ALPHAGO ZERO, OU COMMENT MAITRISER LE GO EN 3 JOURS

Umberto Grandi
IRIT - Université Toulouse I Capitole

MAITRISER LE JEU DE GO

Après l'introduction au jeu de go de ce matin, êtes vous prêts à gagner contre le champion du monde **avec 3 jours d'entrainement**?

C'est possible! Il vous faut juste **une équipe de bons programmeurs** et environs **25 millions d'euros** de hardware!

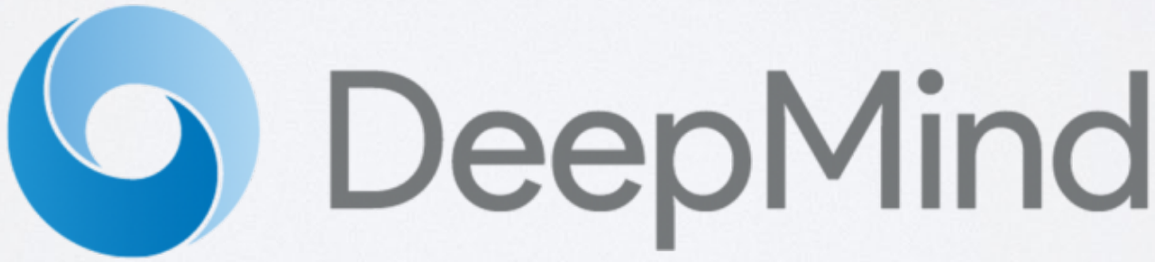
ALPHAGO ZERO

Une fois rassemblé l'argent et les programmeurs,
suivez les instructions dans ce cours!

- Découvrir **l'histoire** d'AlphaGo (Lee, Master, Zero..)
- Comprendre pourquoi go est (était) un jeu **difficile pour un ordinateur**
- Avoir un aperçu de l'**architecture** d'AlphaGo Zero…
- …des **techniques** utilisées (réseaux de neurons profonds, apprentissage par renforcement profond, Monte Carlo Tree Search)…
- …et de ses impressionantes **performances**

UN PEU D'HISTOIRE

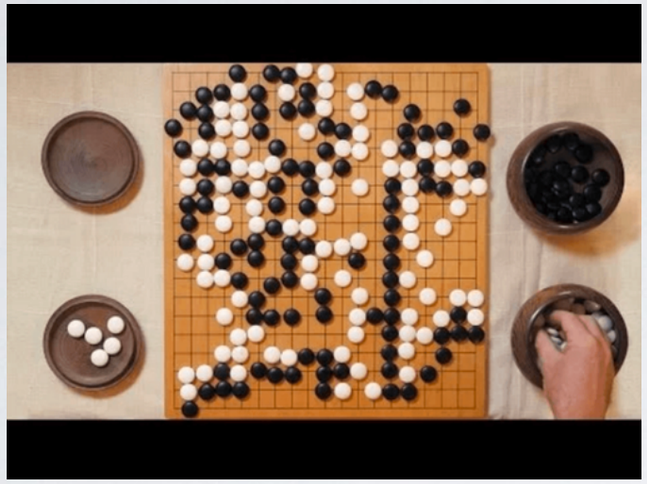
- Octobre 2015: **AlphaGo Fan** bat le champion européen de go, utilisant de l'apprentissage supervisé avec connaissance d'experts, puis apprentissage par renforcement
- Mars 2016: **AlphaGo Lee** bat Lee Sedol, vainqueur de 18 tournois internationaux (meme techniques)
- Octobre 2017: **AlphaGo Zero** gagne contre AlphaGo Lee sans avoir recours à des connaissances humaines!



Qui est le concepteur d'AlphaGo? DeepMind, une ex-startup fondée à Londre en 2010 et racheté par Google en 2014

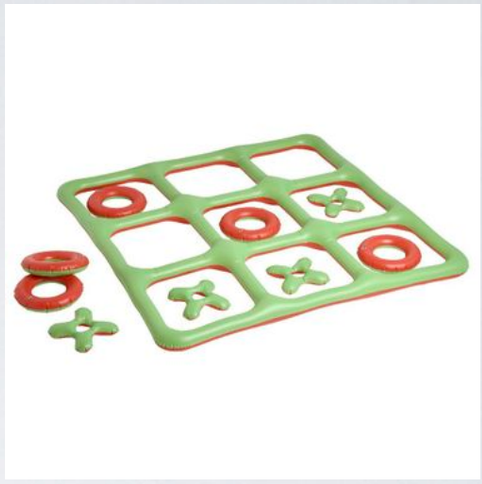
QUELQUES REFLEXIONS AUTOUR DU JEU DE GO

Quelle sont les caractéristiques communes de ces trois jeux:



Go

Echecs

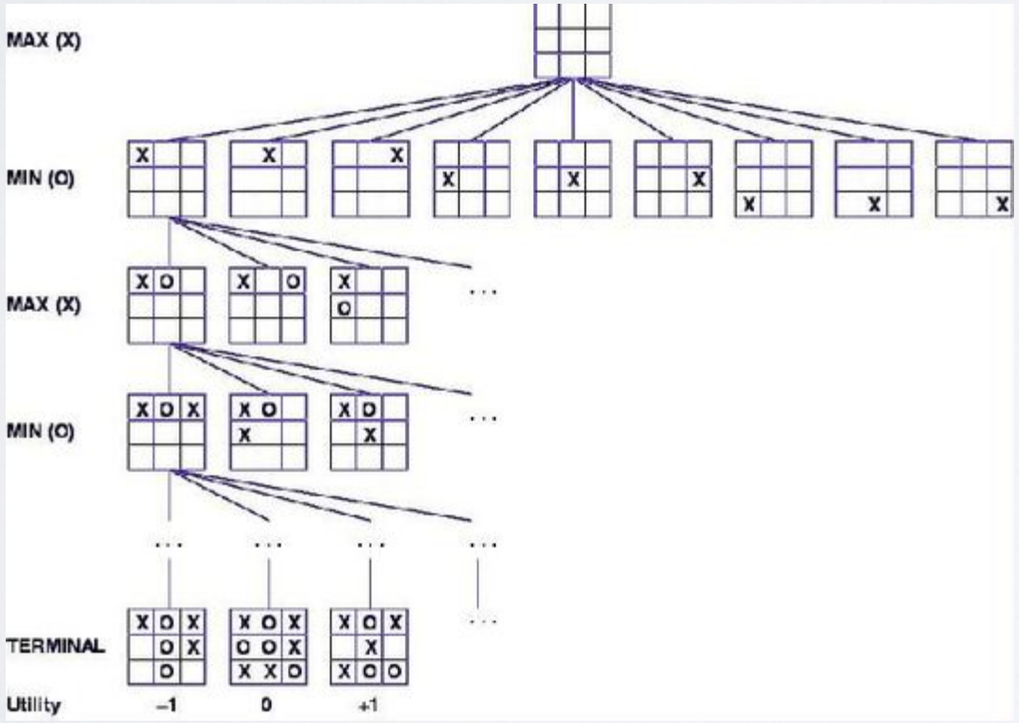


Morpion

- 2 joueurs
- sans hazard
- finis
- à tour
- à information parfaite (tous les coup d'un joueur peuvent être observé par l'autre)

LE THÉORÈME DE ZERMELO

Dans tout jeu à tour, fini, à 2 joueurs, à information parfaite, et sans hazard, **un de deux joueurs a une stratégie pour gagner ou faire match nul**.



L'arbre de jeu du morpion

Donc le premier joueur peut toujours gagner à go!

Mais pourquoi donc on continue à y jouer?

L'ARBRE DE JEUX DE GO

10^170 configurations possibles
10^360 complexité de l'arbre de jeu

(pour comparer, le nombre d'atoms dans l'univers est estimé à 10^80)

Il est impossible d'écrire et encore pire faire des calculs avec meme peu de niveaux de l'arbre de jeu de go!!

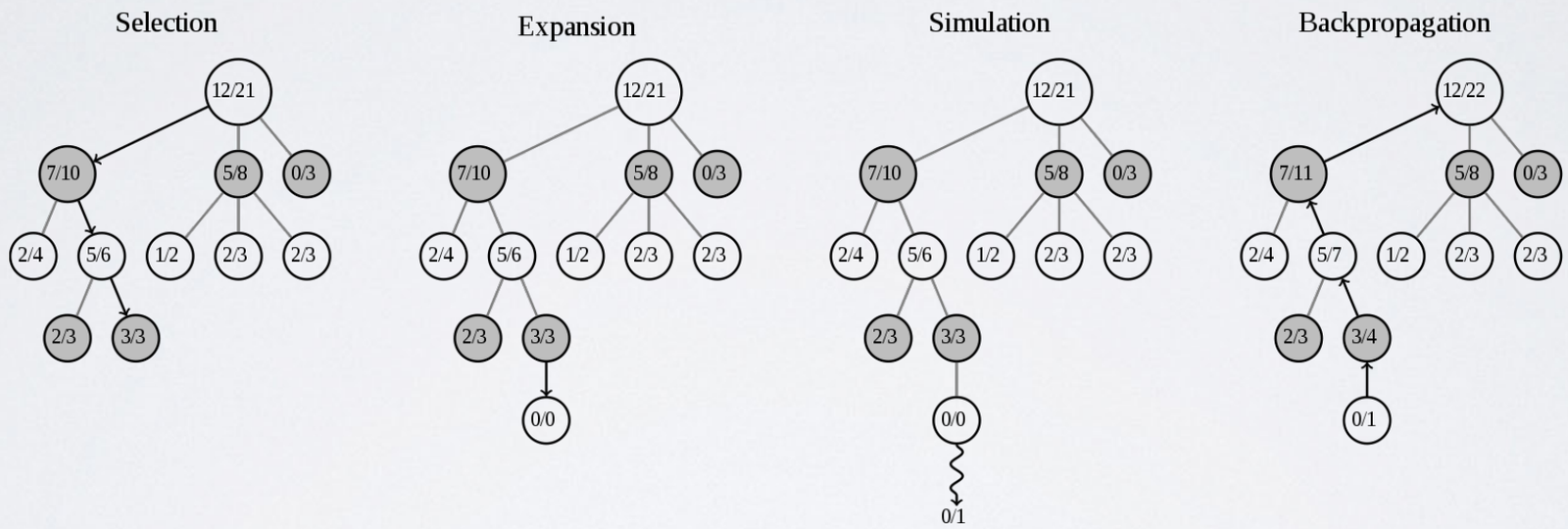
Les algos "standard" de recherche en IA ne peuvent pas gérer une telle complexité

QUELQUES TECHNIQUES UTILES

MONTE CARLO TREE SEARCH

MONTE CARLO TREE SEARCH

Mimer le raisonnement d'un joueur, mais avec des probabilités et la puissance d'un ordinateur.

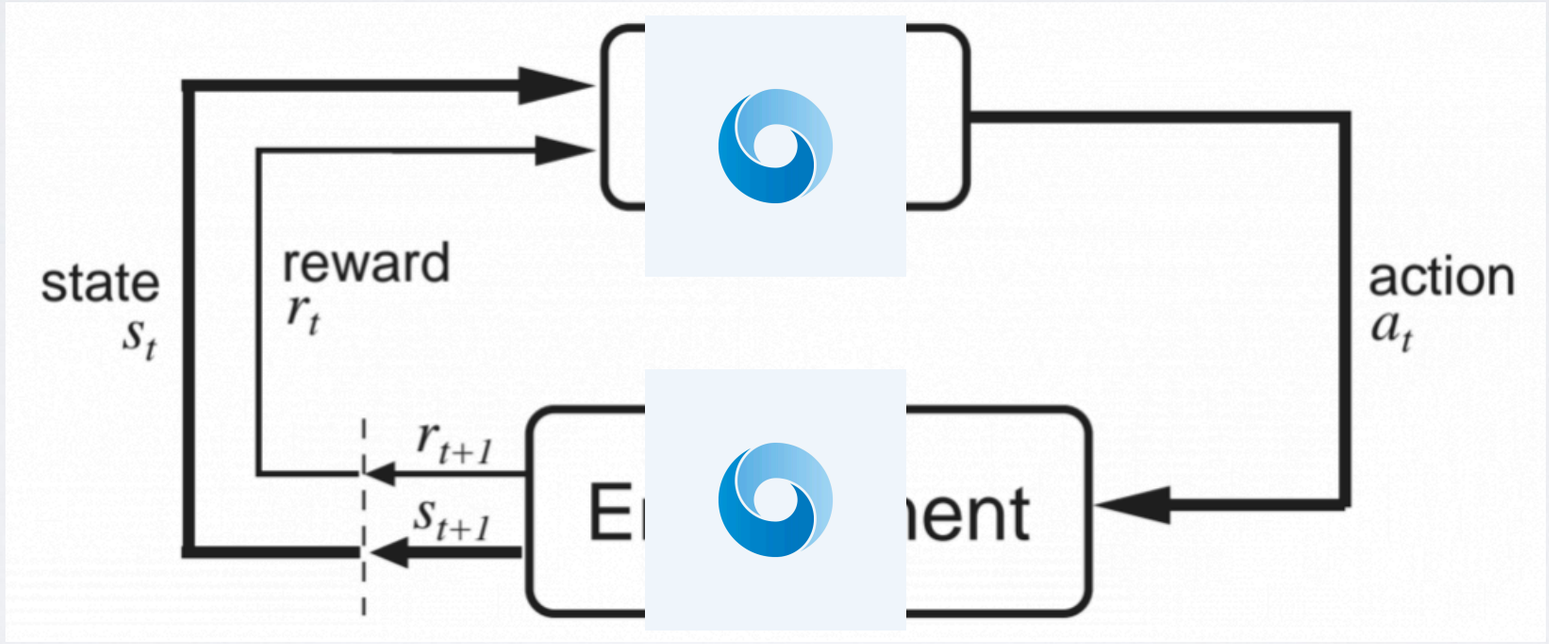


1. **Simuler des jeux au hasard** à partir d'une configuration, et calculer combien de fois on gagne et combien on perd
2. "**Backpropagate**" vers le haut ce résultat, et continuer
3. Choisir l'action avec la **meilleur ratio gagnées/perdues**

APPRENTISSAGE PAR RENFORCEMENT

Quand on apprend à jouer contre un joueur plus expérimenté qui nous indique si le coup qu'on vient de jouer était bon ou pas.

AlphaGo:

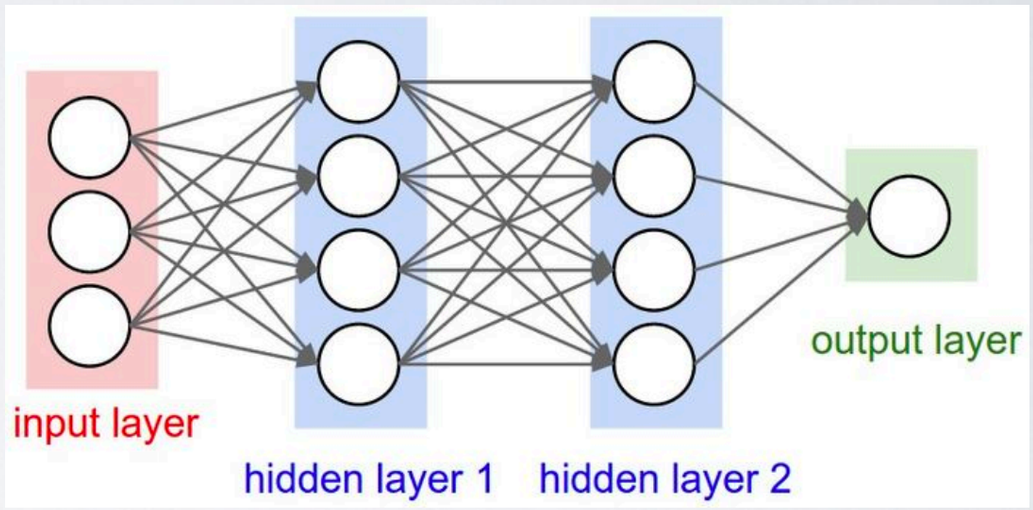


Pour le jeu de go je devrais enregistrer la valeur de chaque action dans chaque état: cela fait >10^170 valeurs…trop!

RÉSEAU DE NEURONES PROFOND

Une boite noire qui **apprend** à reconnaitre des motifs:

- reconnaissance des visages
- reconnaissance de la parole
- traitement du langage…



Quel type de réseau pour le jeu de go?

Input: positions des pierres sur le plateau et historique
Output: probabilité pour jouer le prochain coup (***policy***)
et probabilité de gagner dans l'état courant (***value***)

ARCHITECTURE DE ALPHAGO ZERO

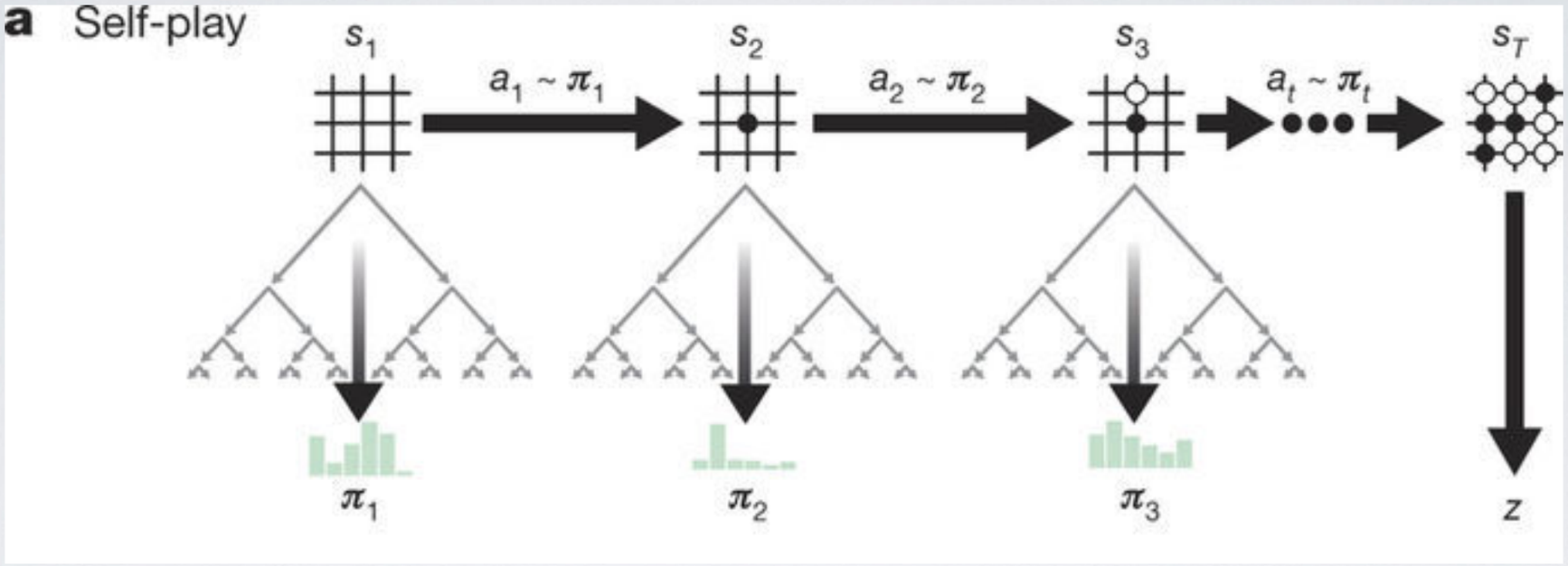
1. Un **RESEAU de neurons profond,** qui prend un historique du jeux en entrée et restitue une stratégie de jeu (probabilités), et la probabilité de gagner

2. Un **algo MCTS** qui à chaque étape utilise les valeurs de RESEAU pour simuler des jeux aléatoires et obtenir une meilleur stratégie de jeu

3. Une période d'apprentissage du réseau, qui utilise de l'**apprentissage par renforcement (profond)**, en jouant des parties contre soi meme

LA PHASE D'APPRENTISSAGE

L'algo qui joue contre soi meme - détails

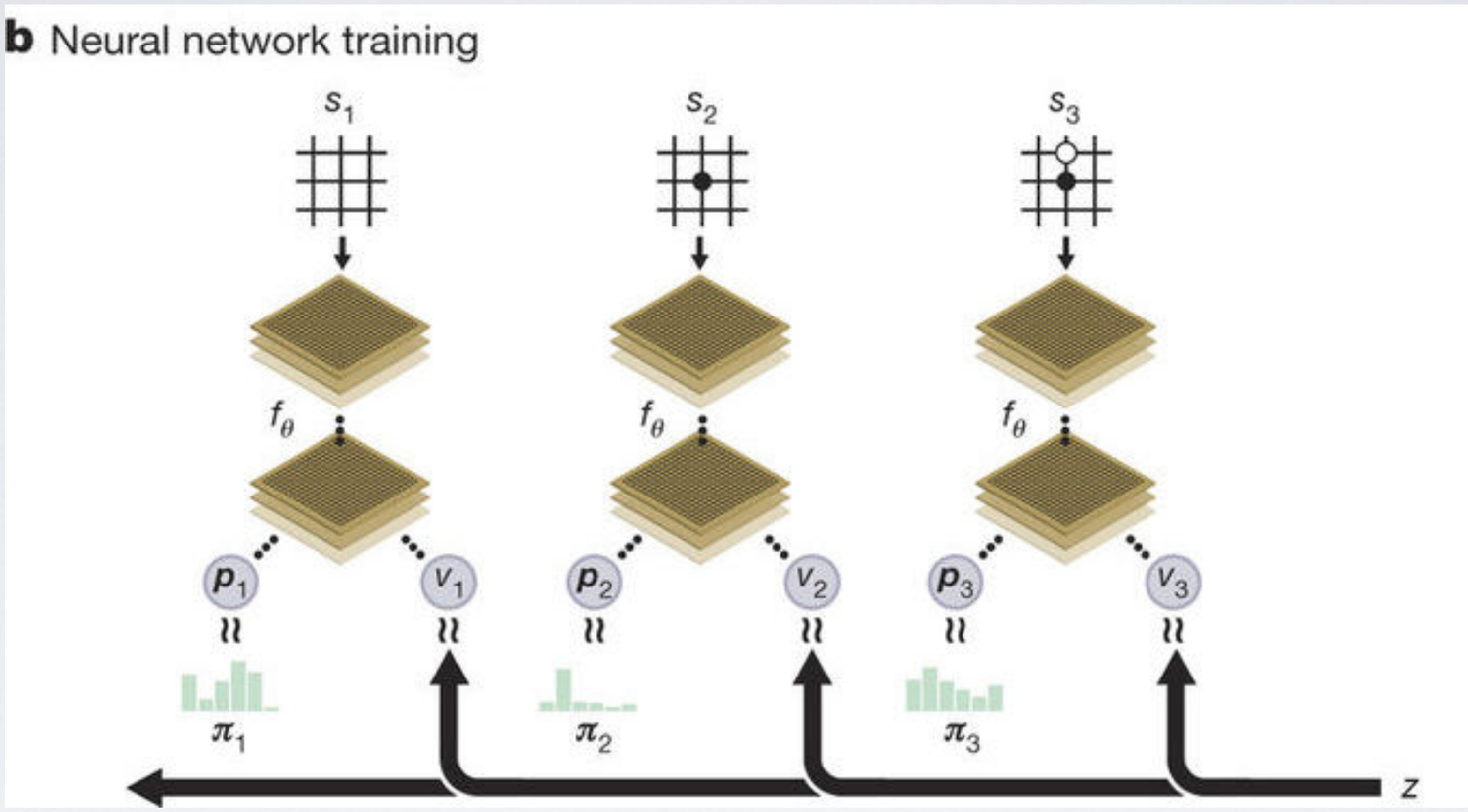


- on commence en S1 (configuration vide)
- on demande au RESEAU la stratégie de jeu (une probabilités par chaque coup possible, qu'on appellera P)

- on demande à MCTS d'utiliser P pour des simulations de jeux et de retourner une meilleur stratégie de jeu π
- on utilise π pour choisir une action et aller en S2

- on entraine le RESEAU pour faire que P se rapproche de π (et plus tard aussi pour prédire la valeur du vrais gagnant du jeu z)
- on recommence en S2!

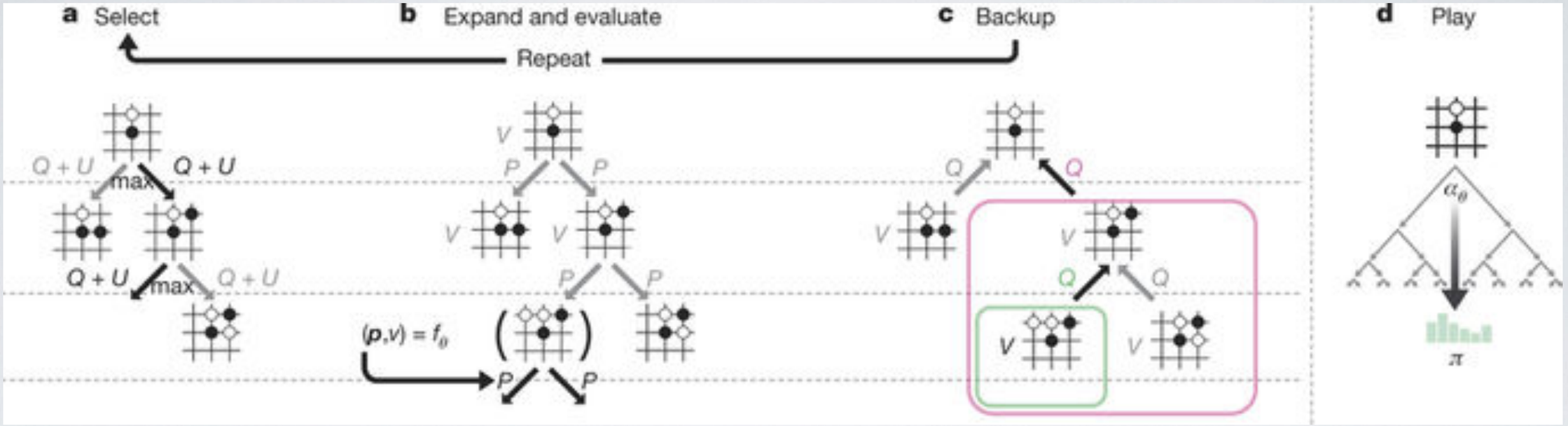
L’apprentissage du réseau de neurones - détails



Rappelez vous que le réseau sert à obtenir une stratégie de jeu et une valeur de la configuration courante.

1. La **stratégie de jeu P** sert à jouer et paramétrer le MCTS: elle est **entrainé** sur la stratégie amélioré du MCTS
2. La **valeur v** sert à paramétrer le MCTS: elle est **entrainé** sur le résultat de la partie joué contre soi meme

Monte Carlo Tree Search - détails



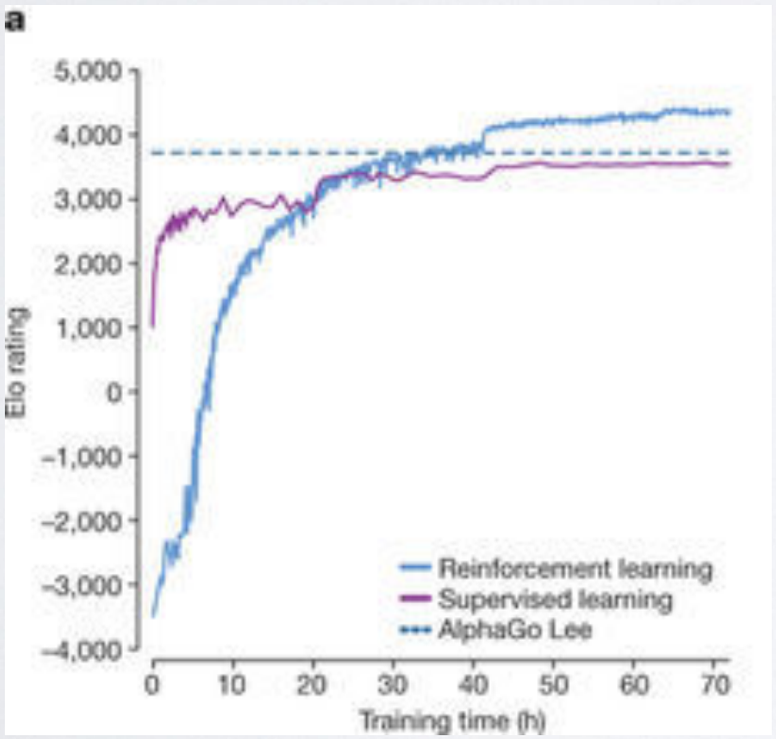
A chaque étape MCTS je demande P et V au réseau:

- J'utilise la stratégie de jeu P comme **heuristique** (pour éviter de devoir considérer des coup "irrationnels" ou des parties "impossibles")
- J'utilise la valeur V pour estimer si le résultat d'une branche est bon ou pas (pas de "rollouts")
- Je fais tout remonter à l'état initial qui nous donne une nouvelle probabilité pour jouer notre coup!

LES RÉSULTATS

APRES 3 JOURS

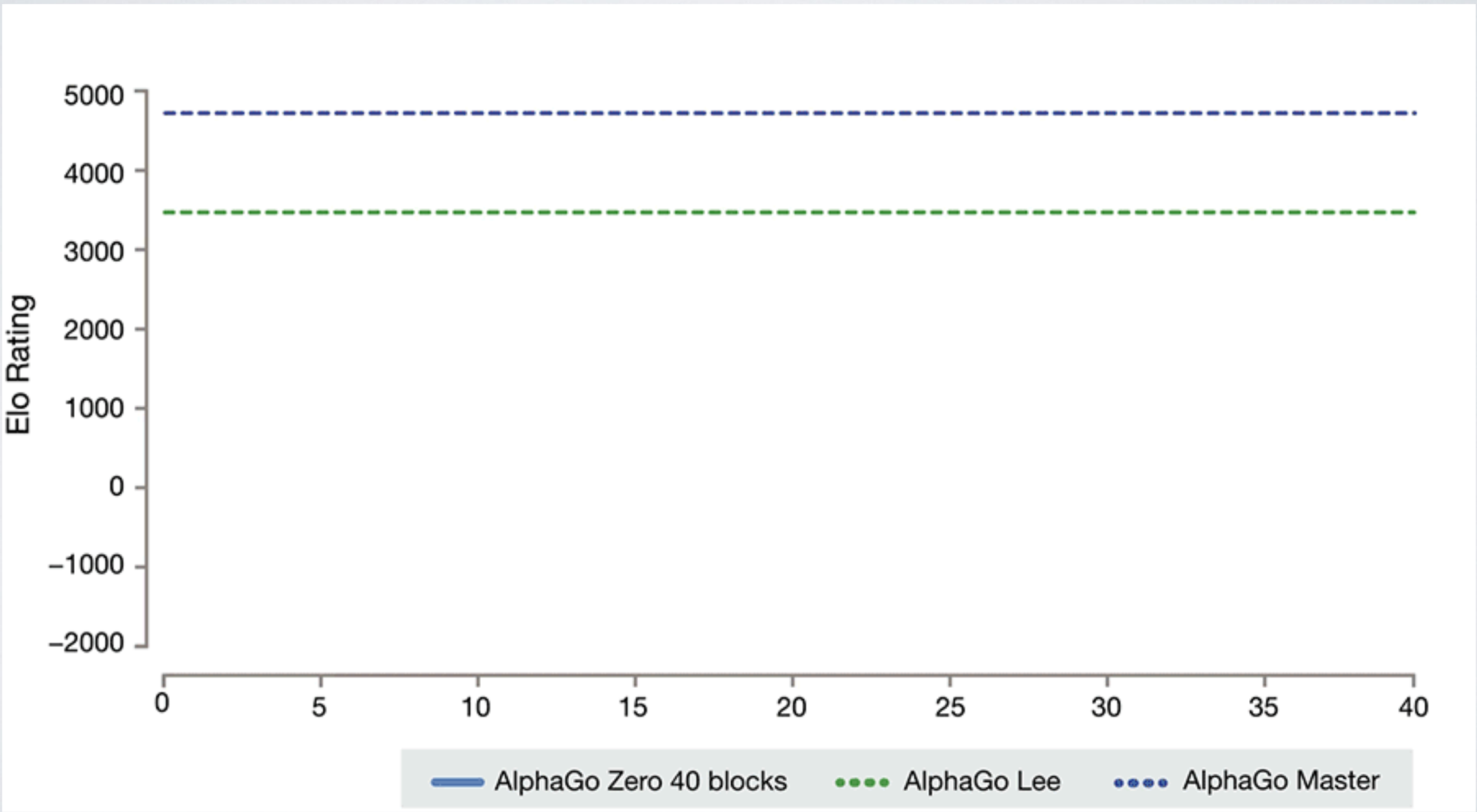
Après 3 jours d'apprentissage, 4.9 millions de "jeux contre soi meme", 1600 simulation MCTS par état, 0.4s pour réfléchir à chaque action…



AlphaGo Zero:

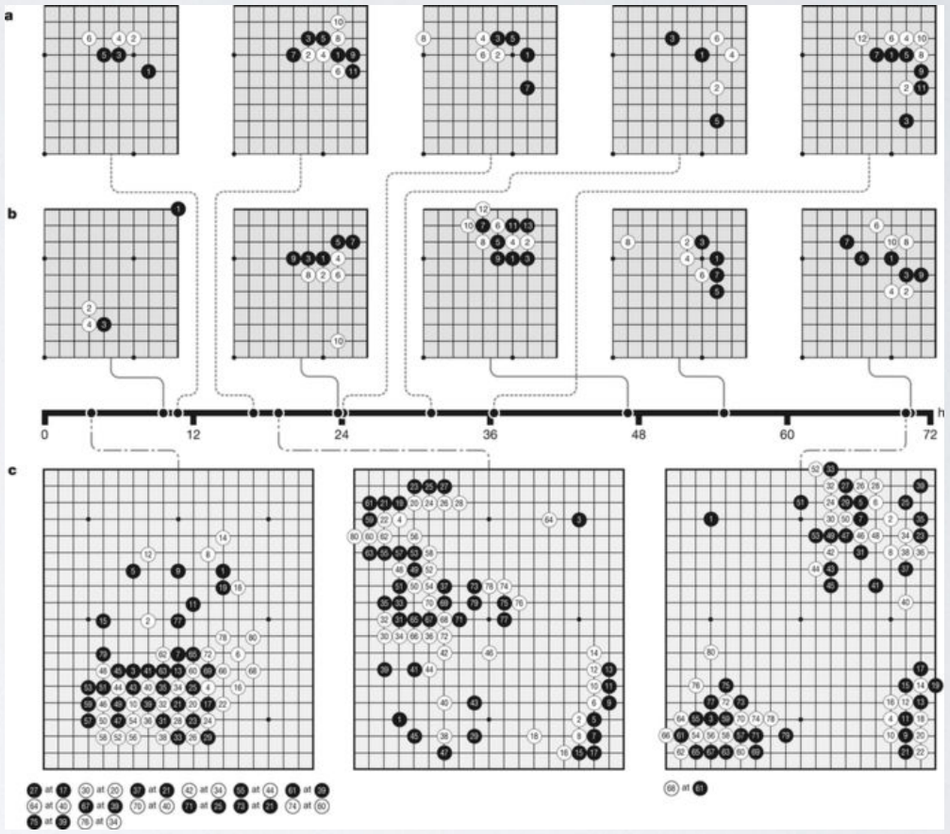
- Gagne contre AlphaGo Lee (celui qui a battu Lee Sedol!) qui avait une phase d'apprentissage de plusieurs mois
- Après 72h et avec seulement 4TPU gagne contre AlphaGo Lee "pro" qui utilise 48TPU!

APRÈS 40 JOURS…



POUR LES “PRO” DE GO

Vous pouvez vous amuser à observer comment AlphaGo Zero a appris des positions du plateau standard ou quelles sont ses stratégies préférées.



Vous pouvez lire plus et télécharger les parties d’AlphaGo Zero sur:
https://deepmind.com/blog/article/alphago-zero-starting-scratch

C'EST LA FIN DU GO?

Non! On continue à jouer aux échecs, et on continuera à s'amuser à jouer à go (seulement contrôlez bien que votre adversaire est un humain!)

Depuis le lancement d'AlphaGo, des joueurs professionnels humains étudient les coups d'AlphaGo pour améliorer leurs stratégies…
…voilà ce que l'IA peut nous apporter!

Toute les figures "techniques" sont prises de l'article suivant:
Silver et al. Mastering the Game of Go without Human Knowledge. Nature, 2017.