Amplitude Uncertainties Everywhere All at Once (The Return of the Ensemble)

Nina Elmer

ACAT 2025

arXiv:2509.00155

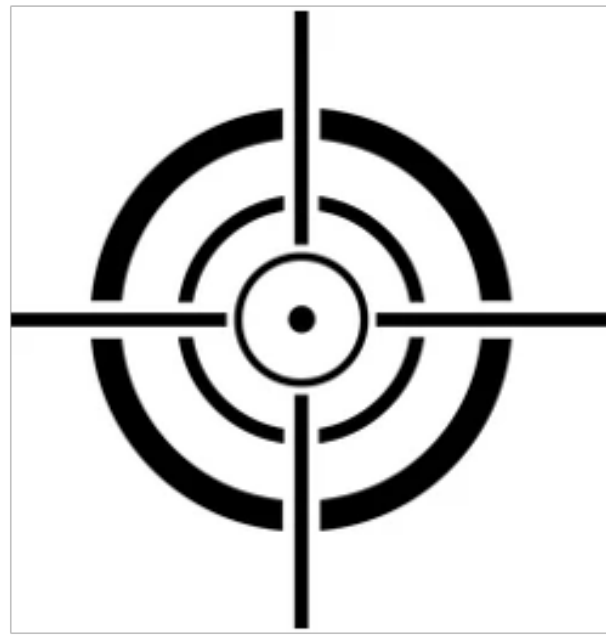
with H. Bahl, R. Winterhalder and T. Plehn

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Collaborative Research Center TRR 257
Particle Physics Phenomenology after the Higgs Discovery

IMPRS for Precision Tests of Fundamental Symmetries
INTERNATIONAL MAX PLANCK RESEARCH SCHOOL

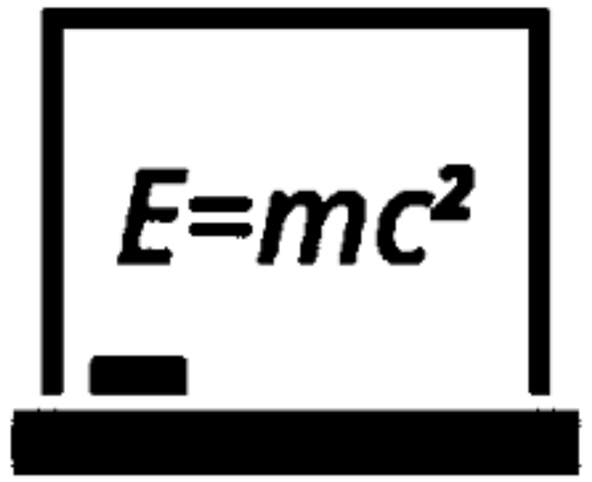
The four horsemen of particle physics



precision

control

speed



physics knowledge

ML and amplitude surrogates

Exact physics computations:

- ✓ Precision
- ✓ Full control of uncertainties
- ✓ Physics Knowledge
- ✗ Computational speed

ML approach:

ML and amplitude surrogates

Exact physics computations:

- ✓ Precision
- ✓ Full control of uncertainties
- ✓ Physics Knowledge
- ✗ Computational speed

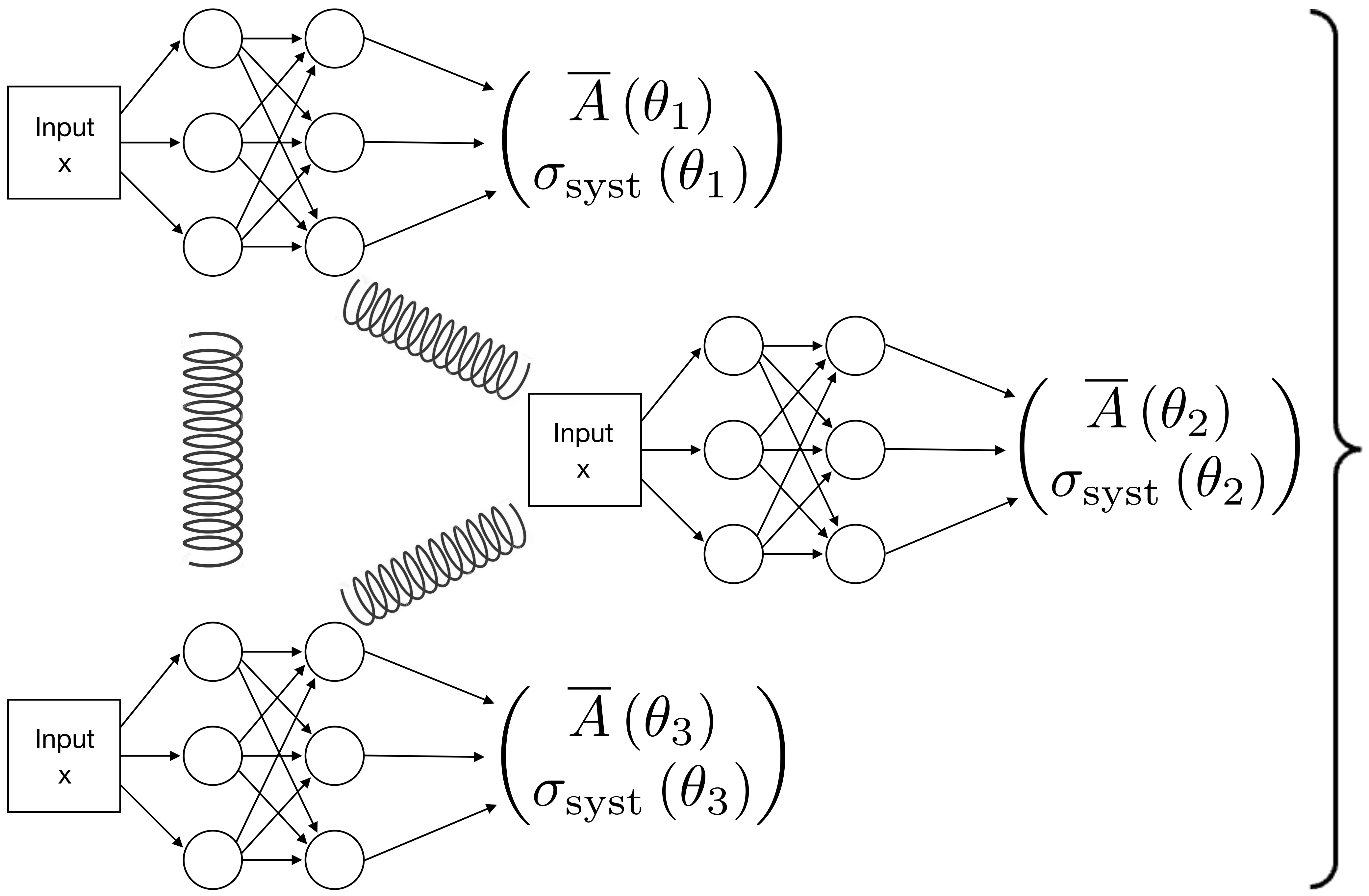
ML approach:

- ✓ Speed
- ~ Precision → Previous work (2412.12069) ✓
- ~ Physics knowledge ✓
  → Include Invariants, L-GATr
  (arXiv: 2411.00446 and 2412.12069)
- ~ Control of uncertainties → This presentation

→ See also Javier Mariño Villadamigo's and Lorenz Vogel's talk

Repulsive ensemble (RE)

Ensemble of networks **Output**



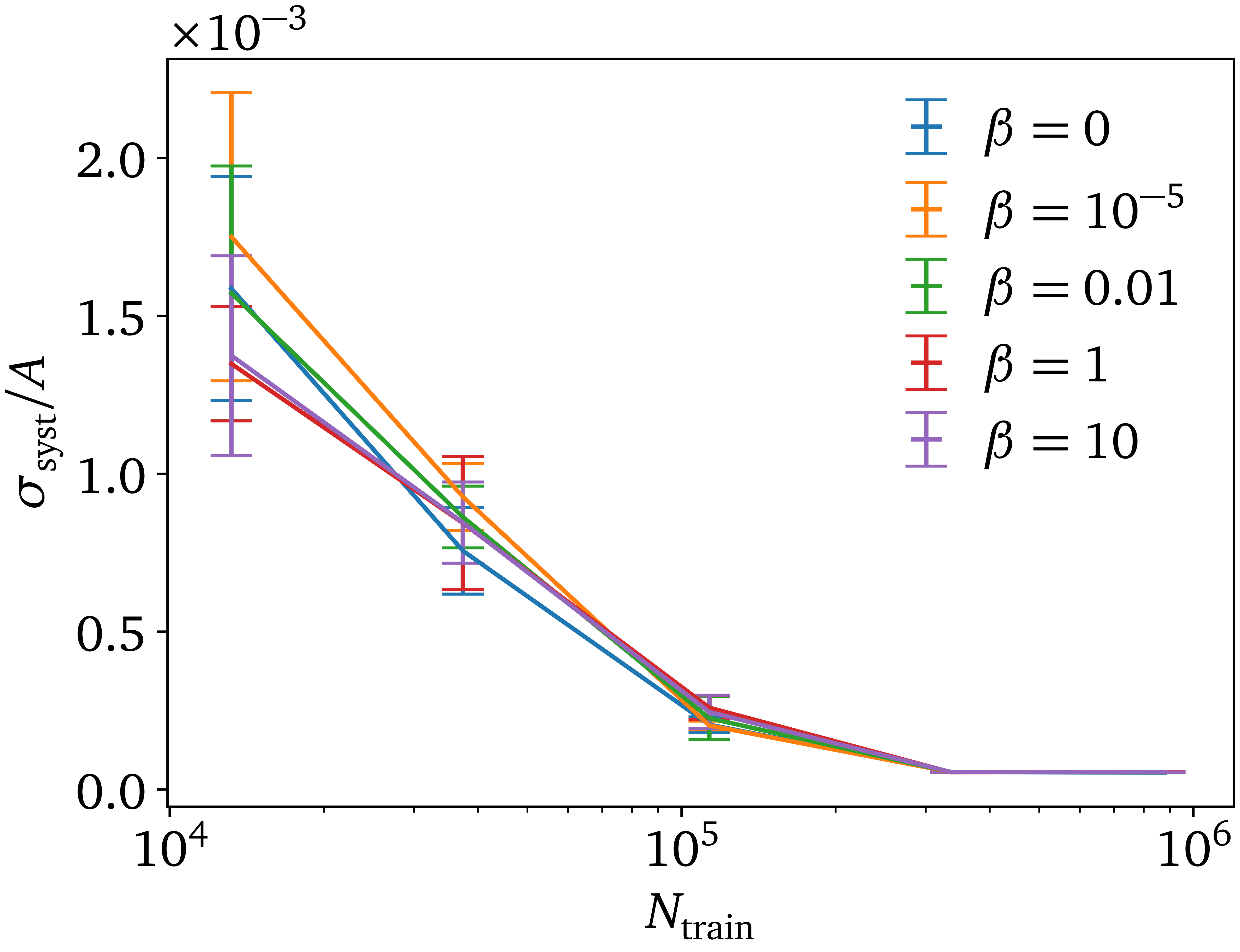
$$\langle A \rangle = \frac{1}{N} \sum_i^N \overline{A}(\theta_i)$$

$$\sigma^2_{\text{syst}} = \frac{1}{N} \sum_i^N \sigma^2_{\text{syst}}(\theta_i)$$

$$\sigma^2_{\text{stat}} = \frac{1}{N} \sum_i^N \left(\langle A \rangle - \overline{A}(\theta_i) \right)^2$$

- Repulsive term: Cover full posterior distribution
- Systematic uncertainty from **heteroscedastic loss** part
- Ensemble members **trained simultaneously**

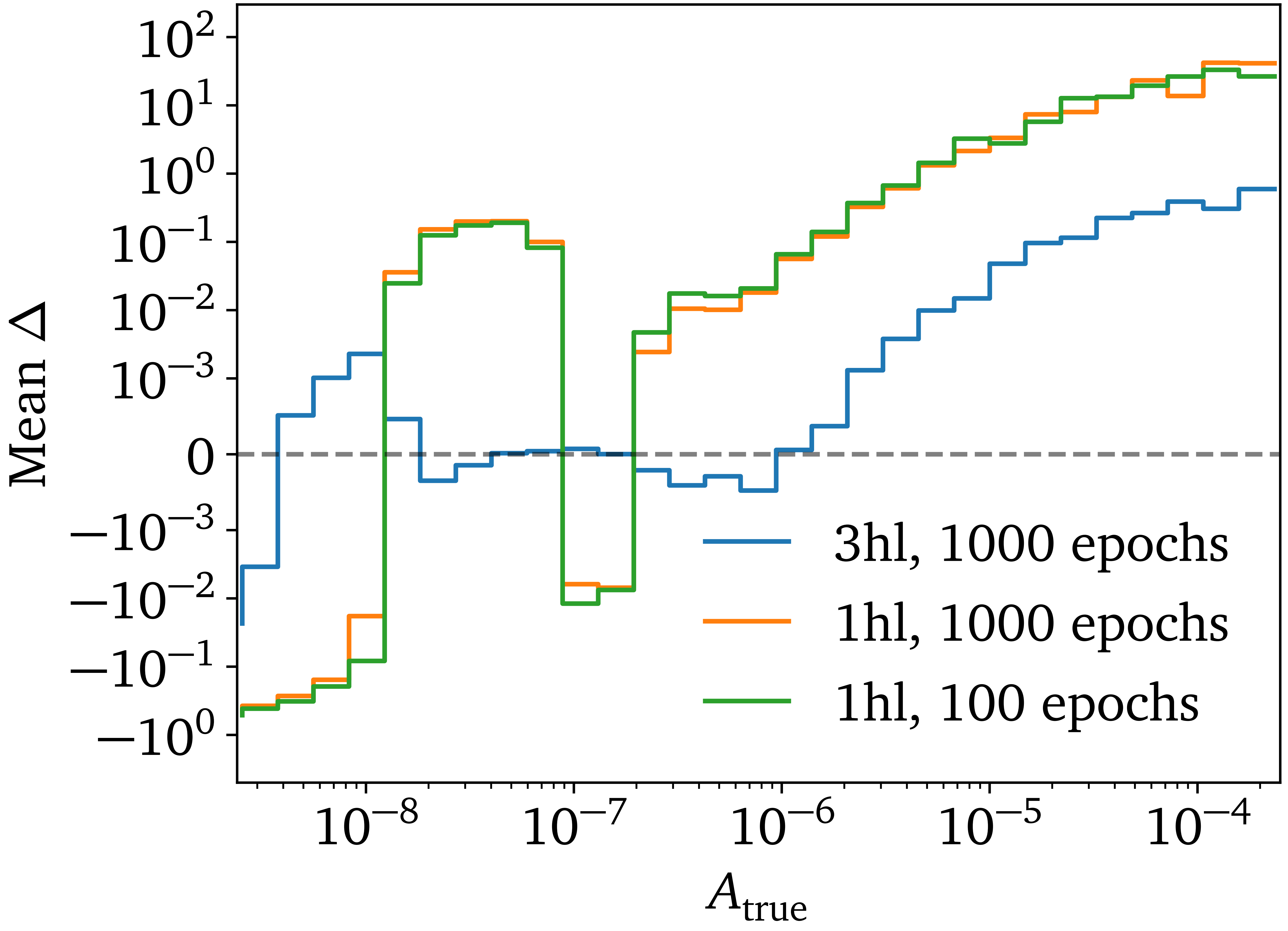
How repulsive are repulsive ensembles?



- Spread vanishes with more training data
- Similar behaviour for statistical uncertainty

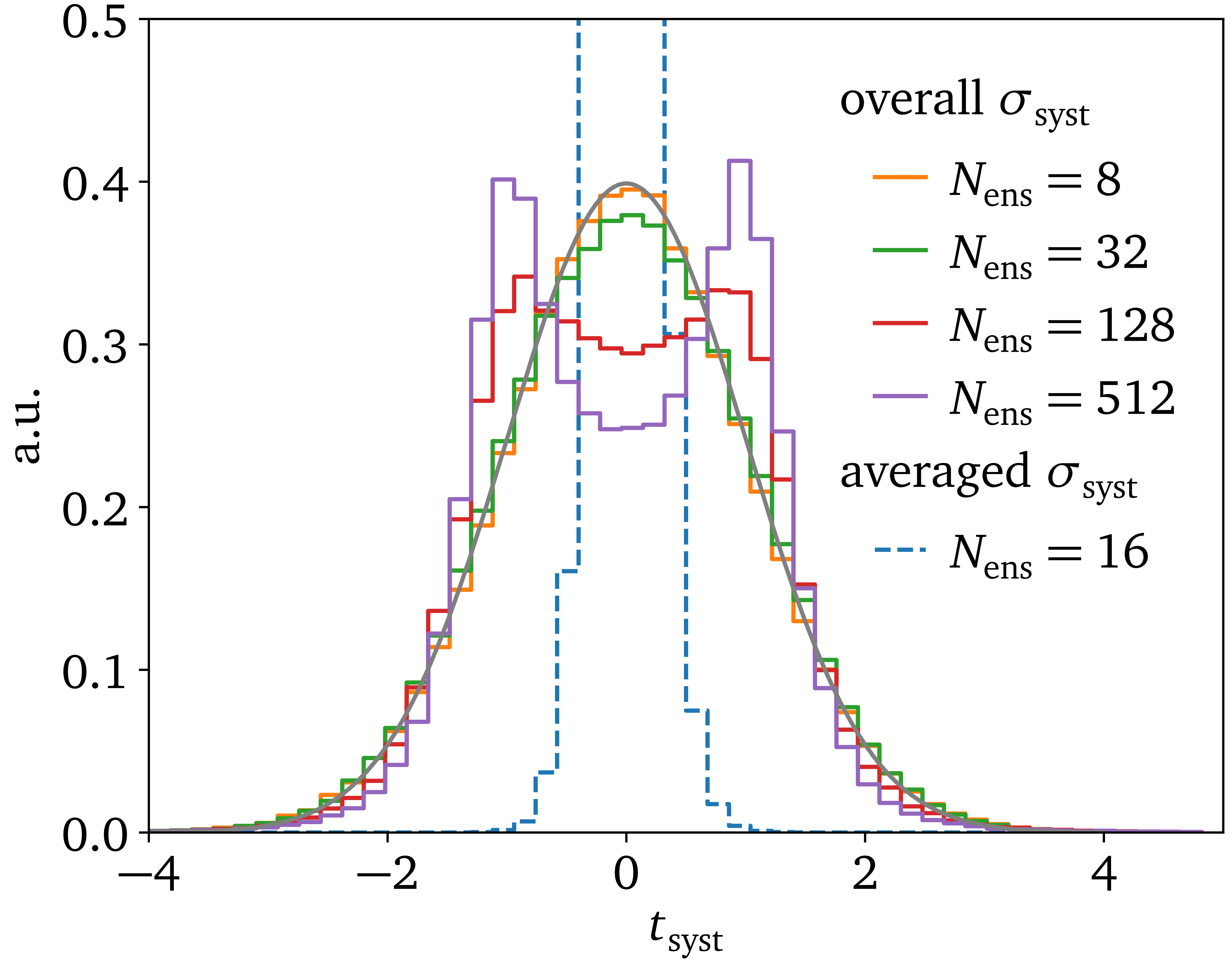
➡ Repulsive kernel impactful for **small training data** sets

Ensembles against biases?



- Training duration: No impact
- **More expressivity**: Smaller bias
- Does ensembling help with this bias?

Global uncertainty estimation as different approach



- Individually σ_{syst} per member: Underconfident
- Introduce **global systematic unc**:
  Train additional NN to predict global σ_{syst}
- Large N_{ens}: Reduced noise, discover bias

➡ Two-mode structure

How to: Evidential regression

- Previously: Systematic uncertainty from loss: $\mathscr{L}_{\text{heteroscedastic}} = \sum_i \frac{|f(x_i) - f_\theta(x_i)|^2}{2\sigma(x_i)^2} + \log \sigma(x_i) + \ldots$

- Different approach: Introduce conjugate prior $p(\lambda \,|\, m)$, λ likelihood parameter
  - Likelihood $\lambda = (f_\theta(x_i), \sigma)$ and evidential parameters $m(x, \theta) = \{\gamma, v, \alpha, \beta\}(x, \theta)$

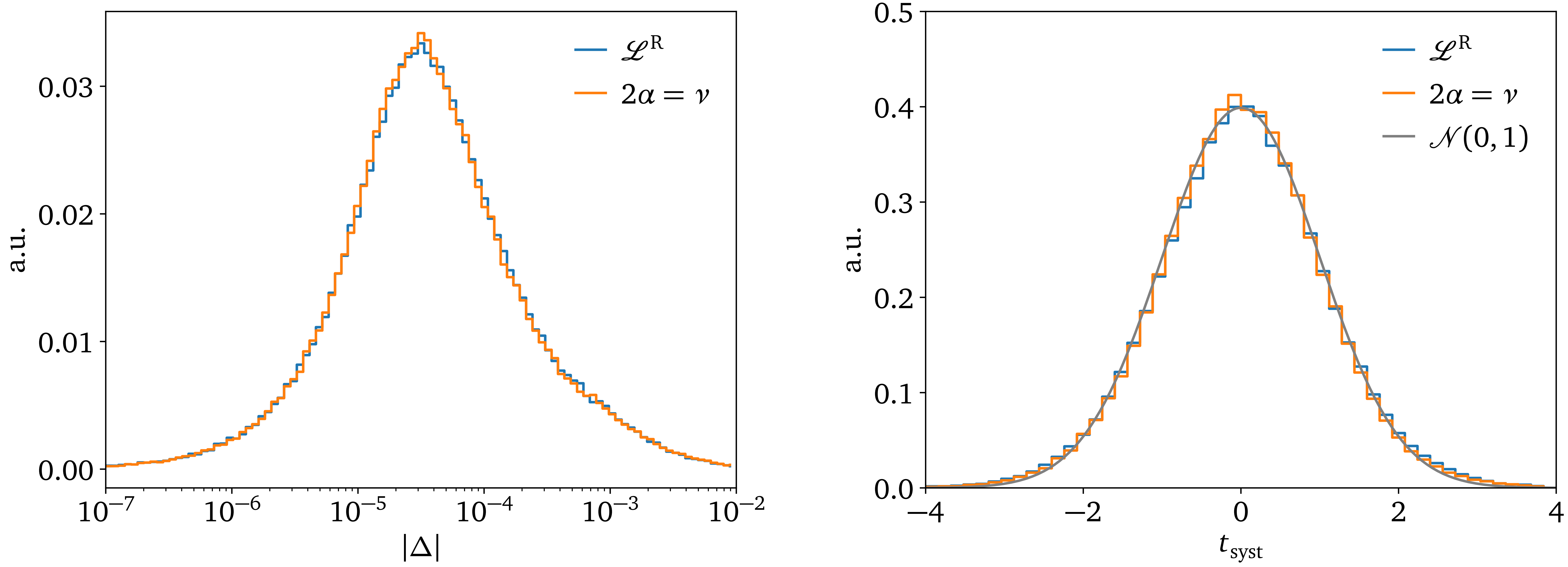
- New likelihood: $p(A|x) \approx \int \mathrm{d}\lambda \; p(A|x, \lambda) p(\lambda|m)$

  → follows Normal-Inverse-Gamma distribution

- Analytic likelihood: $p(A|x, m) = \mathrm{St}\left(A \middle| \gamma, \frac{\beta(1+v)}{v\alpha}, 2\alpha\right)$

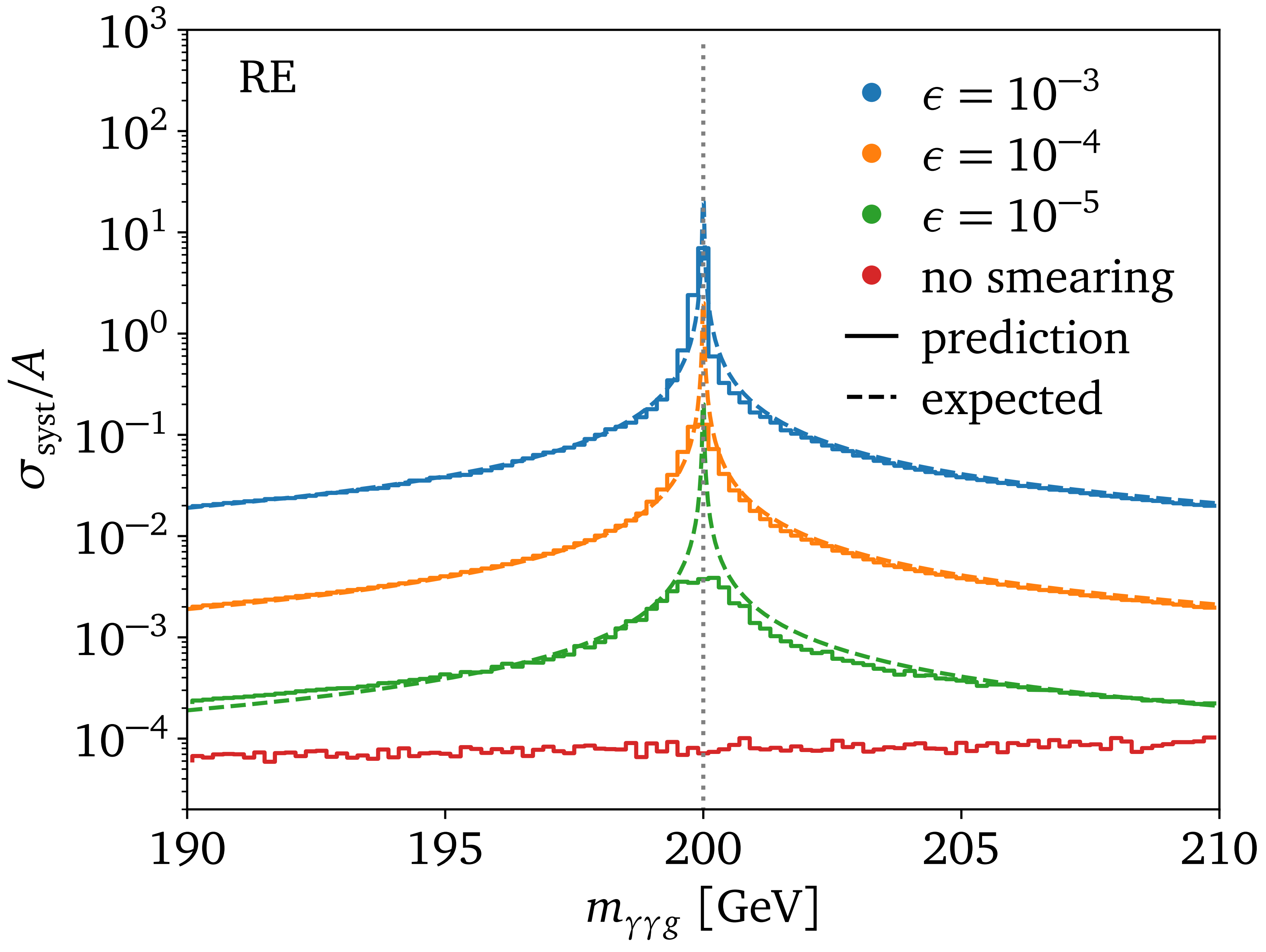
Evidential regression

- Mean and uncertainties: Averaging over $p(\lambda \mid m)$ → Analytical **closed form**
- No MC sampling necessary
- Integrals **directly solvable** → Uncertainties and mean without sampling accessible



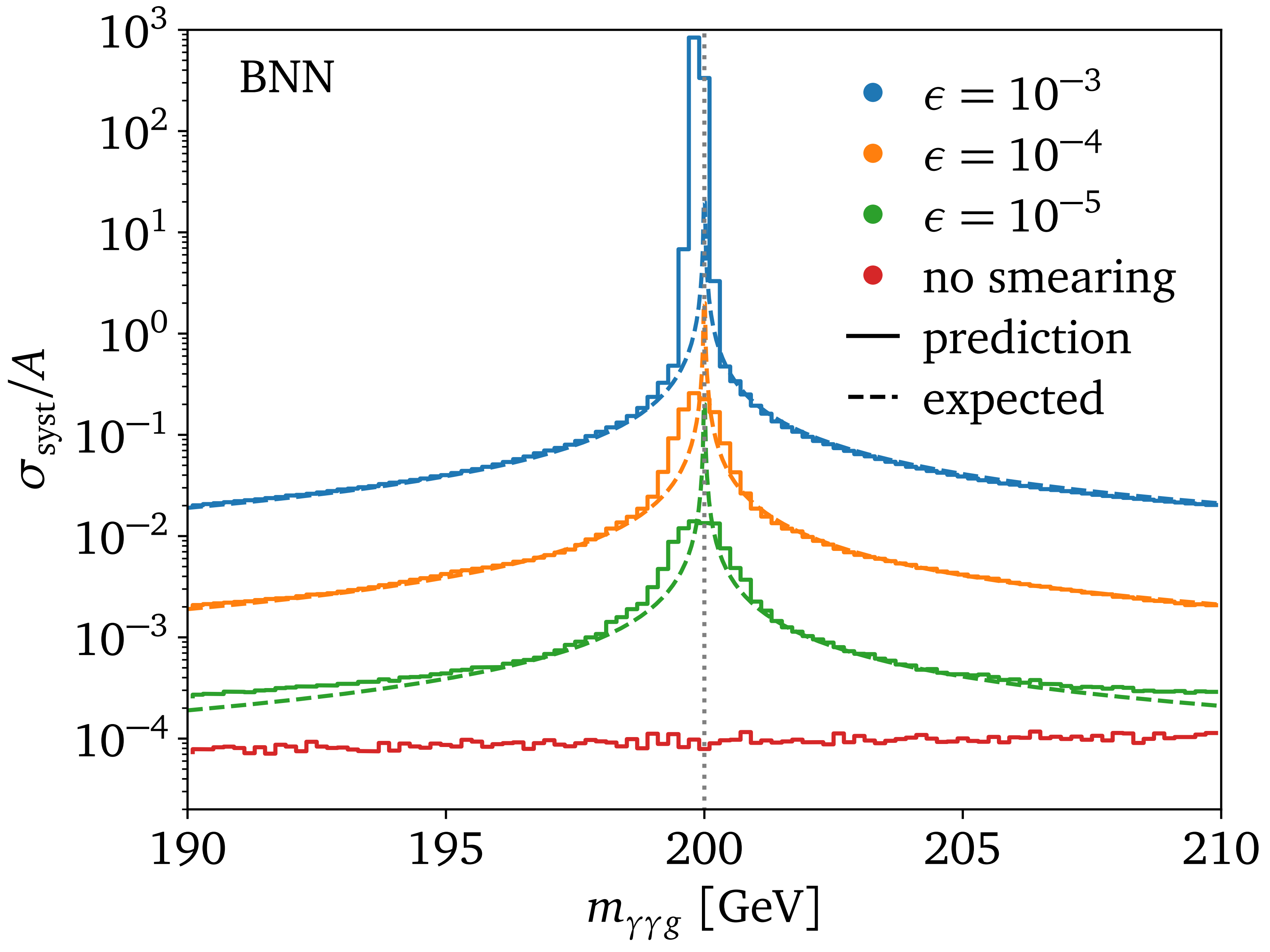
Closer to reality: Peaked threshold smearing

Repulsive ensemble



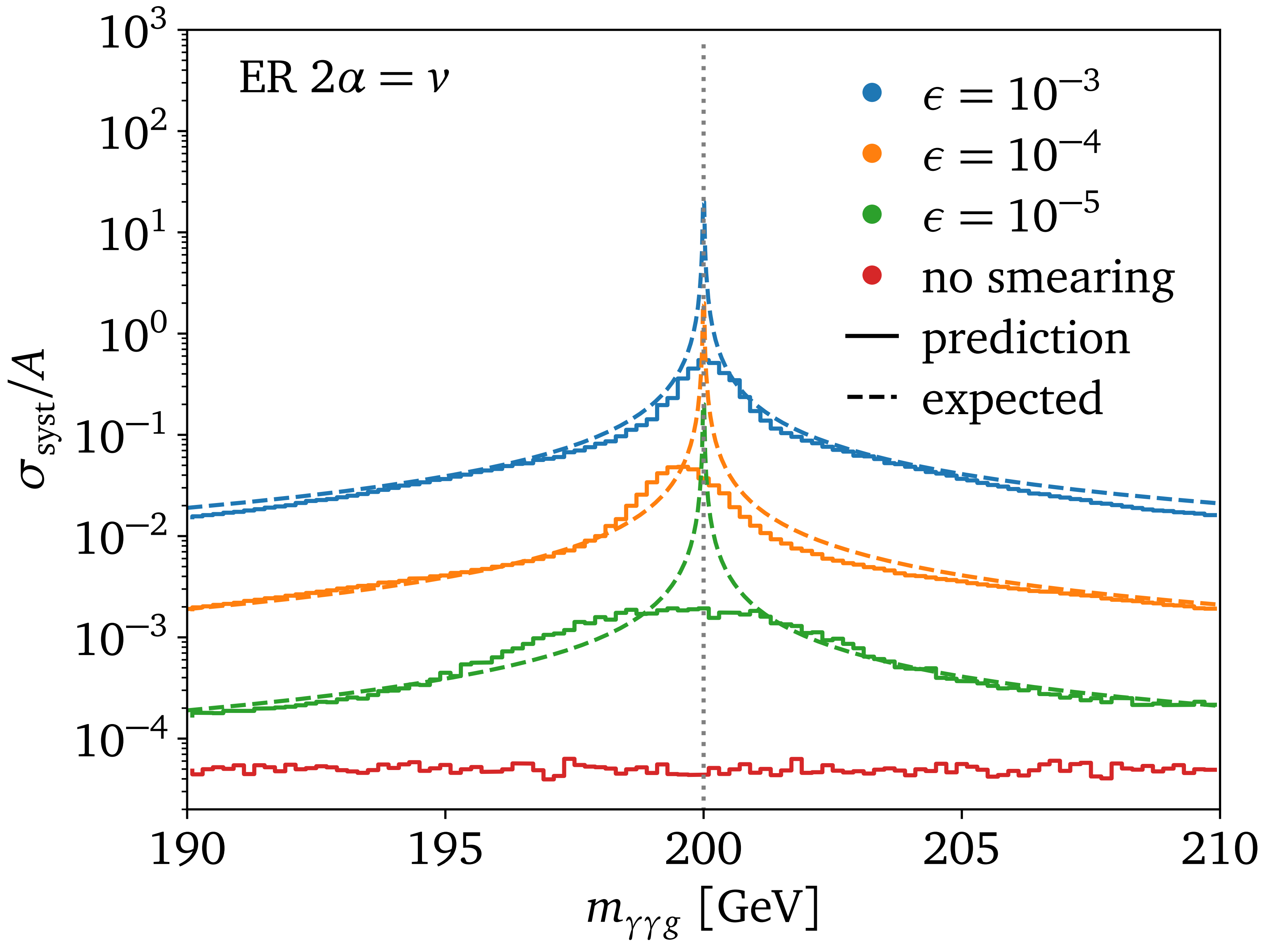
- Captures noise perfectly
- Underestimation for small ϵ

BNN



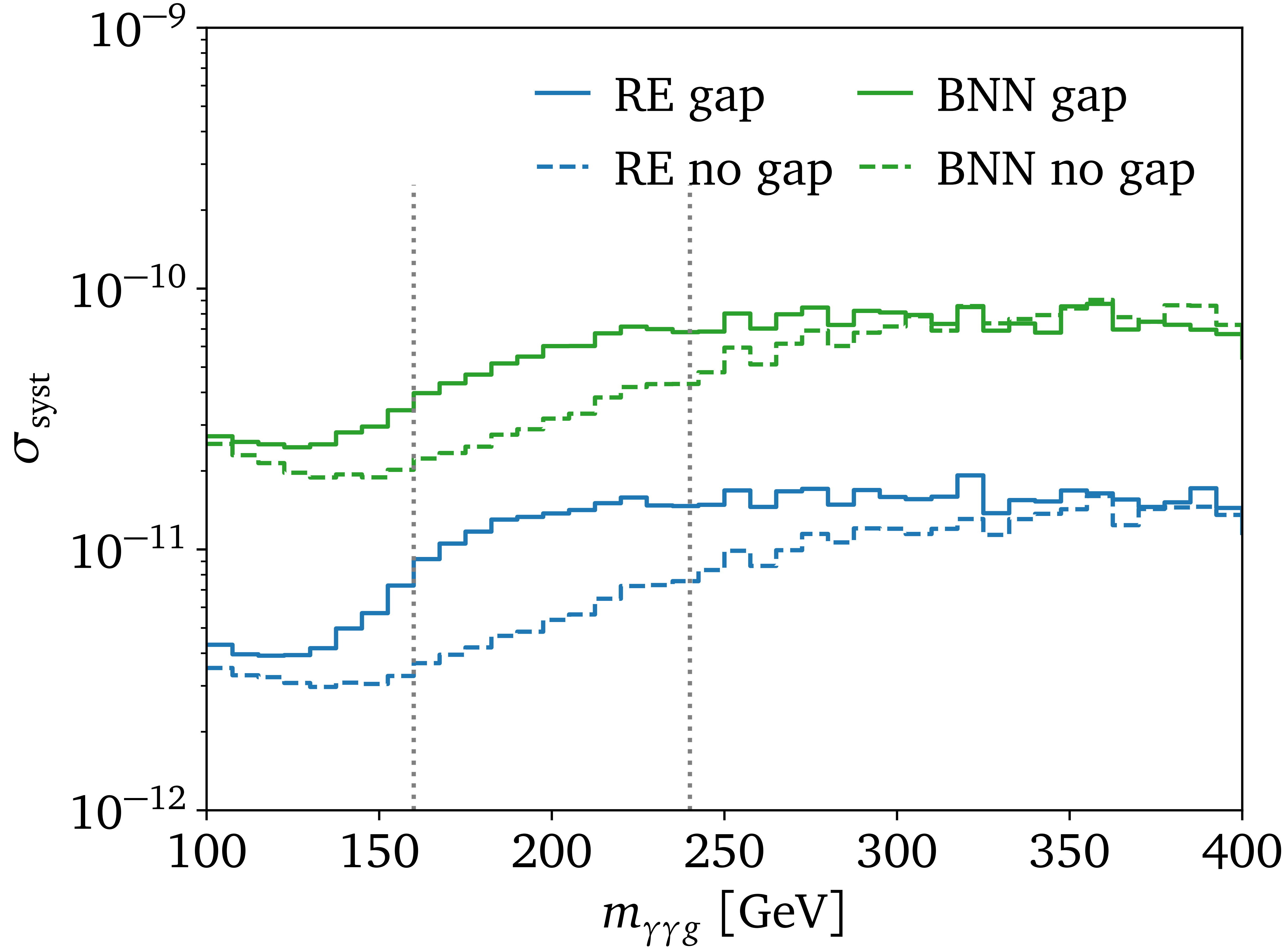
- Overestimates unc. for large ϵ
- Struggle close to threshold

Evidential regression



- Underestimates unc.
- Struggle close to threshold

Removing phase space: Threshold Gap



- Architectures behave individually as expected
- Both show different uncertainties

➡ Different implicit bias

Conclusion and Outlook

1. We can learn calibrated uncertainties

2. BNNs, (repulsive) ensembles, and evidential regression all work as uncertainty estimators

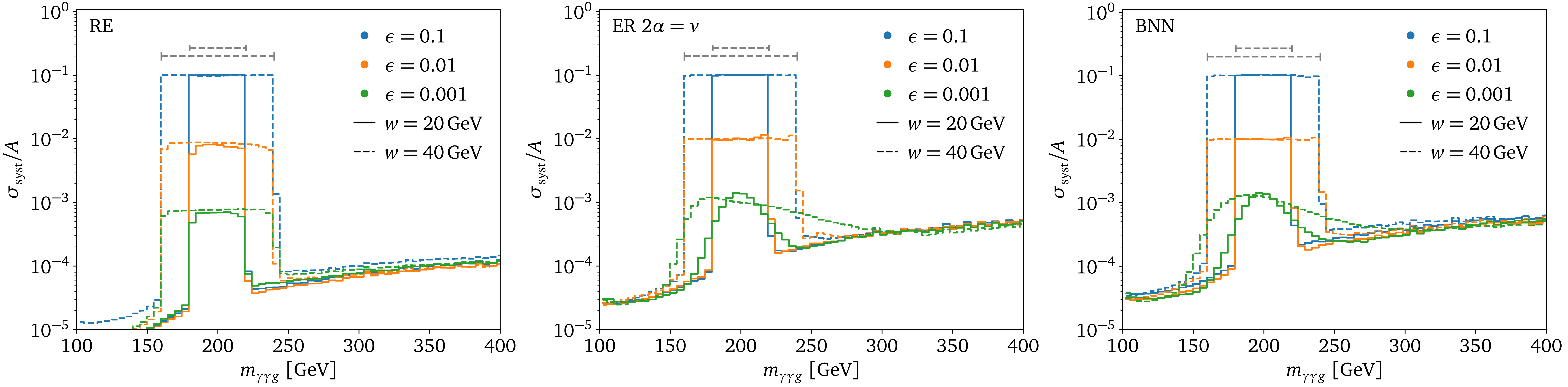
3. Global systematic uncertainty fixes initial calibration problem for RE

Outlook: Can we propagate these uncertainties through simulations and analyses?

Thank you for your attention!

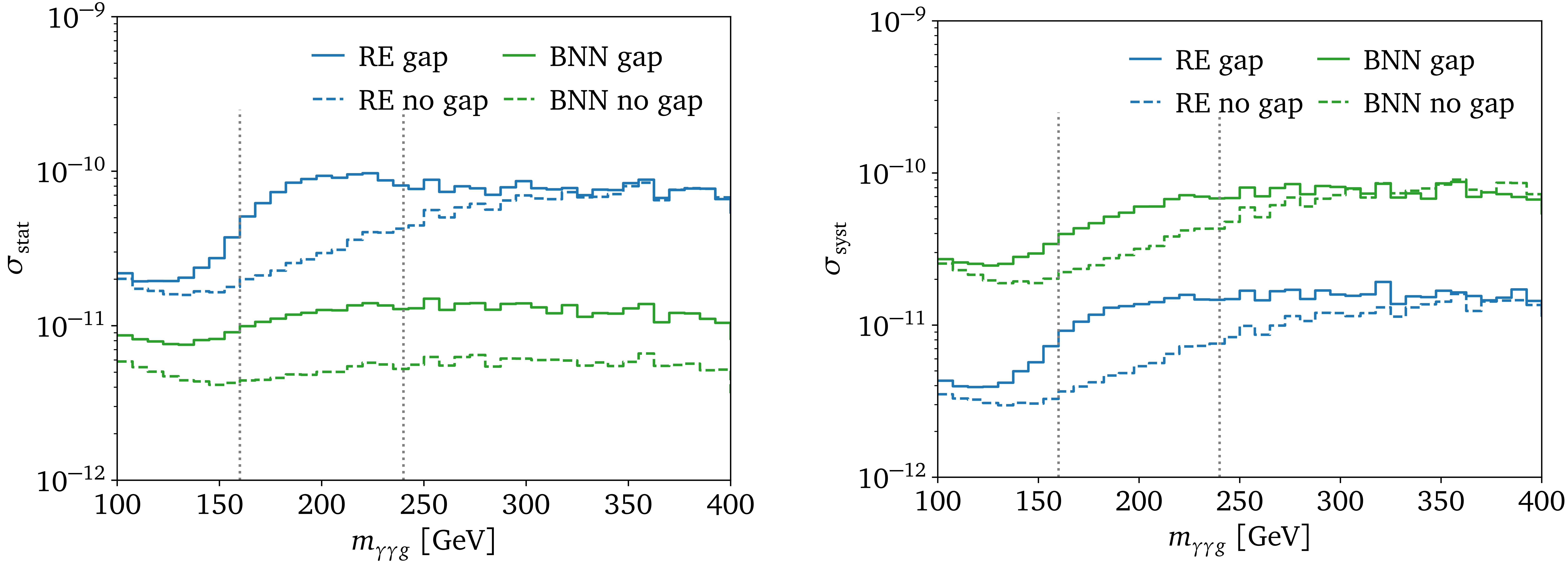
Back up / Additional material

Flat Box Smearing as challenge

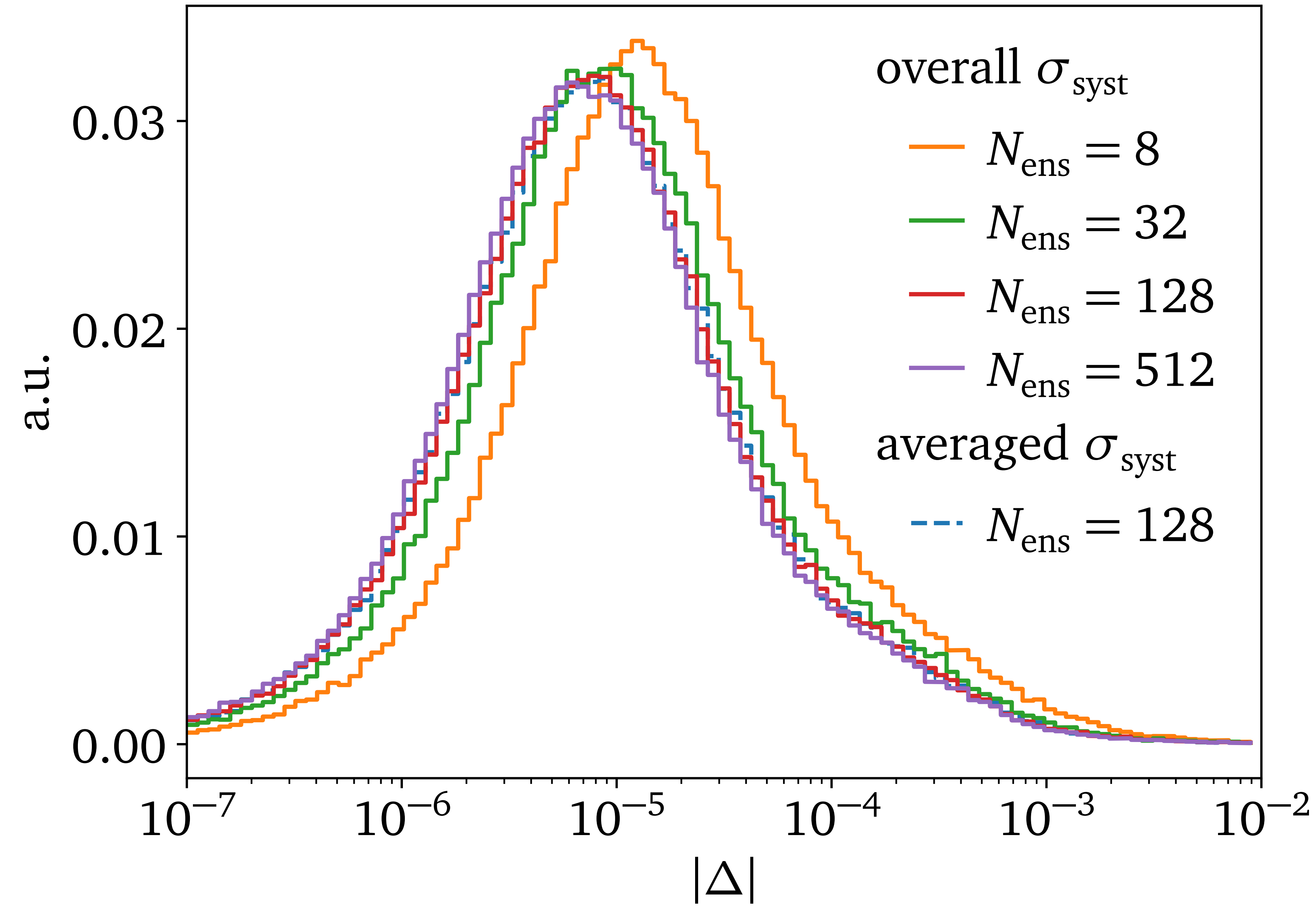


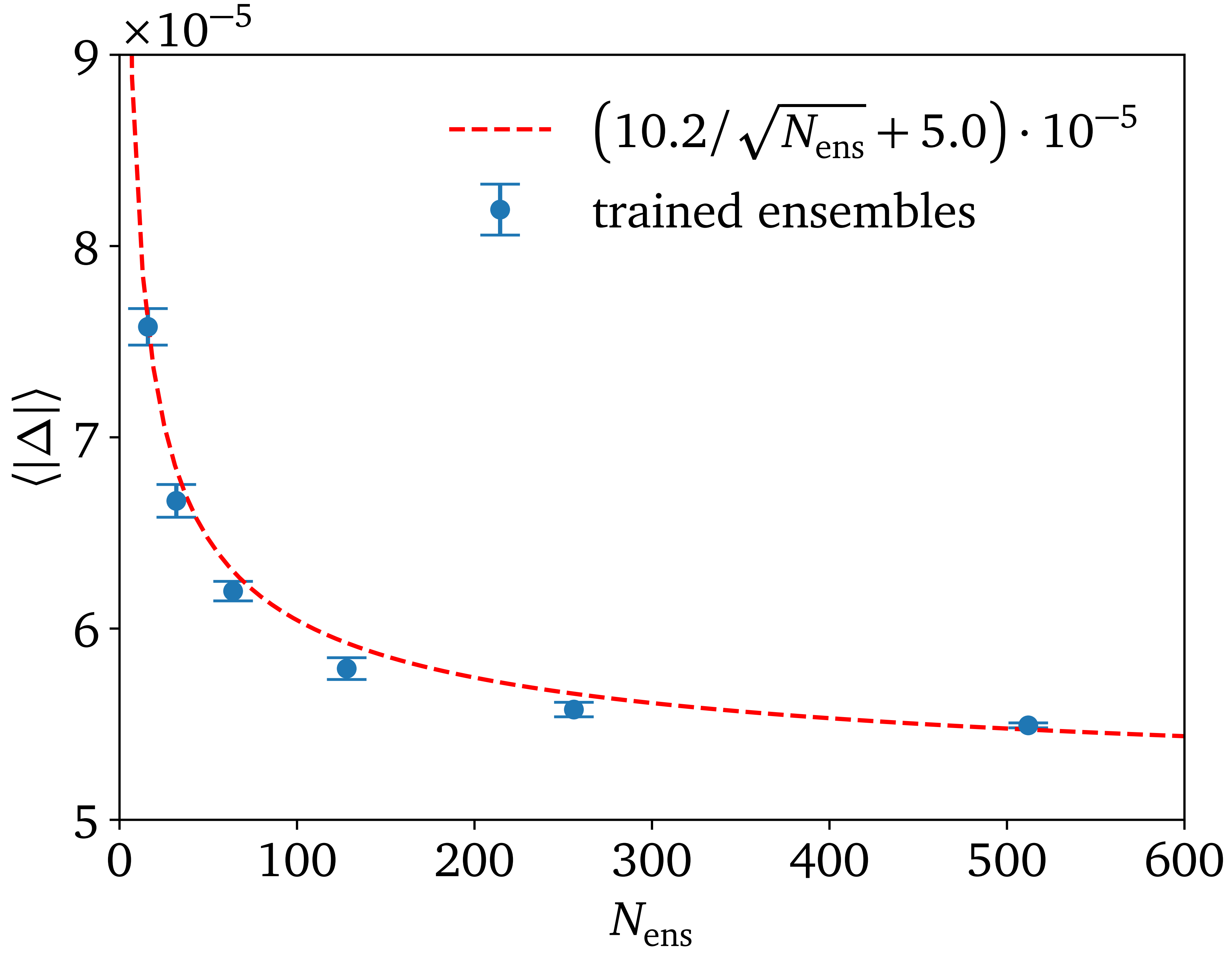
- Box with width w and smearing strength ϵ
- Small ϵ: ER + BNN fail to recover sharp edges
- RE: follows predicted box shape almost perfectly

Removing phase space: Threshold Gap



Global systematic uncertainties for the RE





Peaked threshold smearing - stat unc

Predicted amplitudes for gap study