Imperial College London

The DLAdvocate: playing the *devil's advocate* with hidden systematic uncertainties

Andrei Golutvin[1], Aleksander Iniukhin[2], Andrea Mauri[1], Patrick Owen[3], Nicola Serra[3], Andrey Ustyuzhanin[4]

based on arXiv:2303.15956
published in Eur. Phys. J. C 83, 779 (2023)

[1] Imperial College London
[2] Yandex School of Data Analysis
[3] Universität Zürich
[4] National University Singapore

6th IML workshop, CERN, 29 January 2024

Motivation

Measurement = Central value ± Statistical uncertainty ± Systematic uncertainty ± Hidden systematic uncertainty

- How to alleviate the risk of *hidden systematic uncertainties*
  - independent confirmation from a different experiment

Under which condition one can claim a physics discovery in an experiment which has unique physics sensitivity and therefore no direct competitors?

- *Deep Learning (DL) Advocate* to quantitatively address the unknown unknowns

Motivation

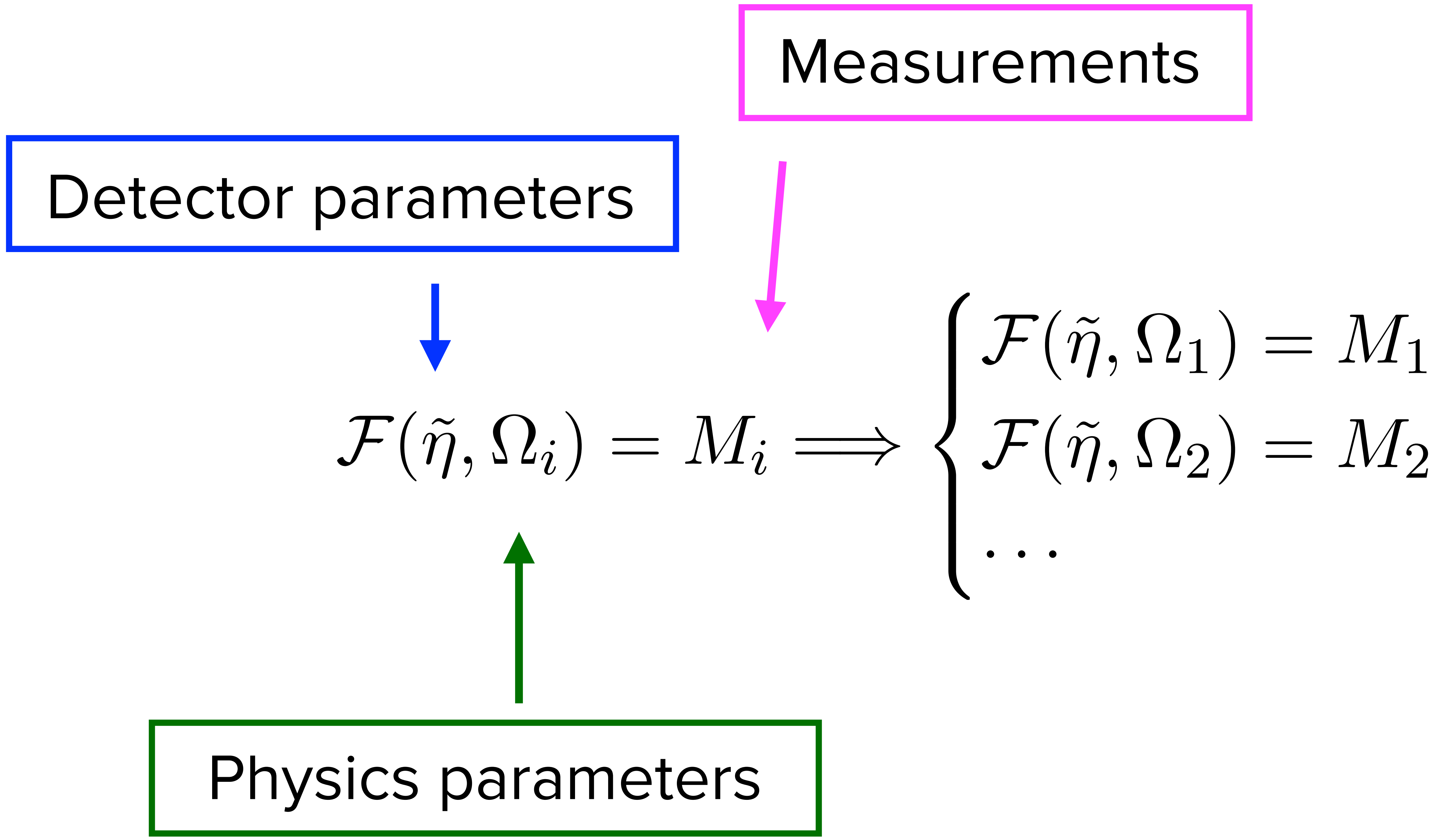
Measurement Central value ± Statistical uncertainty ± Systematic uncertainty ± Hidden systematic uncertainty

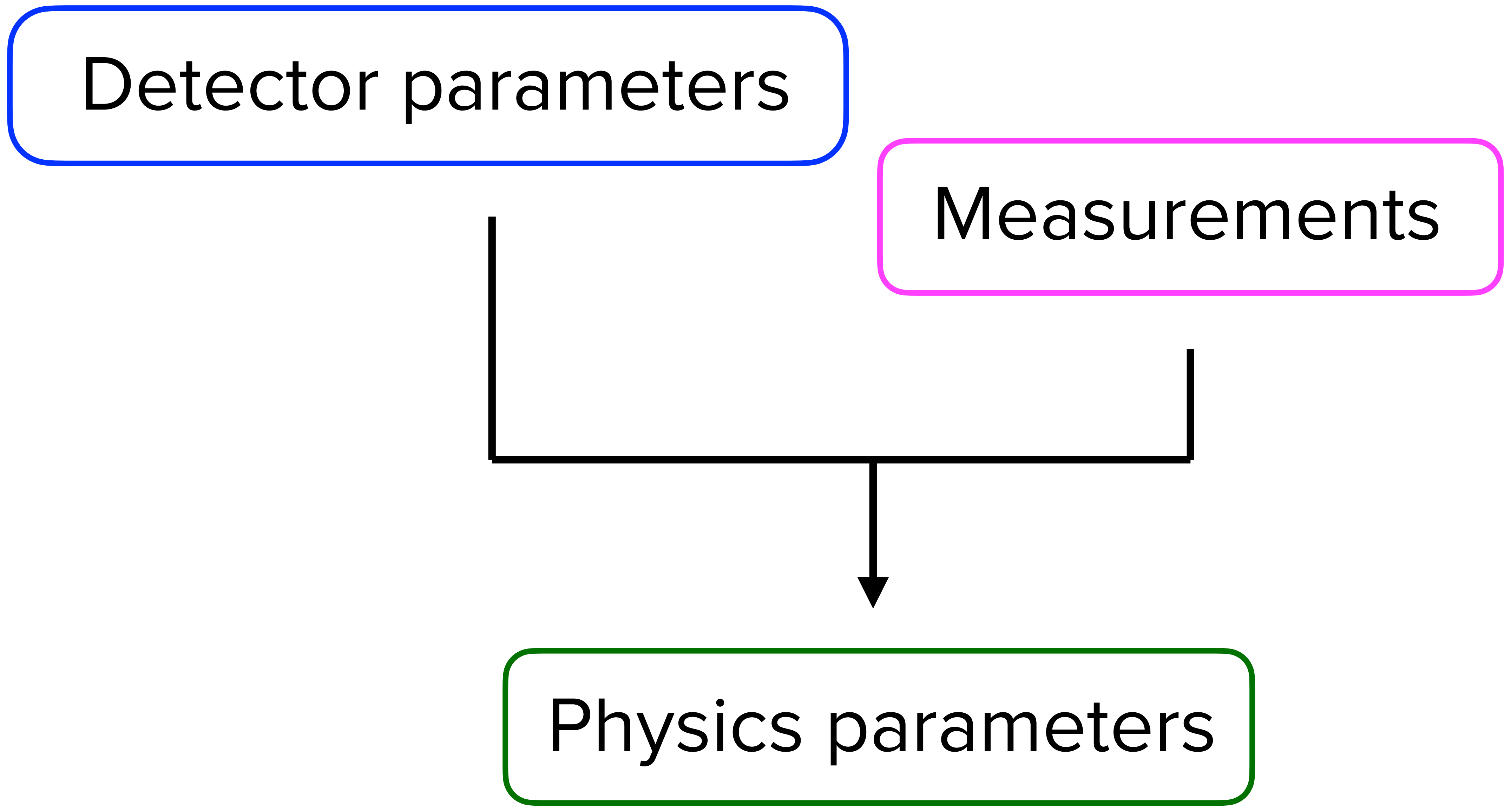
- How to alleviate the risk of *hidden systematic uncertainties*
  - independent confirmation from a different experiment

Under which condition one can claim a physics discovery in an experiment which has unique physics sensitivity and therefore no direct competitors?

- *Deep Learning (DL) Advocate* to quantitatively address the unknown unknowns

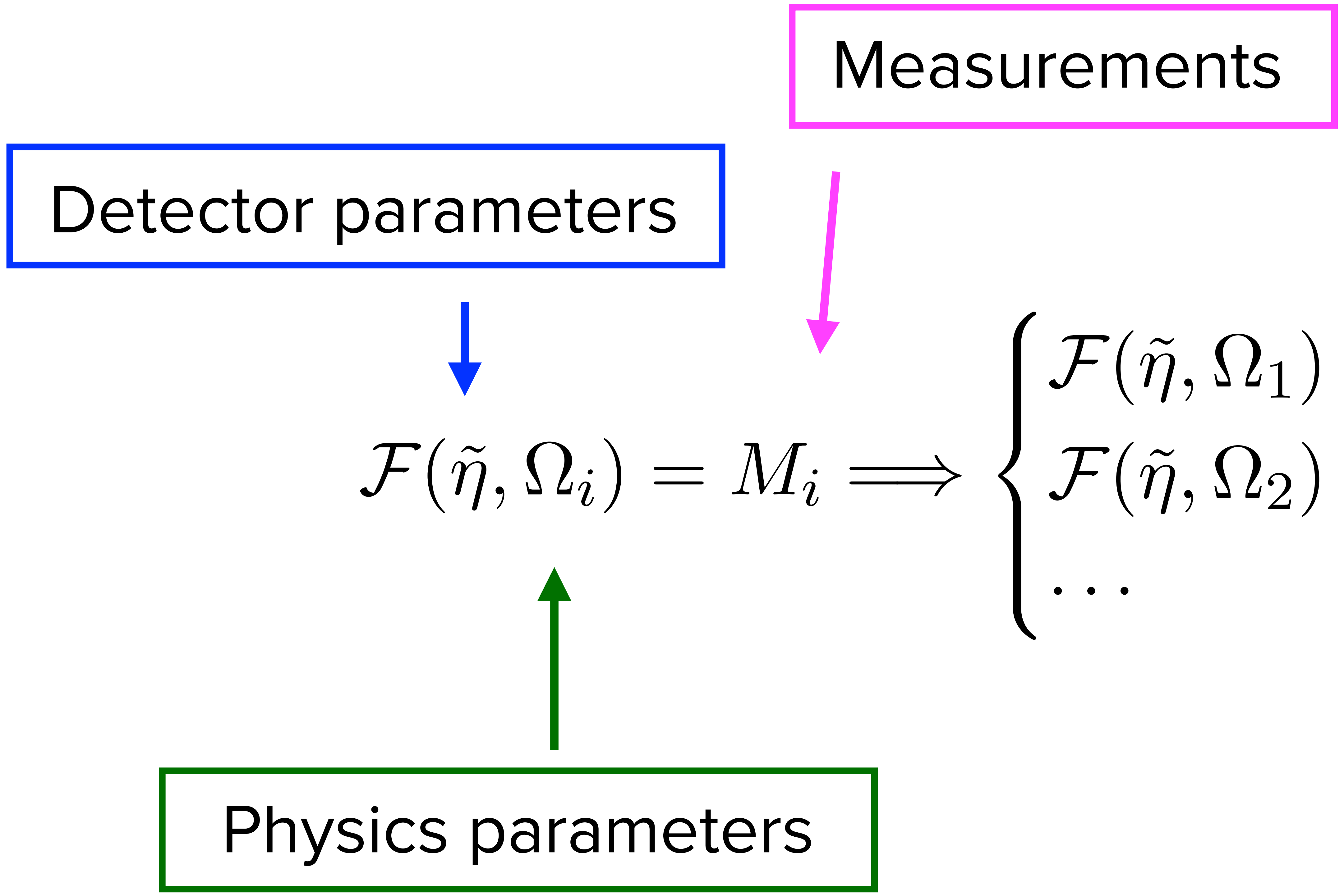
The traditional logic flow of a measurement



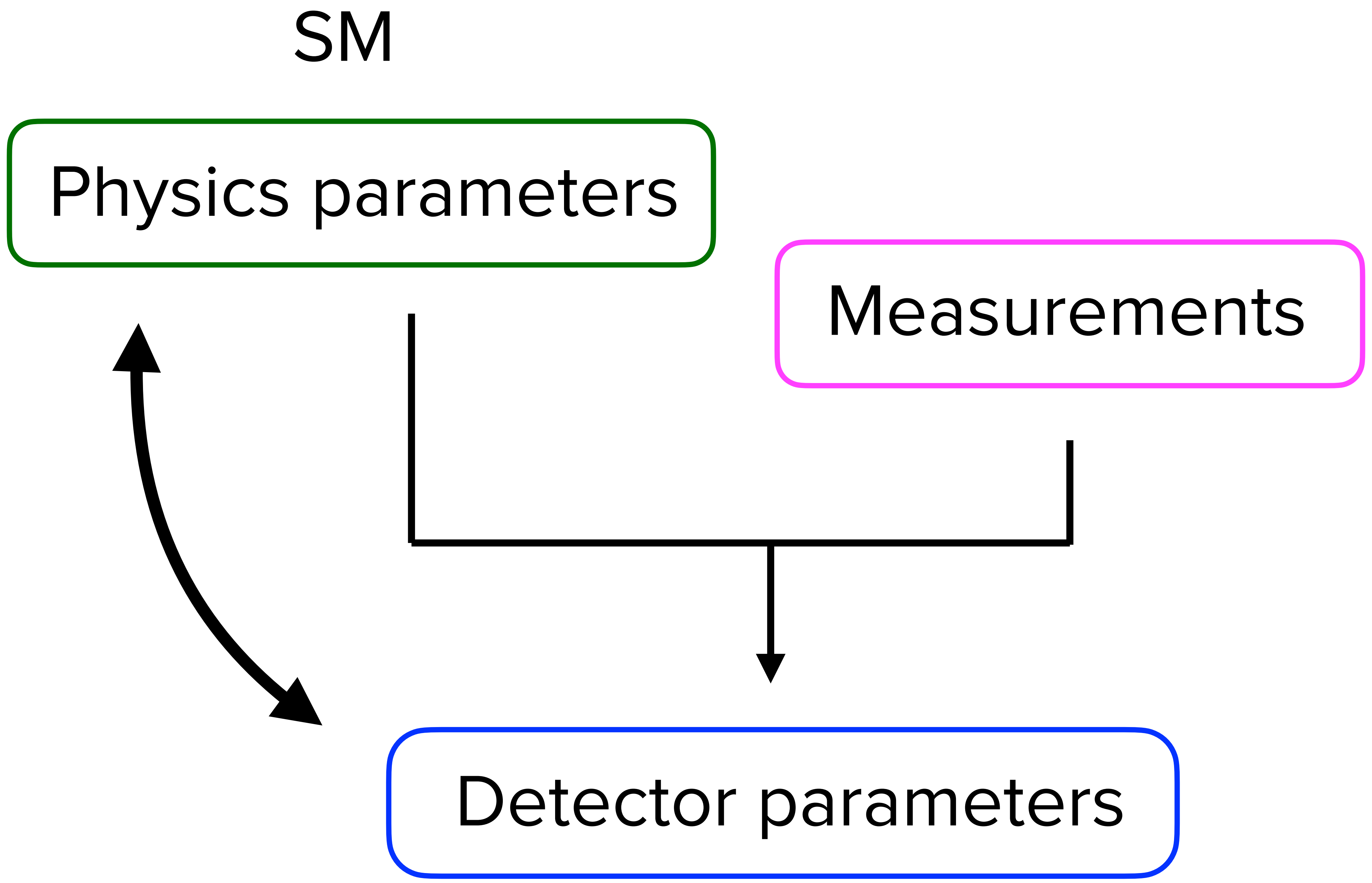


SM or NP?

The DLAdvocate logic flow



$$\mathcal{F}(\tilde{\eta}, \Omega_i) = M_i \implies \begin{cases} \mathcal{F}(\tilde{\eta}, \Omega_1) = M_1 \\ \mathcal{F}(\tilde{\eta}, \Omega_2) = M_2 \\ \dots \end{cases}$$



Can I explain an anomaly I see in the data by modifying the detector parameters?

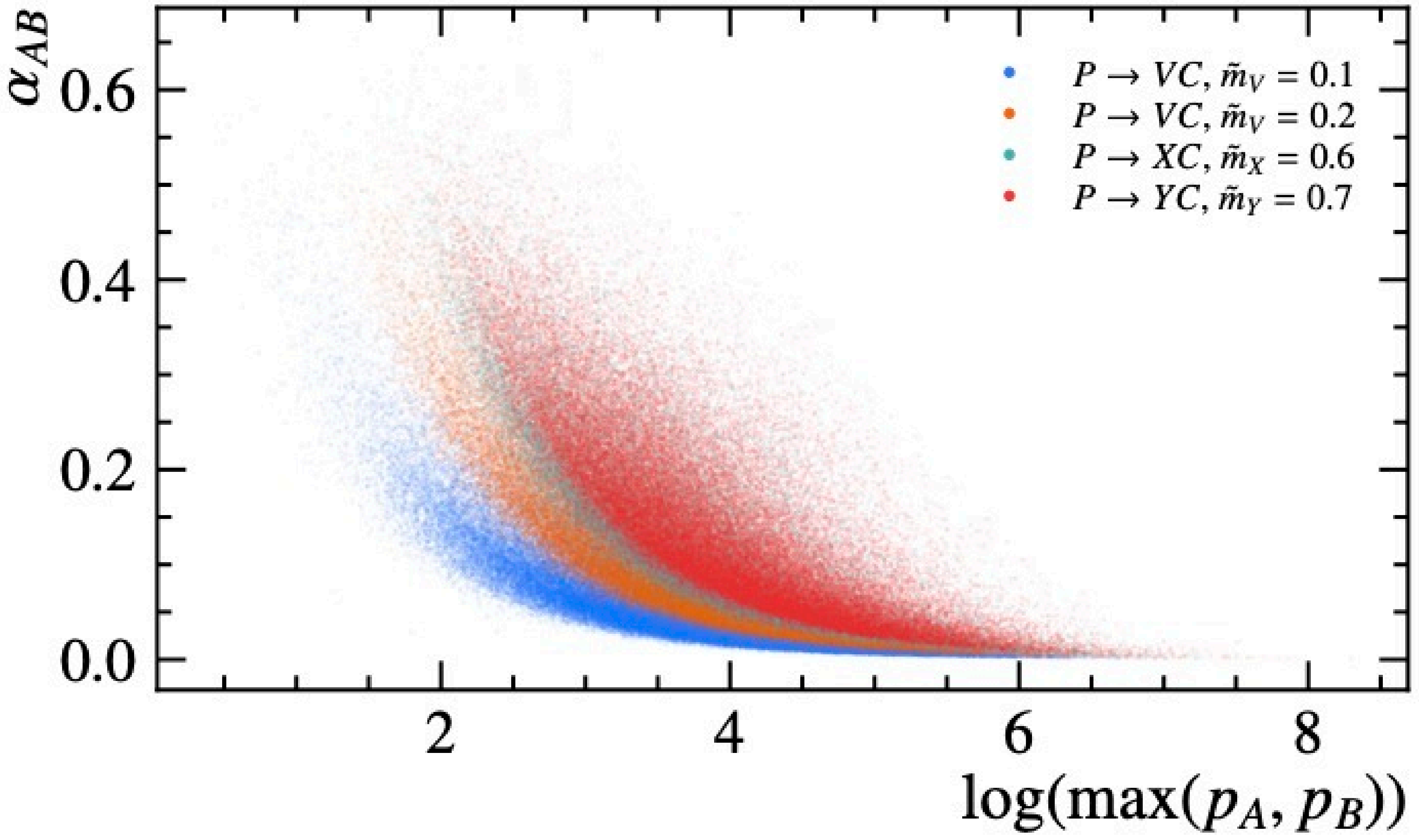
Playing the DL advocate: employ Deep Learning to systematically check all[*] possible effects

[*] For the moment we will focus on the detector efficiency

A simple example: a BR measurement

- Signal mode:
  - $P \to V(\to AB)C$ with mass m_V
- Control channel(s):
  - $P \to X(\to AB)C$
  - $P \to Y(\to AB)C$

  with known masses $m_{X(Y)}$ and known BR



Different masses → different kinematic!

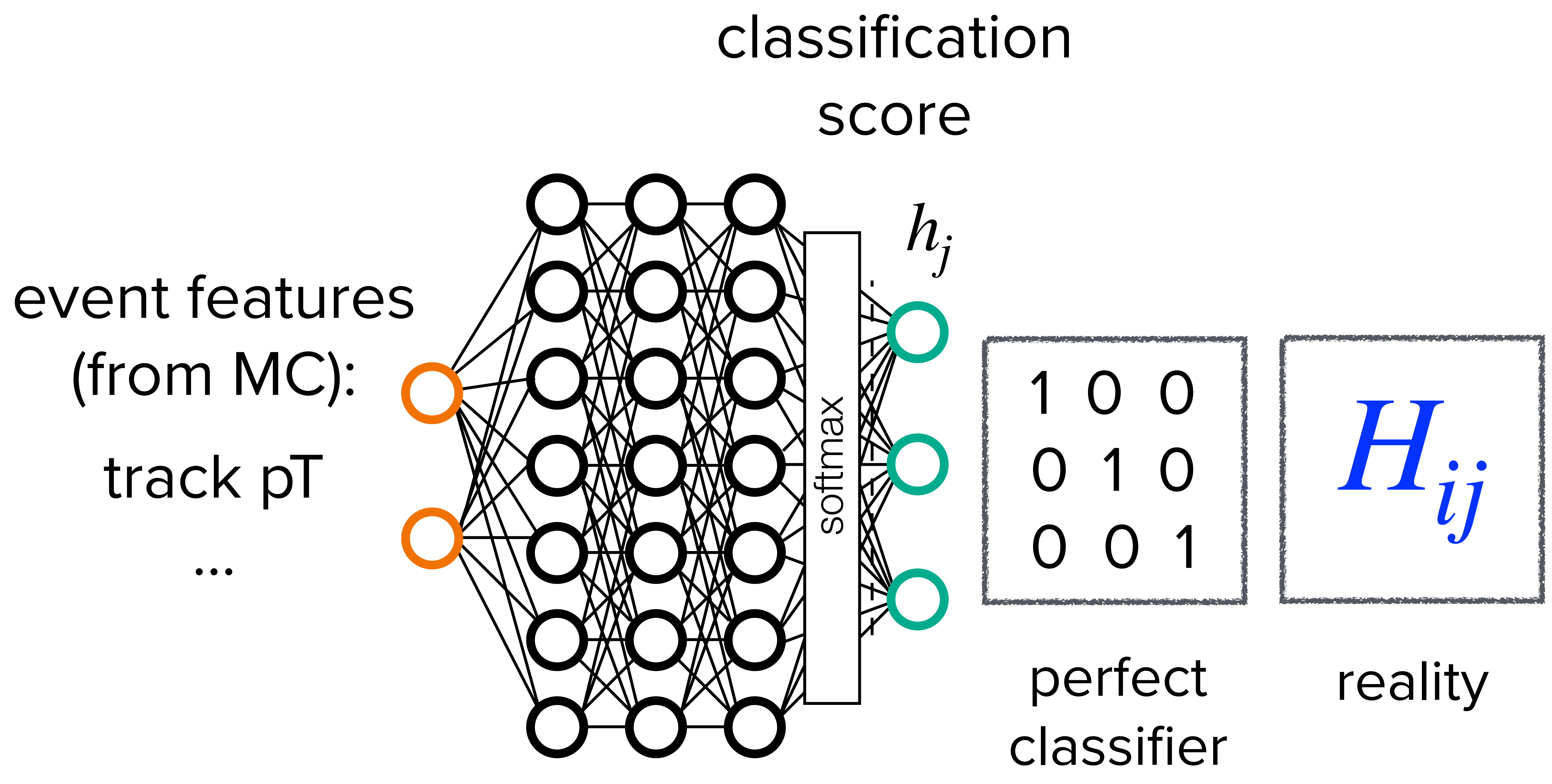
- Detector efficiency typically depends on kinematics (e.g. pT)
- A mismodelling of the efficiency will affect differently signal and control channels

How much a mismodelling of the efficiency can bias the signal given the constraints provided by the control channels?

Key idea - step 1

- Train a classifier to distinguish the different channels

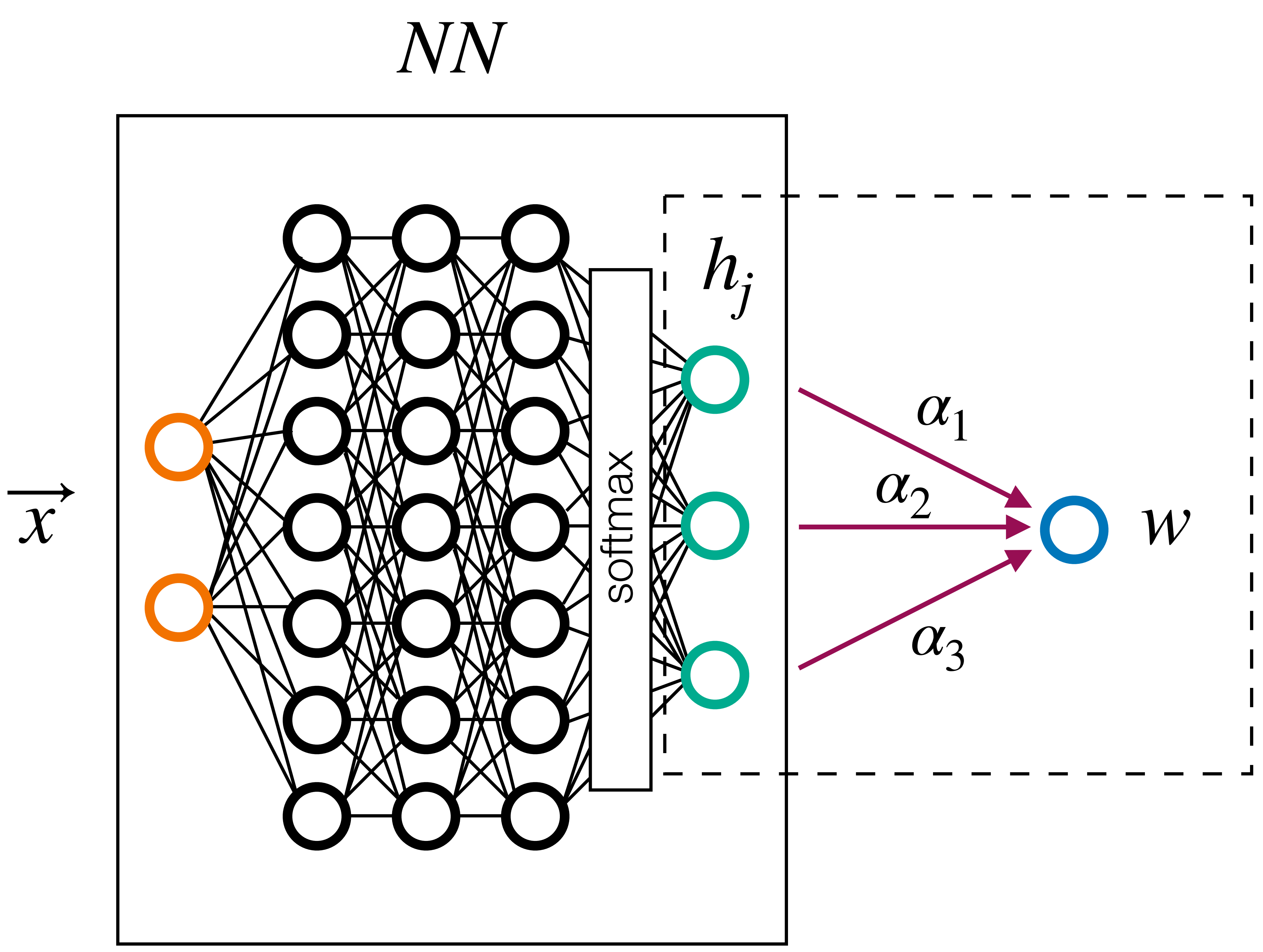
  - The “perfect” classifier would be able to completely separate the phasespace of the different channels
    - I can arbitrarily modify the efficiency to bias the signal without touching the control channels
    - control channels impose no constraints on the signal
  - Overlapping response will give the level of constraints provided by the different channels



Key idea - step 2

- Linear combination of NN output nodes to determine mismodelling weight as function of the input detector features

$$w(x_i) \begin{cases} = 1 & \text{perfect modelling of the efficiency} \\ < 1 & \text{efficiency over-estimated} \\ > 1 & \text{efficiency under-estimated} \end{cases}$$



Channel (mis-modelled) efficiency

$$e_i = \frac{1}{n_i} \sum_k w(x_{k,i})$$

Evaluated on MC sample

Key idea - step 2

- Line…
  to de…
  func…

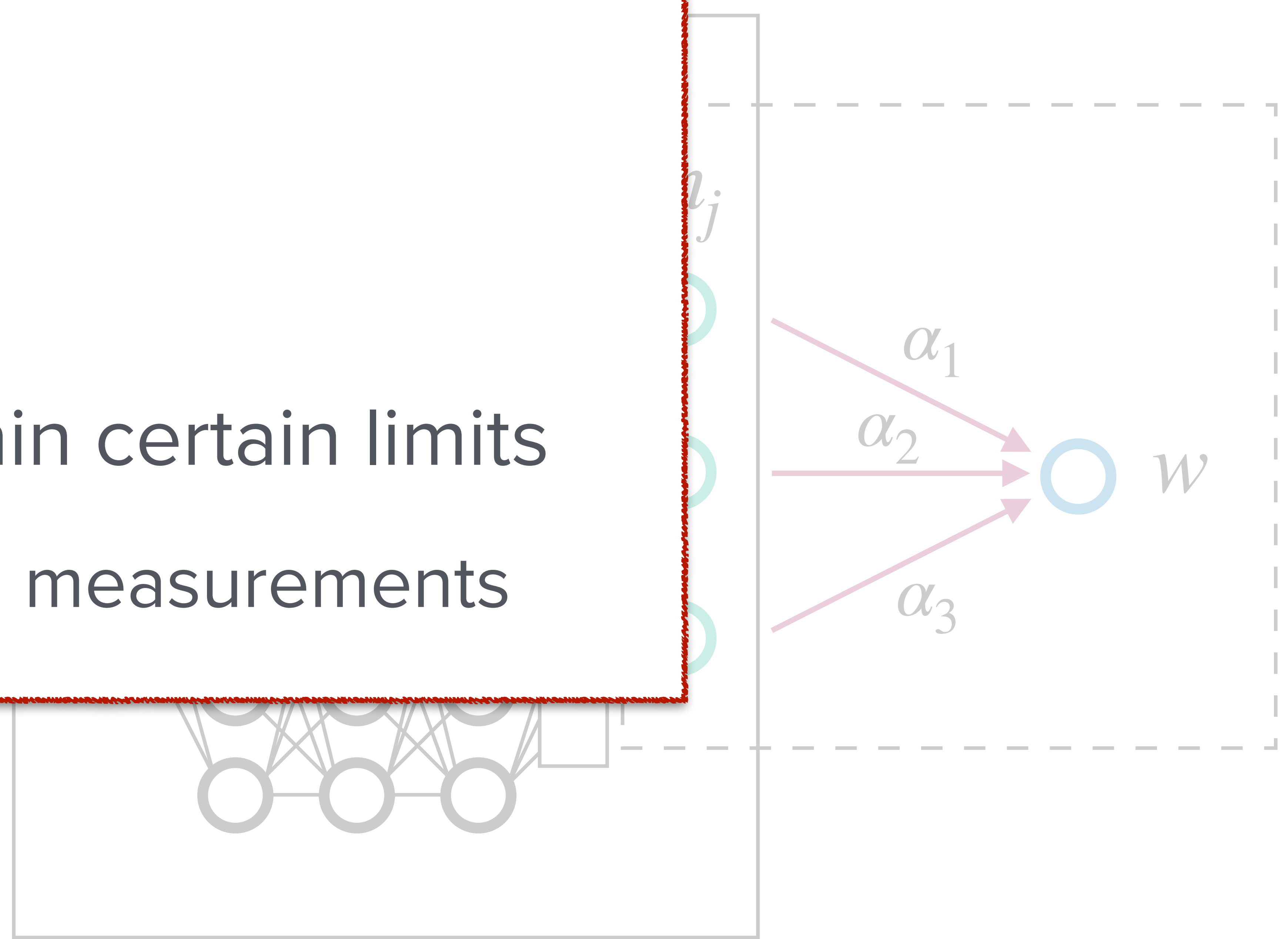
Goal of the algorithm:

Check how biased can be the signal efficiency

$$e_s \to \min$$

while keeping the control channel efficiency within certain limits

$$e_i \in [V_i^{low}; V_i^{high}]$$ ← from measurements



Channel efficiency

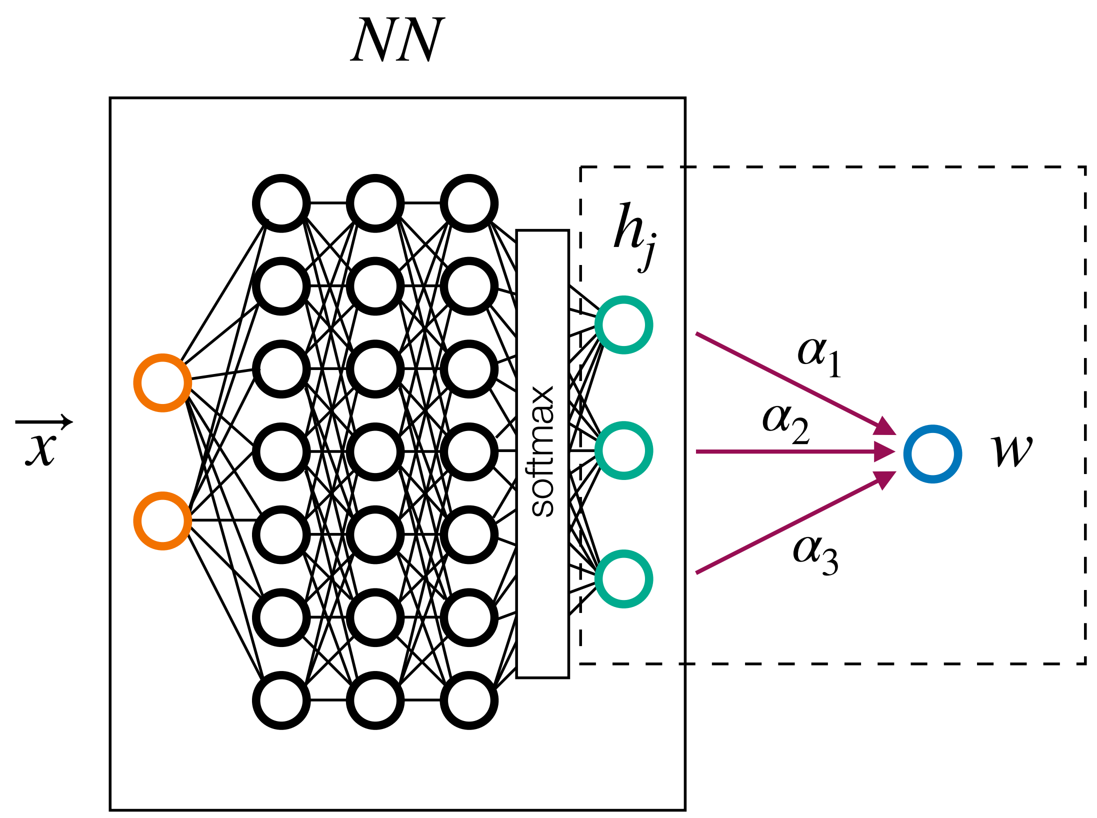
$$e_i = \frac{1}{n_i} \sum_k \vec{\alpha} \cdot \vec{h}(x_{k,i})$$

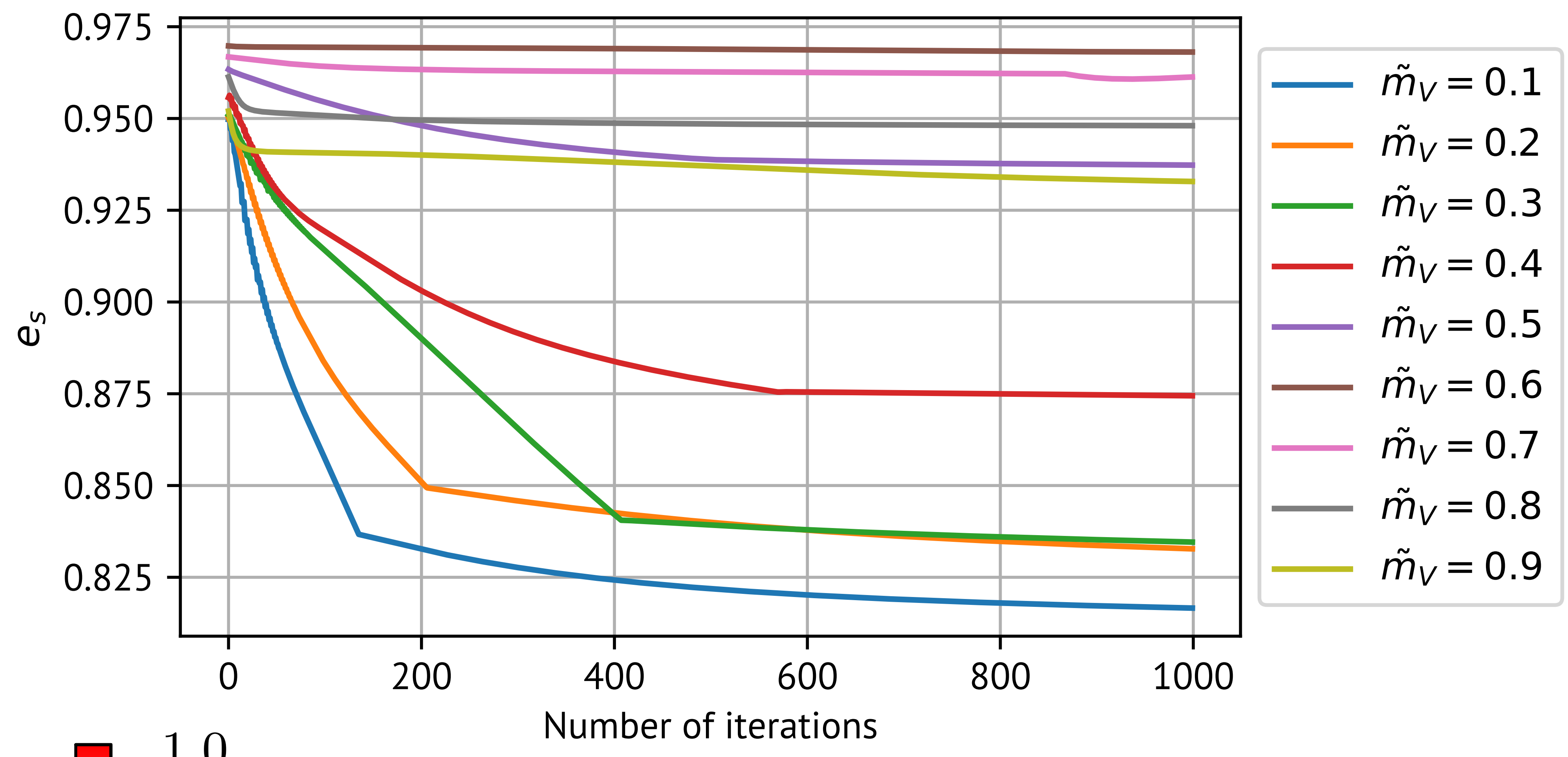
Evaluated on MC sample

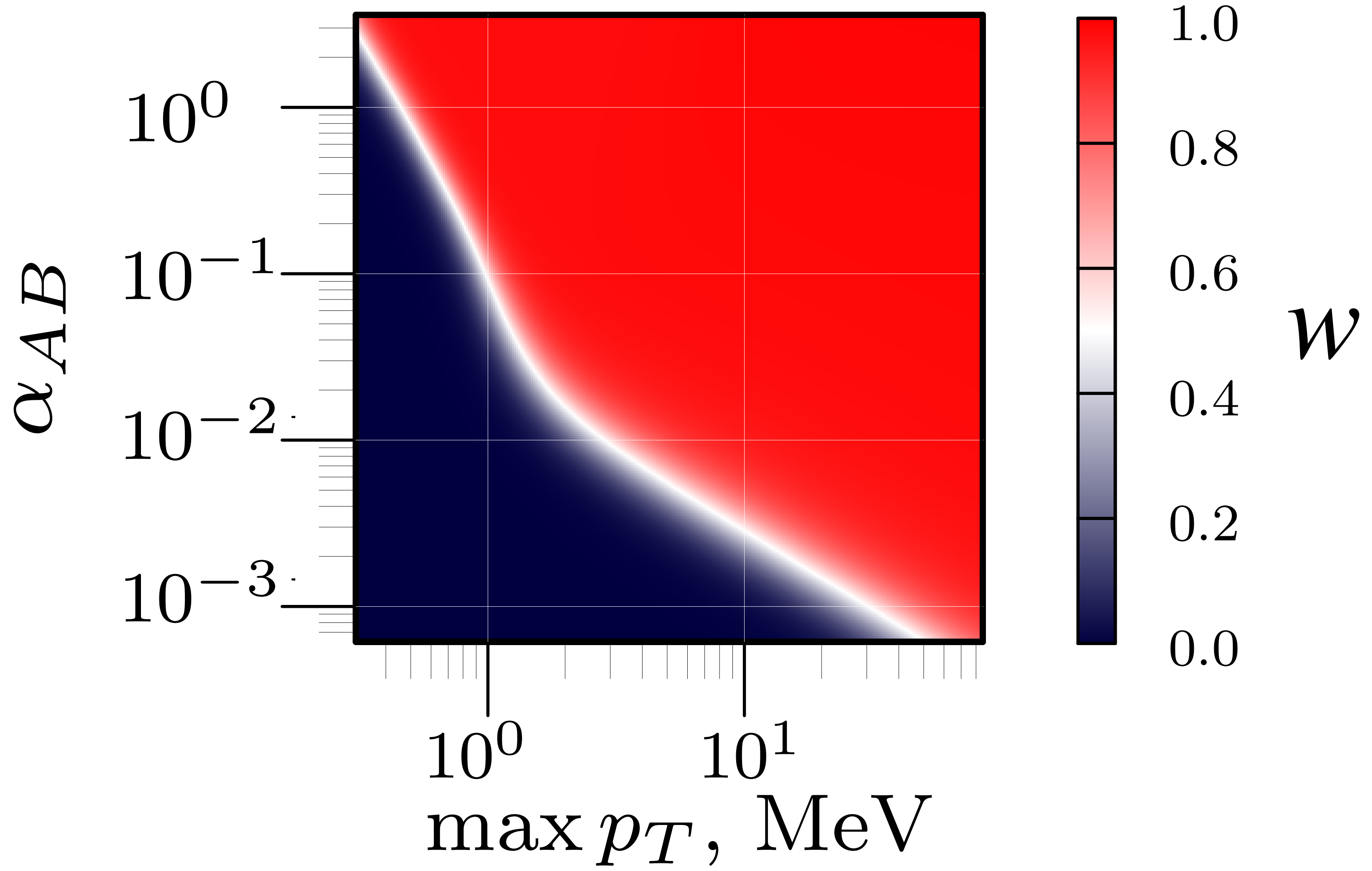
Training

- Iterative procedure:
  0. NN pretrained as a pure classifier
  1. update $\overrightarrow{\alpha}$
     - simple minimization with constraints
  2. update NN parameters
     - $\ell(\theta) = e_s - \log \left|\det(H)\right|$

       keeps matrix invertible







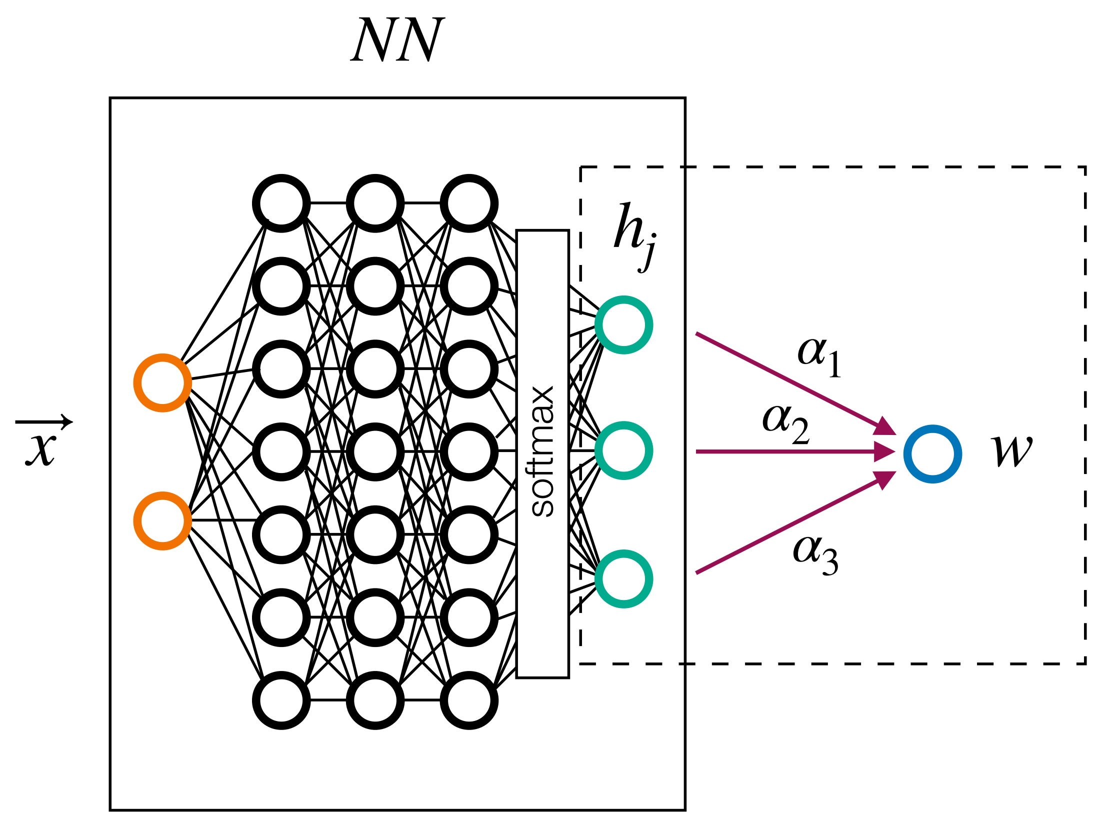
Training

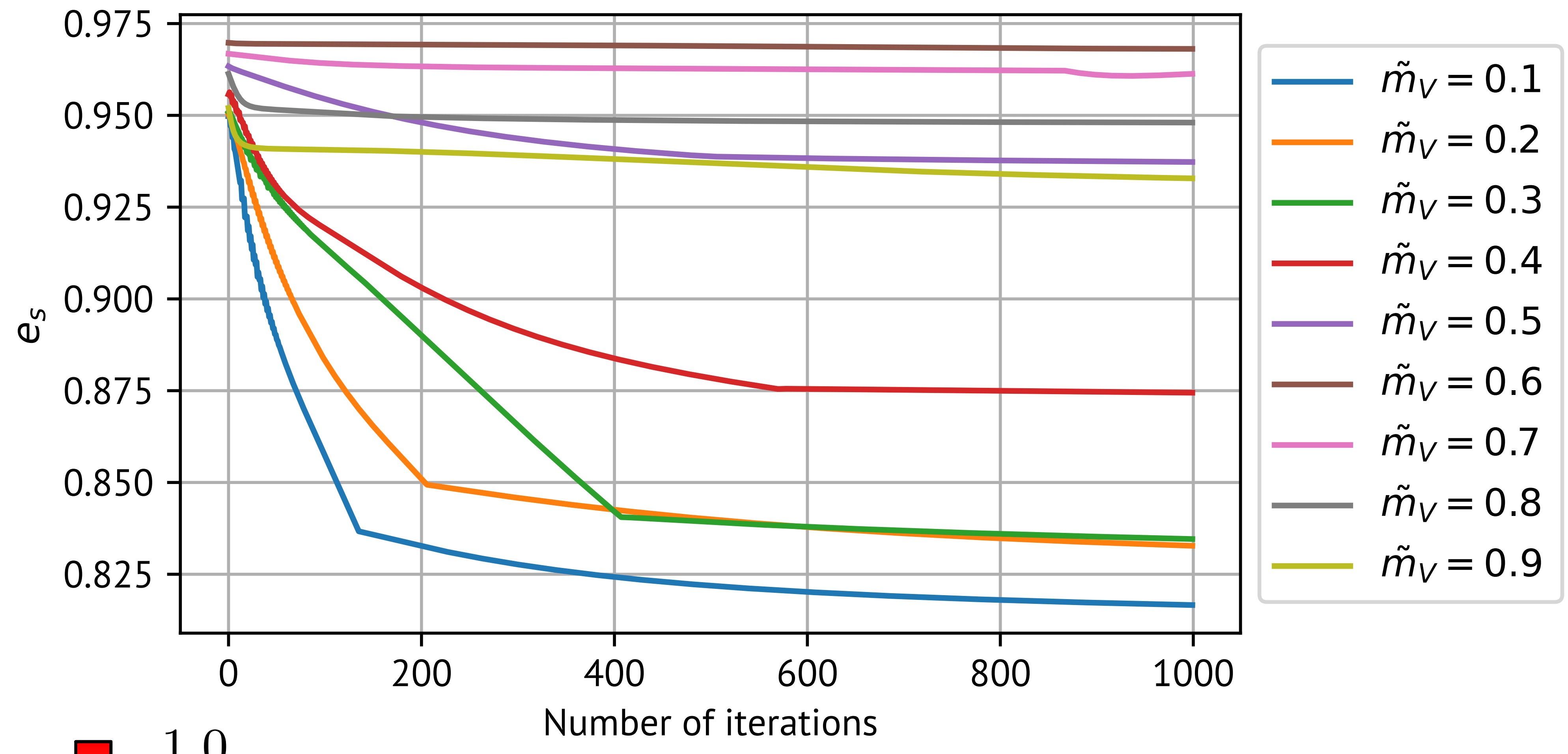
- Iterative procedure:
  0. NN pretrained as a pure classifier
  1. update $\overrightarrow{\alpha}$
     - simple minimization with constraints
  2. update NN parameters
     - $\ell(\theta) = e_s - \log \left|\det(H)\right| + \ell_g$

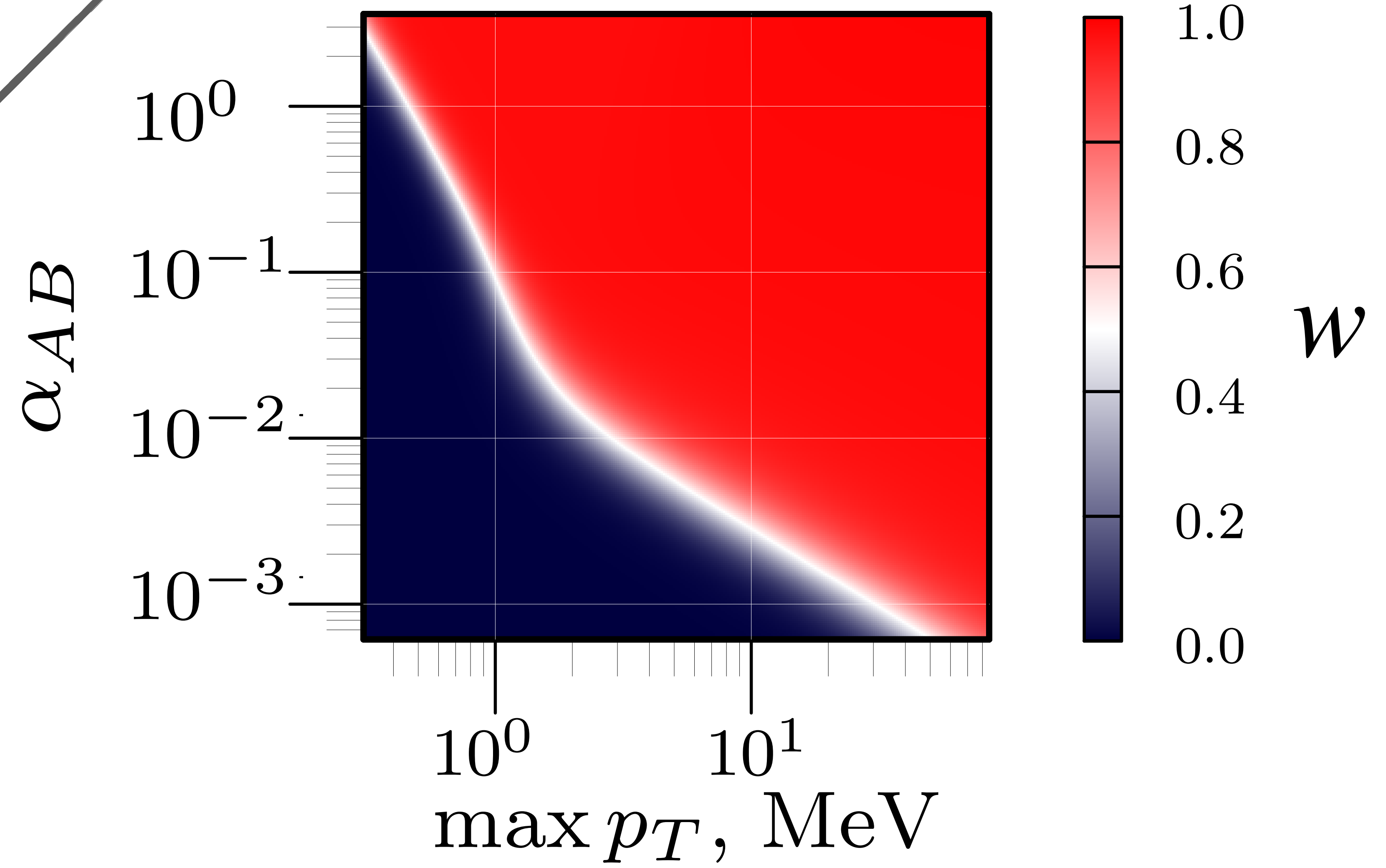
keeps matrix invertible

regulariser

$$\ell_g(\theta) = \sum \left[\max\left(\frac{||\nabla \overrightarrow{h}(x_k, \theta)||}{p} - 1, 0 \right) \right]^2$$





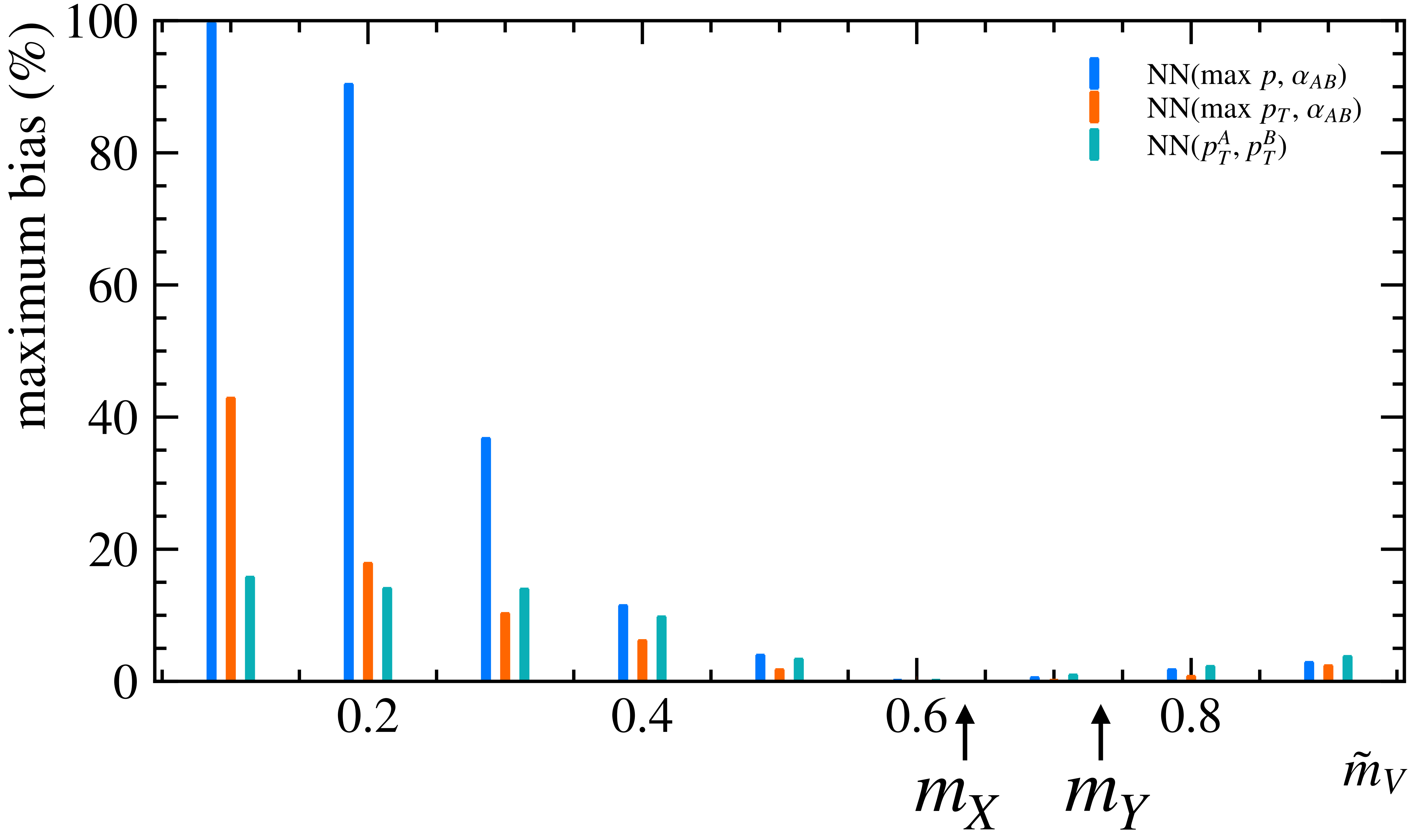


A simple example: results

- Target measurement of $\mathcal{B}(P \to VC)$ as function of m_V
- Control channels:

$$\mathcal{B}(P \to XC) \propto e_{P\to XC} \in [-3\%, 3\%]\,,$$

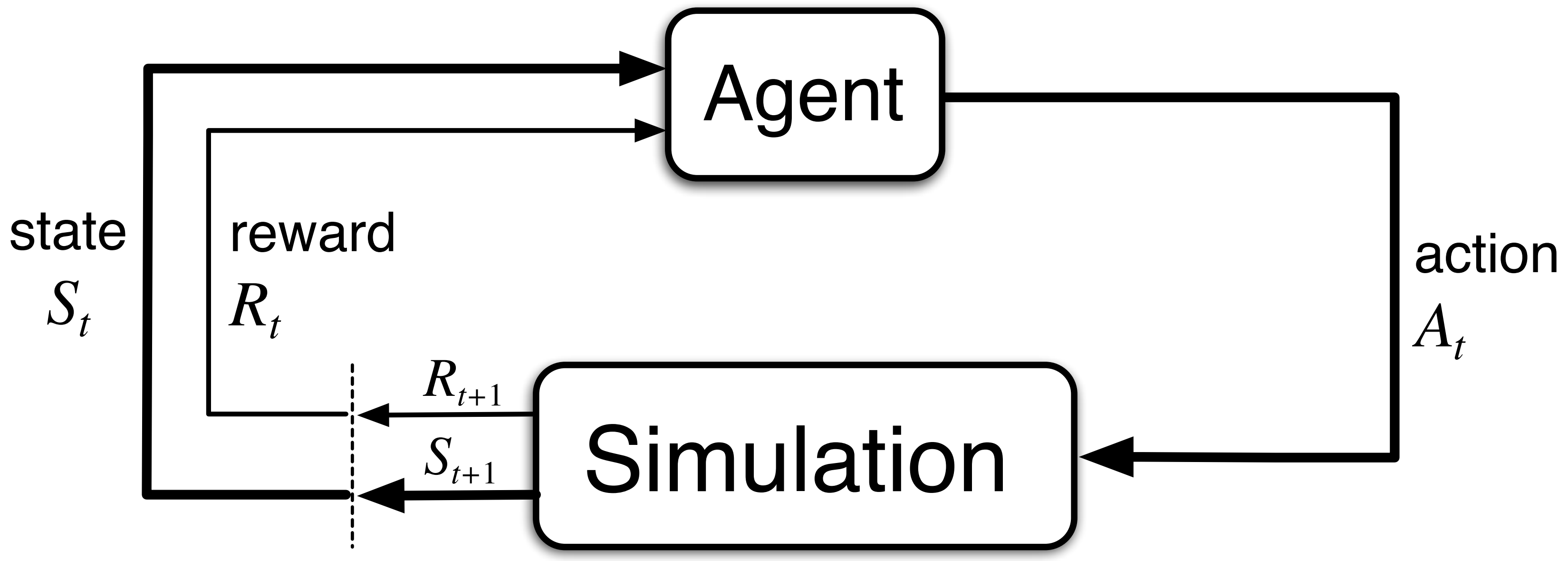
$$\frac{\mathcal{B}(P \to YC)}{\mathcal{B}(P \to XC)} \propto \frac{e_{B\to YC}}{e_{P\to XC}} \in [-1\%, 1\%]$$



- As expected, maximum allowed bias depends on the mass (kinematic overlap) between signal and control channels → But quantifiable now!

Going low level…

- So far, only considered reconstructed quantities (high levels)
- However, everything that happens in the detector happens at low level
  - Hits, energy deposit, material interaction, etc.
- MC simulation cannot be described in a parametric way
  - Requires a different formulation of the problem
  - Interactive tuning of the simulation → RL ?
    - Tested (with high level quantities) on an other example of flavour physics (angular analysis of rare B decay)



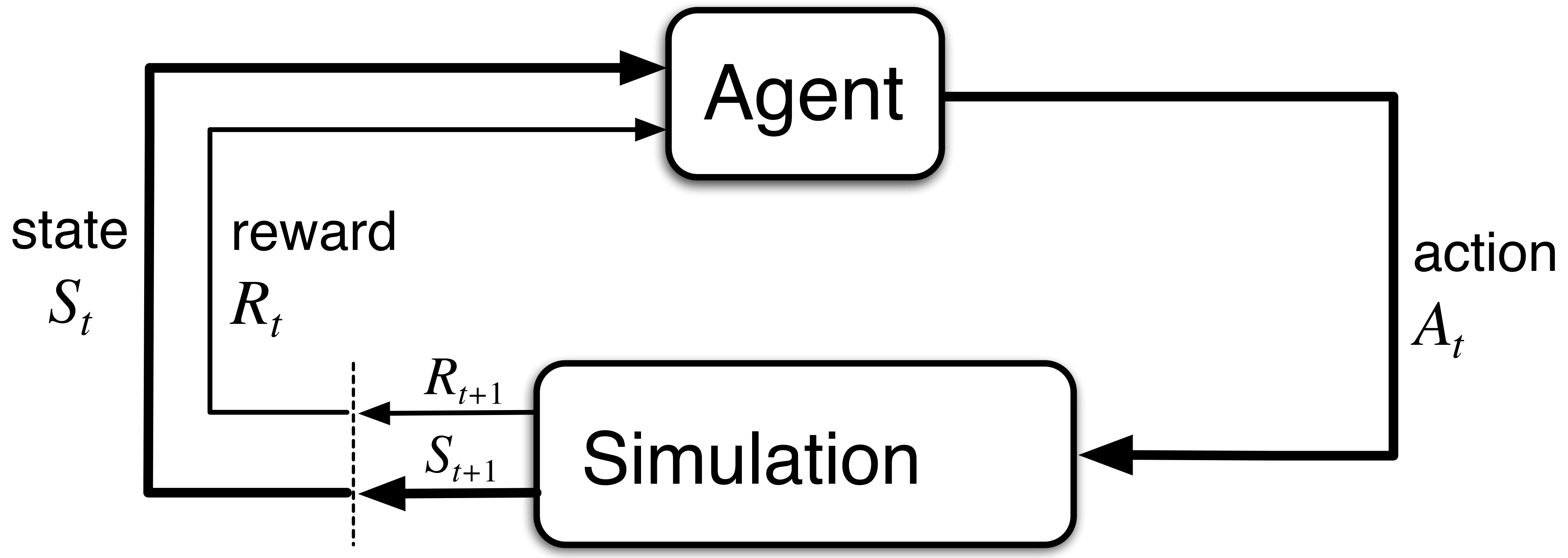
Conclusions & future work

- Presented method to systematically investigate potentially hidden systematics
- Focused on the efficiency aspect of a measurement
  - Tested on a simple example
  - Fully general: can be extended to any measurement that relies on simulation!
  - Full potential when applied to low lever features
- Expand to all aspects of a physics analysis
  - Background contamination (work in progress…see Guillermo's poster on Thursday)

Thank you!

Back up

RL approach



$$r = 0.01 \times \begin{cases} -\chi^2/N_{\text{meas}} & \text{if } \chi^2/N_{\text{meas}} > 3, \\ -\chi^2/N_{\text{meas}} + 10\,(3 - \chi^2/N_{\text{meas}}) & \text{if } \chi^2/N_{\text{meas}} \in [0.1,\ 3], \\ 10^3 & \text{if } \chi^2/N_{\text{meas}} < 0.1, \end{cases}$$

$$\chi^2 \equiv \sum_{i=1}^{N_{\text{meas}}} \left(\frac{M_i - \mu_i}{\sigma_i}\right)^2$$