But sample efficiency is not all
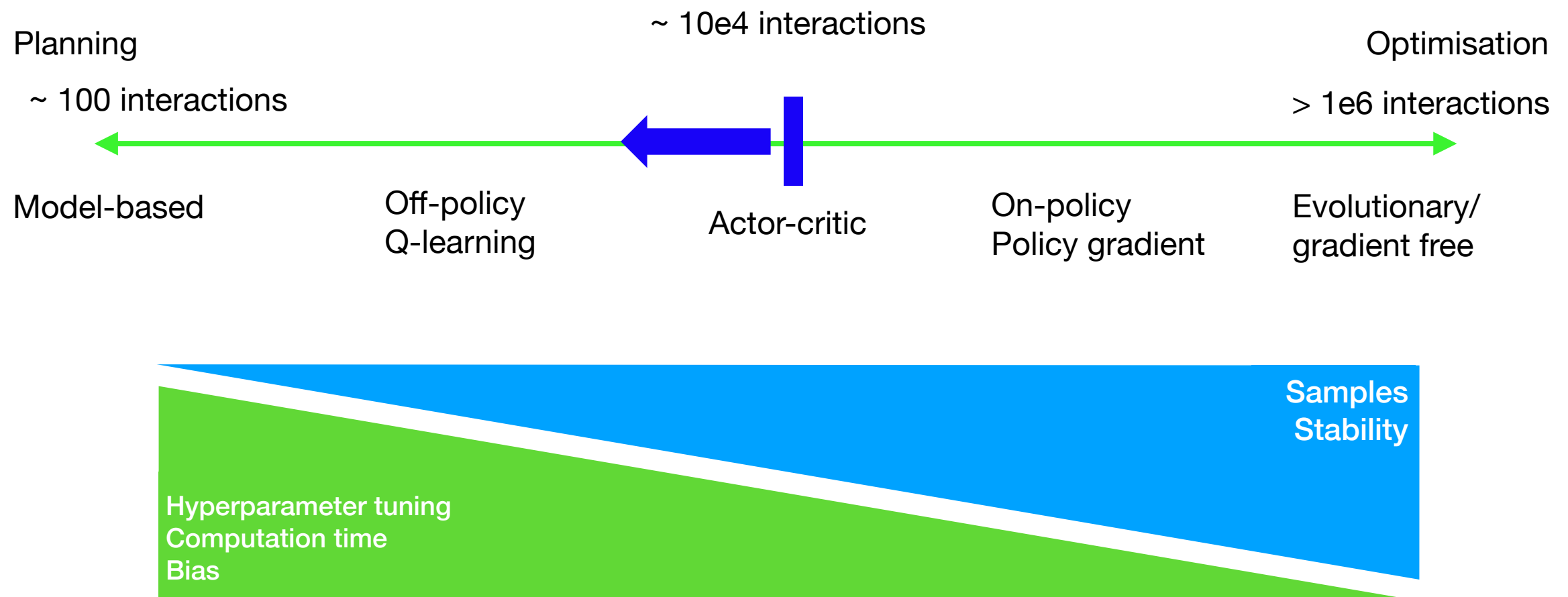
~ 10e4 interactions

Planning

Optimisation

~ 100 interactions

> 1e6 interactions

Model-based

Off-policy
Q-learning

Actor-critic

On-policy
Policy gradient

Evolutionary/
gradient free

Samples
Stability

Hyperparameter tuning
Computation time
Bias

IDALAB

PARIS
LODRON
UNIVERSITÄT
SALZBURG

Scenarios

How to make RL work on real world problems?

Exception

Find good hyper-parameters

Sufficient Data

Direct model-free RL

Simulation-based RL

High fidelity model

Exception

Model-based RL

Adapt to real world

Policy to reach target

IDALAB

PARIS LODRON UNIVERSITÄT SALZBURG