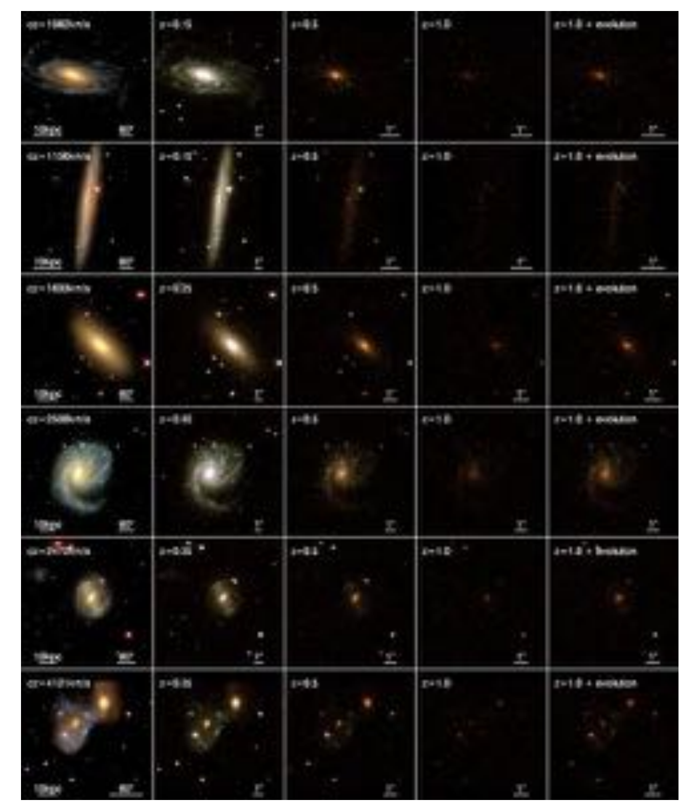
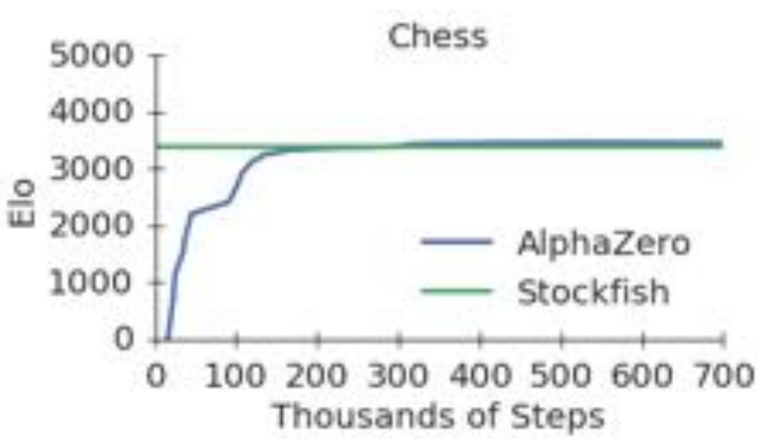


Fig. 3. The three colorful blocks represent Inception (V3) towers up to the second-last layer (PreLogit). *Single scale* utilizes one tower with input images at 40X magnification; *multi-scale* utilizes multiple (*e.g.*,2) input magnifications that are input to separate towers and merged.

[Liu et al. `17]



[Zooniverse `18; picture from Barden et al `08]



[Silver et al. `17]

Motivation - ML in String Theory

- Possible applications of ML in string theory
  - Find string models in the landscape
  - Find generic / common features of string-derived model and extract string theory predictions from the landscape [Patrick's talk] [Gary's talk]
  - Find patterns in mathematics of string theory [Jim's talk] [Sven's talk]
  - Use machine learning / AI to perform computation intensive work [FR'17]
  - …
- Can we use machine learning to study the landscape?
  [He'17; Krefl, Seong'17; FR'17; Carifio, Halverson, Krioukov, Nelson'17; Wang, Zhang `18; Hashimoto, Sugishita, Tanaka, Tomiya `18]

Motivation - ML in String Theory

- Possible applications of ML in string theory
  - Find string models in the landscape
  - Find generic / common features of string-derived model and extract string theory predictions from the landscape [Patrick's talk] [Gary's talk]
  - Find patterns in mathematics of string theory [Jim's talk] [Sven's talk]
  - Use machine learning / AI to perform computation intensive work [FR'17]
  - …
- Can we use machine learning to study the landscape?
  [He'17; Krefl, Seong'17; FR'17; Carifio, Halverson, Krioukov, Nelson'17; Wang, Zhang `18; Hashimoto, Sugishita, Tanaka, Tomiya `18]

Motivation - ML in String Theory

4D string theories highly non-unique

- Different choices lead to 10^{500} to 10^{755} or more string vacua (Go has 10^{177} states)
  [Douglas `03; Douglas, Sen `04; Halverson, Long, Sung `17; Taylor, Wang `15-`17]
- Number huge but seems finite
  [Reid `87; Douglas, Taylor `07; Buchbinder, Constantin, Lukas `14; Groot Nibbelink, Loukas, FR, Vaudrevange `15; Di Cerbo, Svaldi `16]
- Most of these vacua do not correspond to our universe
- Problem: We know the phenomenological properties a string theory that describes our universe has to have, but we lack a vacuum selection mechanism

Motivation - ML in String Theory

When choosing a string background (geometry, flux):

- Need to ensure mathematical/physical consistency
  - Tadpole and anomaly cancellation
  - Solution is actual vacuum (D- and F-flat)
- Need to ensure physically desirable features
  - Gauge algebra of the SM: $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)_Y$
  - Three families of quarks and leptons, one Higgs pair
  - Absence of exotics, realistic Yukawas
  - Realistic cosmological constant

Motivation - ML in String Theory

When choosing a string background (geometry, flux):

- Need to ensure mathematical/physical consistency
  - Tadpole and anomaly cancellation
  - Solution is actual vacuum (D- and F-flat)
- Need to ensure physically desirable features
  - Gauge algebra of the SM: $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)_Y$
  - Three families of quarks and leptons, one Higgs pair
  - Absence of exotics, realistic Yukawas
  - Realistic cosmological constant

Motivation - ML in String Theory

When choosing a string background (geometry, flux):

- Need to ensure mathematical/physical consistency
  - Tadpole and anomaly cancellation
  - Solution is actual vacuum (D- and F-flat)
- Need to ensure physically desirable features
  - Gauge algebra of the SM: $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)_Y$
  - Three families of quarks and leptons, one Higgs pair
  - Absence of exotics, realistic Yukawas
  - Realistic cosmological constant [Jim's talk]

Motivation - ML in String Theory

- Mathematical constraints: Often collection of non-linear, coupled Diophantic equations
- Physical constraints: Further constrains Diophantic solutions in non-obvious way
- Upshot:
  - For a given configuration we can check its viability easily, but we have no idea how to find a good configuration in the first place
- To traverse vacua: Use Reinforcement Learning, a semi-supervised approach to Machine Learning

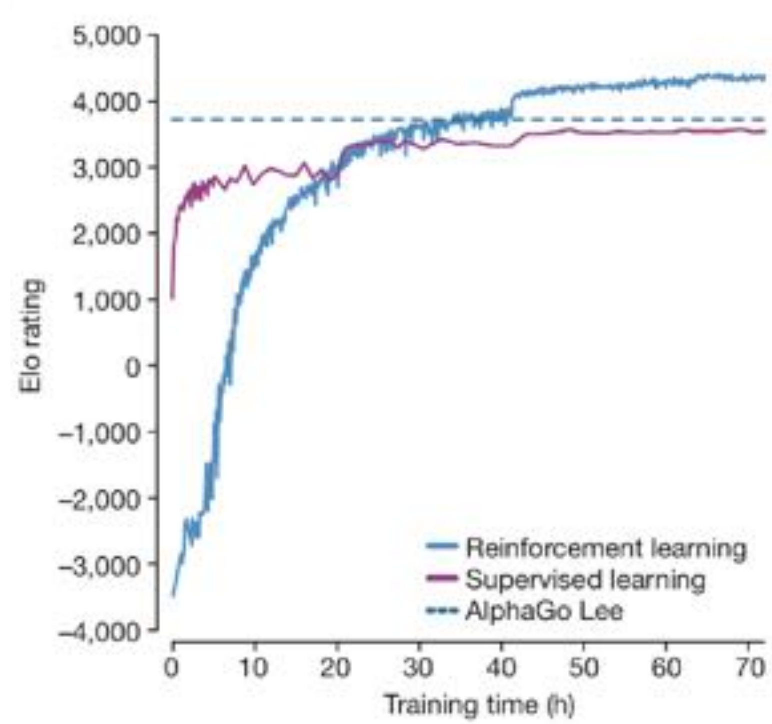
Outline

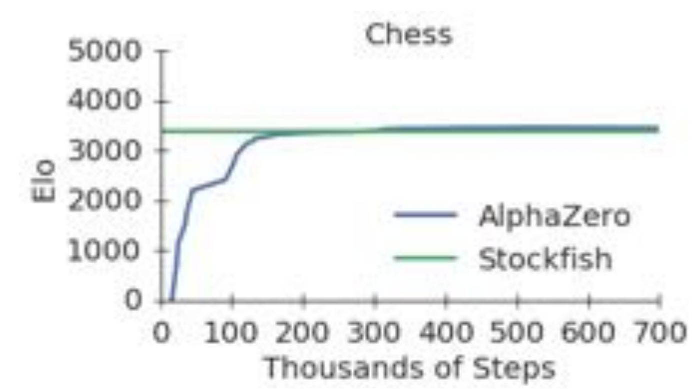
- Reinforcement Learning (RL)
  - Introduction to RL
  - Introduction to NNs + Tree searches
  - Implementation
- Example applications
  - Finding vacua in Type IIA/B intersecting brane models
  - Finding vacua in Heterotic $E_8 \times E_8$
- Conclusion

Reinforcement learning

Reinforcement Learning - Idea

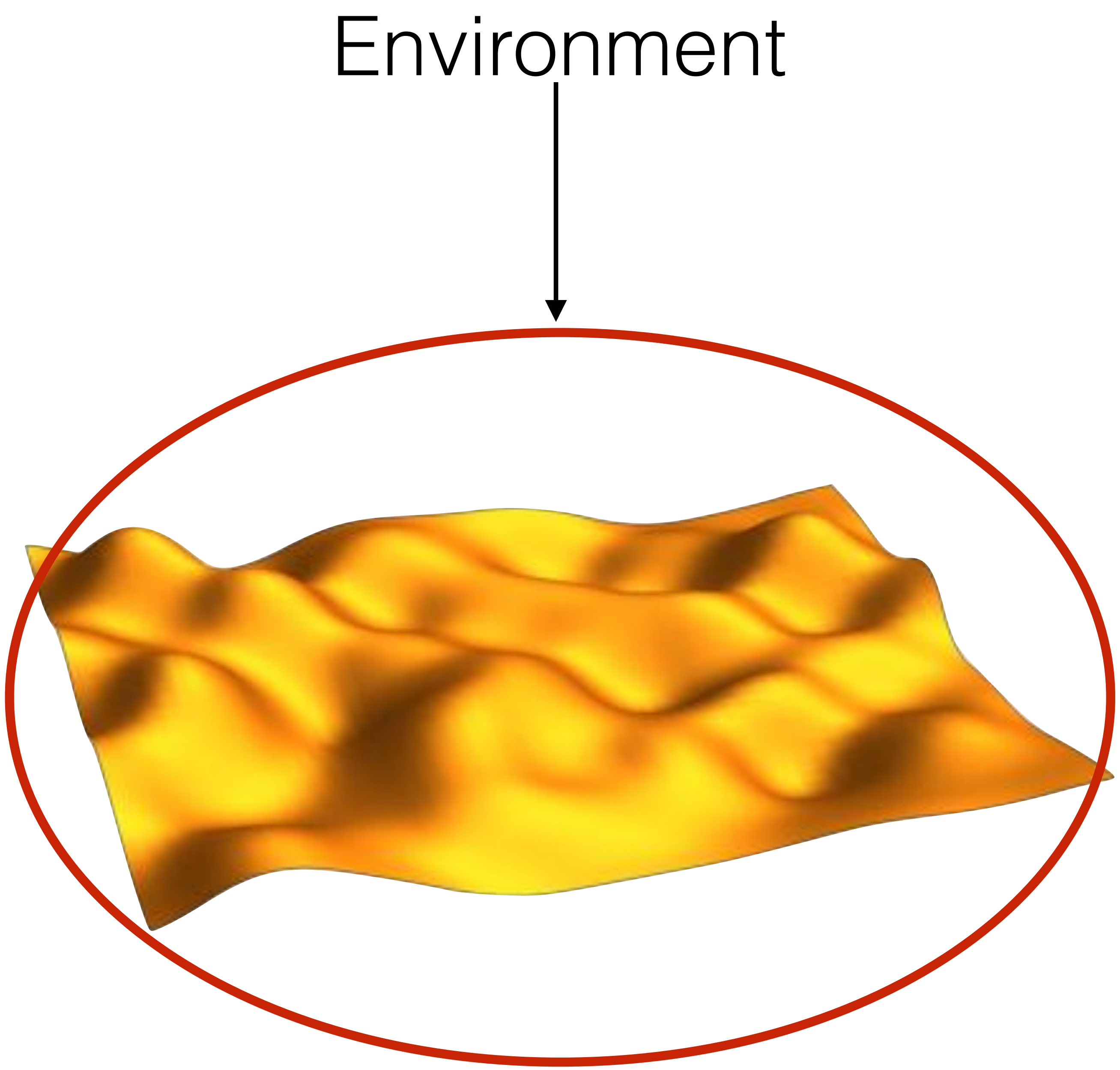
- Basic textbooks/literature [Barton, Sutton ’98 ‘17]
- Based on behavioural psychology: train individual by
  - Rewarding “good” behavior
  - Punishing “bad” behavior
- Used e.g. in Go (Note: Go has 10^{177} states) [Silver et. al. ’16 ‘17]





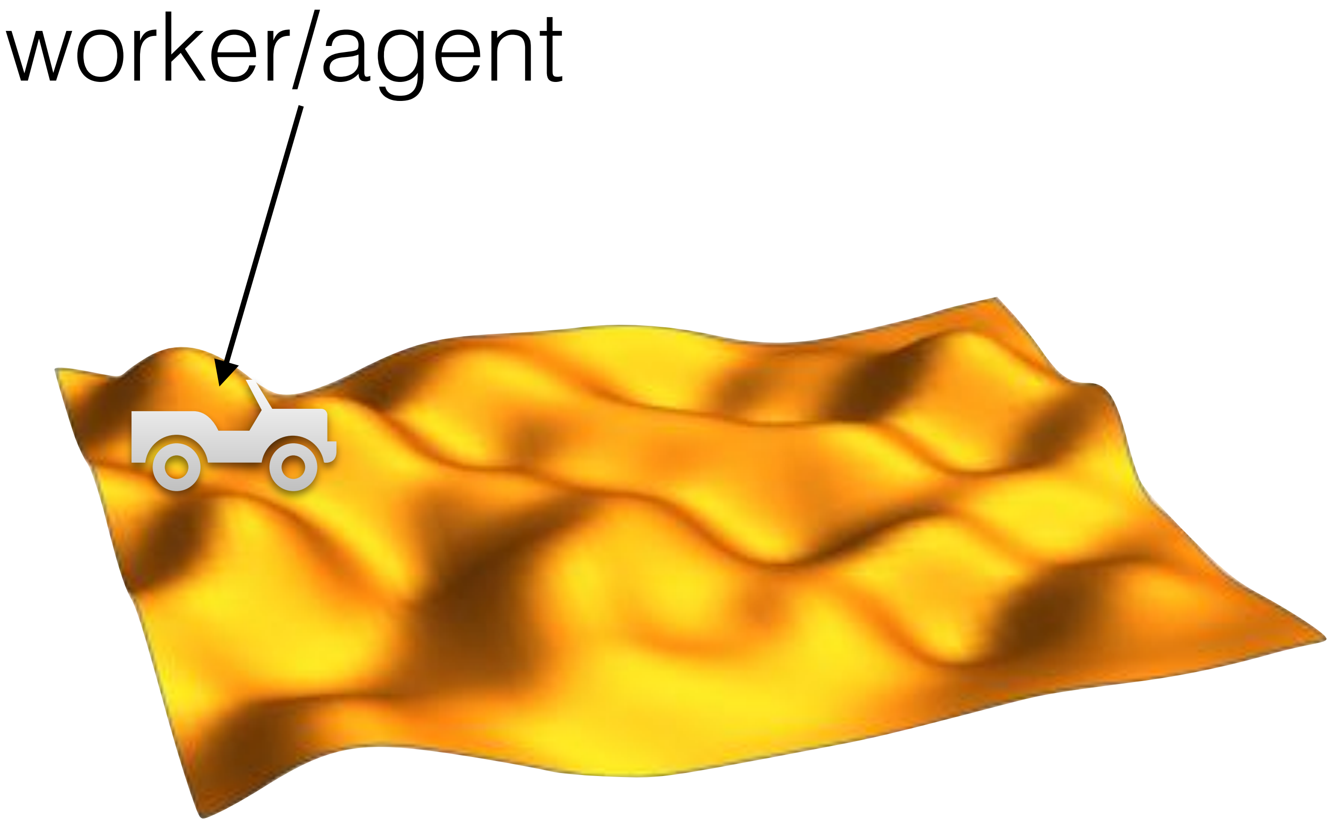
Reinforcement Learning - Vocabulary

- Want to explore the string landscape ("environment")



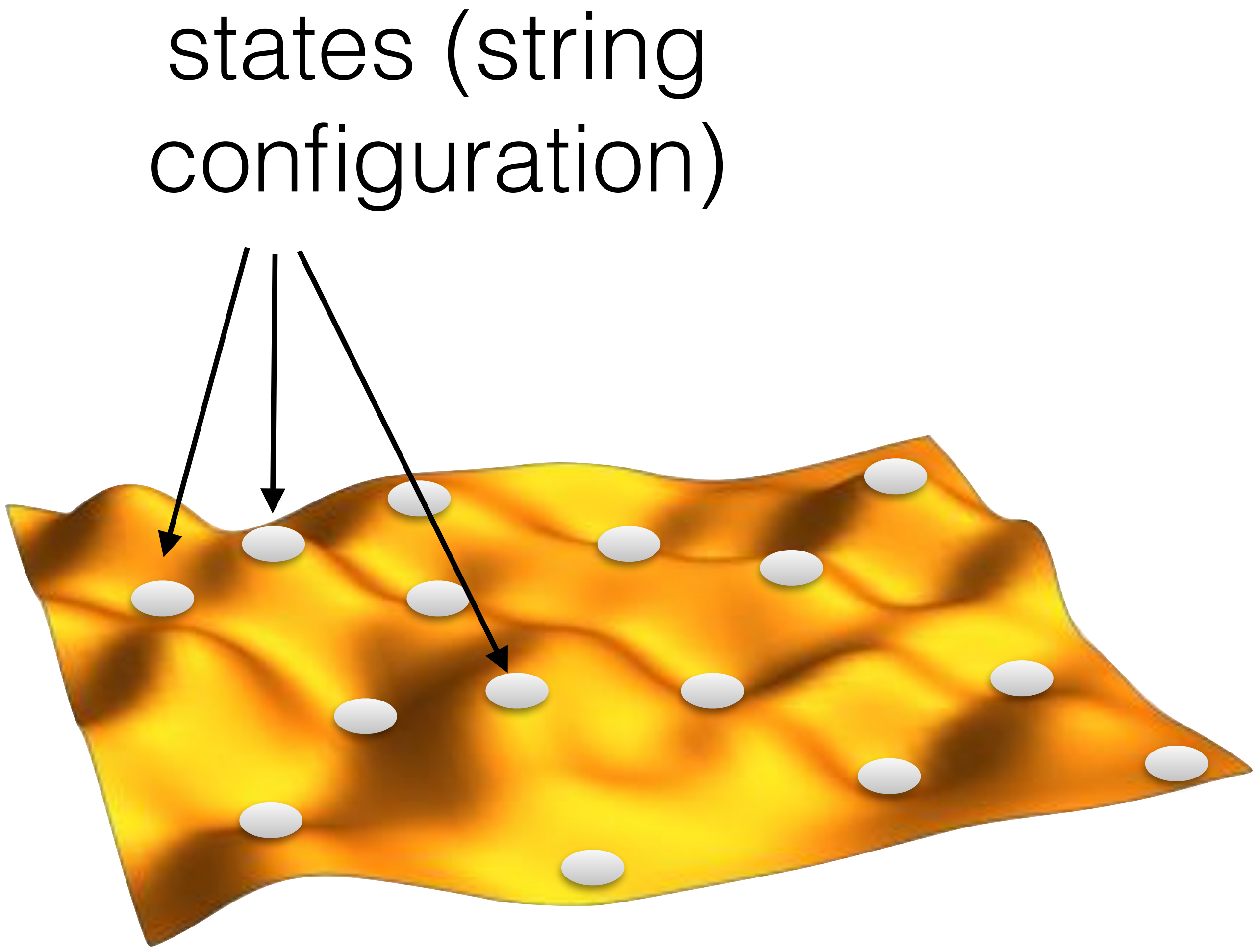
Reinforcement Learning - Vocabulary

- Want to explore the string landscape ("environment")
- Done by "workers" that are conditioned



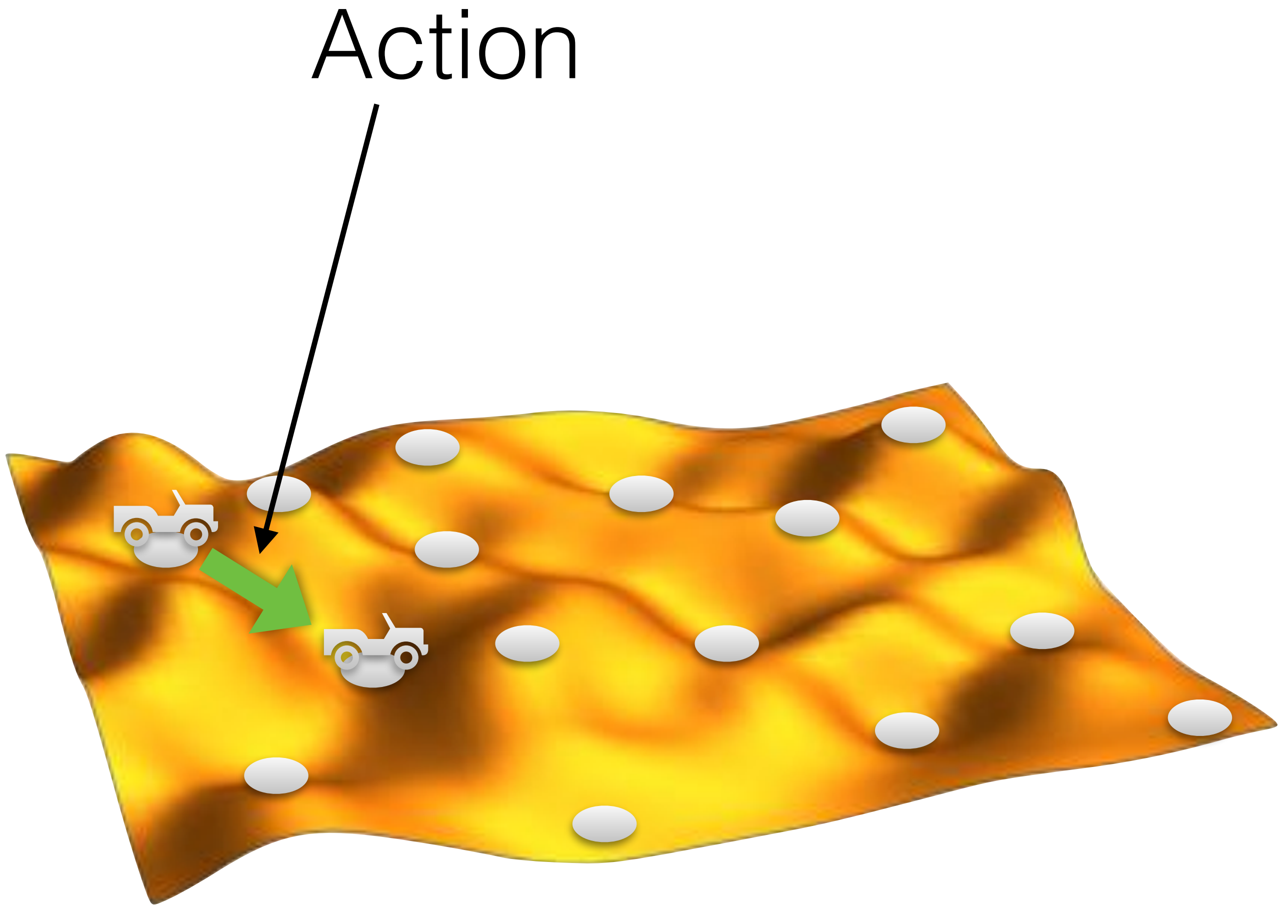
Reinforcement Learning - Vocabulary

- Want to explore the string landscape (“environment”)
- Done by “workers” that are conditioned
- At any given moment, a worker is in a specific string configuration (“state”) defined by discrete topological data (branes, flux, cycles, …)



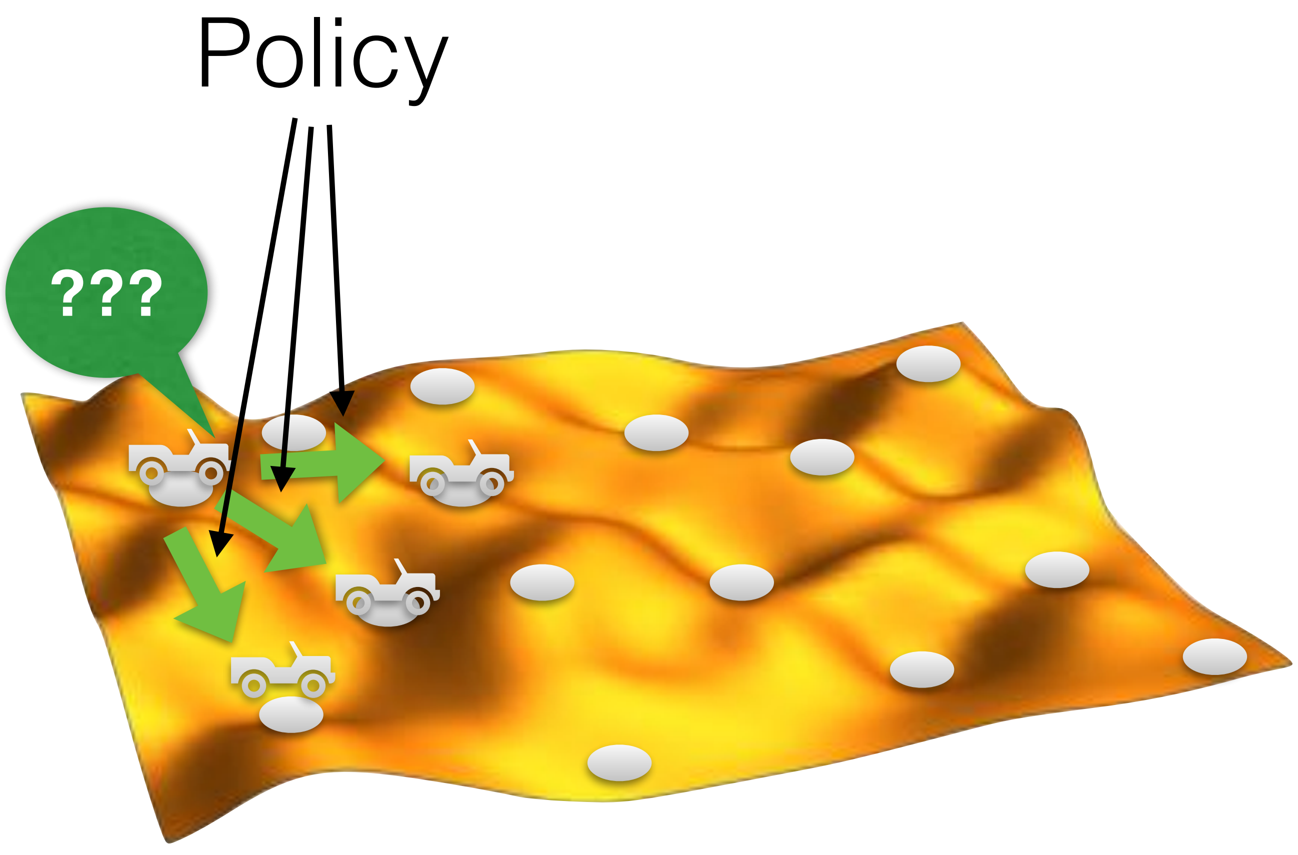
Reinforcement Learning - Vocabulary

- Want to explore the string landscape ("environment")
- Done by "workers" that are conditioned
- At any given moment, a worker is in a specific string configuration ("state") defined by discrete topological data (branes, flux, cycles, …)
- Workers change state by taking "actions" to reach new states ("elements of the environment")



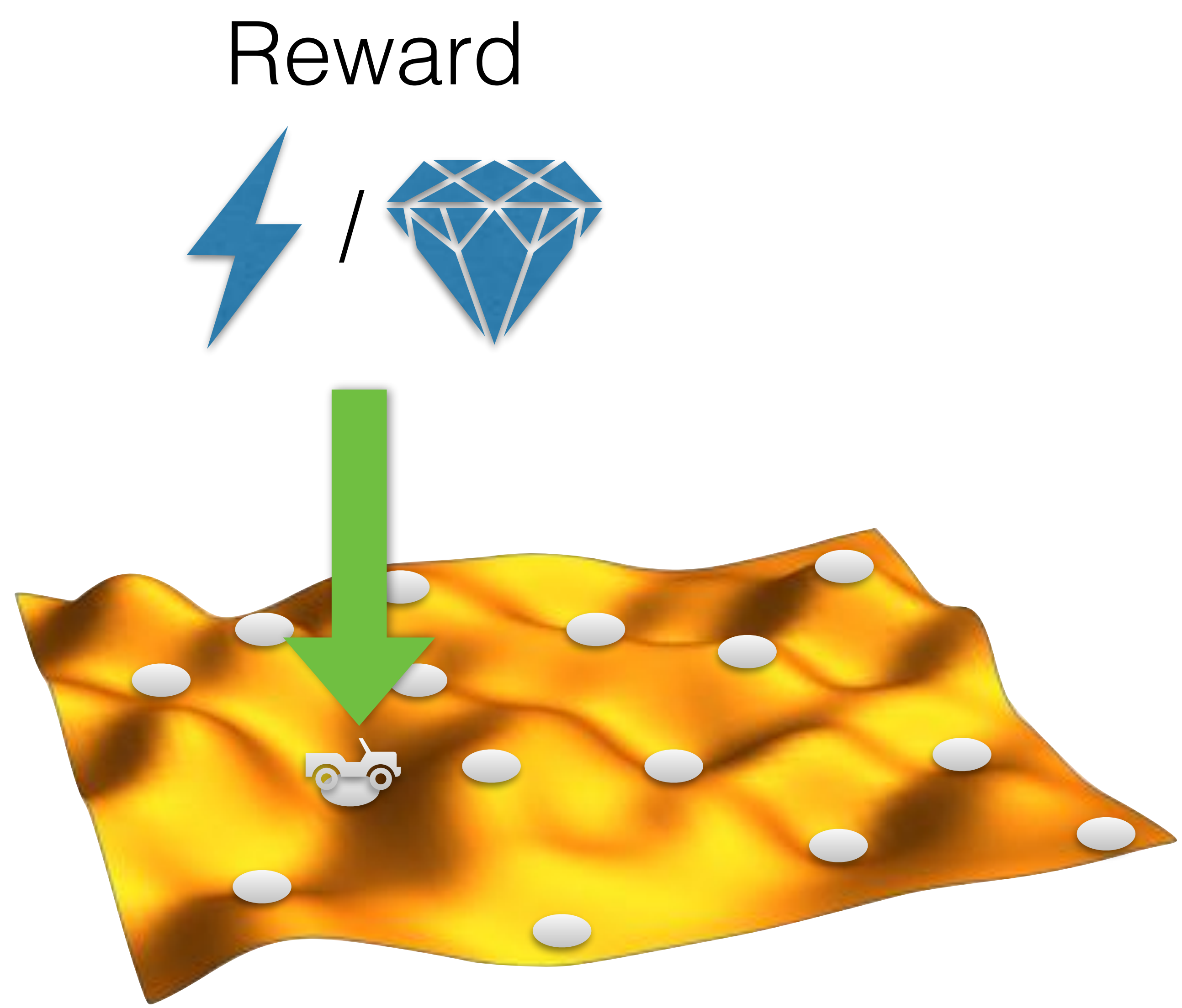
Reinforcement Learning - Vocabulary

- Want to explore the string landscape ("environment")
- Done by "workers" that are conditioned
- At any given moment, a worker is in a specific string configuration ("state") defined by discrete topological data (branes, flux, cycles, …)
- Workers change state by taking "actions" to reach new states ("elements of the environment")
- They select these actions via some "policy"



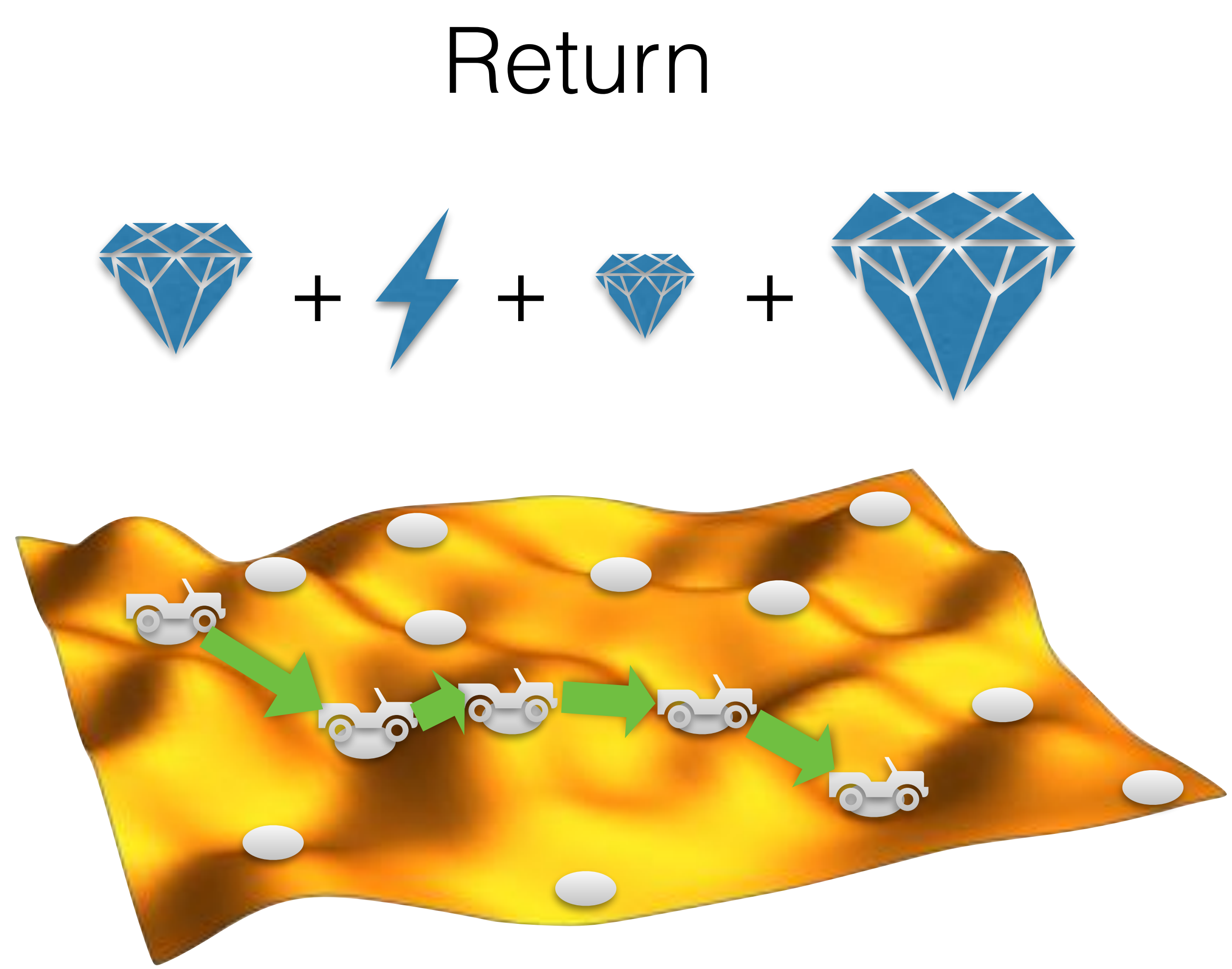
Reinforcement Learning - Vocabulary

- Want to explore the string landscape (“environment”)
- Done by “workers” that are conditioned
- At any given moment, a worker is in a specific string configuration (“state”) defined by discrete topological data (branes, flux, cycles, …)
- Workers change state by taking “actions” to reach new states (“elements of the environment”)
- They select these actions via some “policy”
- Depending on the chosen action they receive a pos/neg “reward”



Reinforcement Learning - Vocabulary

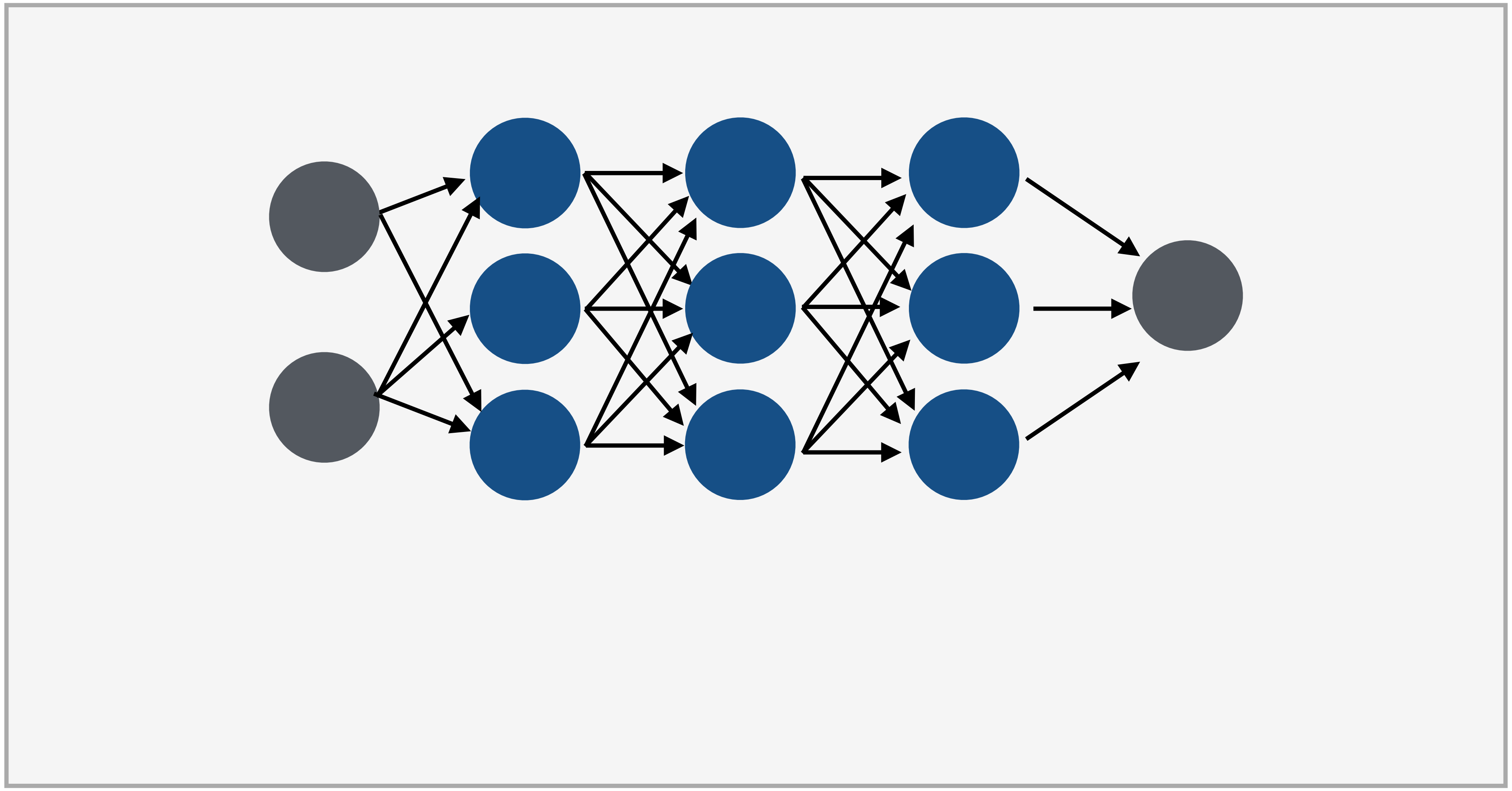
- Want to explore the string landscape ("environment")
- Done by "workers" that are conditioned
- At any given moment, a worker is in a specific string configuration ("state") defined by discrete topological data (branes, flux, cycles, …)
- Workers change state by taking "actions" to reach new states ("elements of the environ-ment")
- They select these actions via some "policy"
- Depending on the chosen action they receive a pos/neg "reward"
- Via this reinforcement, the agent learns a policy that, given a state, selects an action that maximises its "return" (accumulated long-term reward)



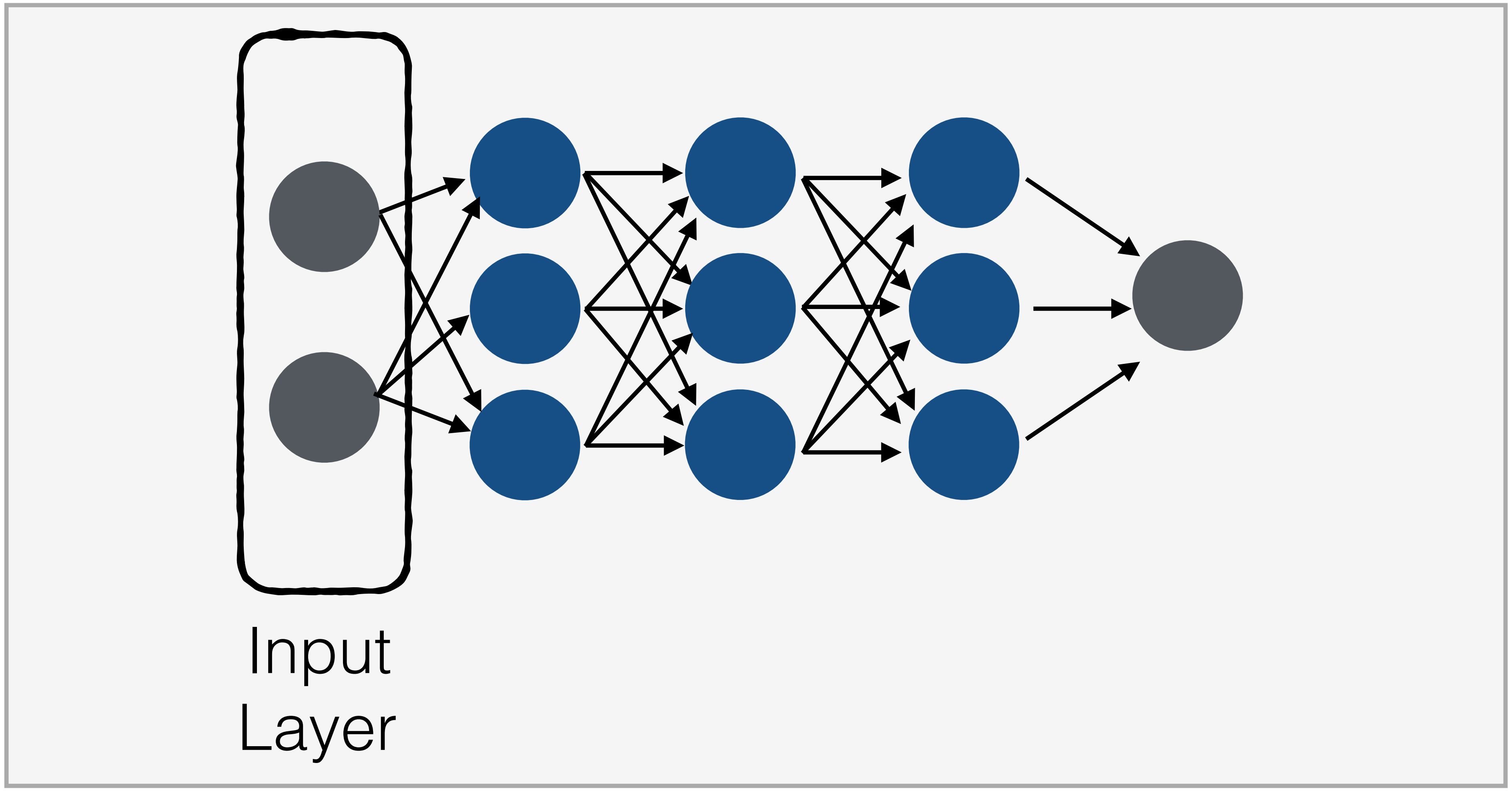
Reinforcement Learning - Prediction Problem

- In order to maximize long-term return, we need to predict:
    1. how beneficial is a given state
    2. how high will the reward of future actions be
- In order to predict this, we use neural networks that learn to make good predictions based on previous experience

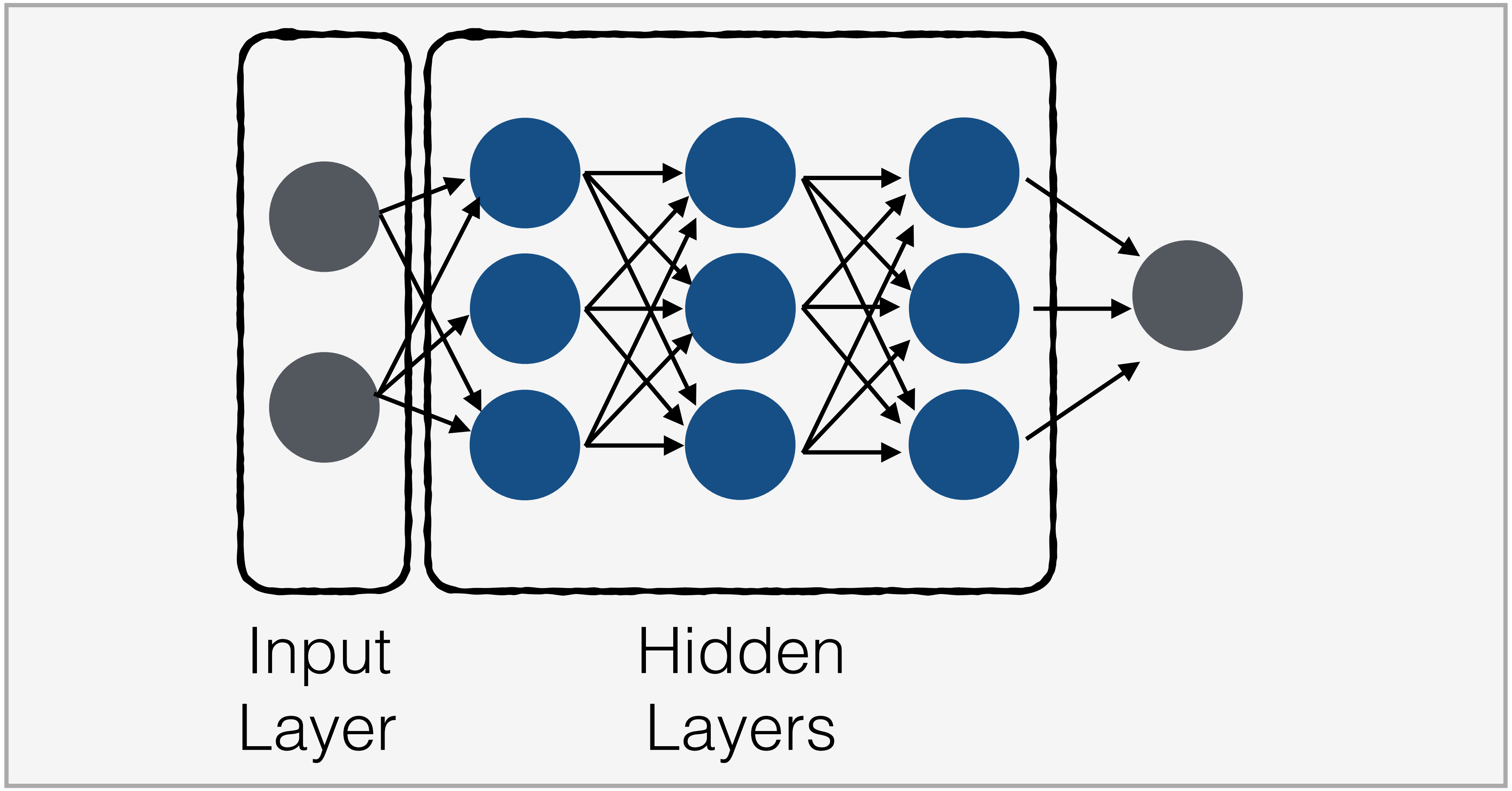
Neural Networks 101



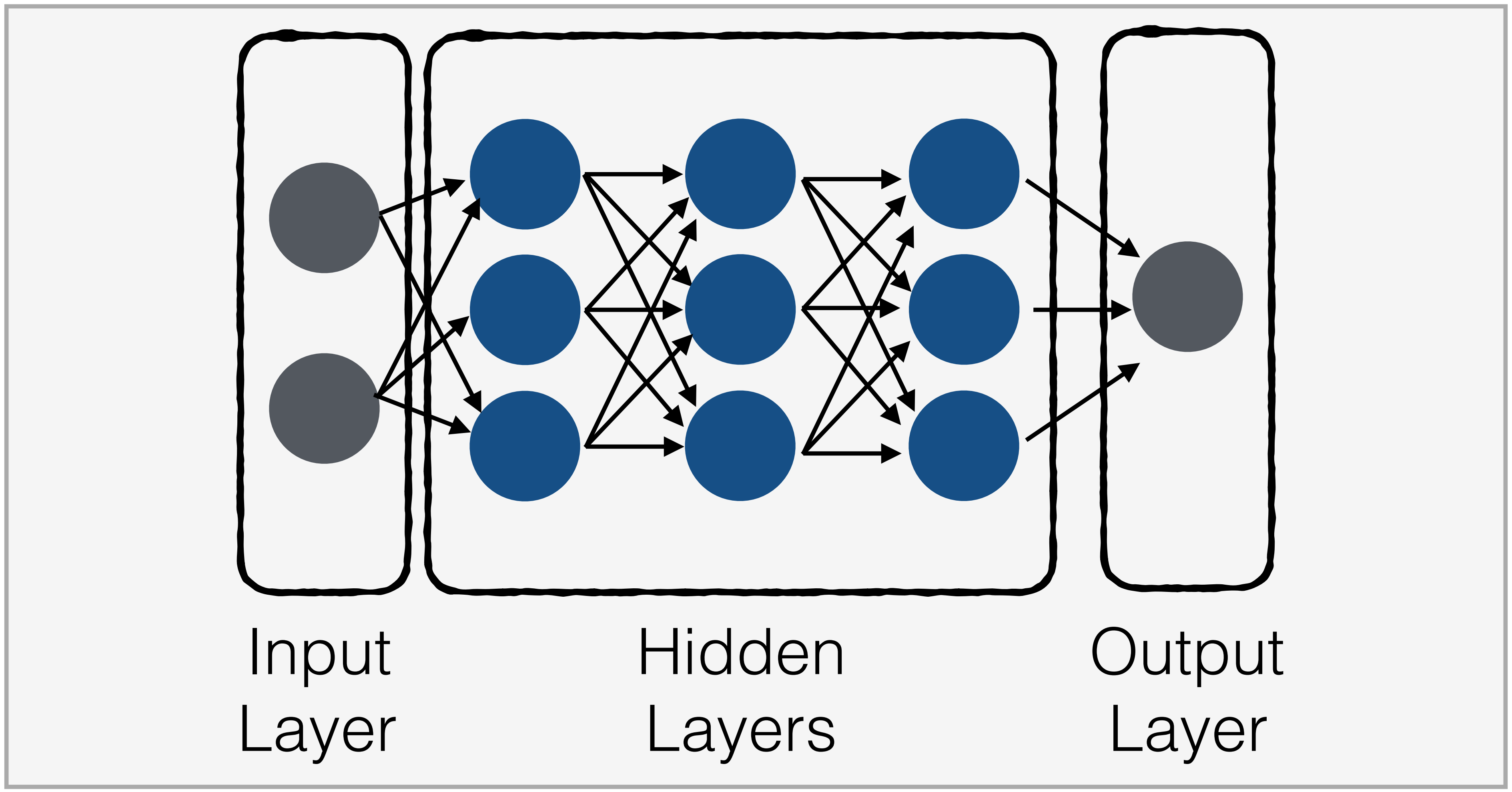
Neural Networks 101



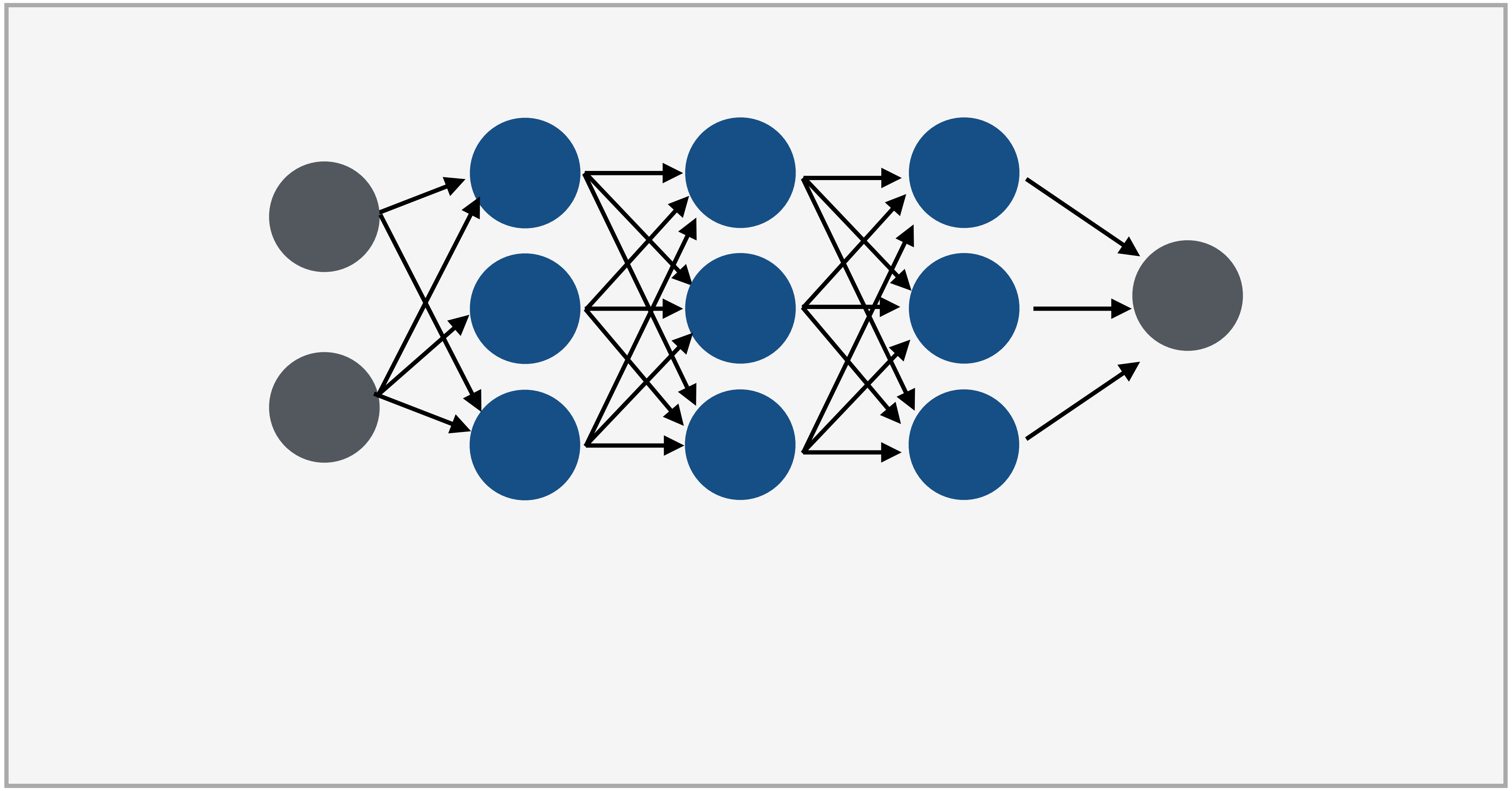
Neural Networks 101



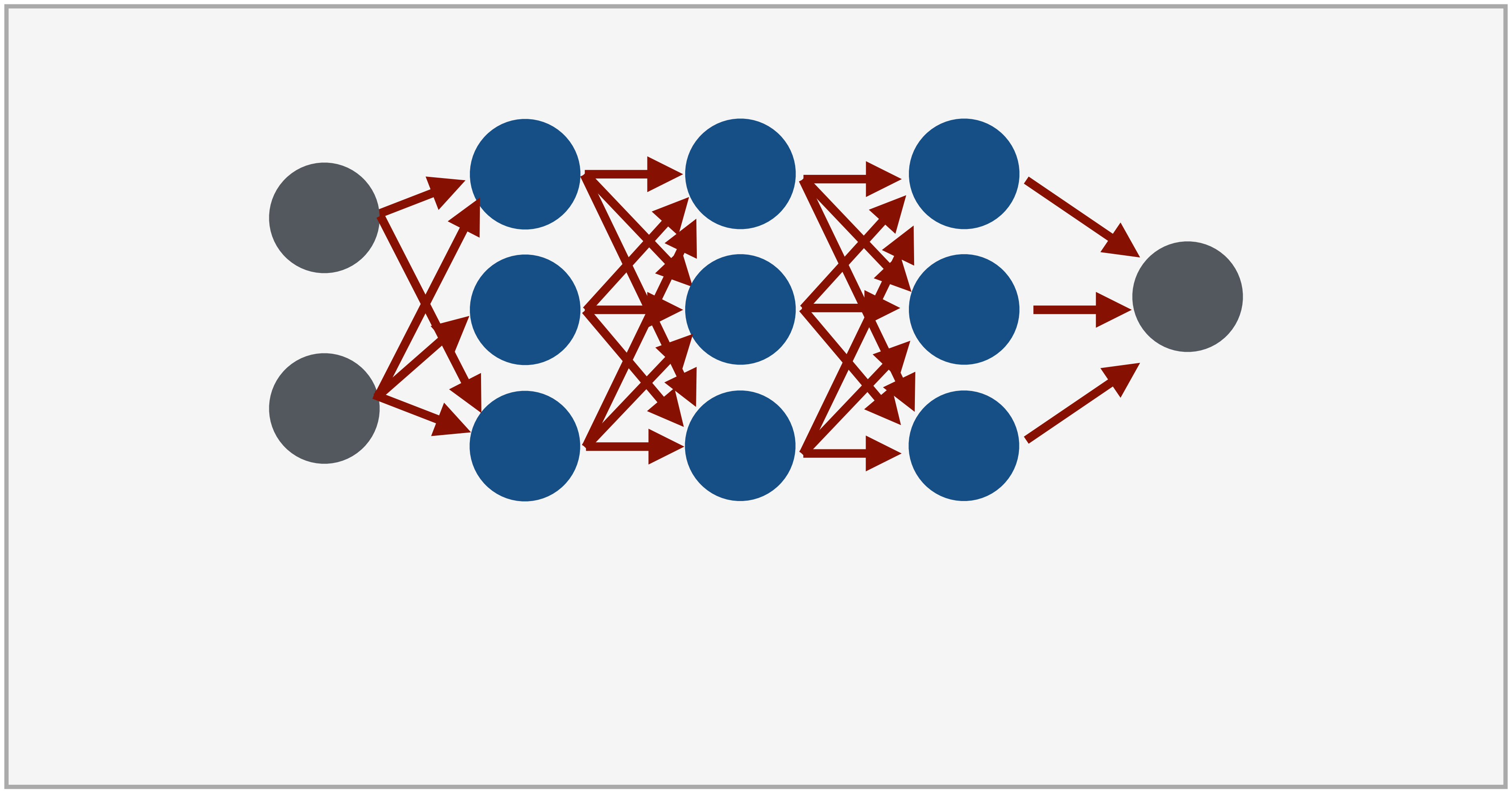
Neural Networks 101



Neural Networks 101

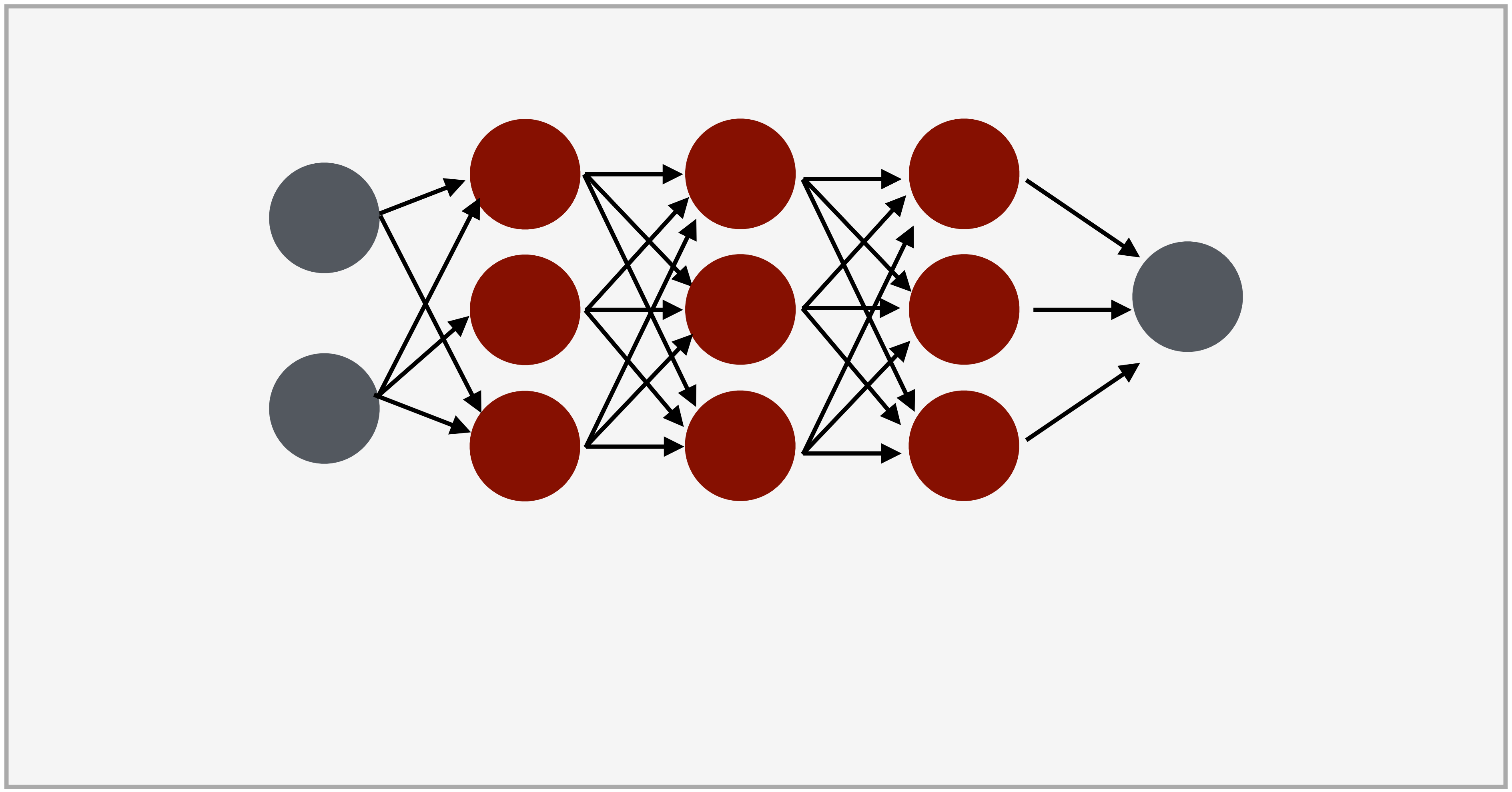


Neural Networks 101



- **Connections**: Matrix Multiplication
- Nodes: Apply some activation function f

Neural Networks 101

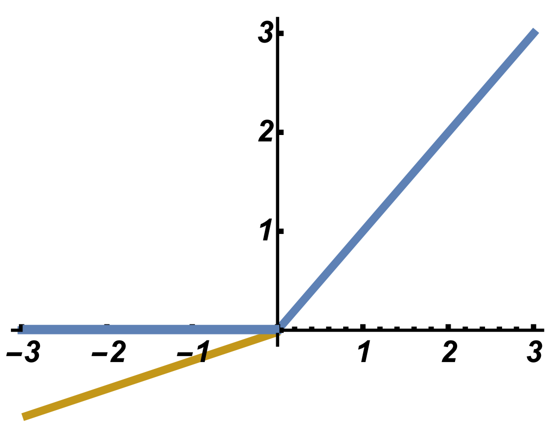


- Connections: Matrix Multiplication
- **Nodes**: Apply some activation function f

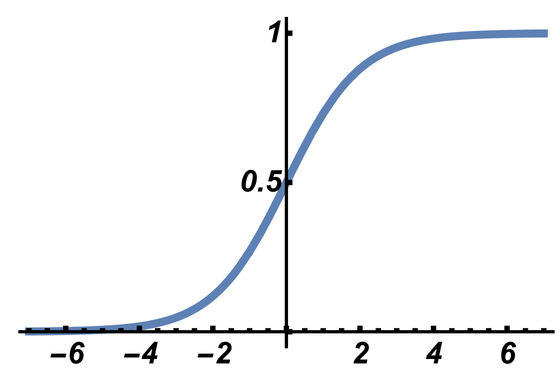
Neural Networks 101

- Connection between layers : Linear transformations L_i:
  Matrix multiplication $v^i_{\mathrm{out}} = A^i v^i_{\mathrm{in}} + b^i$
- Each layer applies a function (activation function) to its input to compute its output. Common choices are

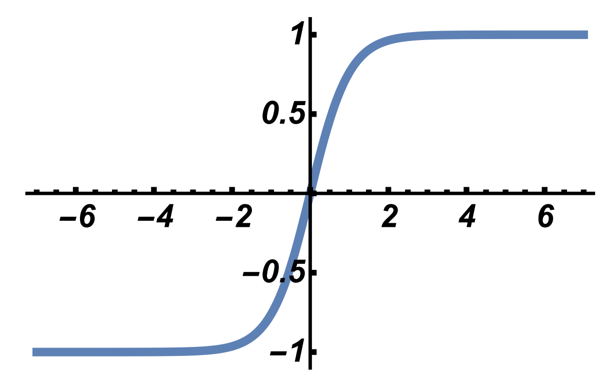
(leaky) ReLu



Logistic Sigmoid



Tanh

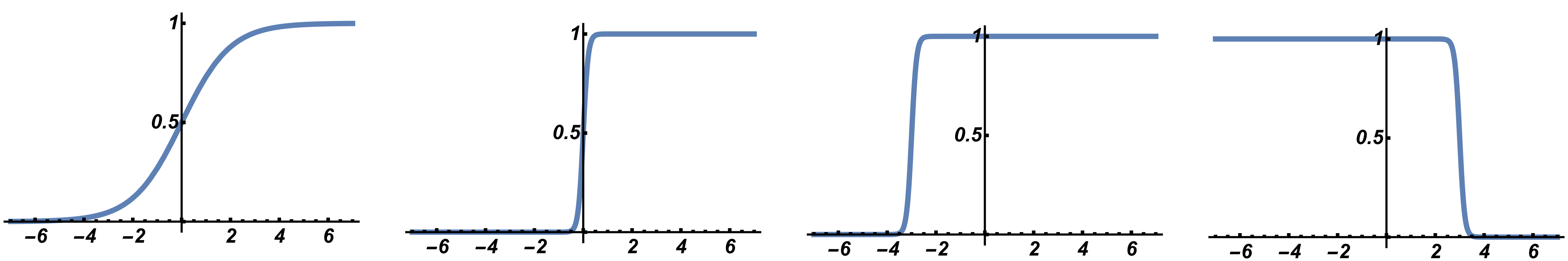


- Typical NN: $\mathbb{R}^M \to \mathbb{R}^N$

$$v \mapsto f_n \circ L_n \circ \ldots \circ f_0 \circ L_0$$

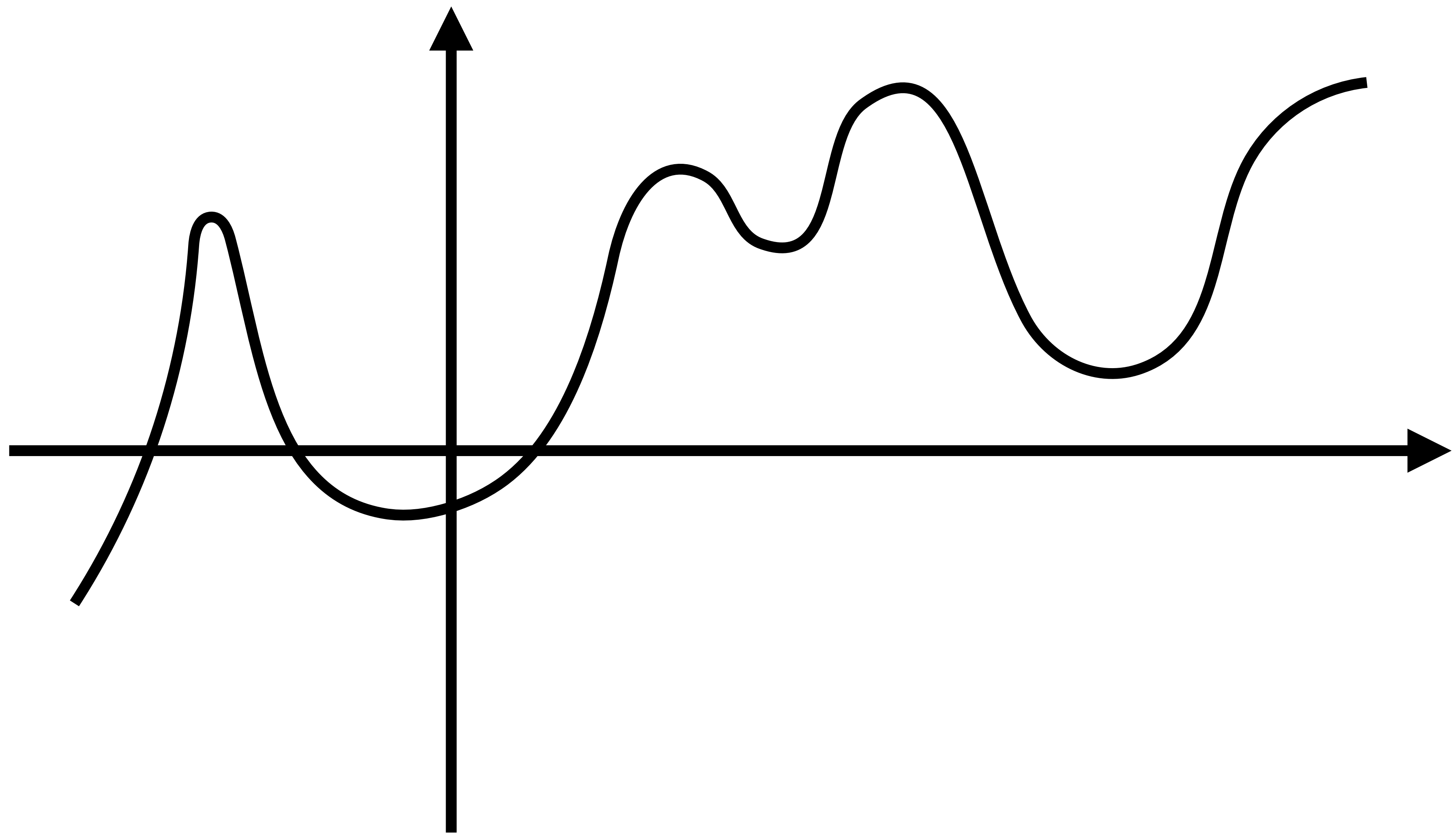
Neural Networks 101

- Look at simplest case: 1 layer, 1 node, logistic sigma function $x_{\text{out}} = (1 + \exp(a x_{\text{in}} + b))^{-1}$
  - a : Steepness of step (step function for $a \to \infty$)
  - b : Position of step: (intersects y-axis at $y = 1/2$ for $b = 0$)

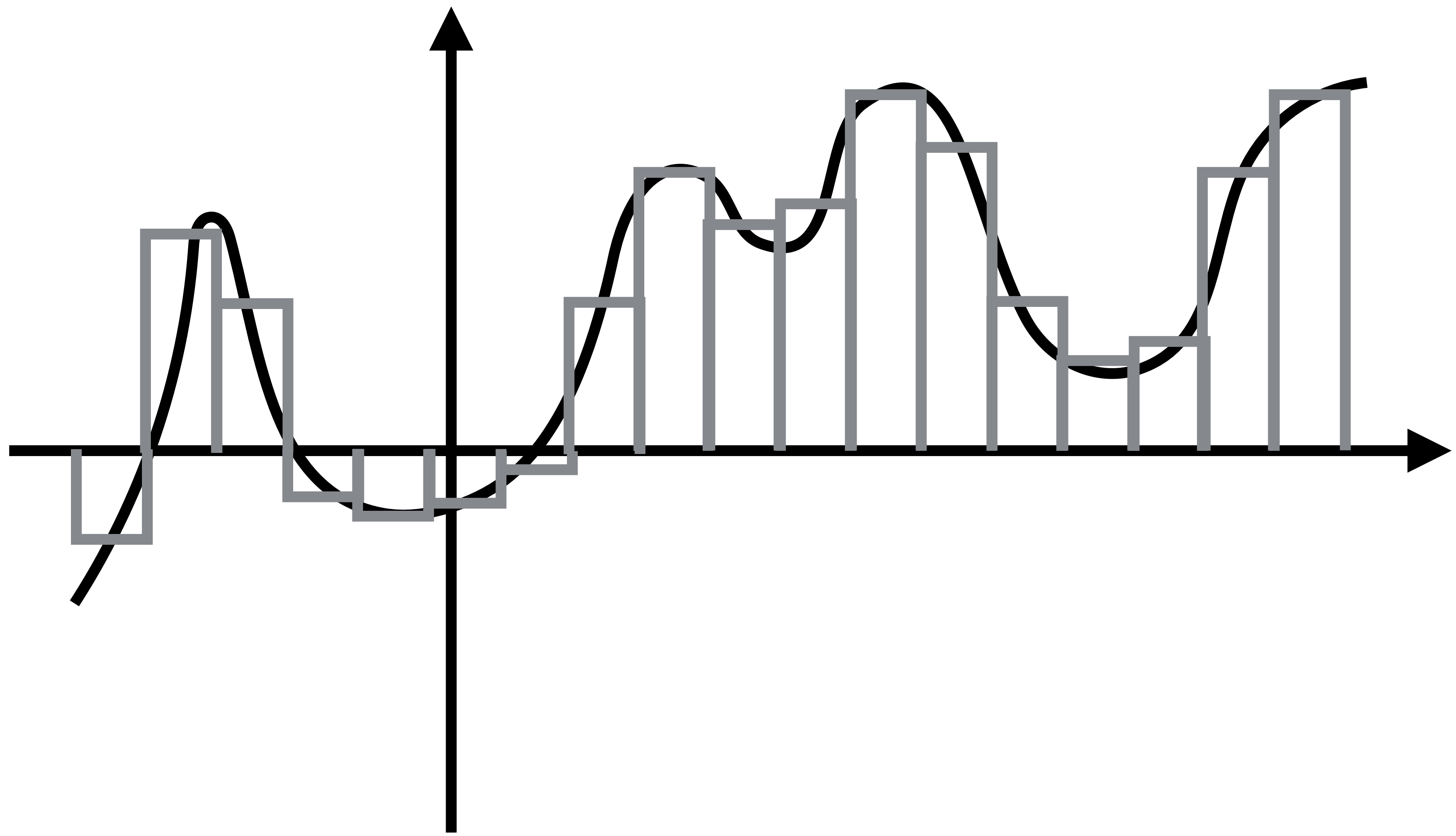


$a = 1, b = 0$ $\quad$ $a = 10, b = 0$ $\quad$ $a = 10, b = -30$ $\quad$ $a = -10, b = 30$

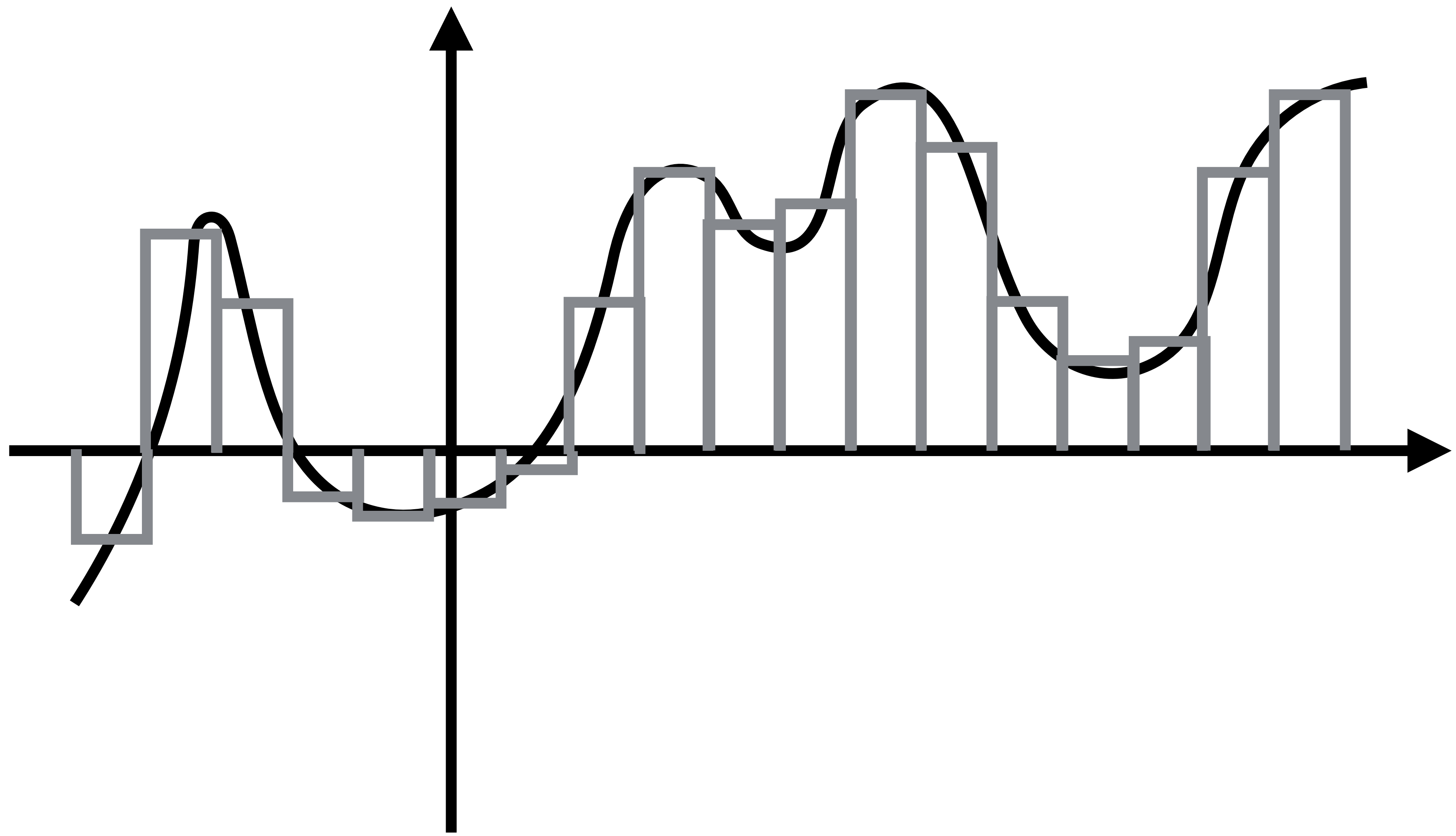
(B) Using NN to approximate functions



(B) Using NN to approximate functions



(B) Using NN to approximate functions



- More nodes ⇒ more steps ⇒ approximate any function (with one layer) [Cybenko ’89; Hornik ’91; Nielsen‘15]

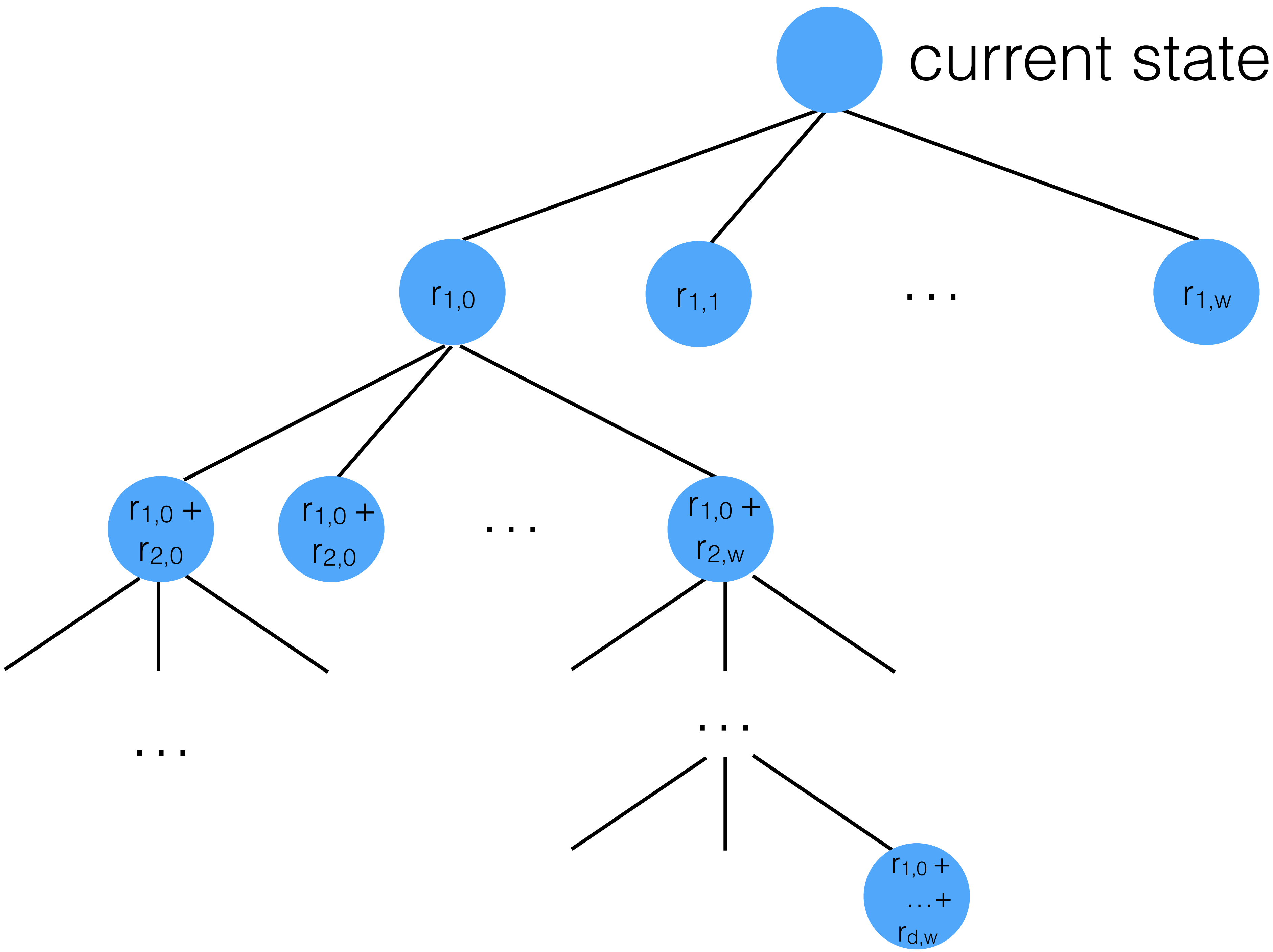
Reinforcement Learning - Details

- Commonly used policies:
  - Greedy: Choose the action that maximizes the action value function: $\pi'(s) = \text{argmax } q(s,a)$
  - Draw next action from probability distribution $\pi'(s) = \text{argmax}[\log(q(s,a)) + \text{gumbel}(q(s,a))]$
  - Perform tree search
- We use ChainerRL implementation of A3C [Mnih et al '16] (Asynchronous advantage actor-critic) possibly combined with tree search

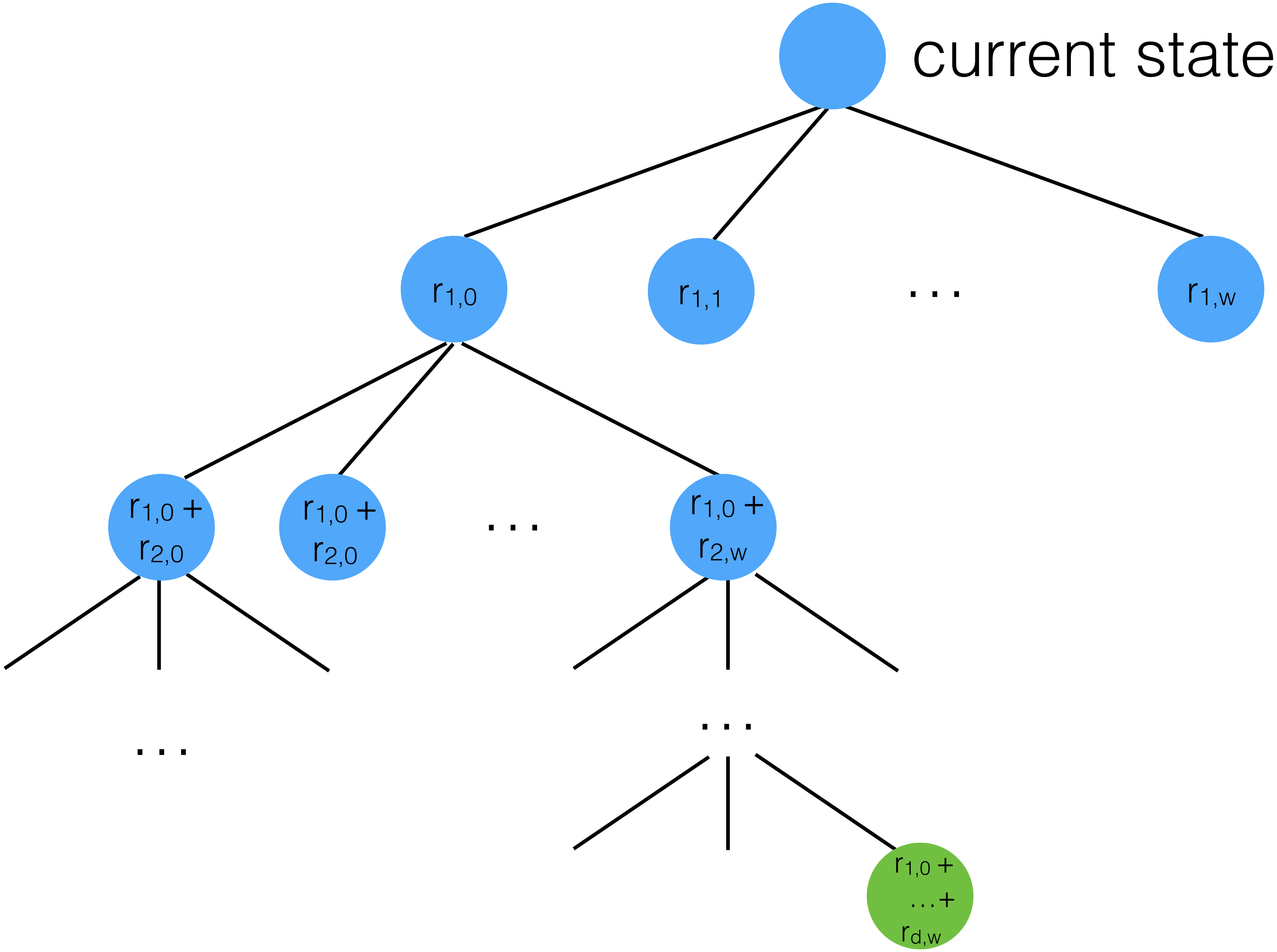
Reinforcement Learning - A3C

- **Asynchronous**: Have *n* workers explore the environment simultaneously and asynchronously
  - improves training stability (experience of workers separated)
  - improves exploration
- **Advantage**: Use advantage to update policy
- **Actor-critic**: To maximize return need to know state or action value **and** optimize policy (use neural network for estimate).
  - Actor-critic
    - “critic”: update action value
    - “actor”: update policy based on action value estimate (i.e. on the critic)

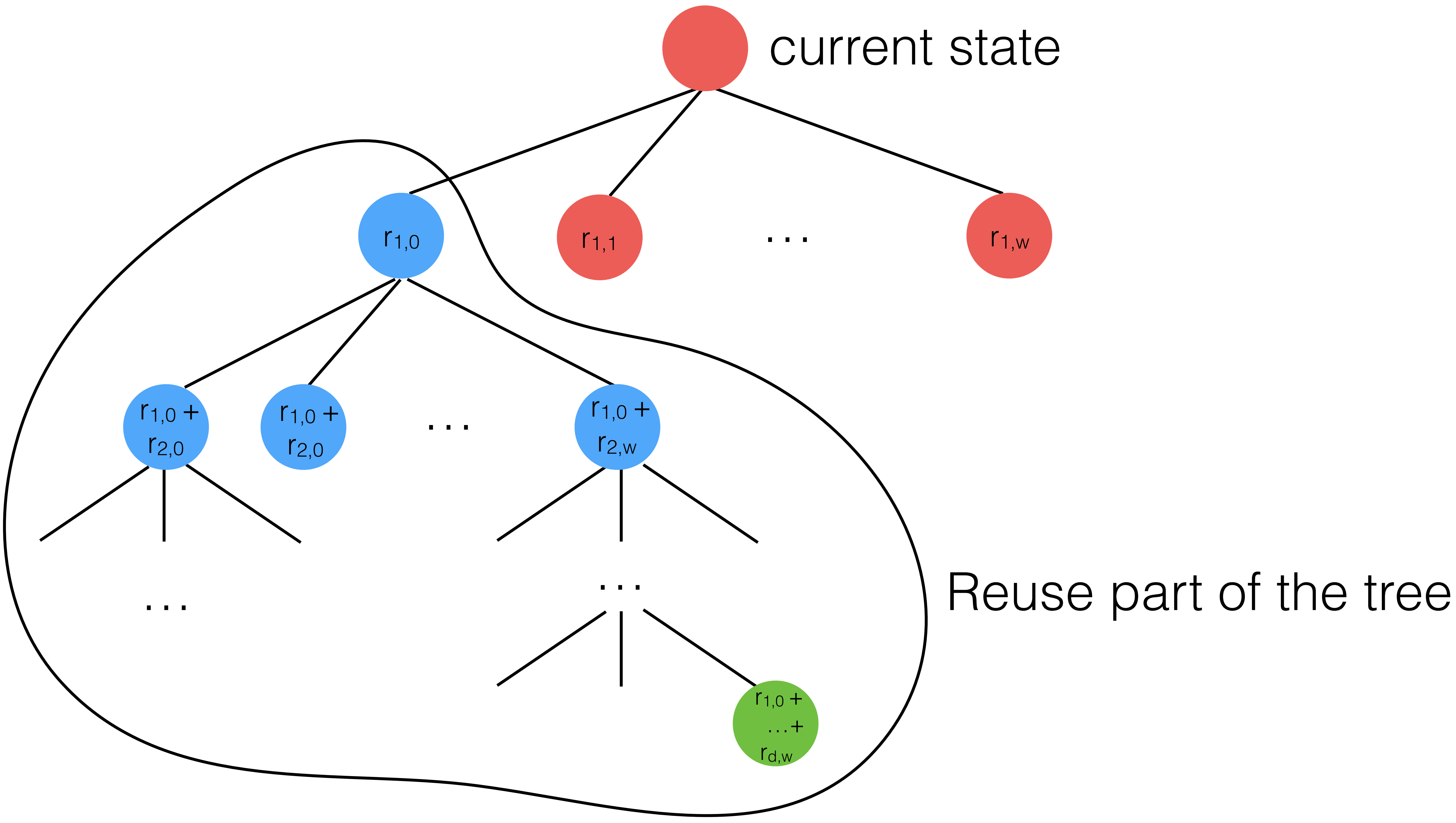
Reinforcement Learning - Tree search



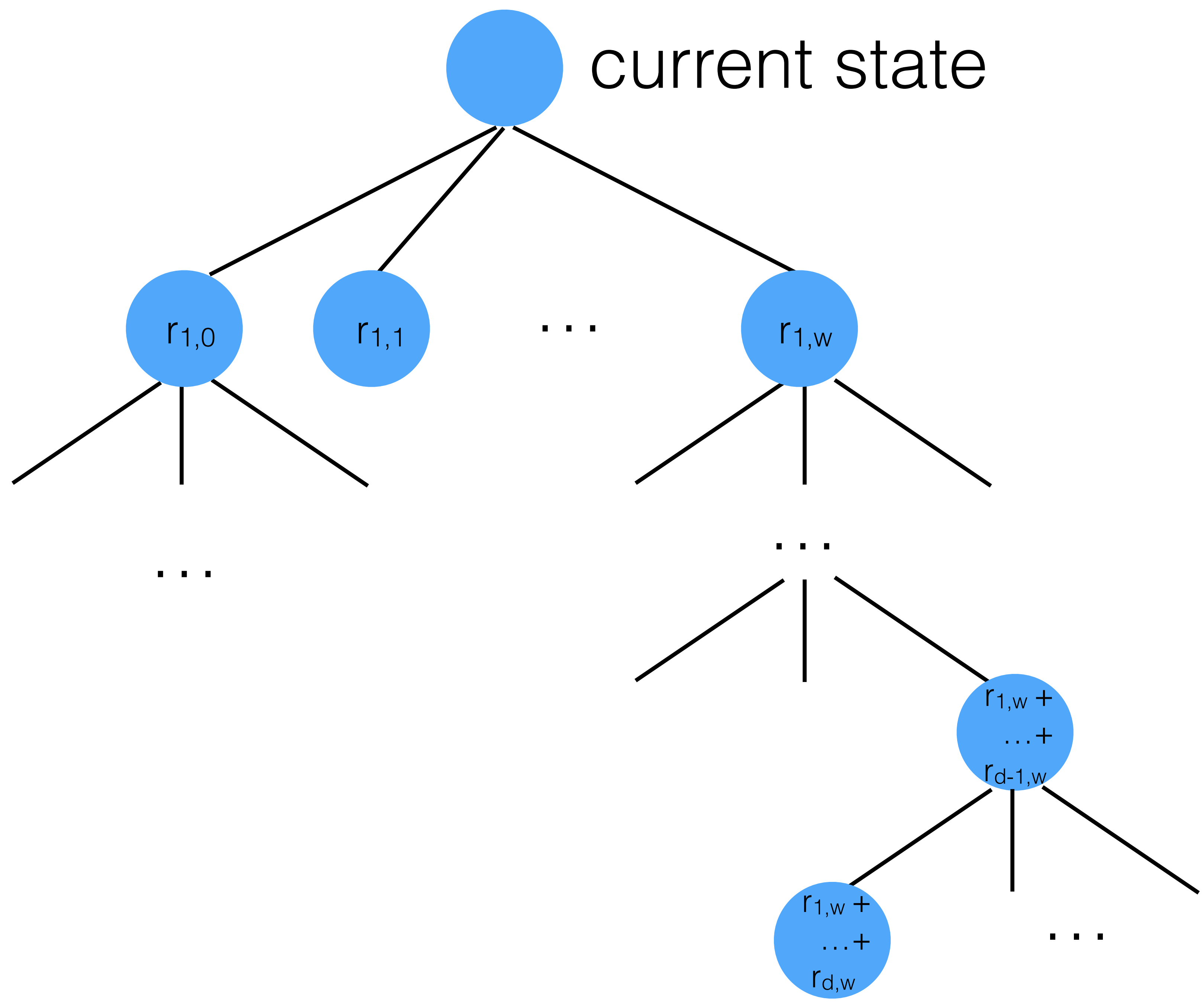
Reinforcement Learning - Tree search



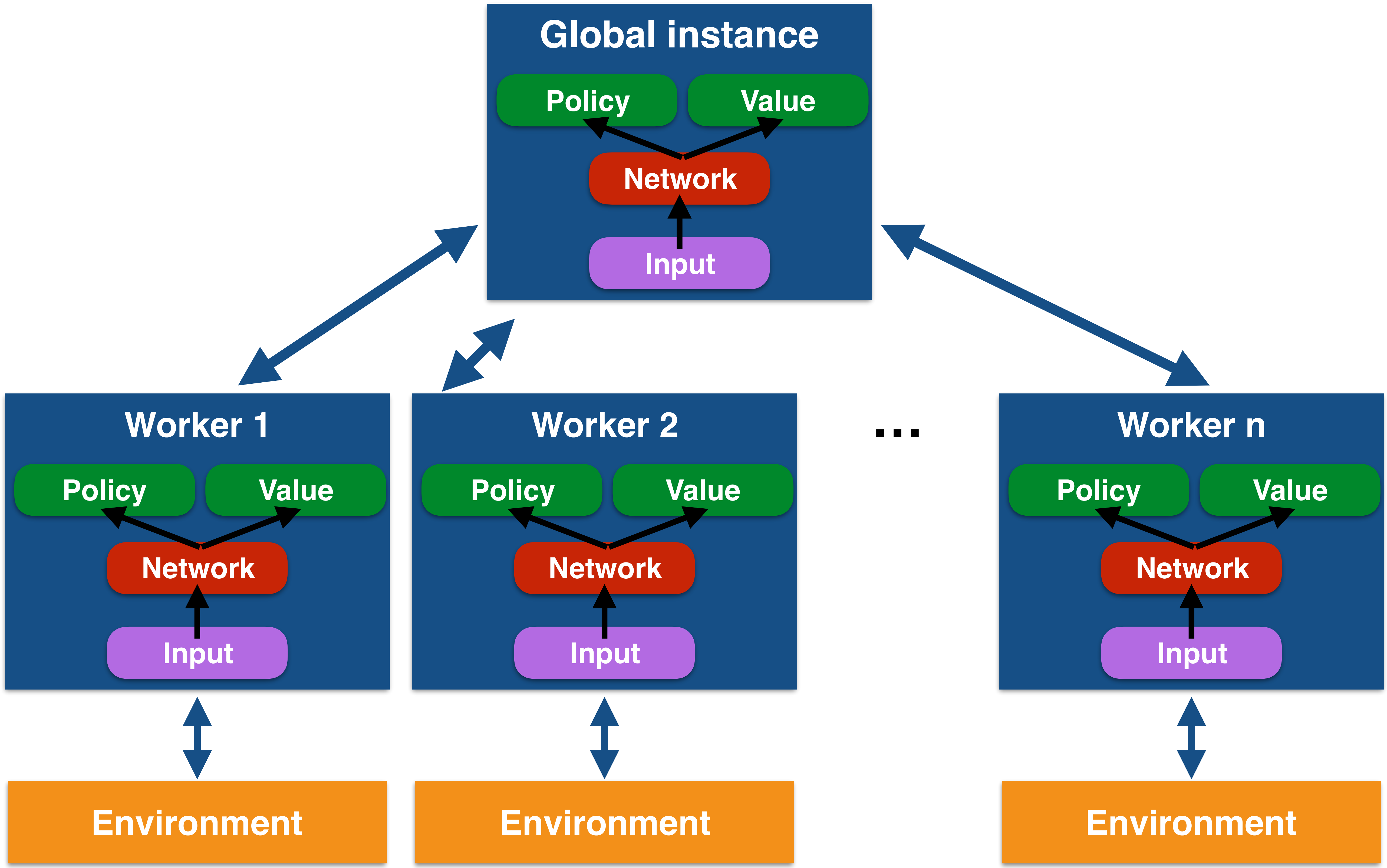
Reinforcement Learning - Tree search



Reinforcement Learning - Tree search

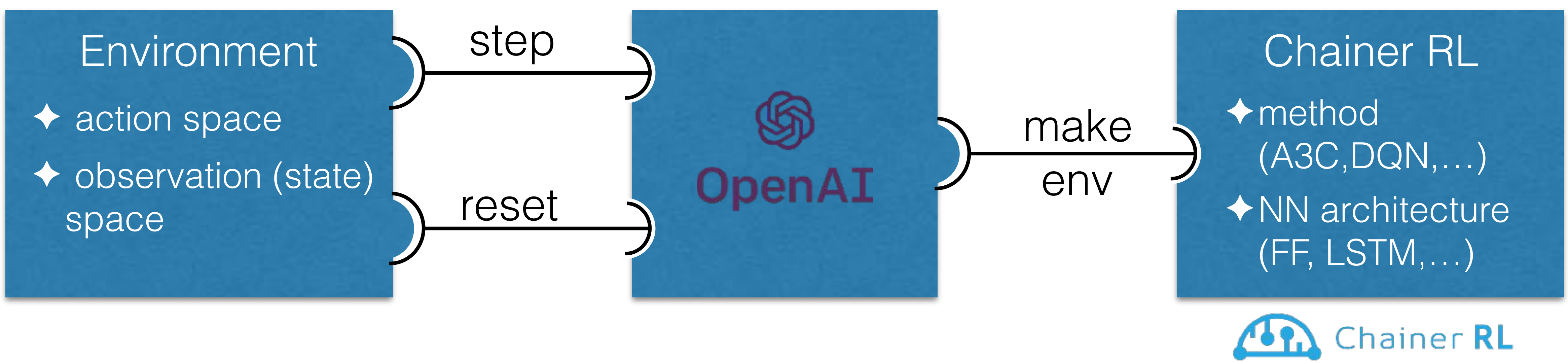


Reinforcement Learning - A3C



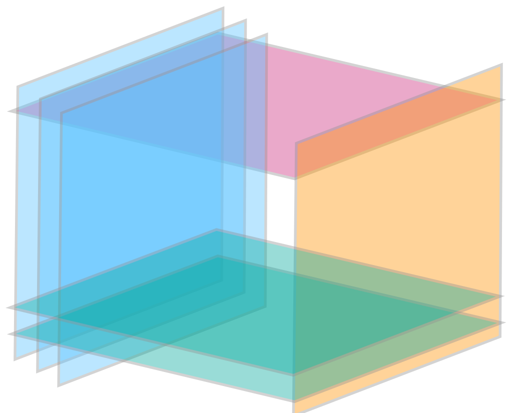
Reinforcement Learning - Implementation

- Open AI Gym: Interface between agent (RL) and environment (string landscape) [Brockman et al '16]
  - We provide the environment
  - We use ChainerRL's implementation of A3C for the agent

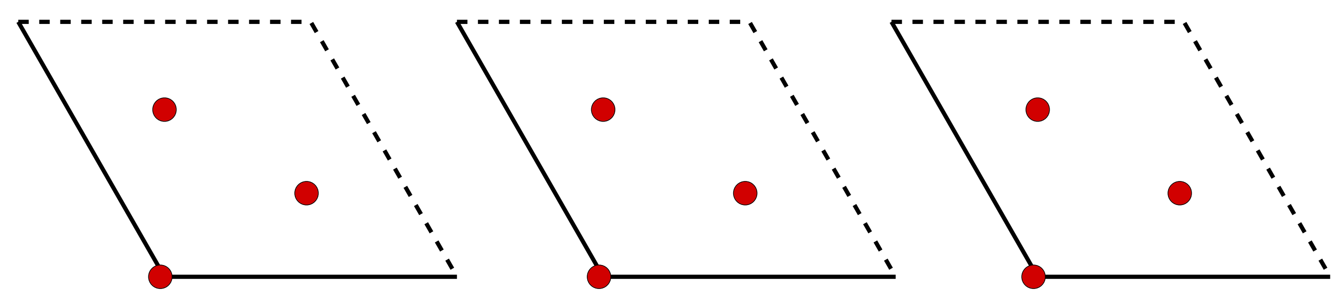


- step:
  - go to new state
  - return (new_state, reward, done, comment)
- reset:
  - reset episode
  - return start_state

- make environment
- specify RL method (A3C)
- specify policy NN (FF,LSTM)

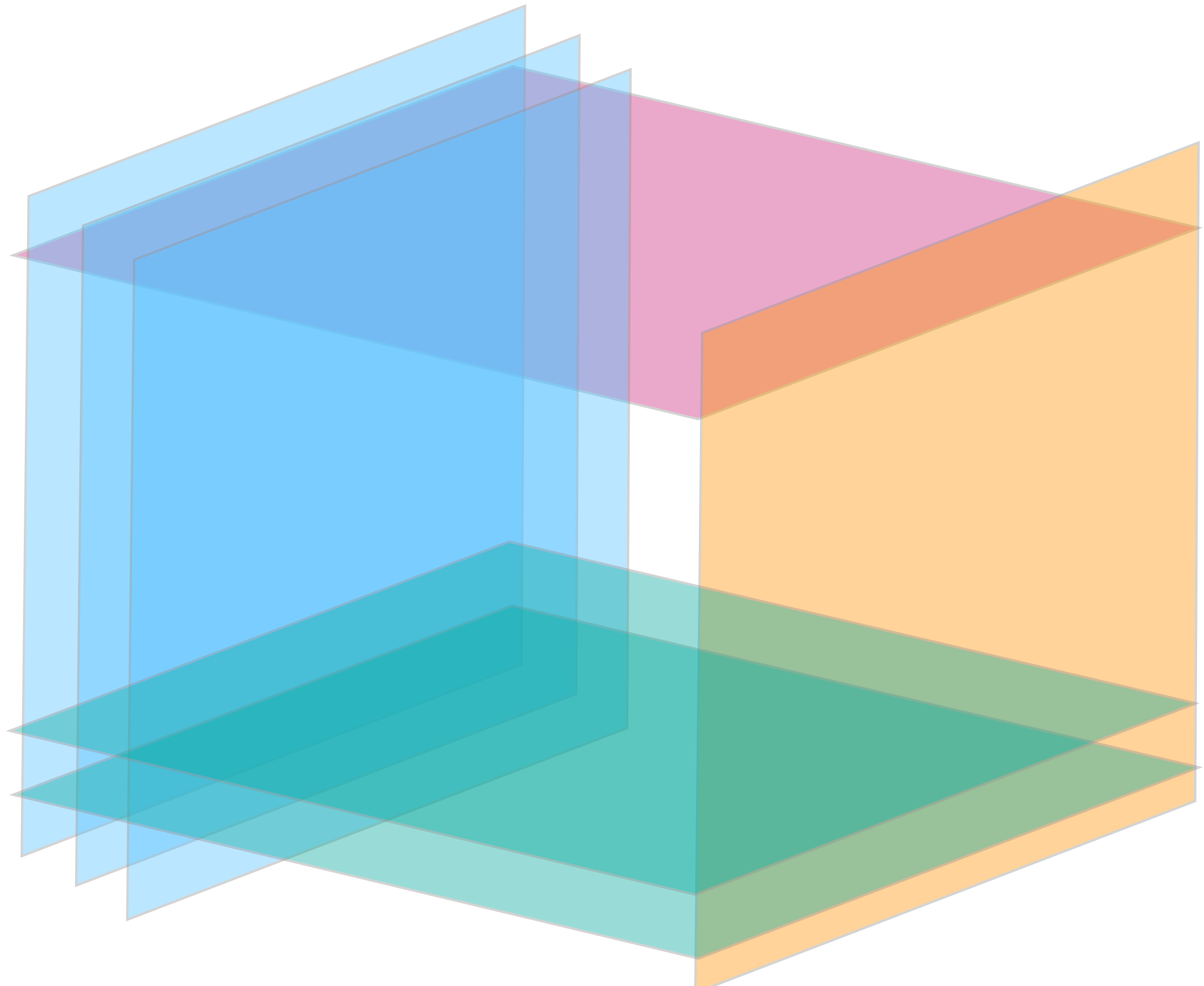


Type II Intersecting branes
Orientifolds of toroidal orbifolds



Heterotic $\mathrm{E}_8 \times \mathrm{E}_8$ string theory
on orbifolds

Example applications

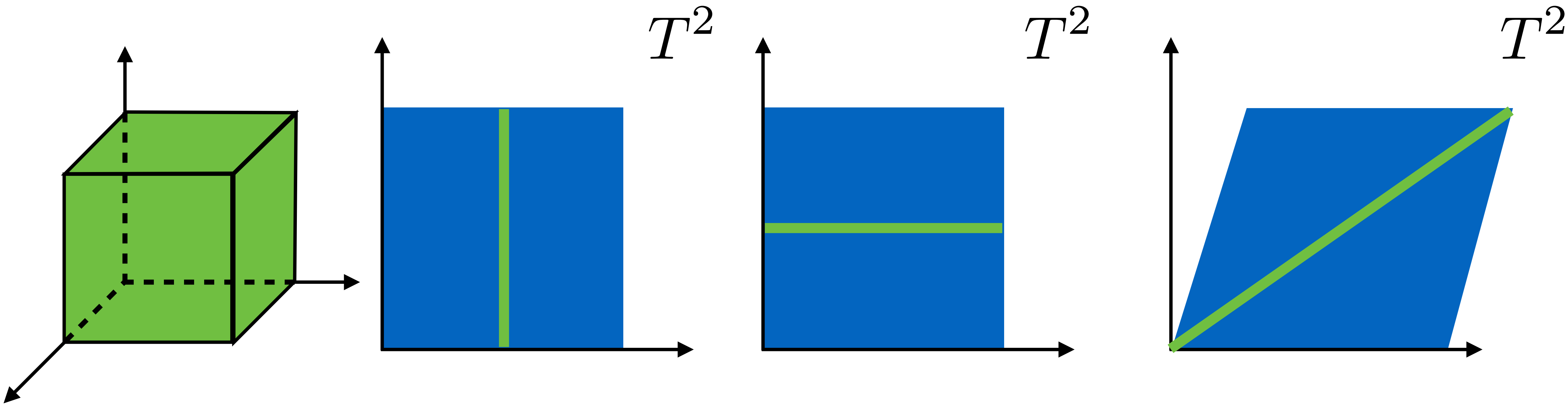


Type II Orientifolds

IIA Orientifolds

- Why this setup?
  - Well studied [Blumenhagen,Gmeiner,Honecker,Lust,Weigand '04'05; Douglas, Taylor '07, ...]
  - Comparatively simple
  - Number of (well-defined) solutions known to be finite: [Douglas, Taylor '07]
    - Use symmetries to relate different vacua
    - Combine consistency conditions to rule out combinations
  - BUT: Number of possibilities so large that not a single "interesting" solution could be found despite enormous random scans (estimated to $1{:}10^9$)
  - Interesting to study with big data / AI methods

D6 branes



- Can (have to for three generations) tilt torus (2 different complex structure choices compatible with orientifold)
- D6 brane: 4D Minkowski + a line on each torus
- Can stack multiple D6 branes on top of each other
- Brane stacks $\Leftrightarrow$ Tuple: $(N, n_1, m_1, n_2, m_2, n_3, m_3)$

D6 Branes - Consistency Conditions

- Tadpole cancellation: Balance D6 / O6 charges:

$$\sum_{a=1}^{\#\text{stacks}} \begin{pmatrix} N^a n_1^a\, n_2^a\, n_3^a \\ -N^a n_1^a\, m_2^a m_3^a \\ -N^a m_1^a\, n_2^a\, m_3^a \\ -N^a m_1^a m_2^a\, n_3^a \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \\ 4 \\ 8 \end{pmatrix}$$

- K-Theory: Global consistency constraint:

$$\sum_{a=1}^{\#\text{stacks}} \begin{pmatrix} 2N^a m_1^a m_2^a m_3^a \\ -N^a m_1^a\, n_2^a\, n_3^a \\ -N^a n_1^a\, m_2^a\, n_3^a \\ -2N^a n_1^a\, n_2^a\, m_3^a \end{pmatrix} \mod \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

D6 Branes - Consistency Conditions

- SUSY: $\forall a = 1, \ldots, \#\ \text{stacks}$

$$m_1^a m_2^a m_3^a - j\, m_1^a n_2^a n_3^a - k\, n_1^a m_2^a n_3^a - \ell n_1^a n_2^a m_3^a = 0$$

$$n_1^a n_2^a n_3^a - j\, n_1^a m_2^a m_3^a - k\, m_1^a n_2^a m_3^a - \ell m_1^a m_2^a n_3^a > 0$$

- Pheno: $SU(3) \times SU(2) \times U(1)$ + MSSM particles

- Massless $U(1)$'s: $T_r \in \ker(\{N^k m_i^k\})$

$$i = 1, 2, 3 \quad \text{(three tori)}$$

$$k = 1, \ldots, \#U \text{ brane stacks}$$

$$r = 1, \ldots, \dim(\ker(\{N^k m_i^k\}))$$

$= k - 3$ (generically)

TypeII RL - Model the environment

- State space: $s_t \in S, \quad |S| = N_{\max}^{N_S} \binom{N_B}{N_S}$

$$s_t = [(N^1, n_1^1, m_1^1, n_2^1, m_2^1, n_3^1, m_3^1), (N^2, n_1^2, \ldots), \ldots]$$

- Action space: Two approaches

  - Construct collection of winding number 6-tuples. Actions can add/remove branes from the brane stacks or exchange entire 6-tuples from pool of constructed stacks

    $$A = \{N^a \to N^a \pm 1,\ \text{add stack } (N, n_1, \ldots),\ \text{remove stack } (N, n_1, \ldots)\}$$

  - Start with all winding numbers zero. Actions can add/remove branes from the brane stacks or add ± 1 to any winding number in any stack

    $$A = \{N^a \to N^a \pm 1,\ n_i^a \to n_i^a \pm 1,\ m_i^a \to n_i^a \pm 1\}$$

TypeII RL - Model the environment

- Reward R: Need a notion of "how good a state is"
  1. By how much does a set of stacks violate the tadpole?
  2. Is a set of stacks fully consistent (Tadpole, K-Theory, SUSY) (Note: the latter two are binary, hard to define distance)
  3. How far is the state from the Standard Model
     - Missing a group factor of $SU(3) \times SU(2) \times U(1)$?
     - Too few Standard model particles $(Q, u, d, L, H_u, H_d, e)$?
     - Extra exotics (particles charged under the Standard Model but not observed so far)

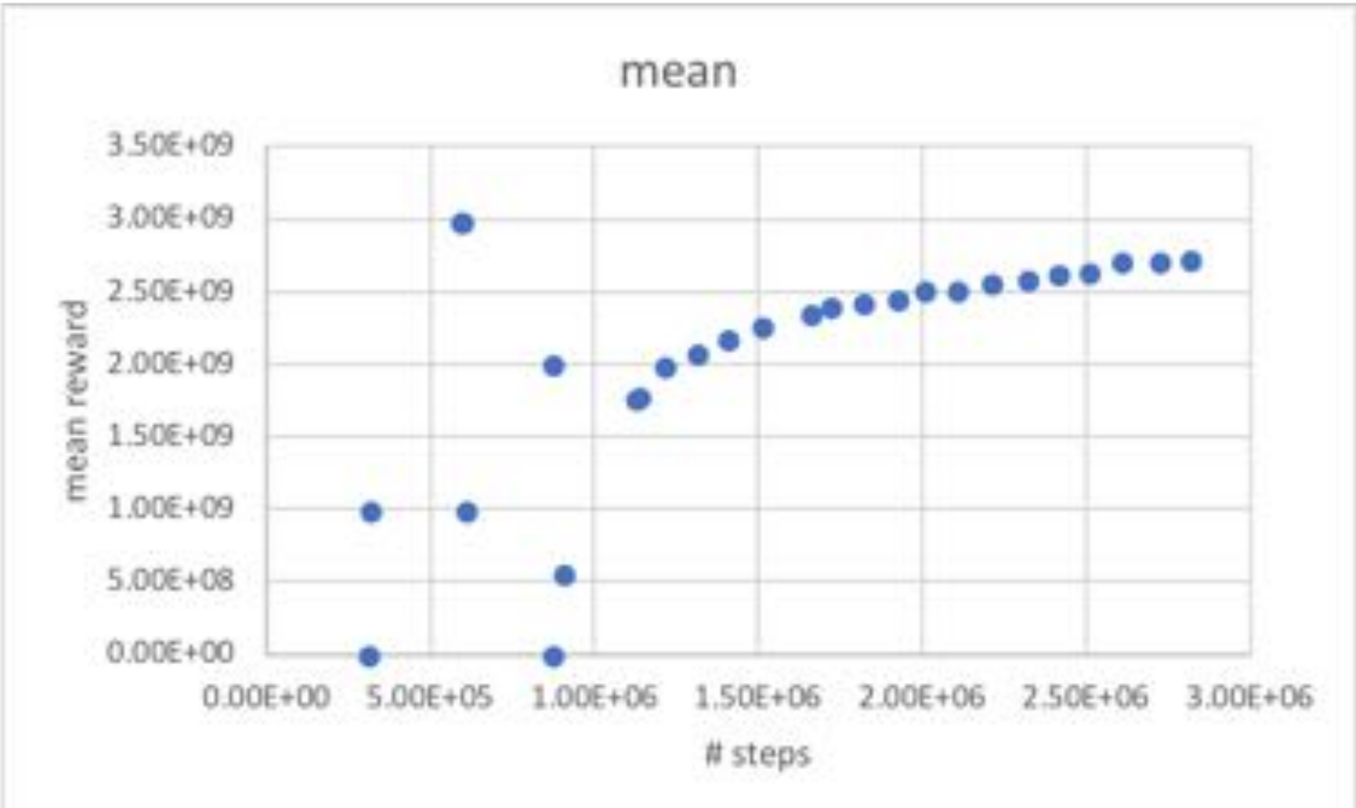
  Note: Only works if good states are "close by" in this sense…

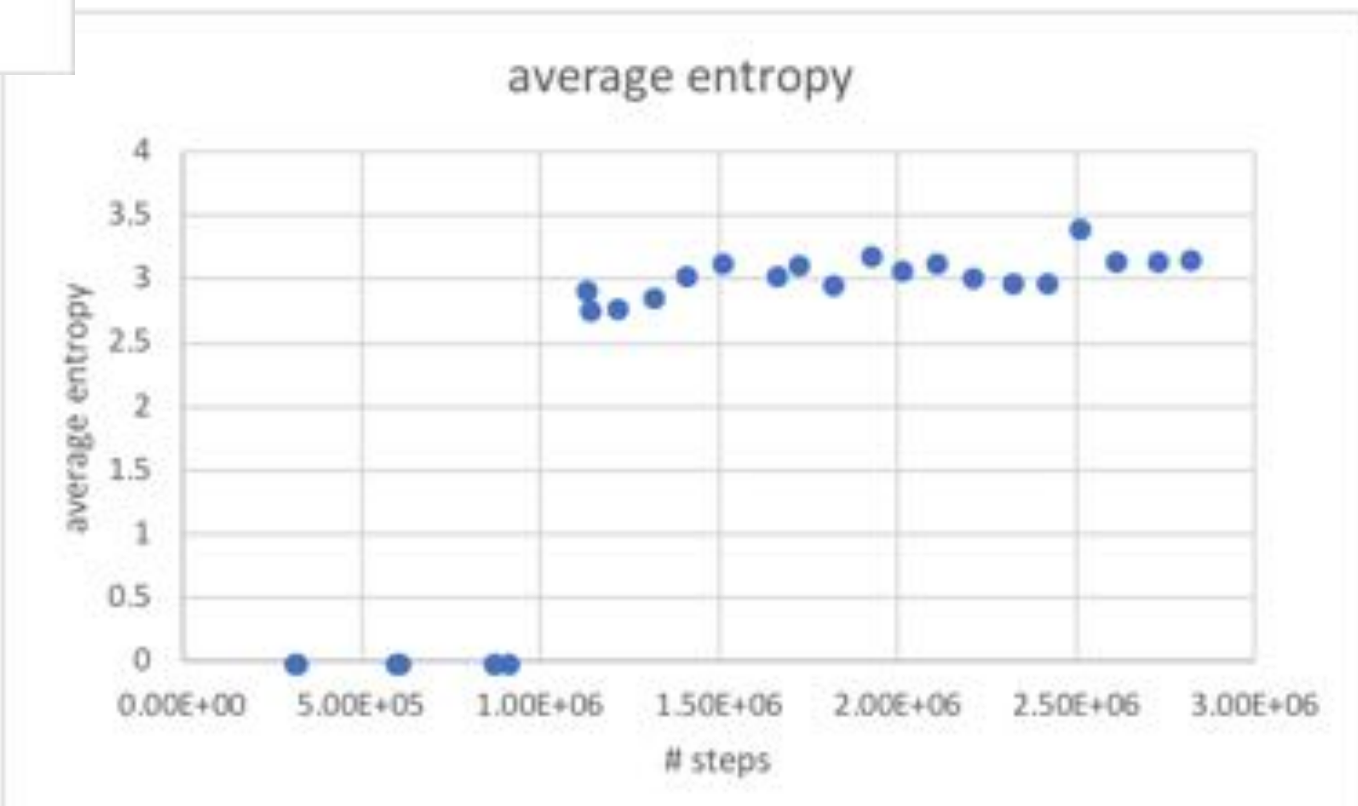
- Need multi-task RL:
  - Check properties consecutively/simultaneously and use different reward hierarchies for different tasks
  - Split up async workers and let them prioritise different goals

Preliminary results

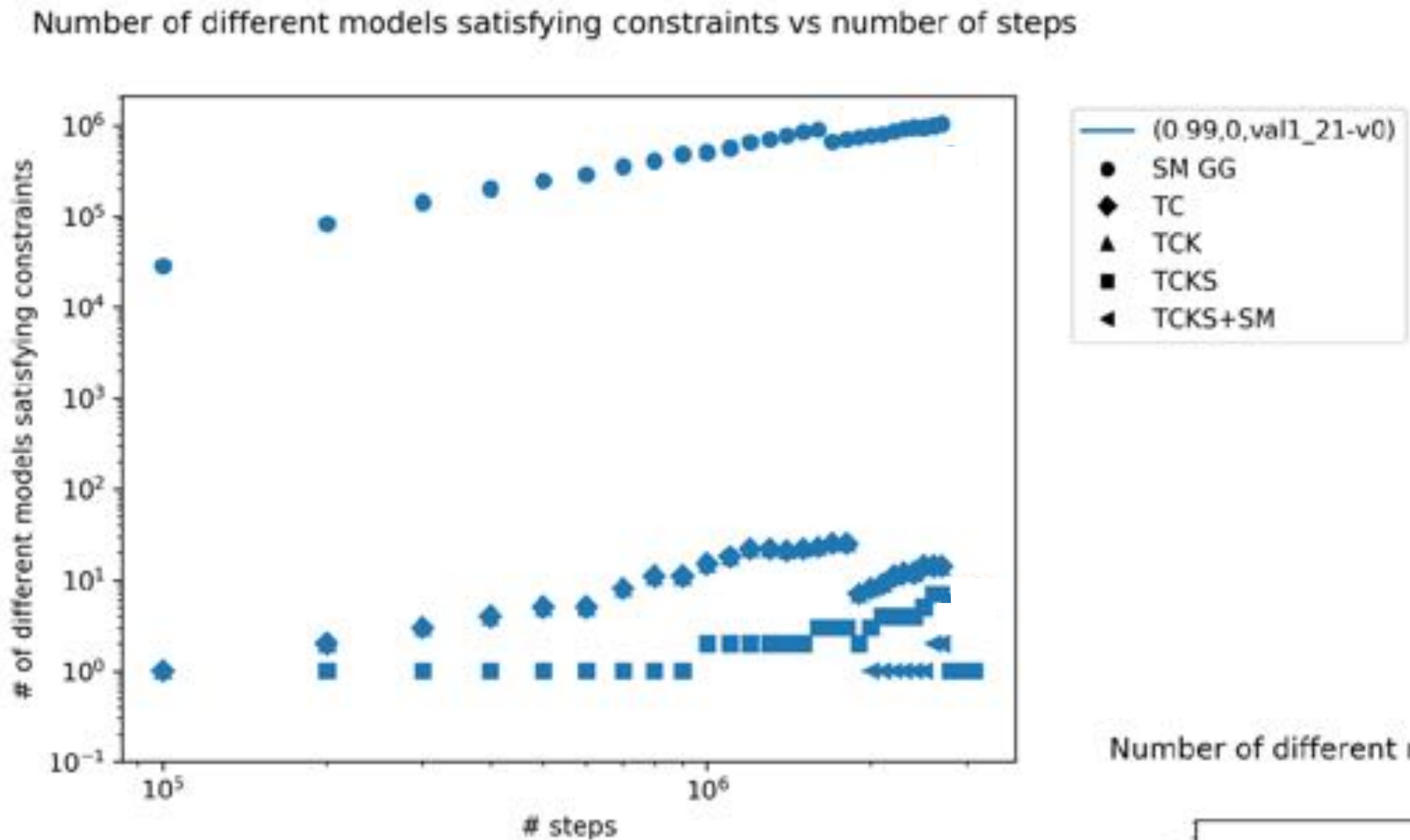
- Parameters:
  - 16 or 32 workers (1 CPU, 16-32 threads, 2.6GHz)
  - Training time of the order few hours to a day
  - Neural network for value and policy evaluation:
    Feed-forward NN with 2 hidden Softmax layers with 200 nodes
  - Initial state: Empty stack
  - Maximal steps per episode: 10,000 - 250,000
  - 10 evaluation runs every 100,000 steps

Preliminary results - Finding models Approach 1

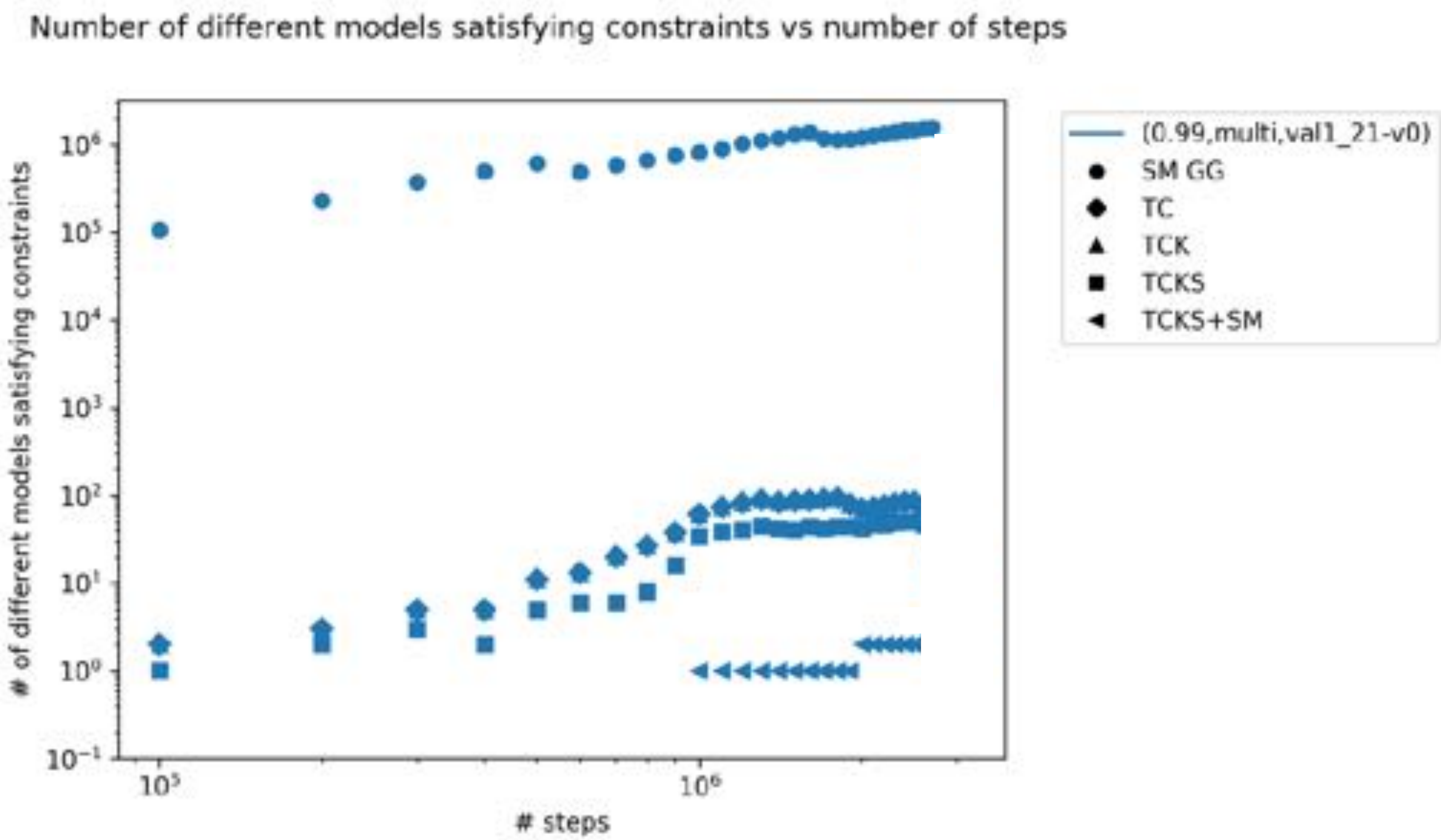




Preliminary results - Finding models Approach 1

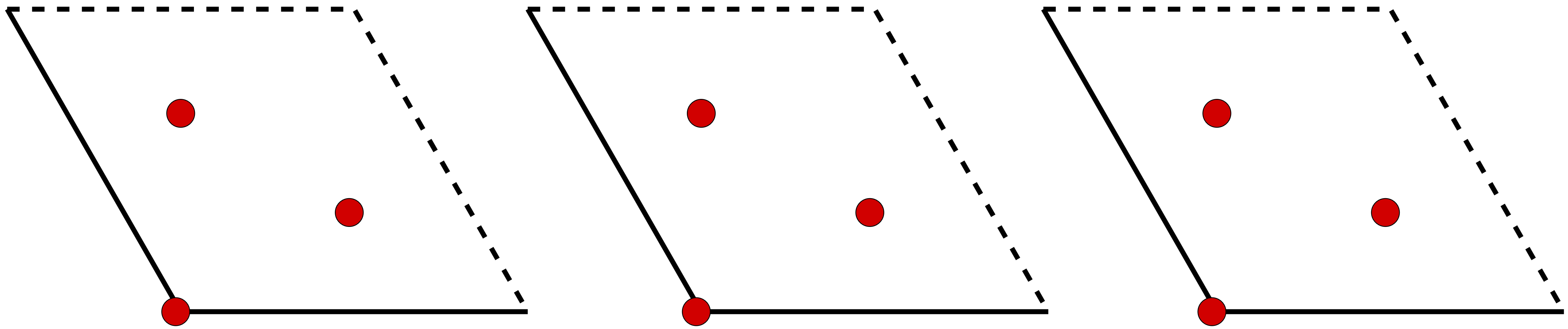


1.) Check consistency
2.) Check particle physics



multitask:

- 16 workers consistency
- 16 workers particle physics



Heterotic Orbifolds

Heterotic Orbifolds

- Why this setup?
    - Consistent models constructed
      [Blaszczyk, Buchmuller, Groot Nibbelink, Hamaguchi, Kim, Kyae, Lebedev, Nilles, Raby, Ramos–Sanchez, Ratz, FR, Trapletti, Vaudrevange, Wingerter, … `06-10]
    - Comparatively simple
    - Phenomenologically promising
    - Well-developed mathematics and computer codes to perform CFT computations for spectrum, couplings, …
      [Dixon, Harvey, Vafa, Witten `86; Gross, Harvey, Martinec, Rohm `86]
      [Nilles, Ramos-Sanchez, Vaudrevange, Wingerter `11]

Heterotic Orbifolds

- Start from constructed models
  - Already identified MSSM gauge group and spectrum
  - … but the vacua of the theory have to be found s.t.
    - D-term induced from an FI parameter of an anomalous U(1) symmetry is canceled
    - No F-terms are induced in the process
    - extra vector-like exotics (order 40) decouple
    - extra gauge symmetries (U(1)s) get broken
  - All achieved by singlet VEVs
  - Encode VEV of singlets in bit string $[s_1, s_2, \ldots, s_n]$
  - Assign $s_i = 0 \,/\, 1$ if singlet i has no VEV / VEV

Heterotic RL - Consistency Conditions

- D-Terms:

$$\begin{pmatrix} q_{1,1} & q_{1,2}\cdots & q_{1,n} & \\ q_{2,1} & q_{2,2}\cdots & q_{2,n} & \\ \vdots & \vdots & \ddots & \vdots \\ q_{r,1} & q_{r,2}\cdots & q_{r,n} & \end{pmatrix} \begin{pmatrix} |s_1|^2 \\ |s_2|^2 \\ \vdots \\ |s_n|^2 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$$

$$r = \#\ \mathrm{U(1)s}$$
$$n = \#\ \mathrm{singlets}$$

- F-terms:

$$F_i = \frac{\partial W}{\partial s_i} = \sum a_d p_d(s) \overset{!}{=} 0$$

$$m_d : \text{polynomials in } s_i \text{ of degree d}$$

Heterotic RL - Consistency Conditions

- Pheno:
  - generate full rank mass matrix for exotics
  - keep one vector-like Higgs pair
  - break additional U(1) gauge groups but not hyper charge
- Massive U(1)s:

$$\ker \begin{pmatrix} q_{1,i_1} & q_{1,i_2} \cdots & q_{1,i_k} & \\ q_{2,i_1} & q_{2,i_2} \cdots & q_{2,i_k} & \\ \vdots & \vdots & \ddots & \vdots \\ q_{r,i_1} & q_{r,i_1} \cdots & q_{r,i_k} & \end{pmatrix} = 0$$

singlets $i_1, \ldots, i_k$ have a VEV

Heterotic RL - Model the environment

- State space: $s_t \in S_{\text{total}}$, $|S_{\text{total}}| = 2^{\#\text{singlets}}$

  $$s_t = [s_1, s_2, \ldots, s_n]$$

- Action space: Two approaches

  - Start with all VEVs off (no F-terms, but D-terms, exotics, U(1)s) and turn VEVs on

    $$A = \{s_i = 1\}$$

  - Start with all VEVs on (no D-terms, exotics, U(1)s, but many F-terms) and turn VEVs off

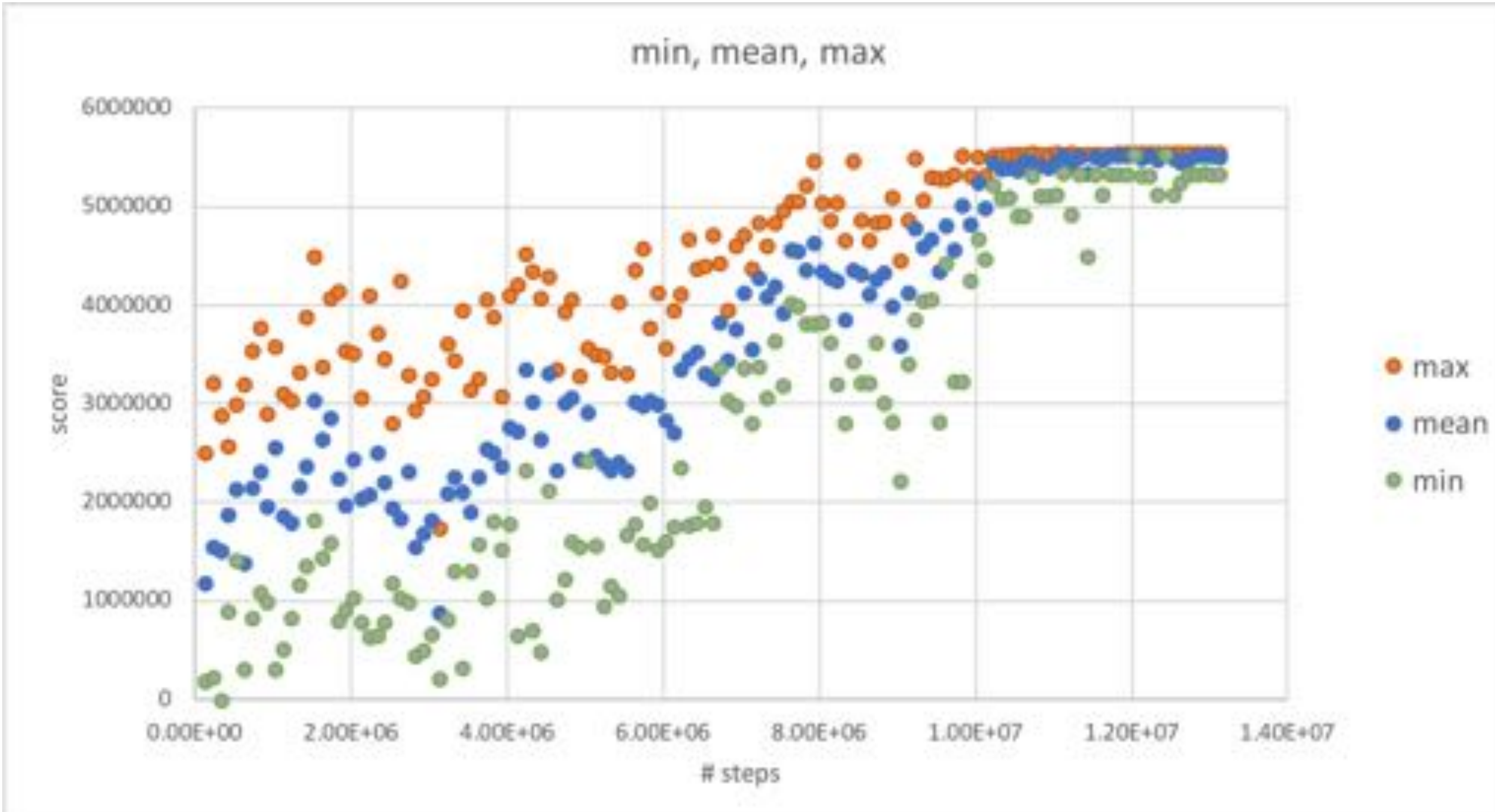
    $$A = \{s_i = 0\}$$

Heterotic RL - Model the environment

- Reward R: Need a notion of “how good a state is”
  1. How many F-terms does a VEV configuration generate?
  2. How many U(1)s are left unbroken?
  3. How many exotics are not decoupled?
  4. Is a Higgs pair kept light?
  5. Are all D-terms cancelled?

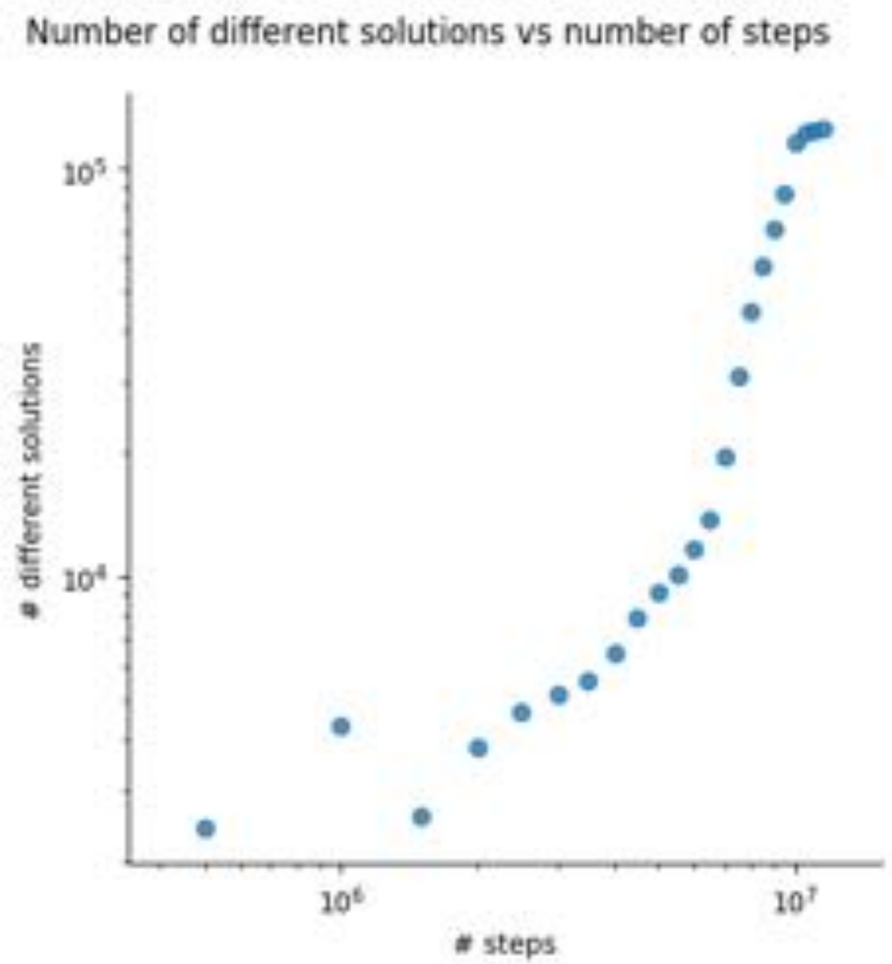
- Note:
  - Approaches require multi-task RL

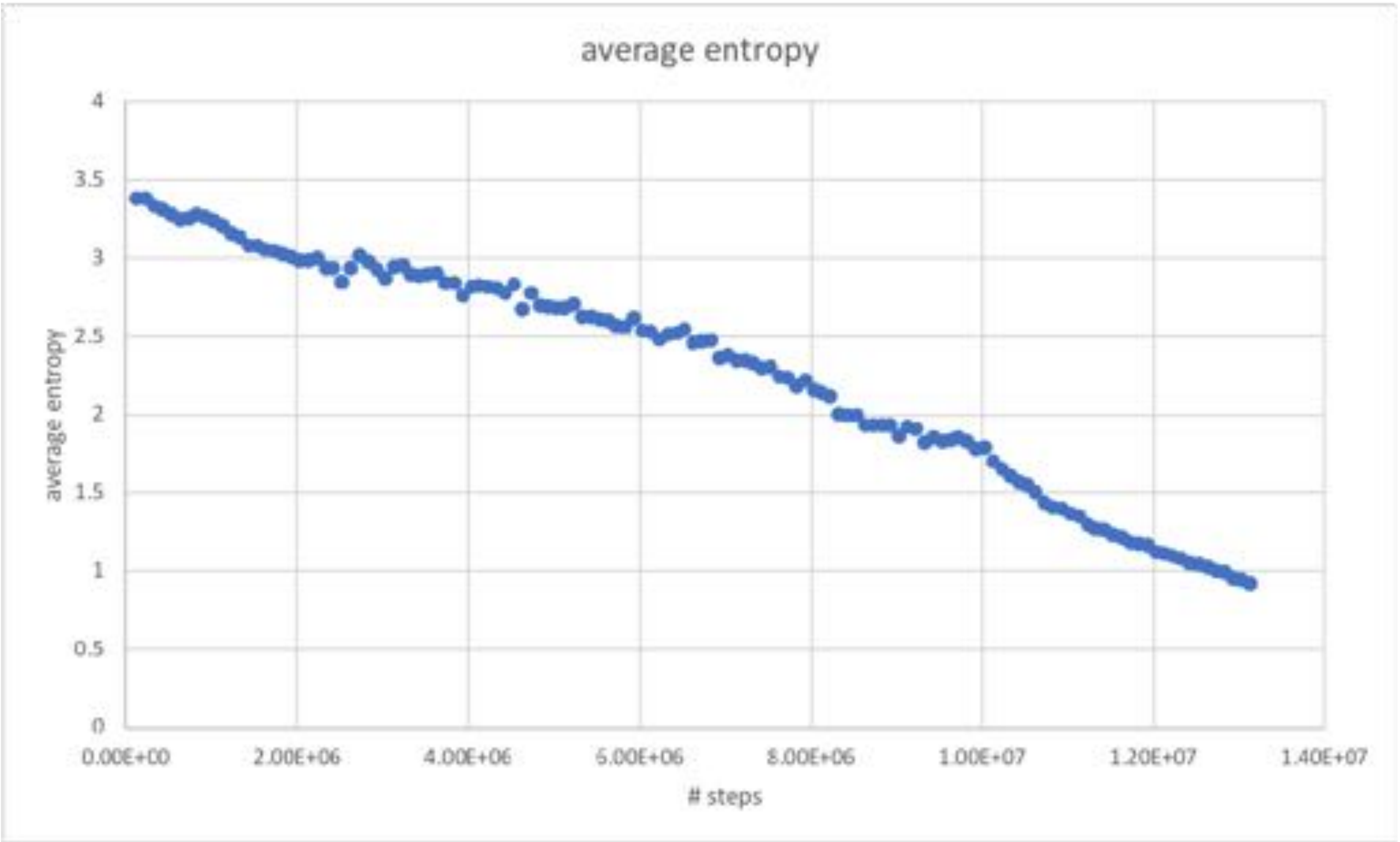
Preliminary results - Heterotic Model Approach 2



- Reward structure:
  - +100 for each F-term that is canceled
  - +10k for keeping Higgs light while decoupling all other exotics
  - end episode if exotics increase, D-term is not canceled, U(1)s become massless
- Best state: 0/6 D-terms, 0/8 U(1)'s, 0/36 exotics, 17/1,124

Preliminary results - Heterotic Model Approach 2





- Reward structure:
  - +100 for each F-term that is canceled
  - +10k for keeping Higgs light while decoupling all other exotics
  - end episode if exotics increase, D-term is not canceled, U(1)s become massless
- Best state: 0/6 D-terms, 0/8 U(1)'s, 0/36 exotics, 17/1,124

Conclusion

- RL well suited for search & explore in the string landscape
- Very versatile applications to string theory:
  - String models in Type II intersecting brane models on toroidal orientifolds
  - Vacuum configurations for Heterotic $\mathrm{E}_8 \times \mathrm{E}_8$ string theory

Thank you for your attention!