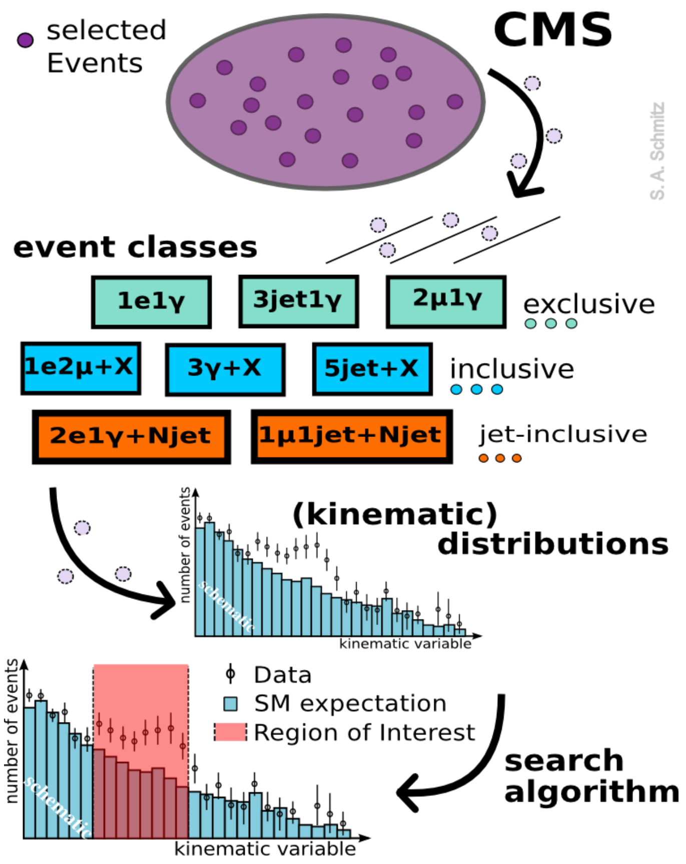
UPDATE: DEVELOPMENT OF A FAST SEARCH ALGORITHM FOR THE MUSIC FRAMEWORK

Jonas Lieb, 18.08.2015

REMINDER: MUSIC (MODEL UNSPECIFIC SEARCH IN CMS)

- Sort events into **event classes** by their physics object content (μ, e, γ, jets, MET)
- Three distributions of interest: $\sum |\overrightarrow{p_T}|, M_{\text{inv}}, \text{MET}$
- Find most significant region (RoI) in each distribution
- Determine look-elsewhere corrected **p-value** ($\tilde{p}$) for each distribution through pseudo-experiments
- Compare distribution of $\tilde{p}$ from data with MC

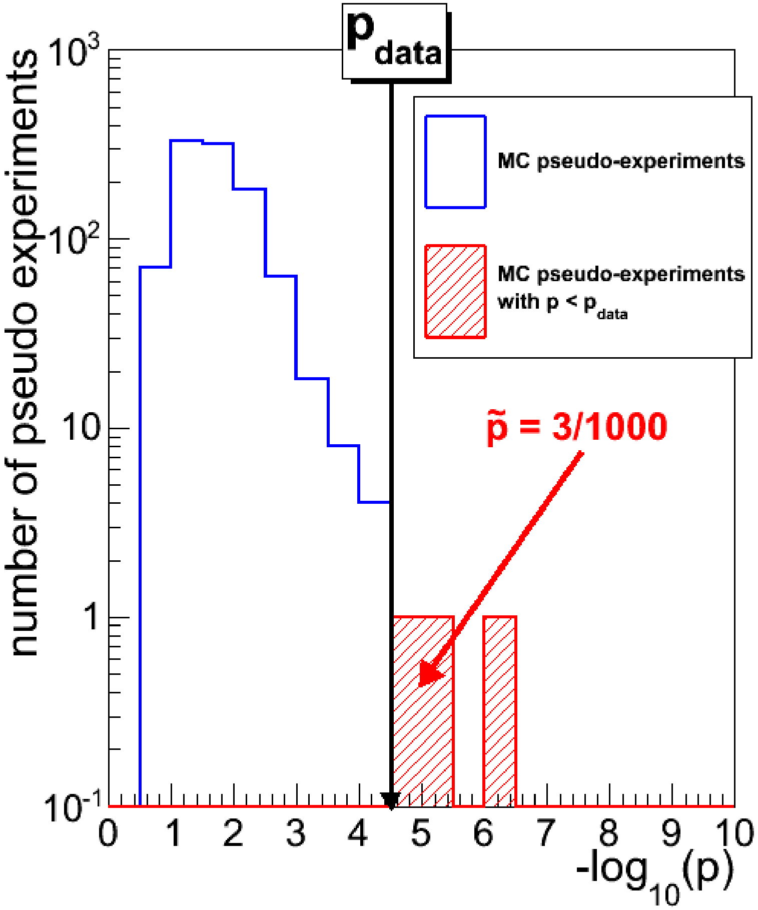


SCANNING

- Construct **connected bin regions** from histogram
- Calculate **p-value** for each region:
  - $$p_{\text{data}} = \begin{cases} \sum_{N=N_{\text{data}}}^{\infty} C \cdot \int_0^{\infty} \mathrm{d}\theta \; \exp\left(-\frac{(\theta - N_{SM})^2}{2\,\sigma_{SM}^2}\right) \frac{e^{-\theta}\theta^N}{N!}, & \text{if } N_{\text{data}} \geq N_{\text{SM}} \\ \sum_{N=0}^{N_{\text{data}}} C \cdot \int_0^{\infty} \mathrm{d}\theta \; \exp\left(-\frac{(\theta - N_{SM})^2}{2\,\sigma_{SM}^2}\right) \frac{e^{-\theta}\theta^N}{N!}, & \text{if } N_{\text{data}} < N_{\text{SM}} \end{cases}$$
- Find **most significant region (smallest p-value)** for each histogram

CALCULATION OF $\tilde{p}$

- Needed to account for **“look-elsewhere-effect”**
- Repeat scanning with **pseudo-experiments,** each mean is shifted within its Standard Model uncertainty



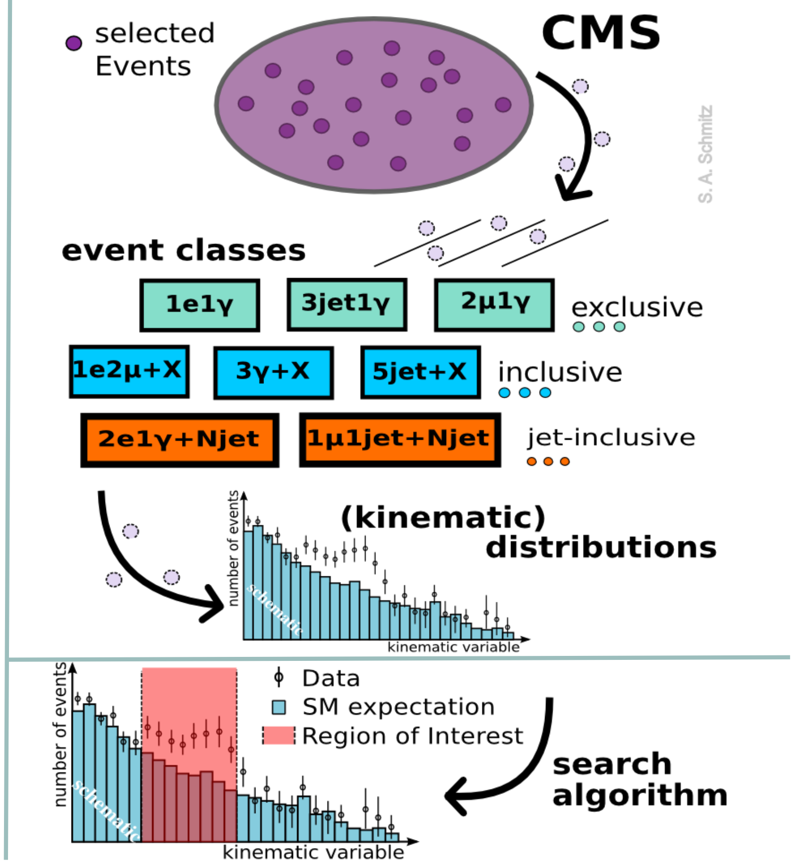
$$\tilde{p} = \frac{\text{number of pseudo experiments with } p_{pseudo} < p_{data}}{\text{total number of pseudo experiments}}$$

QUICKSCAN

- Problem: the p-value is evaluated many times, its calculation is time consuming
- Mitigation: **preselect interesting regions** using a less computation intense algorithm
- Select a certain number of candidate regions, with the maximum

$$\chi = \frac{|N_{obs} - N_{MC}|}{\sqrt{\sigma_{MC}^2 + N_{MC}}}$$

- ~~This estimator does not consider effects depending on the absolute number of events → "vertical" binning by magnitude~~
- To select the most significant region, calculate the p-value integral only for the Quickscan candidates
- ~~Two~~ One parameters:
  - **number of candidates per vertical bin**
  - ~~**magnitude bin size**~~



STATISTICAL TERM IN ESTIMATOR

- Problem: estimator used to be

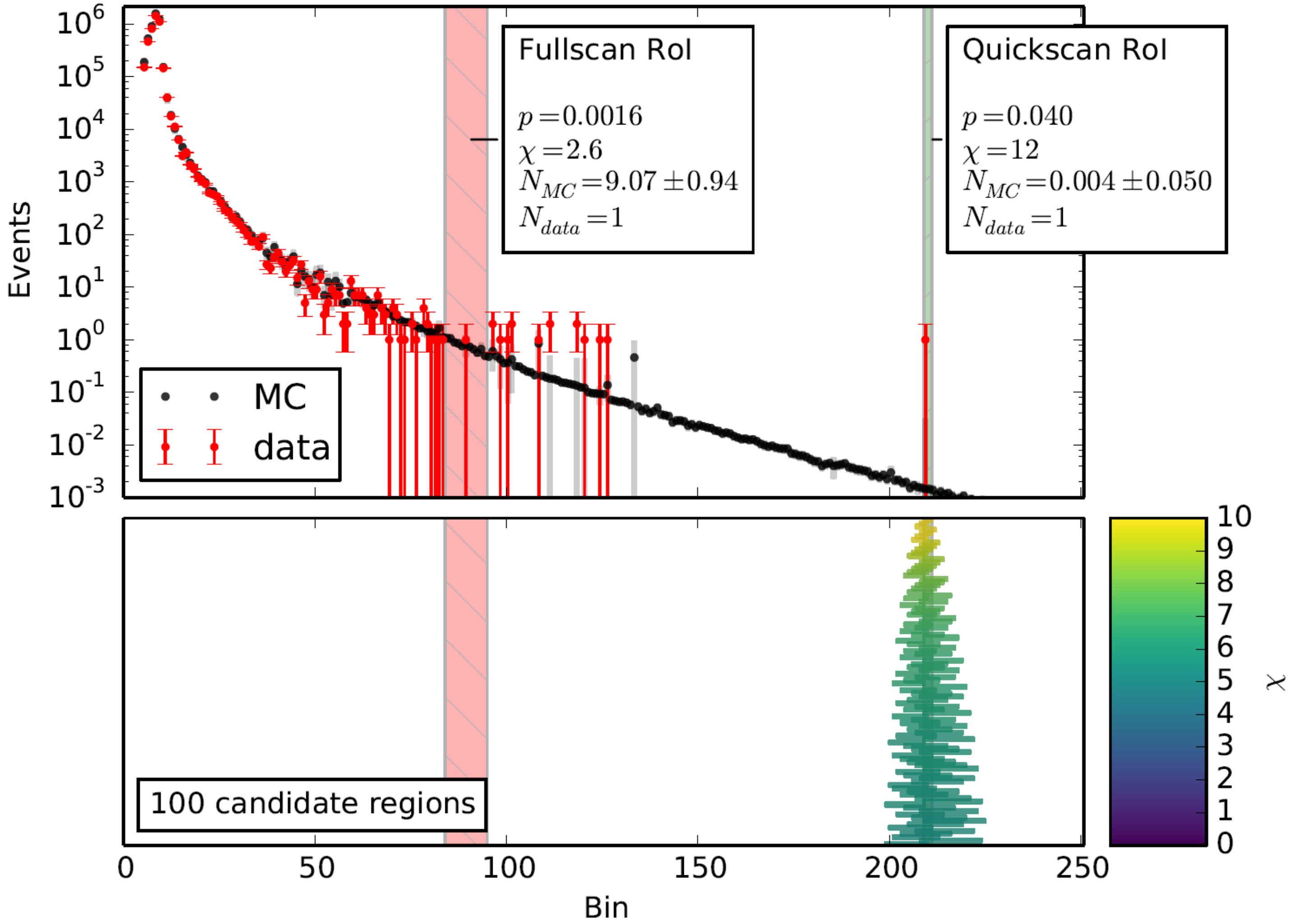
$$\chi = \frac{|N_{obs} \; - \; N_{MC}|}{\sigma_{MC}}$$

- σ_{MC} does not include expected statistical deviation between N_{obs} and N_{MC} , $\sqrt{N_{MC}}$

- Solution: replace $\sigma_{MC} \rightarrow \sqrt{\sigma_{MC}^2 + \sqrt{N_{MC}}^2} = \sqrt{\sigma_{MC}^2 + N_{MC}}$

- Solved a lot of problems

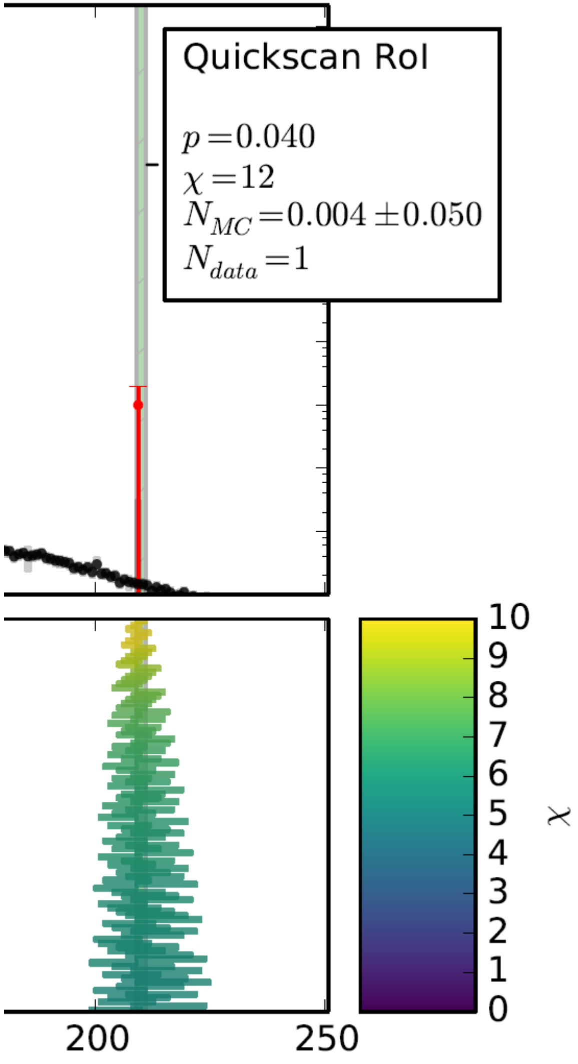
ADDITIONAL PROBLEM IN HIGH ENERGY TAILS



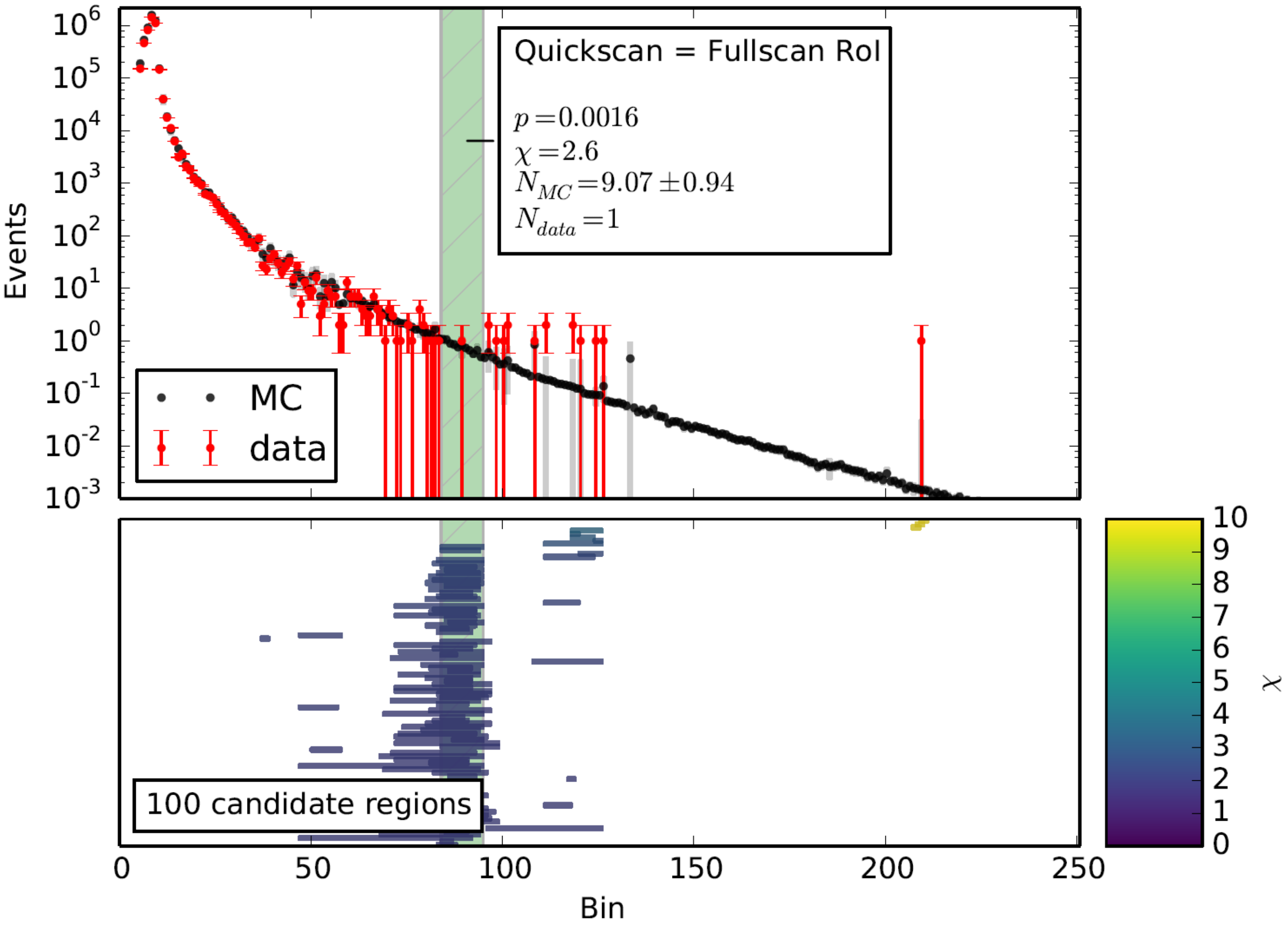
SOLUTION: SPECIAL HANDLING OF NESTED REGIONS

- Region A is nested in region B
- A and B are excesses
- A and B have the same amount of data
- → A is more significant

- Solves (almost) all problematic cases

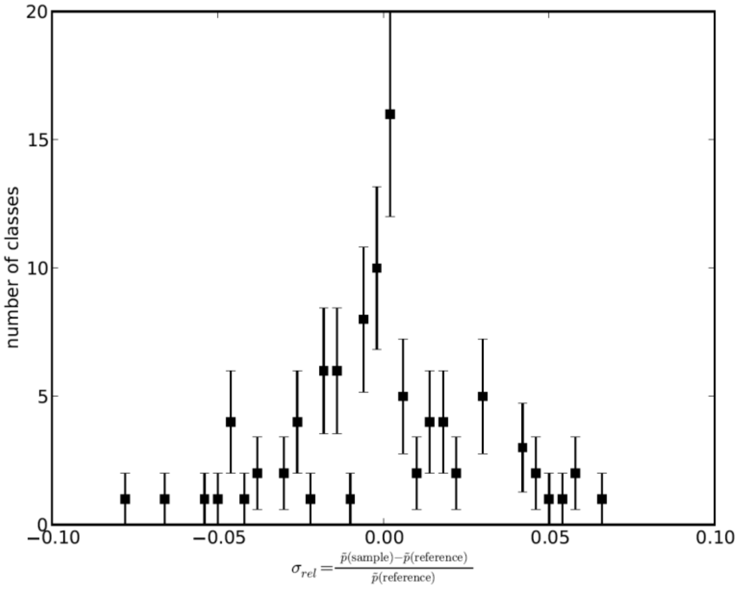


FIXED!


Events
10^6
10^5
10^4
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
Quickscan = Fullscan RoI
$p = 0.0016$
$\chi = 2.6$
$N_{MC} = 9.07 \pm 0.94$
$N_{data} = 1$
MC
data
100 candidate regions
0
50
100
150
200
250
Bin
10
9
8
7
6
5
4
3
2
1
0
χ

PARAMETER OPTIMIZATION

- Optimization of the parameters is performed by measuring their effect on two metrics:
  - **Runtime / Speed-up** $= \frac{T_{classic}}{T_{quickscan}}$
  - **Relative deviation of $\widetilde{p}$:**
    $$\frac{\Delta\tilde{p}}{\tilde{p}(\text{classical})} = \frac{\tilde{p}(\text{quickscan}) - \tilde{p}(\text{classical})}{\tilde{p}(\text{classical})} \quad (\leq 0)$$
- Working on a **subset:** 2012 data, exclusive classes only, max. 2 jets, dicing exactly 1000 rounds
- **Status quo:**
  - Runtime ~ 1h 30min
  - Random $\Delta\tilde{p}$ spread through dicing about 5%

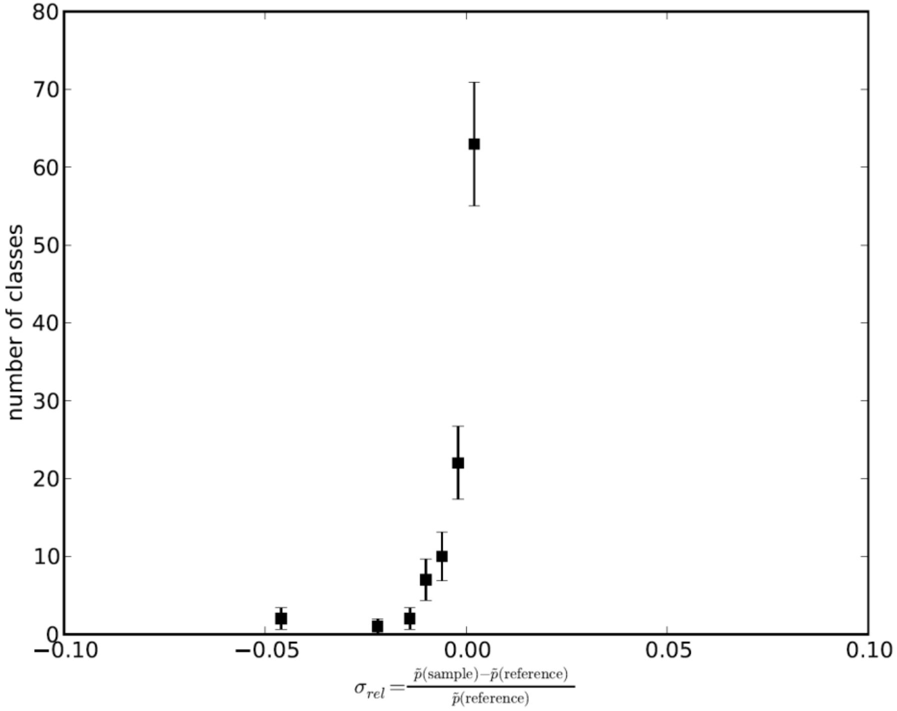


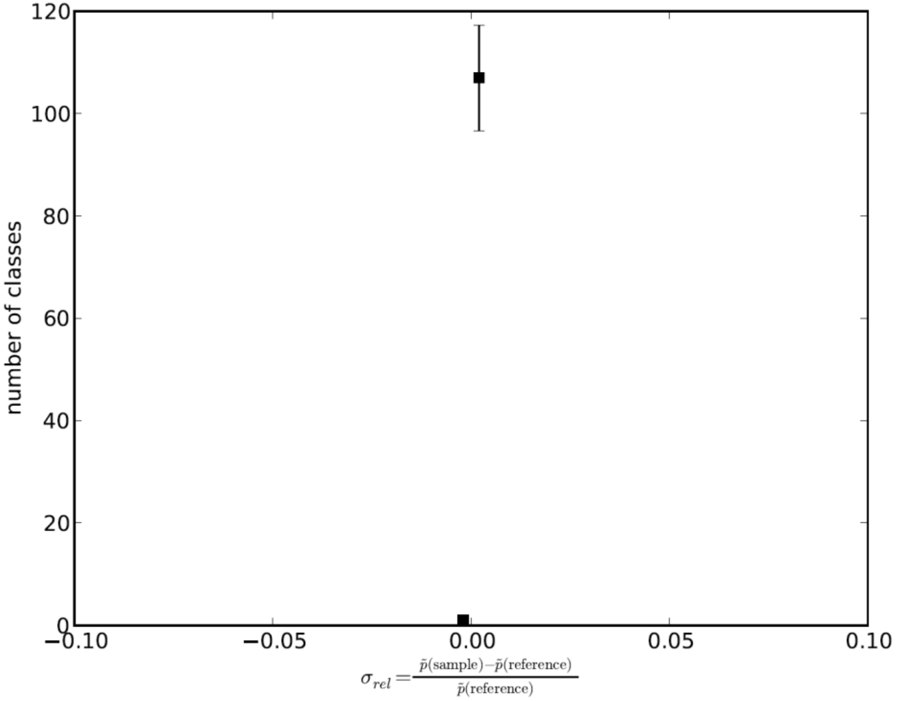
w/o Quickscan vs. w/o Quickscan

SELECTED RESULTS

10 candidates
Runtime: 25 minutes

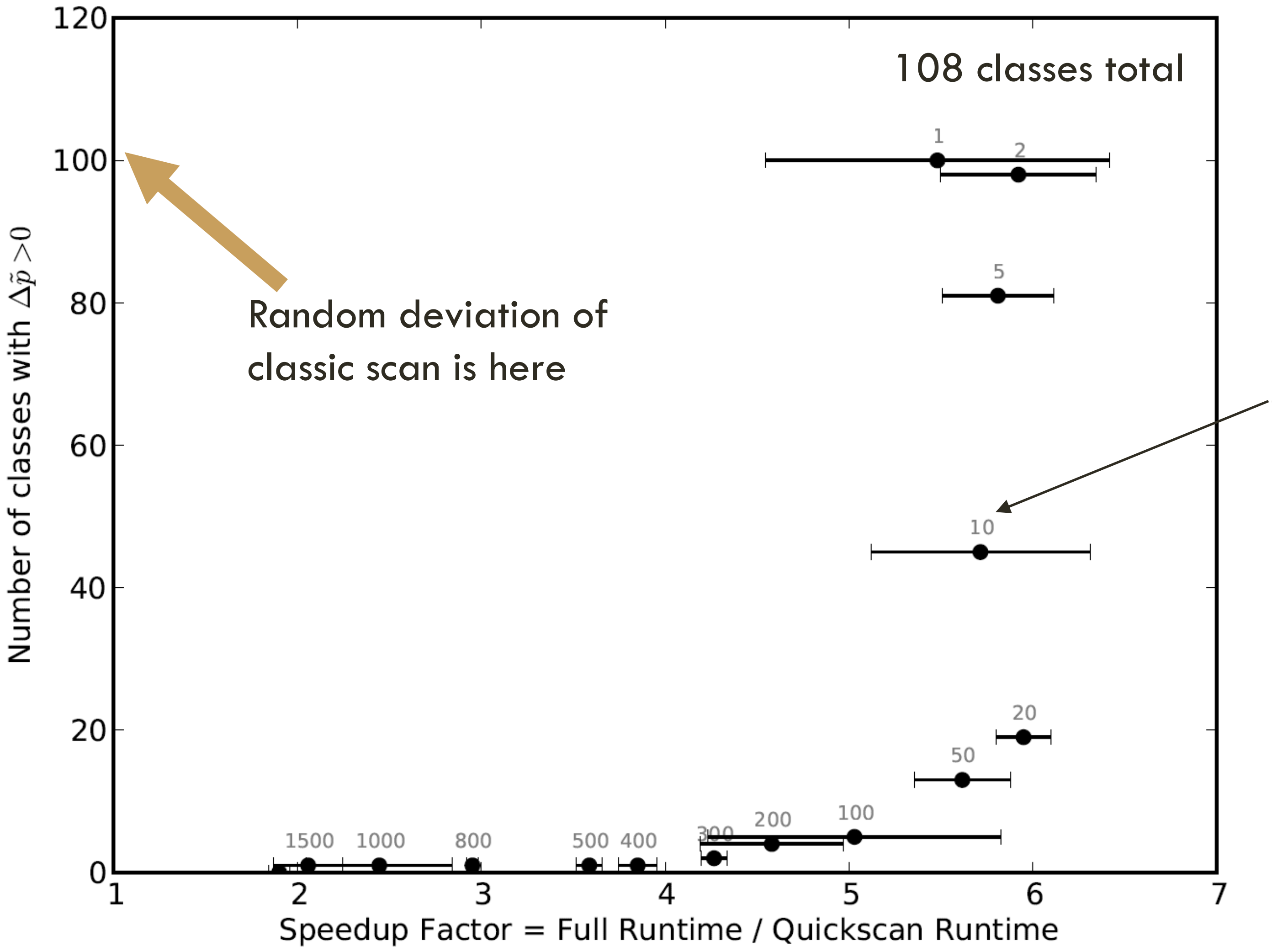
1000 candidates
Runtime: 54 minutes





Quickscan vs. w/o Quickscan

RESULTS FOR THE BIN SIZE



108 classes total

Random deviation of classic scan is here

Number of candidates

RESULTS

- Quickscan seems to work (even better!)
- Magnitude binning not necessary anymore
- Speed-up up to 6 times while keeping very good physics results

OUTLOOK

- Validation Run
- Write-up as Bachelor thesis