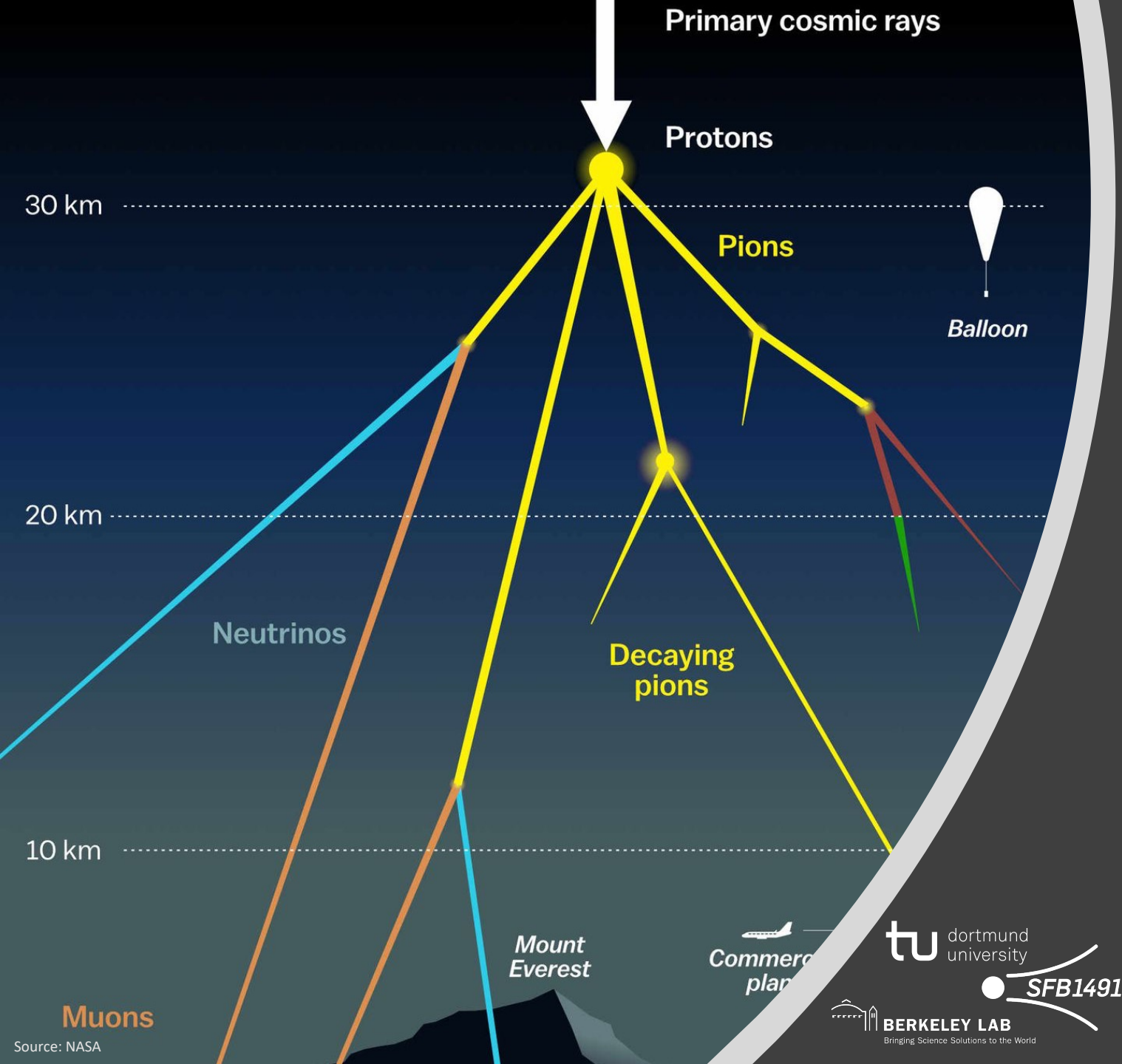
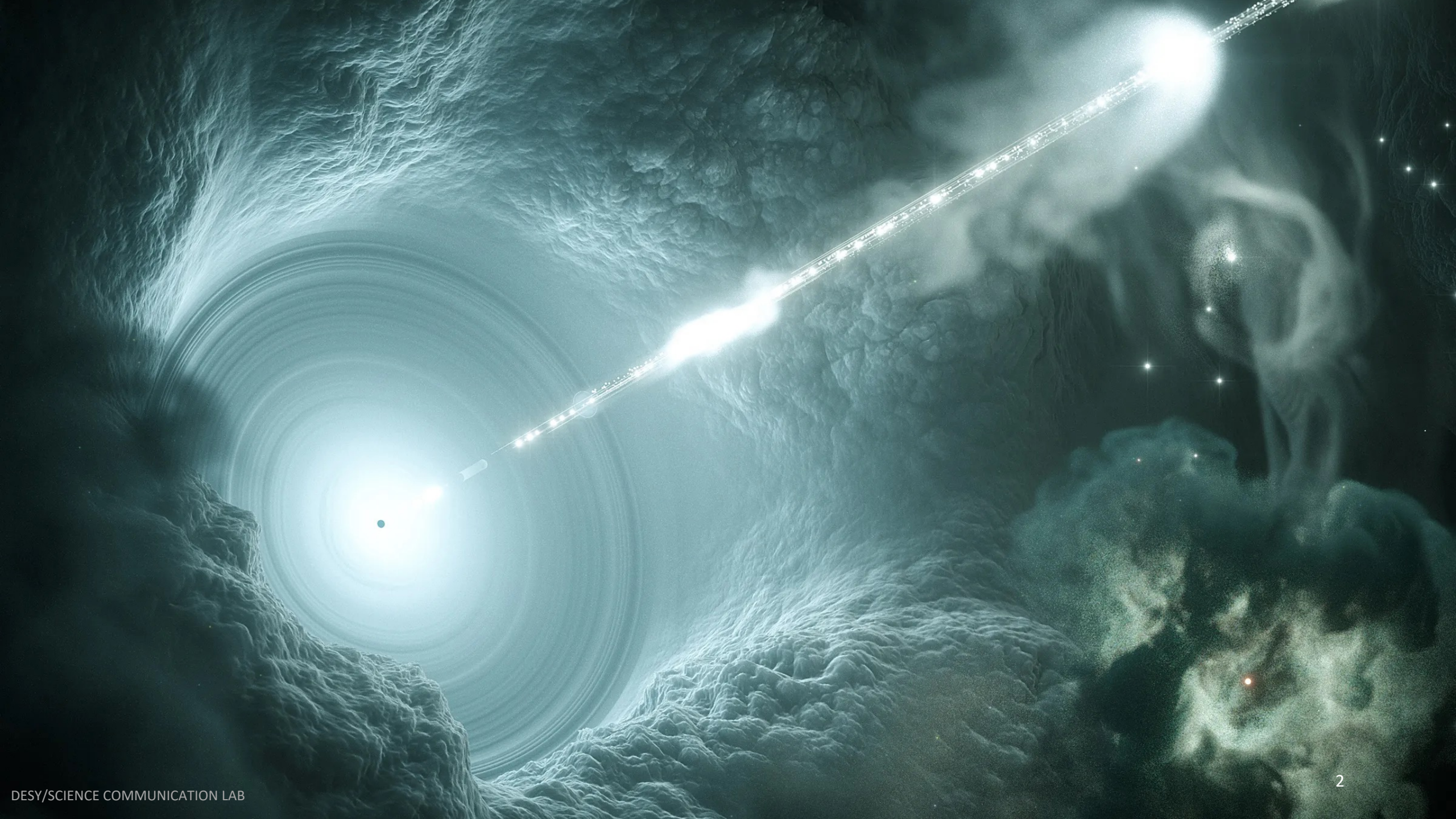


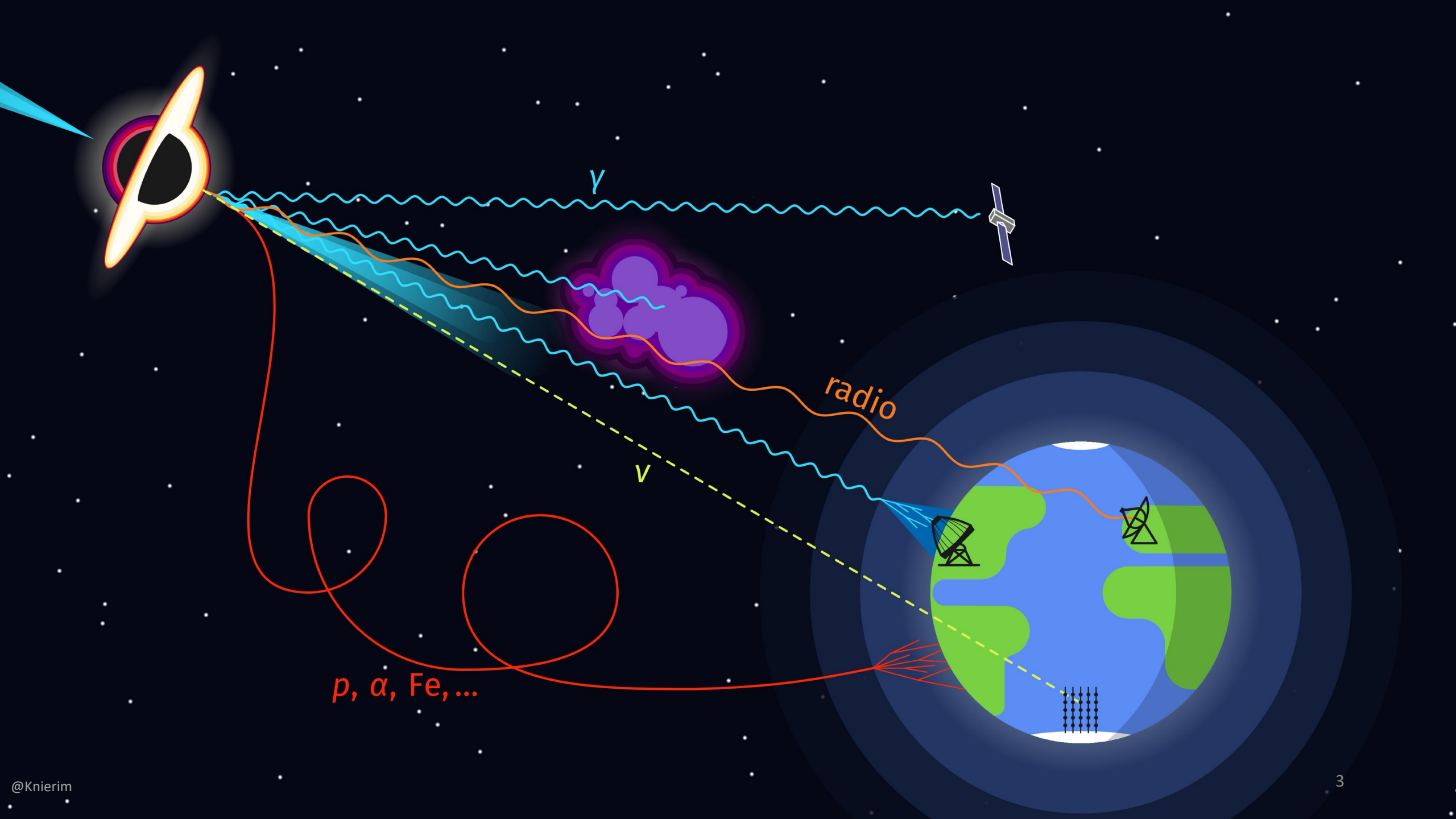
The Muon Puzzle: Linking Air-Shower Physics and the LHC through Atmospheric Muons

Pascal Gutjahr

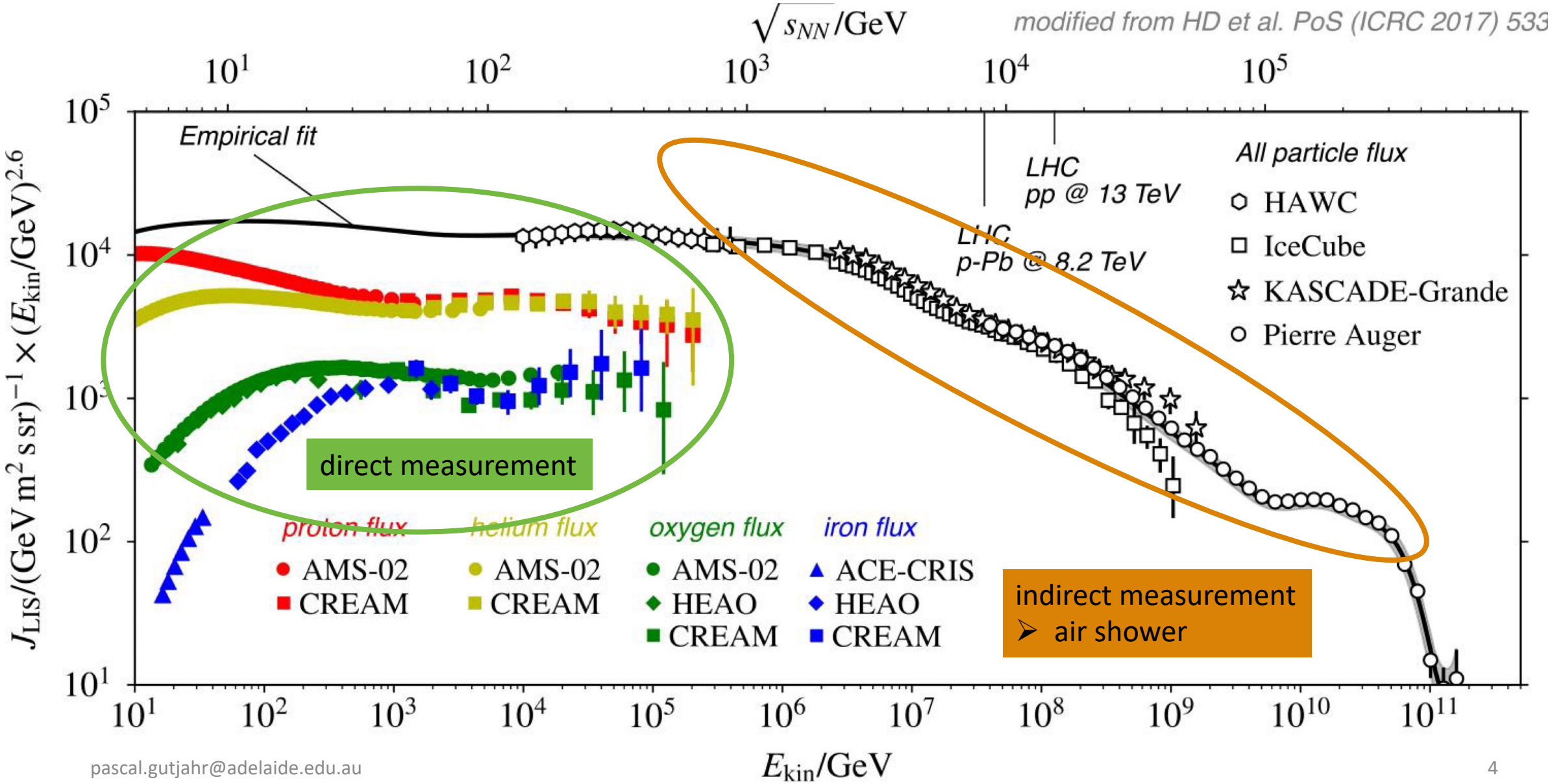
LIP, Lisbon, Portugal

November 5, 2026

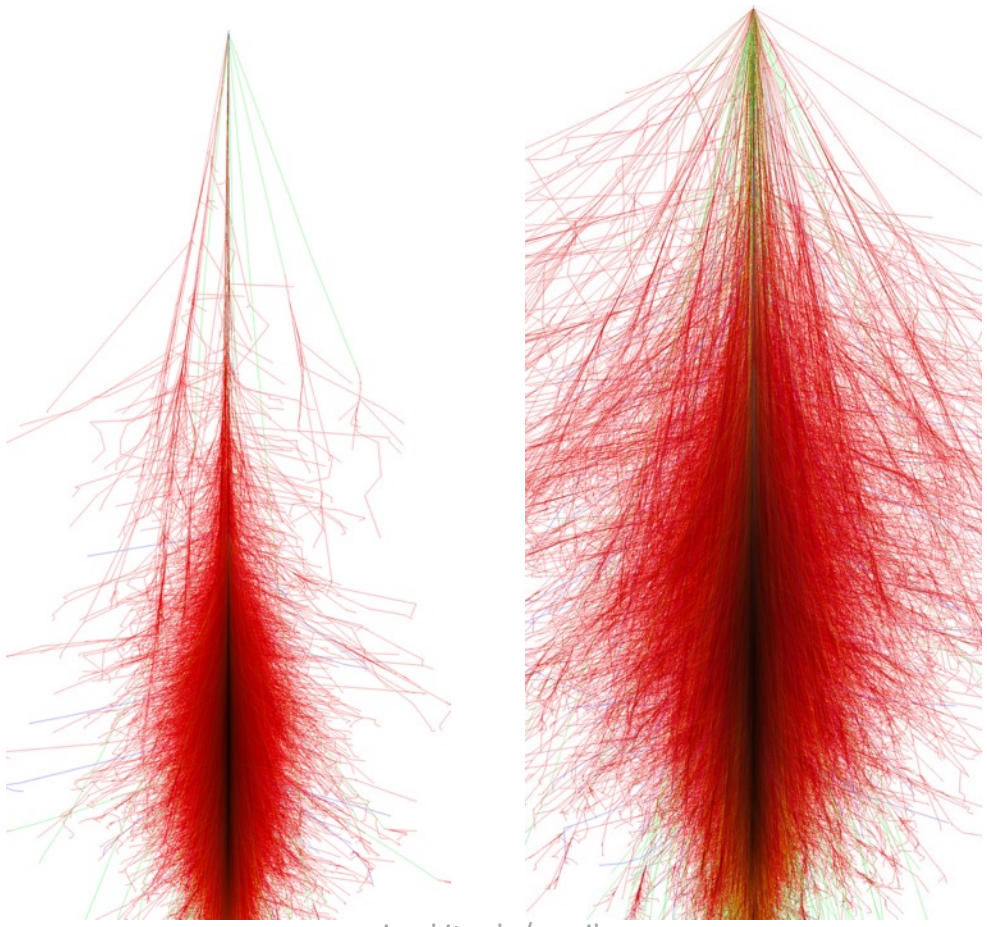



γ
radio
ν
p, α, Fe, …

Cosmic Ray Flux

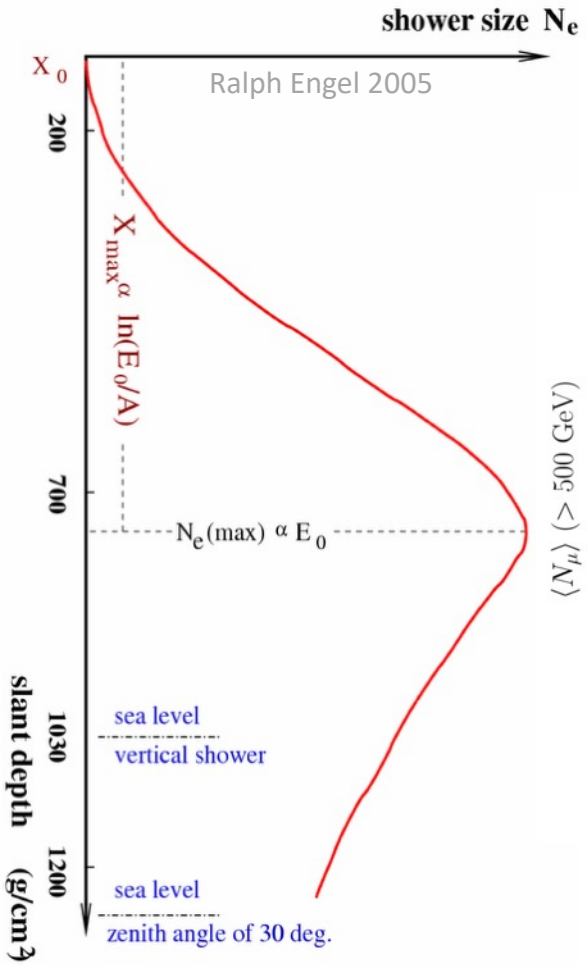


Air shower – 10 TeV

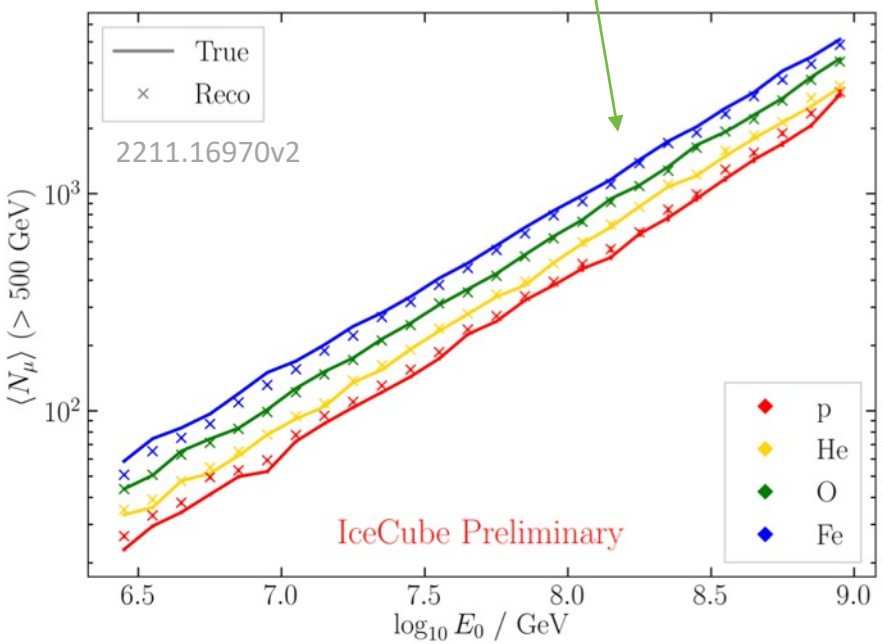


www.iap.kit.edu/corsika

Proton Iron



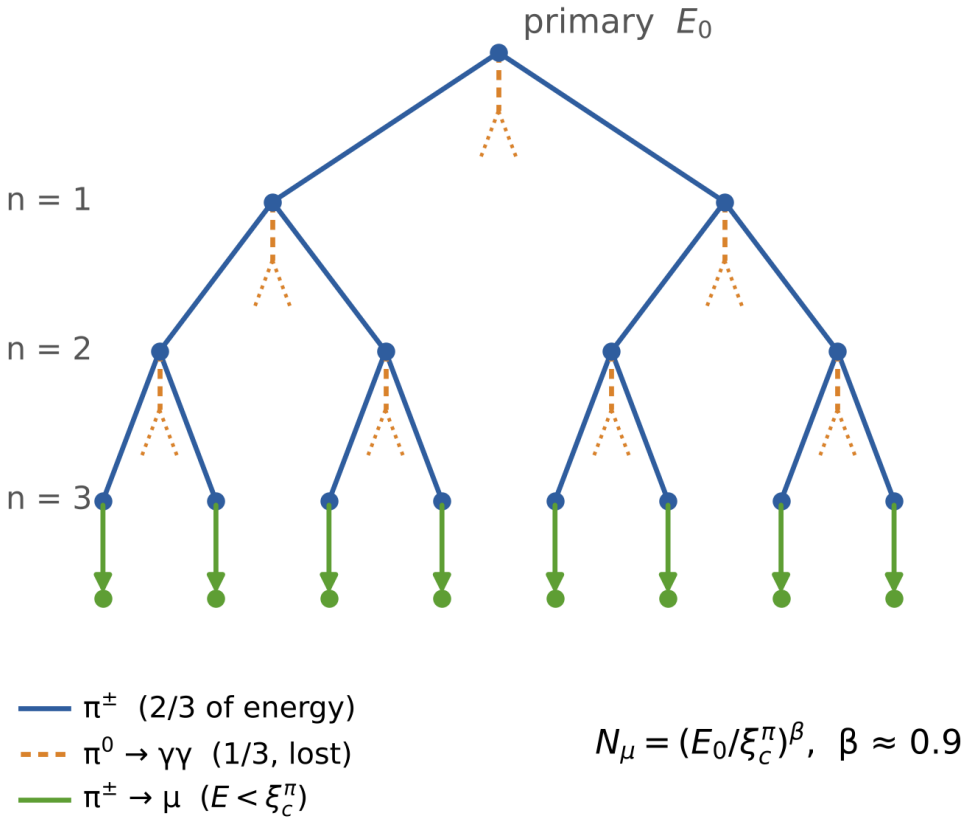
Ralph Engel 2005



Why Muons Probe Hadronic Interactions

- Each hadronic generation sends ≈ ⅓ of its energy into $\pi^0 \rightarrow \gamma\gamma$: lost for muon production
- Charged pions decay into muons once decay wins over interaction (critical energy ξ_c^{π})
- Heitler–Matthews model:
  $\mathbf{N_\mu = A^{1-\beta}\,(E_0/\xi_c)^{\beta}\,,\quad \beta \approx 0.9}$
- Iron showers: ≈ 40 % more muons than proton showers (EeV scale)
- Small changes add up: "a 5 % deviation per generation converts into ~30 % deviation after six generations"

➢ N_μ measures how much energy stays hadronic, but also depends on the mass → needs an independent mass estimate (X_{max})



J. Matthews, Astropart. Phys. 22 (2005) 387; J. Albrecht et al., Astrophys. Space Sci. 367 (2022) 27; quote: Auger, PRL 126 (2021) 152002

The Muon Puzzle

Muons in ultra-high-energy air showers

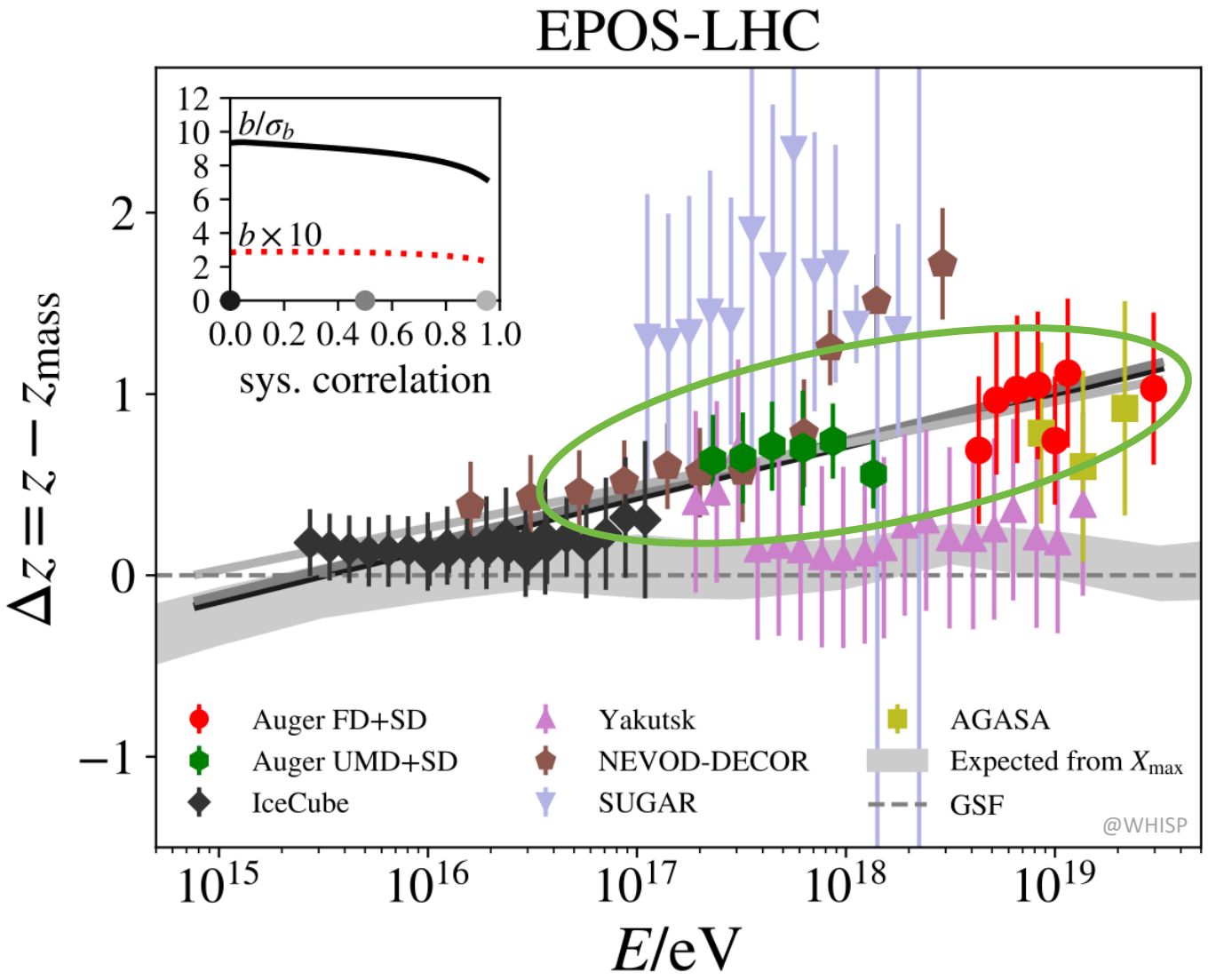
The Muon Puzzle

- Compare measured muons with predicted muons
- "Muon number"

$$z = \frac{\ln\langle N_\mu \rangle - \ln\langle N_\mu \rangle_p}{\ln\langle N_\mu \rangle_{\mathrm{Fe}} - \ln\langle N_\mu \rangle_p}$$

- Hadronic interactions models
  - SIBYLL
  - EPOS-LHC
  - DPMJet
  - ....

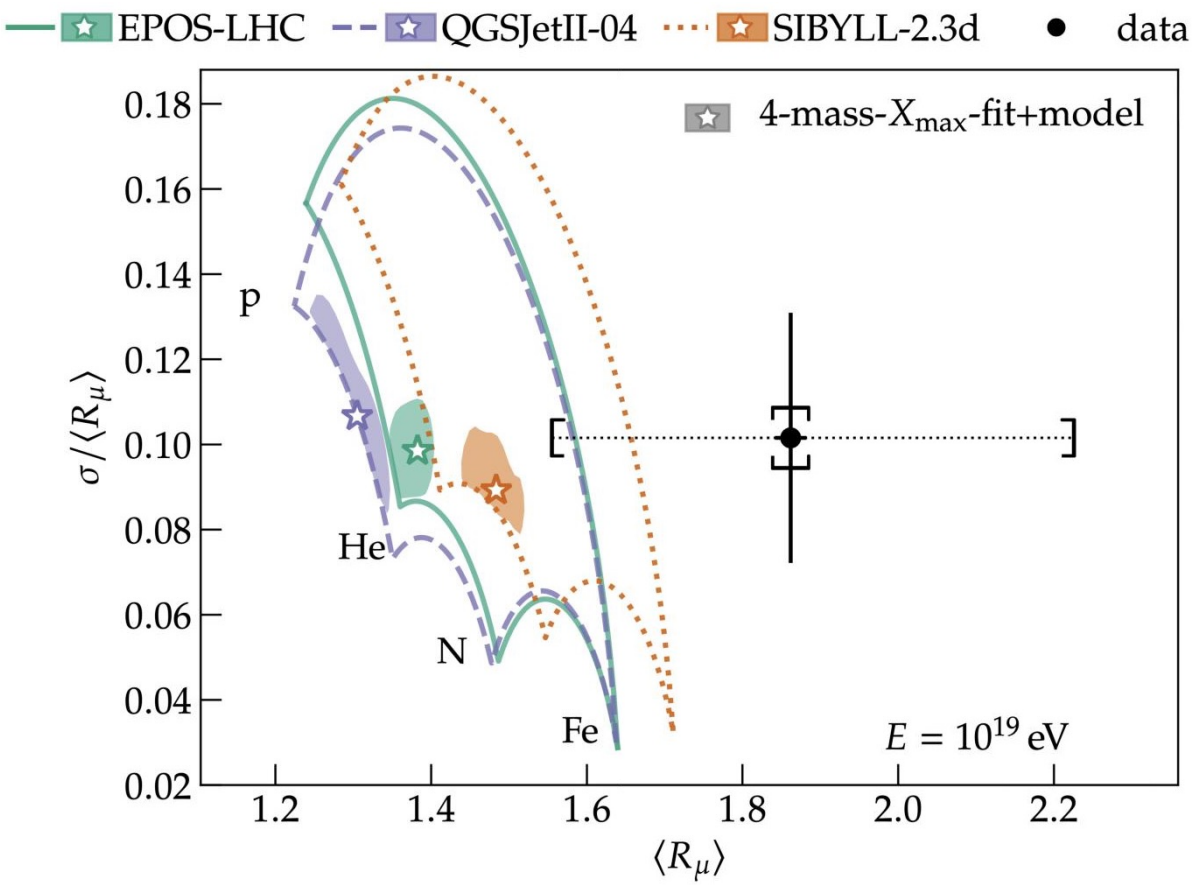
➢ More muons measured than predicted for energies > 40 PeV



Auger: Mean Too Low, Fluctuations Fine

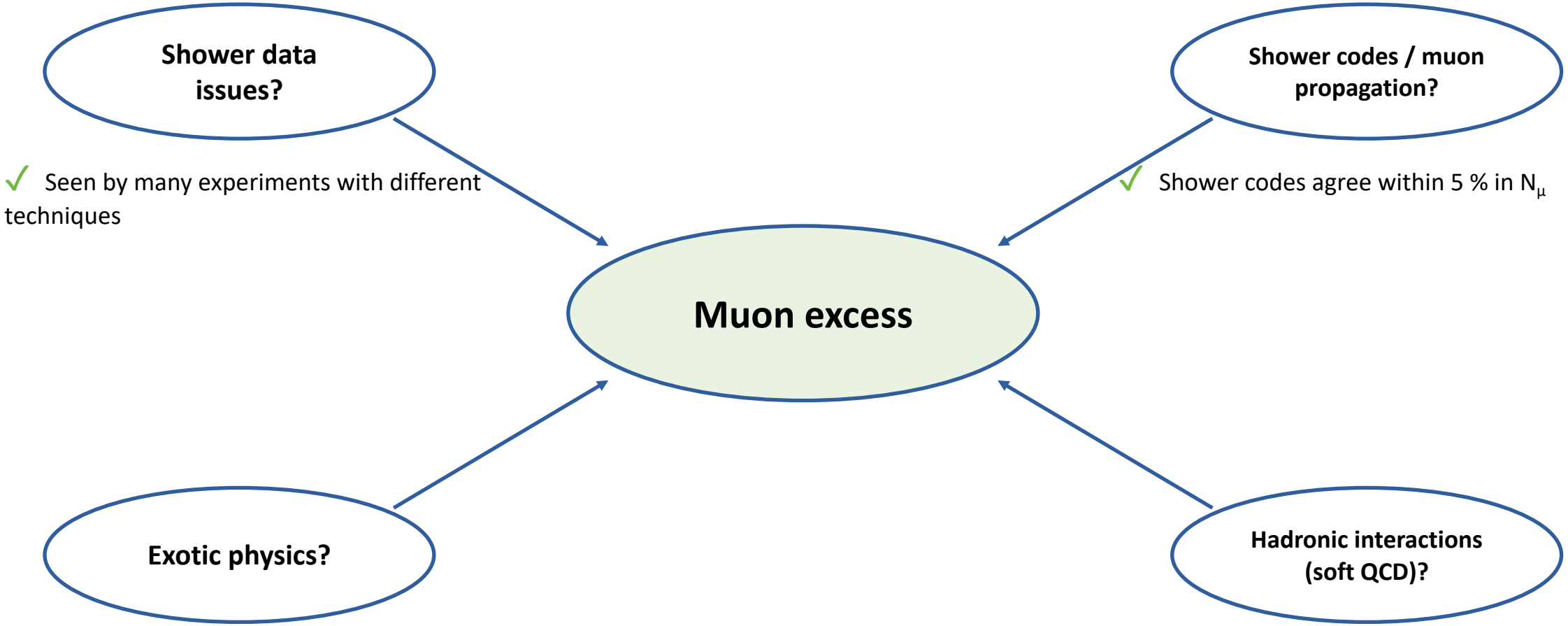
- Inclined showers at 10^{19} eV: energy from fluorescence, muons from the surface detector
- Mean muon number: models need **+43 % (QGSJet-II-04), +35 % (EPOS-LHC), +26 % (Sibyll 2.3d)**
- Muon-number fluctuations: "no rescaling is necessary for any model"
- Fluctuations are dominated by the first interaction

➢ Deficit comes from small deviations at every stage of the shower, "rather than a discrepancy in the first interaction"



Pierre Auger Collab., Phys. Rev. Lett. 126 (2021) 152002, arXiv:2102.07797 [figure: Fig. 3]

What Can Cause the Muon Excess?



✓ Seen by many experiments with different techniques

✓ Shower codes agree within 5 % in N_μ

✗ Fluctuations agree with models (Auger 2021)

→ Main suspect: next section

Diagram after H. Dembinski; J. Albrecht et al., Astrophys. Space Sci. 367 (2022) 27; Auger, PRL 126 (2021) 152002

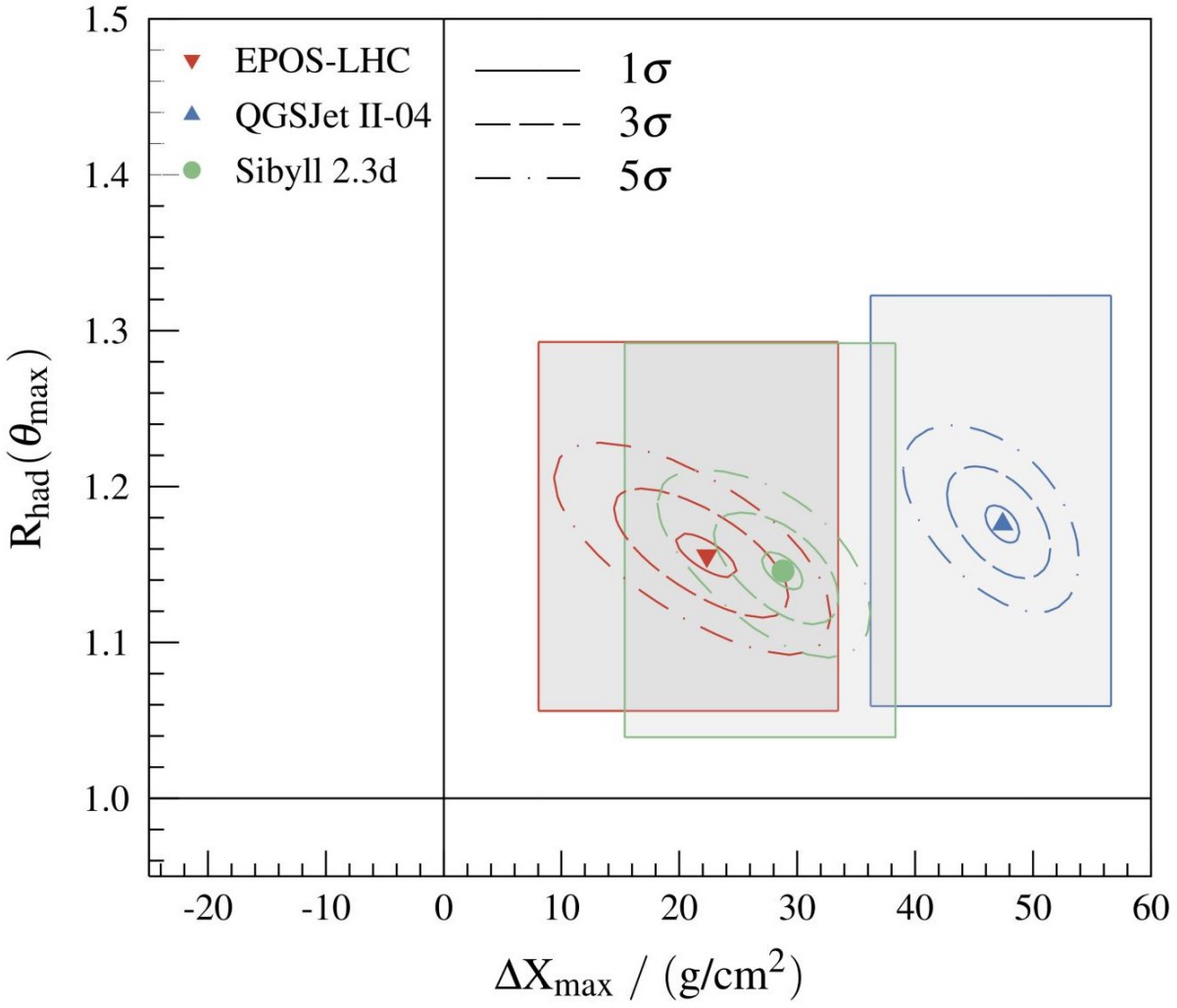
New Results 2024–2026

Air-shower data, models and the LHC

Auger 2024: Two Shifts Needed – X_{max} and Hadronic Signal

- 2239 hybrid events, $10^{18.5}$–10^{19} eV, zenith < 60°
- Fit of measured (X_{max}, S(1000)) distributions with simulated templates
- Free parameters: shift of X_{max}, rescaling of the hadronic ground signal, composition
- Best description: **deeper X_{max} and larger hadronic signal**, improvement > 5σ

➢ Muon puzzle and X_{max} scale are linked: both must be fixed together

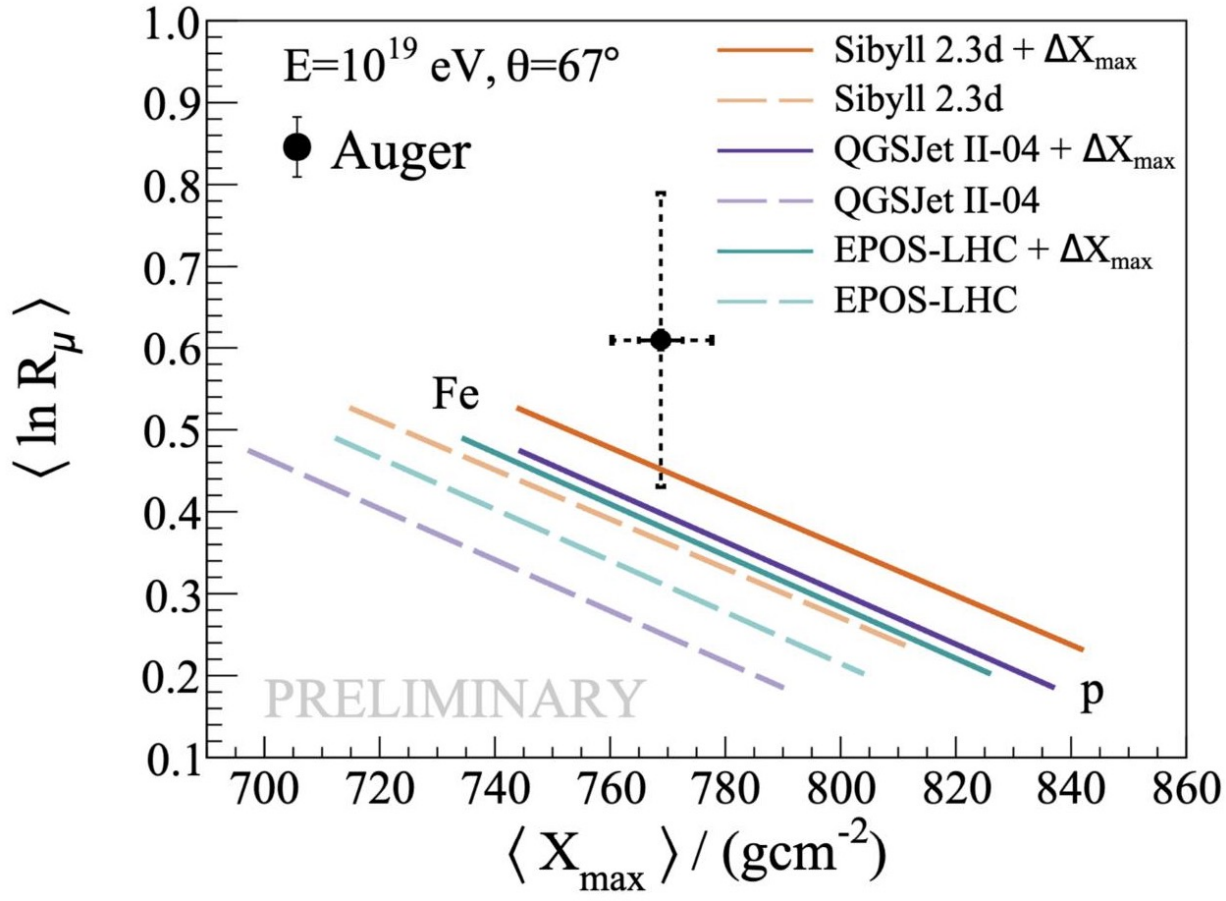


Pierre Auger Collab., Phys. Rev. D 109 (2024) 102001, arXiv:2401.10740 [figure: Fig. 6 (left)]

Muons vs X_{max} at 10 EeV: Data Above All Models

- Inclined showers ($\theta = 67°$) at 10^{19} eV: muon scale R_μ vs $\langle X_{max} \rangle$
- Lines: model predictions from proton to iron
  - dashed: original models; solid: shifted by ΔX_{max}
- Data point lies above all lines
- With the X_{max} shift, the underestimation of the muon scale is "reduced to about 15–25 %"

➢ Even with deeper X_{max} (heavier composition), the older models do not produce enough muons

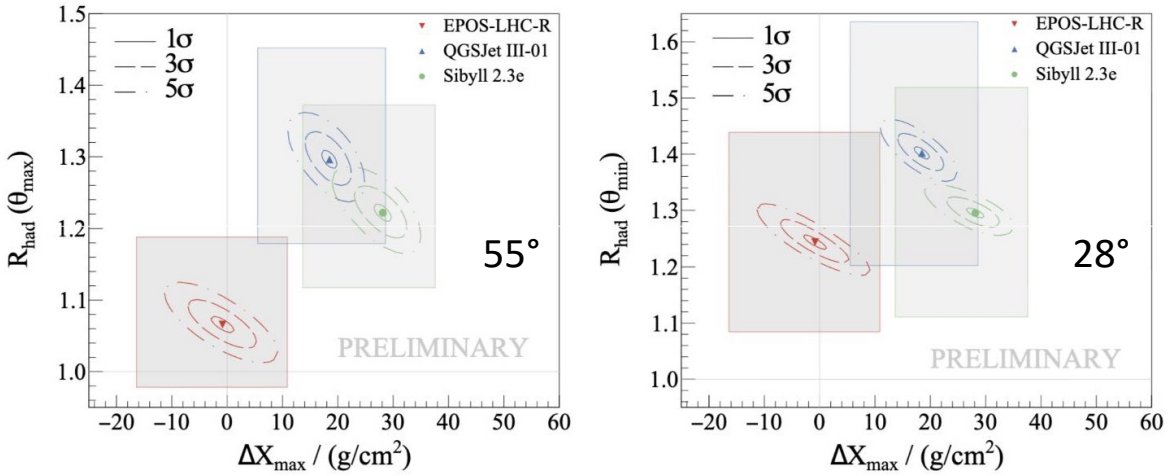


J. Vícha (Auger), ICRC 2025, arXiv:2507.09802 [figure: Fig. 6, preliminary]; data: Auger, PRL 126 (2021) 152002

New Models: Gap Reduced, Not Closed

Model	ΔX_{max} [g/cm²]	Hadronic signal needed
EPOS-LHC	+22	≈ +15–25 %*
QGSJet-II-04	+47	≈ +15–25 %*
Sibyll 2.3d	+29	≈ +15–25 %*
EPOS LHC-R	−1	differs between zenith angles
QGSJet-III-01	+18	≈ +30–40 %
Sibyll 2.3e	+28	5–10 % more than 2.3d †

$10^{18.5}$–10^{19} eV, preliminary. Syst. on ΔX_{max} ≈ +9…12 / −11…16 g/cm². * range quoted for the three older models (PRD 2024).
† mainly due to improved SD reconstruction



J. Vícha (Auger), ICRC 2025, arXiv:2507.09802 [figure: Fig. 1]; T. Pierog & K. Werner, arXiv:2508.07105

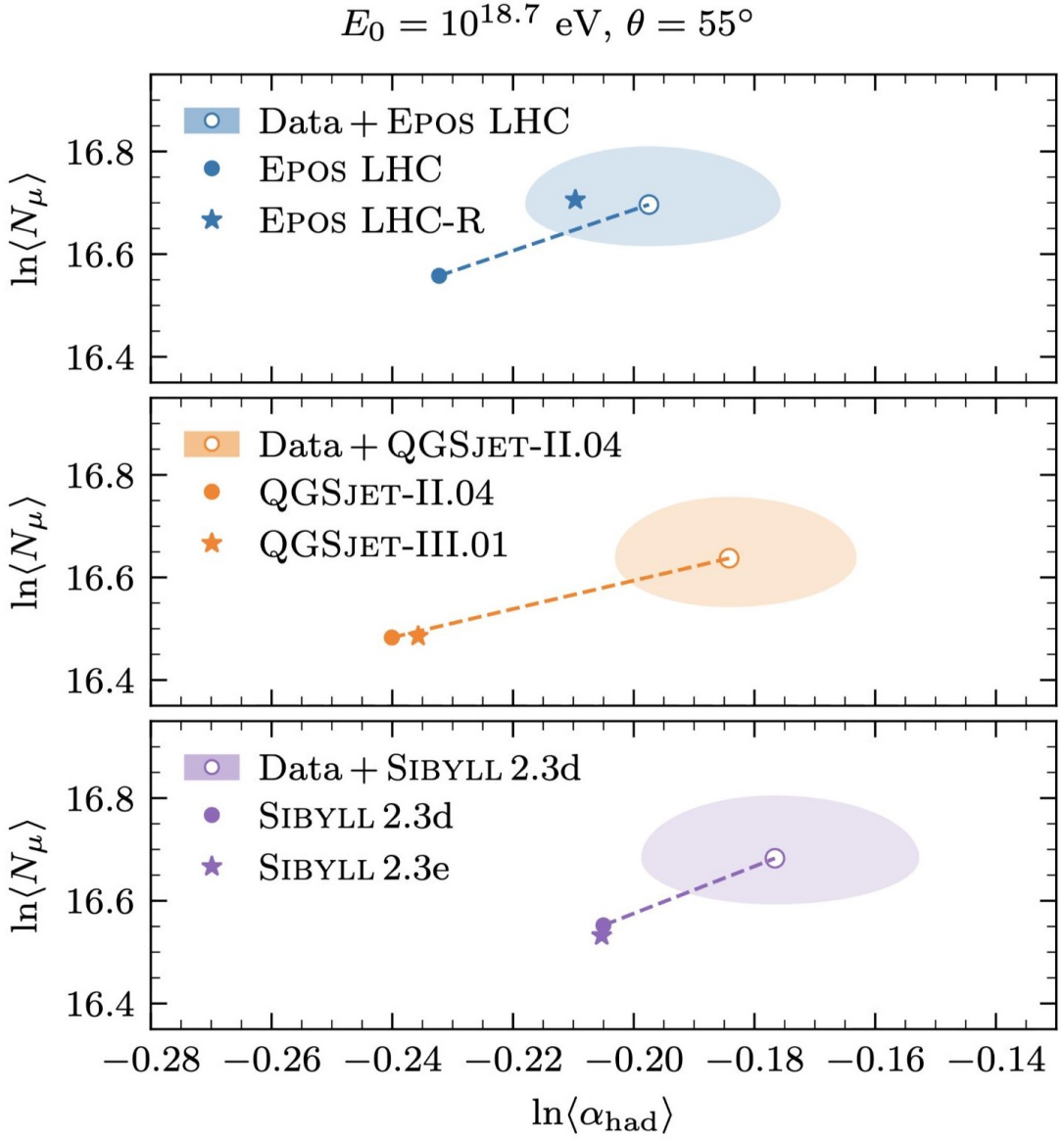
- **EPOS LHC-R**: best-performing model, X_{max} scale compatible
- But rescaling differs between θ ≈ 28° and 55° → muon spectra "too hard"
- "All models seem to predict too hard spectra of muons"

➢ "All models are still unable to describe the measured data satisfactorily well"

Auger 2026: Proton–Air Interactions at √s ≈ 100 TeV

- Models share a common relation between $\langle X_{max} \rangle$ and first-interaction properties
- Map the X_{max} preferred by Auger data onto the first proton–air interaction (√s = 97.7 TeV)
- All models favour **higher elasticity** and a **larger hadronic energy fraction**
- This change "must be amplified by a factor of 2.8 to 4.6 to account for the muon puzzle"

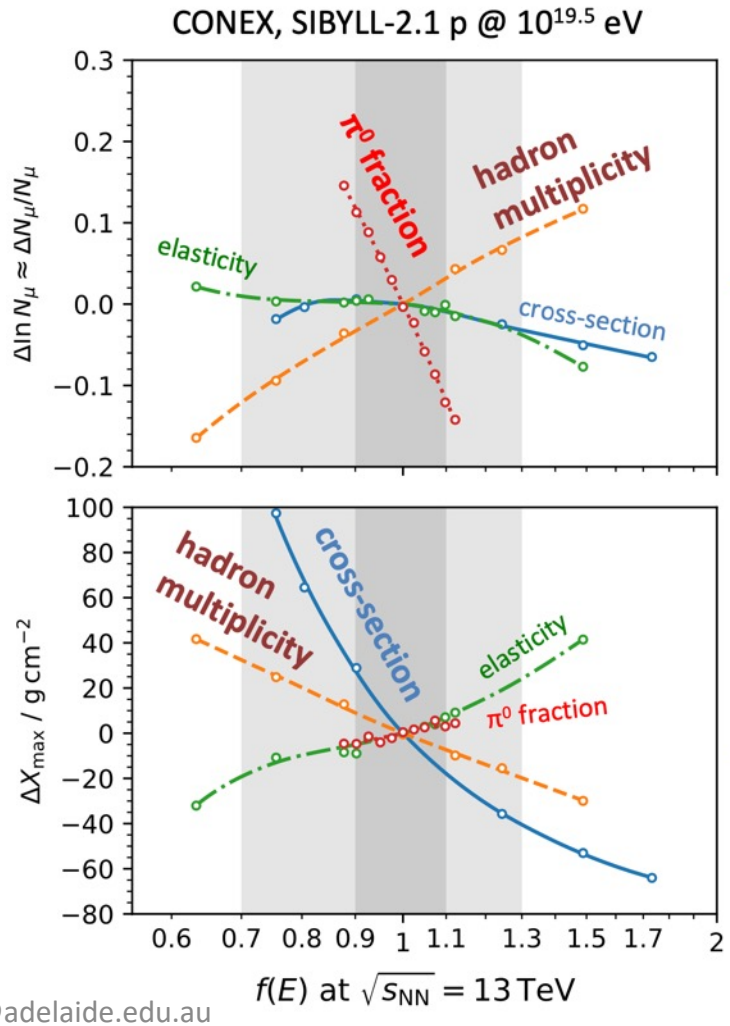
➢ Points to "enhanced far-forward hadronic energy flow", testable at forward LHC experiments



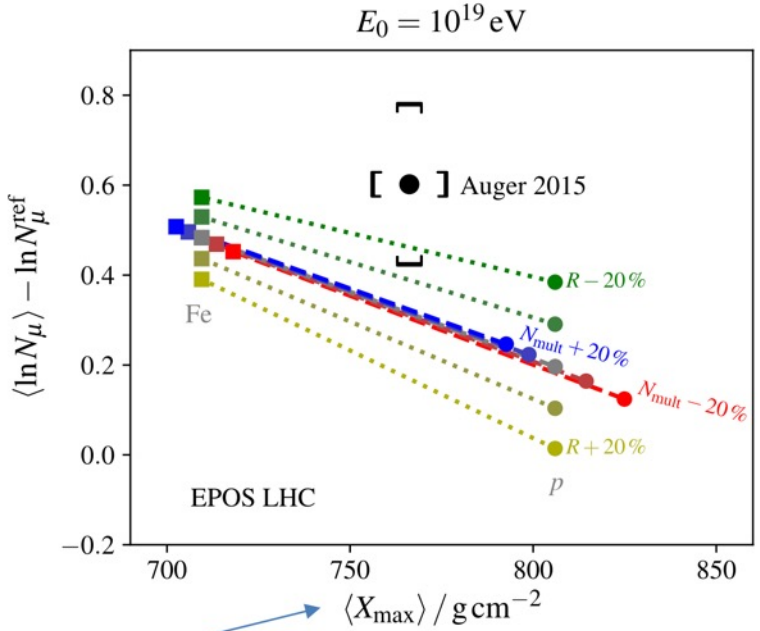
Pierre Auger Collab., arXiv:2607.22323 (July 2026) [figure: Fig. 2]

What Could Fix It? The Energy Ratio R

R. Ulrich, R. Engel, M. Unger, PRD 83 (2011) 054026



S. Baur, HD, M. Perlin, T. Pierog, R. Ulrich, K. Werner, arXiv:1902.09265



$$R = \frac{E_{\pi^0}}{E_{\text{other hadrons}}}$$

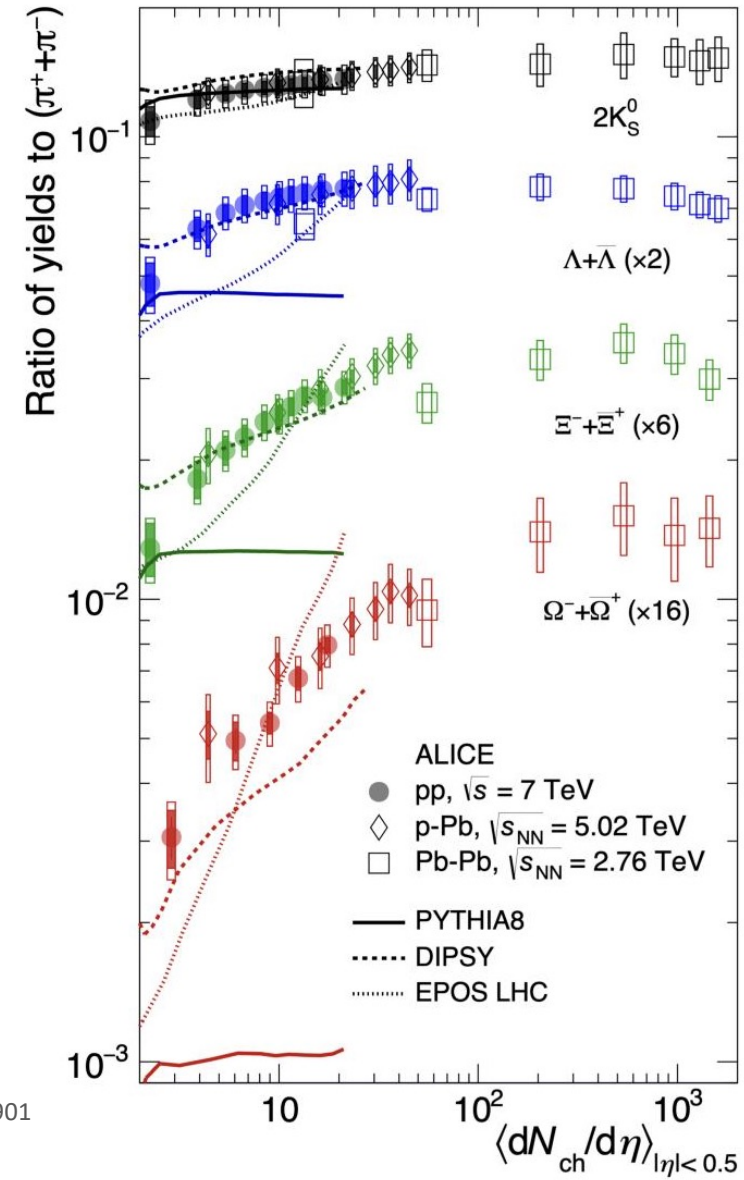
- Only changes to *R* can solve muon puzzle
- Small changes have large effect,
  R needs to be known to about 5 %

Slide taken from Hans Dembinski

Strangeness Enhancement: a Way to Lower R

- ALICE: strange-to-pion ratios rise with event multiplicity, from pp to Pb–Pb
- String fragmentation: R ≈ 0.4
- Statistical hadronization in a dense "core": R ≈ 0.34 → fewer π^0 → more muons
- Core–corona picture implemented in EPOS
- **LHCb** (forward, 2 < y < 4.5): more B_s^0 relative to B^0 at high multiplicity, 3.4σ evidence (b-quark hadronization)

➢ Open question: does it also happen in the very forward region, where air-shower muons come from?

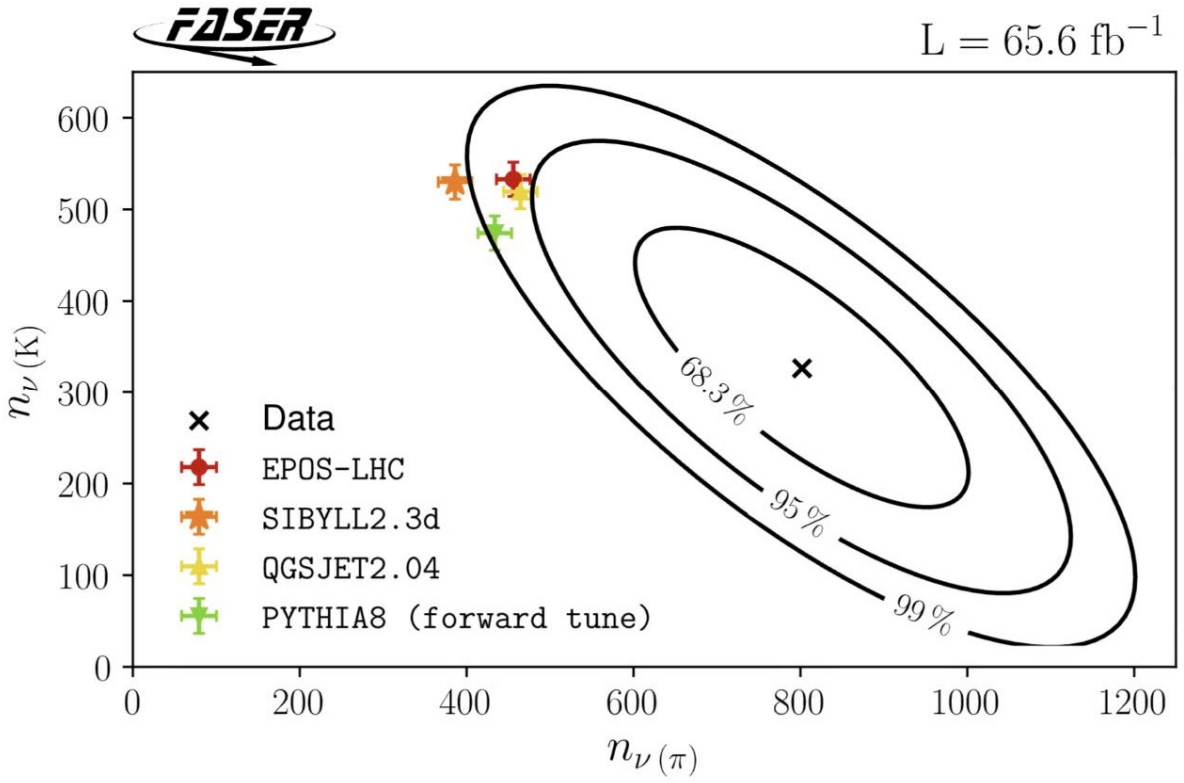


ALICE, Nature Phys. 13 (2017) 535, arXiv:1606.07424 [figure: Fig. 2]; Baur et al., PRD 107 (2023) 094031; LHCb, PRL 131 (2023) 061901

FASER 2025: First Direct Test in the Forward Region

- Neutrinos from forward π and K decays, 480 m downstream of ATLAS
- Energy spectrum separates neutrinos from pions and kaons
- "Larger pion-to-kaon ratio than expected", p-value 5.9 %
- "This disfavors the explanations for the cosmic muon puzzle that involve an enhanced forward strangeness production"
- Run 3 K/π precision (~11 % LHCb, ~8 % FASER) can test the scenario

➢ First hint against forward strangeness enhancement, not yet conclusive

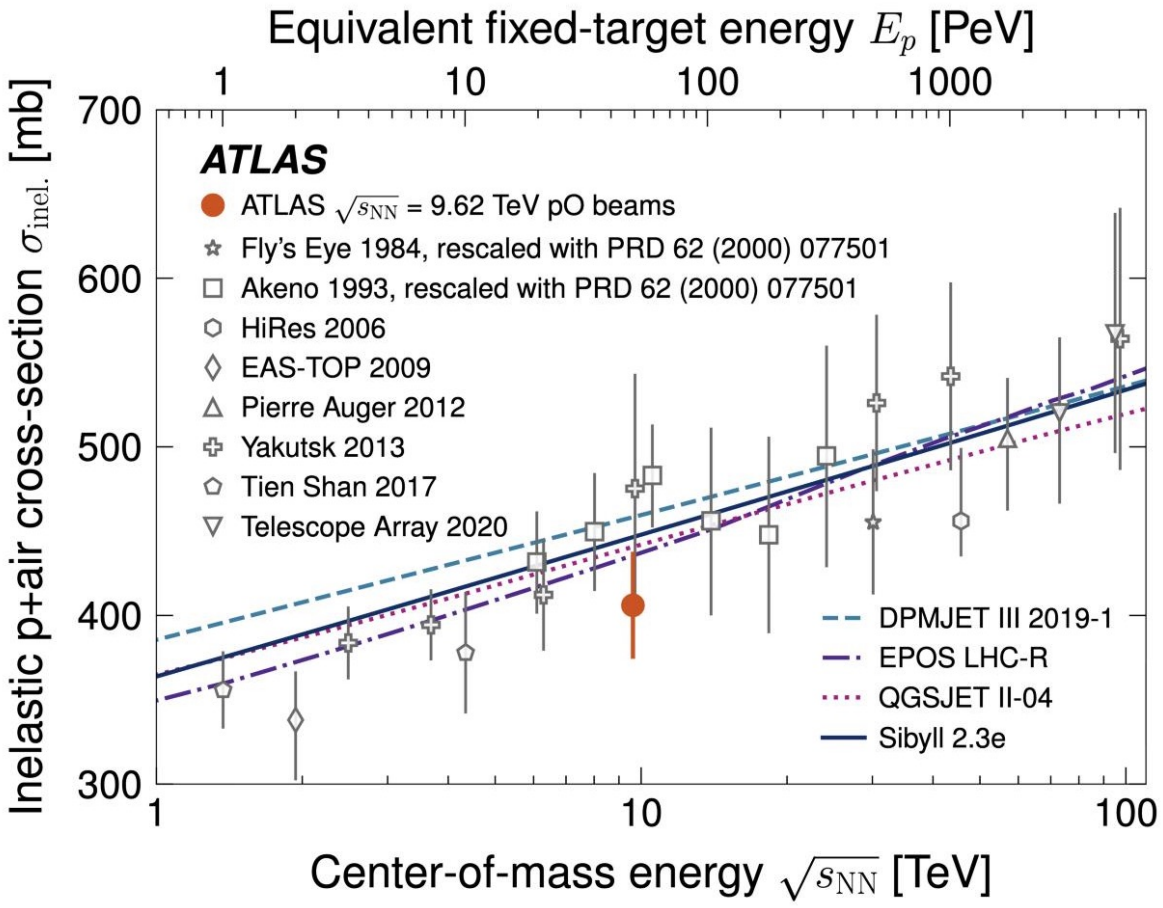


FASER Collab., Phys. Rev. Lett. 134 (2025), arXiv:2412.03186 [figure: Fig. 4]; K. Ohashi, A. Fedynitch, H. Menjo, arXiv:2509.24021

The 2025 LHC Oxygen Run

- Summer 2025: first proton–oxygen and oxygen–oxygen collisions: the system of air showers
- **ATLAS p–O** ($\sqrt{s_{NN}}$ = 9.62 TeV): σ_{inel}(p–O) = 438 ± 6 ± 10 ± 30 mb, extrapolated σ_{inel}(p–air) = 406 mb
  - compatible with EPOS and HIJING, ≥ 3σ below the other models
  - data precision ≈ 10× better than the spread between models
- **LHCb** recorded ≈ 33 nb^{-1} of p–O: results pending

➢ Direct input for air-shower models

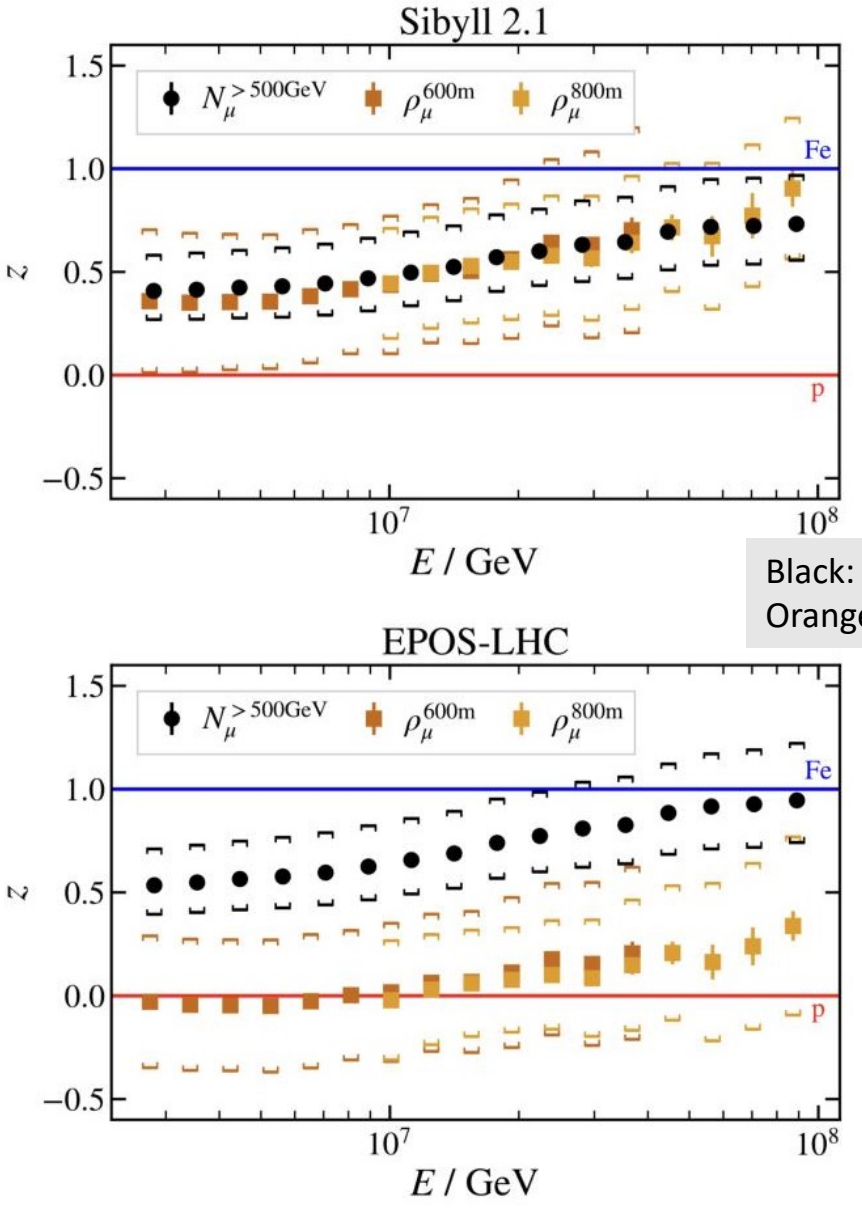


ATLAS Collab., Phys. Rev. Lett. (2026), arXiv:2604.05512 [figure: Fig. 1(b)]; LHCb outreach, Sept. 2025

IceCube: GeV and TeV Muons in the Same Showers

- IceTop (surface) + in-ice array: 2.5–100 PeV primaries, near-vertical showers
- Measure GeV and TeV muons from the same showers
- Mean number of muons above 500 GeV agrees with simulations (Sibyll 2.1, QGSJet-II.04, EPOS-LHC)
- Compared with IceTop GeV muons: inconsistent low- and high-energy muon predictions for EPOS-LHC

➢ Muon energy matters: models must get the muon energy spectrum right



Black: TeV muons
Orange: GeV muons

IceCube Collab., Phys. Rev. D 112 (2025) 082004, arXiv:2506.19241 [figure: Fig. 12, Sibyll 2.1 and EPOS-LHC panels]

Open Questions → What Atmospheric Muons Can Test

Open question (2024 – 2026)	What our atmospheric-muon measurement provides
Forward π/K production (FASER: more pions, fewer kaons than expected)	Conventional muon flux, from forward π and K decays
Muon energy spectra (Auger: too hard in all models; IceCube: GeV vs TeV inconsistent for EPOS-LHC)	The muon energy spectrum itself, ~600 GeV to ~10 PeV
Forward heavy-hadron production (charm, unflavoured mesons)	Prompt muon component above ~1 PeV

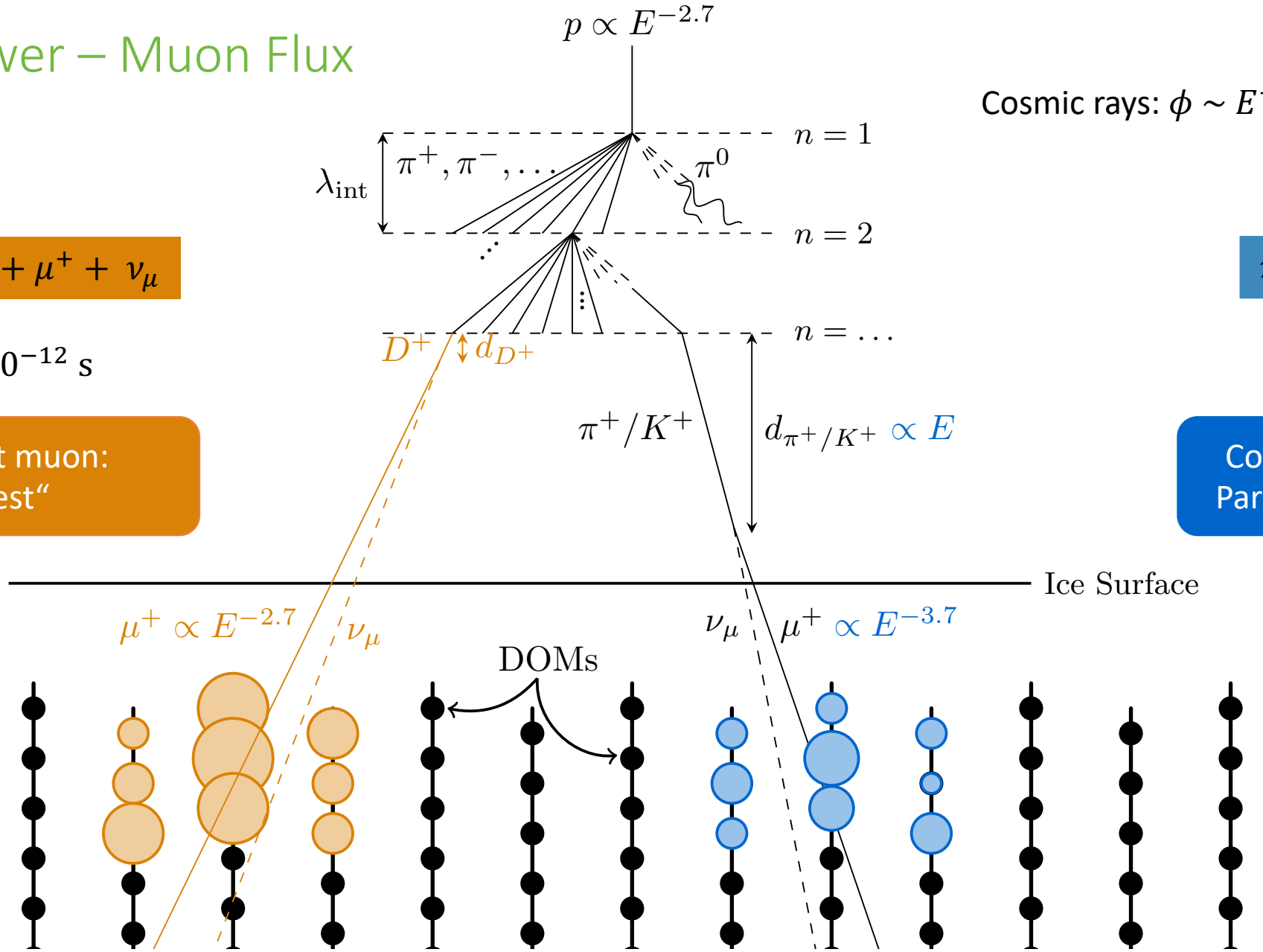
- High statistics, inclusive flux: no shower-by-shower energy and mass reconstruction needed

Atmospheric Muons

Unfolding 12 years of IceCube data

Air Shower – Muon Flux

Cosmic rays: $\phi \sim E^{-\gamma}$



$D^+ \rightarrow \overline{K}^0 + \mu^+ + \nu_\mu$

$\tau \sim 10^{-12}$ s

Prompt muon:
„Rest“

$\pi^+ \rightarrow \mu^+ + \nu_\mu$

$\tau \sim 10^{-8}$ s

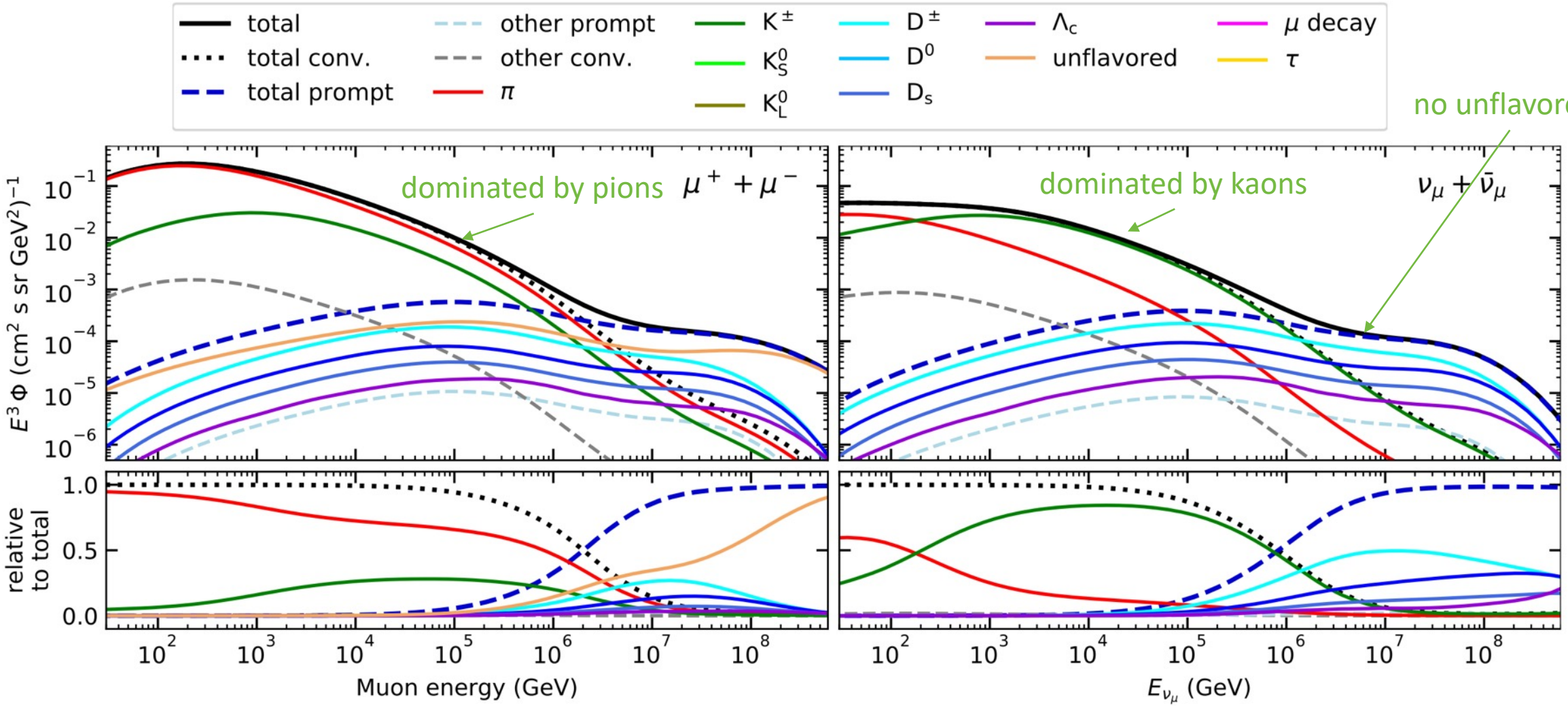
Conventional Muon:
Parent is pion or kaon

pascal.gutjahr@adelaide.edu.au

adapted from L. Neste

Atmospheric Muon and Muon-Neutrino Flux

10.1103/PhysRevD.100.103018



Goal: Measure atmospheric muon flux

- Detect prompt component
- Handle on pion/kaon ratio
- Handle on charmed mesons → prompt neutrino prediction

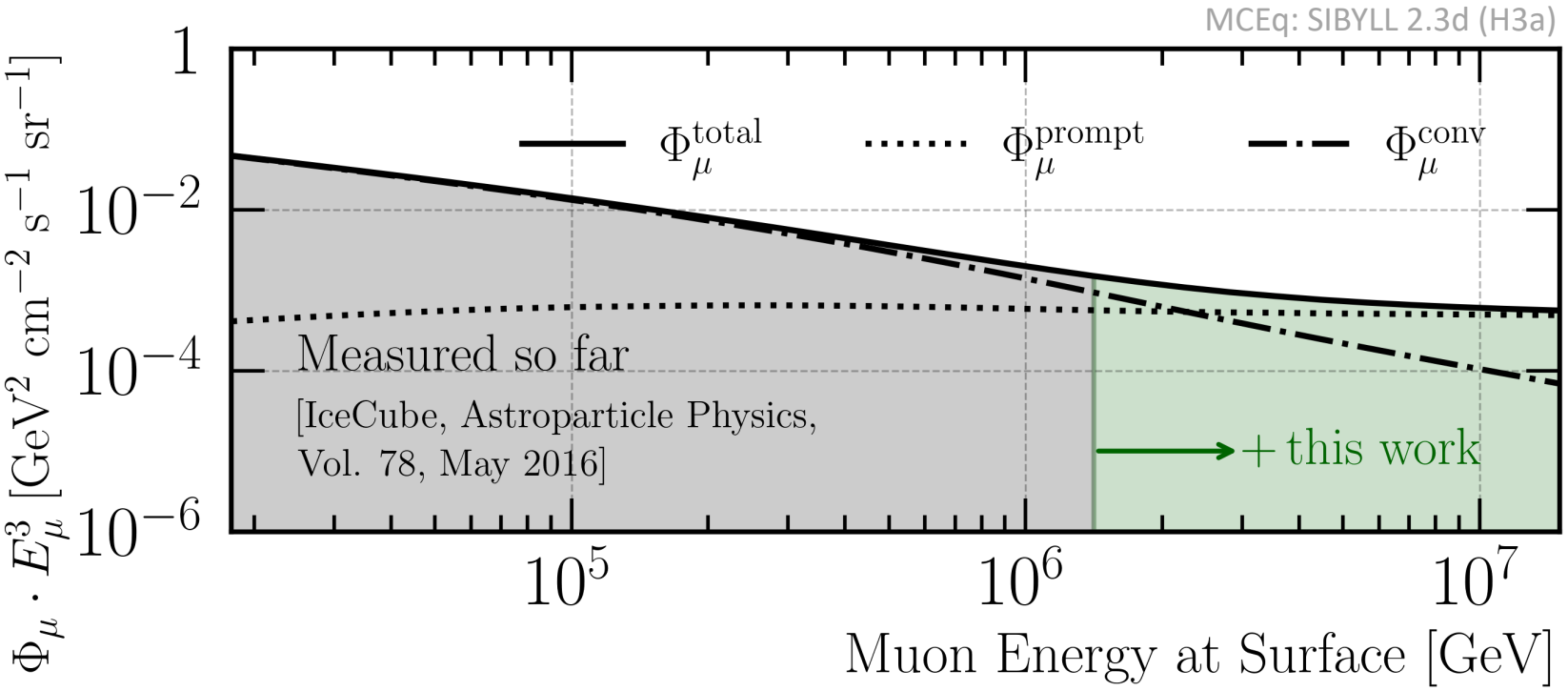
Goal: Measure Muon Flux at Surface

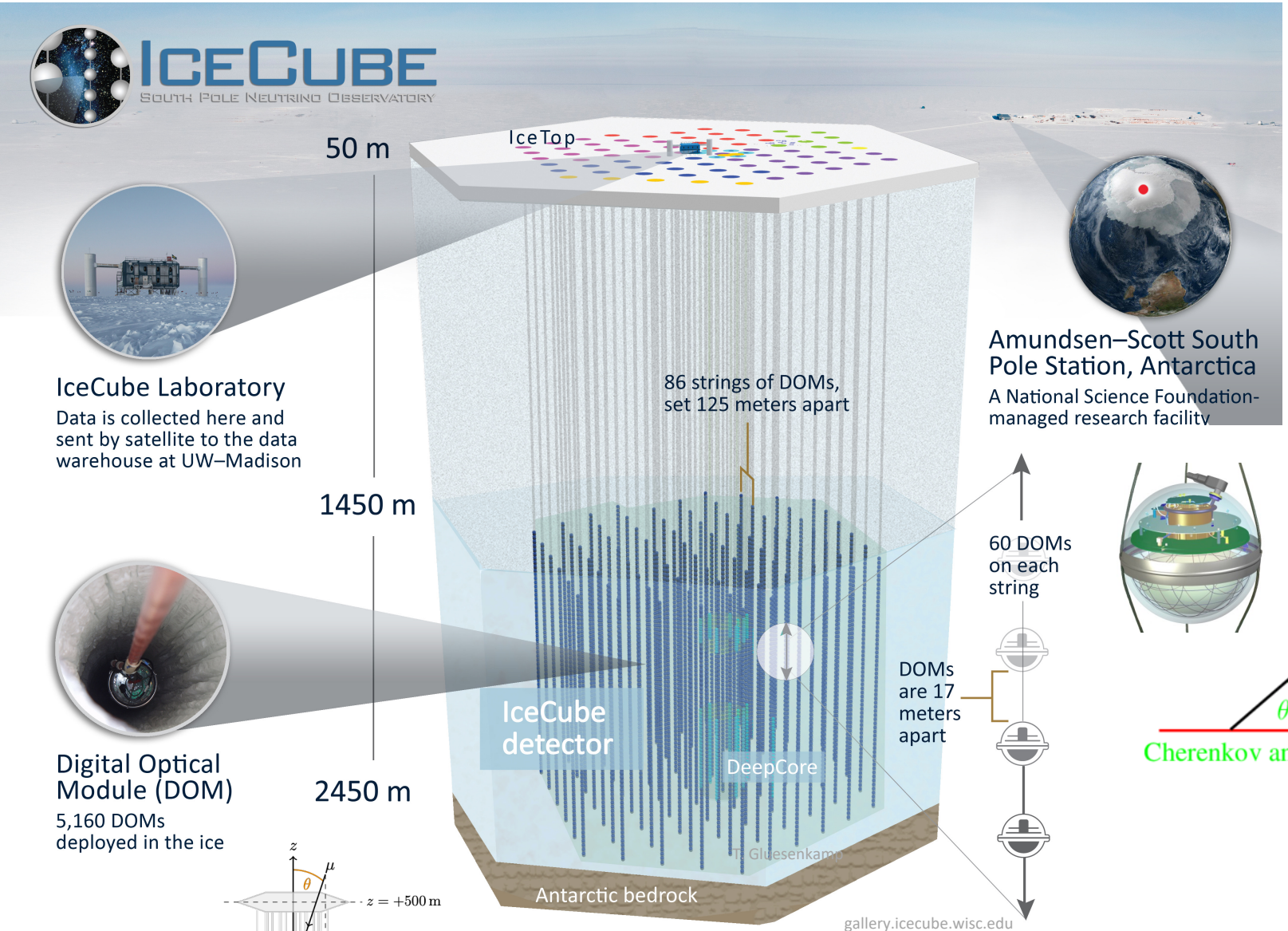
Motivation

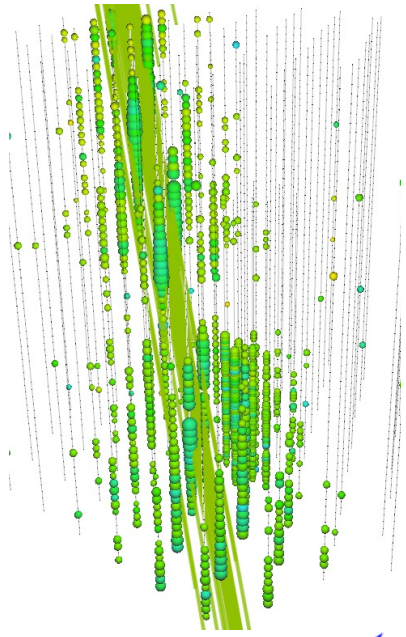
- Measure the muon energy spectrum at the surface: ~600 GeV to ~10 PeV
- Separate conventional (π, K) and prompt (charm, unflavoured) components
- Provide a new tuning input for hadronic interaction models

Long term

- Predict the prompt atmospheric neutrino flux (background for neutrino astronomy)

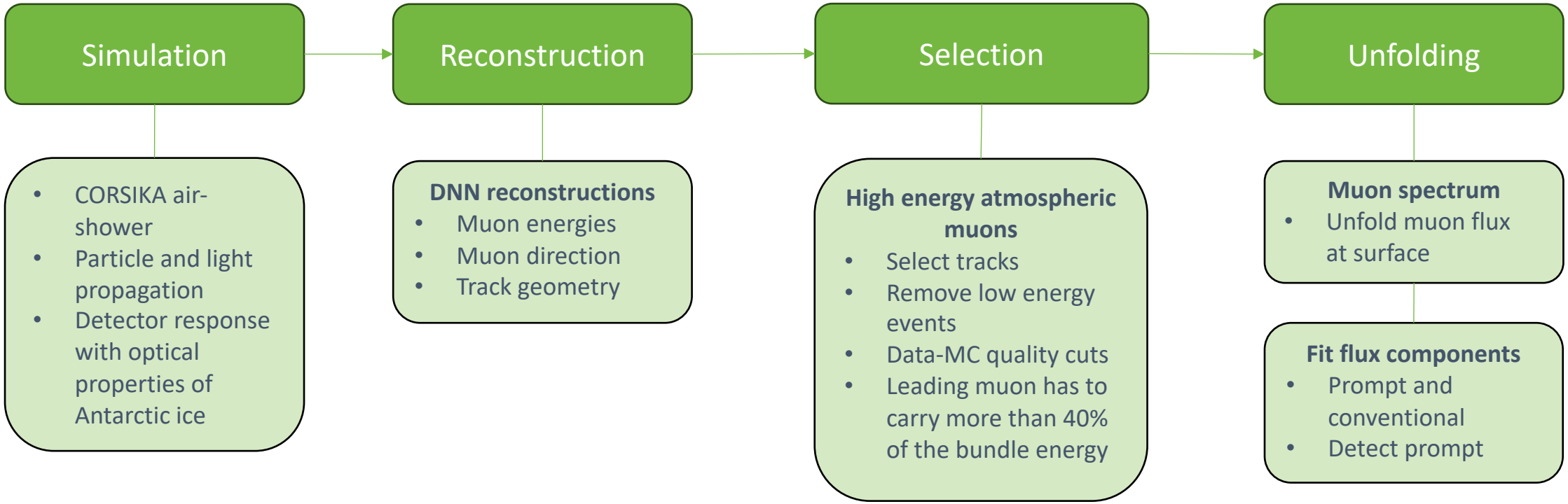






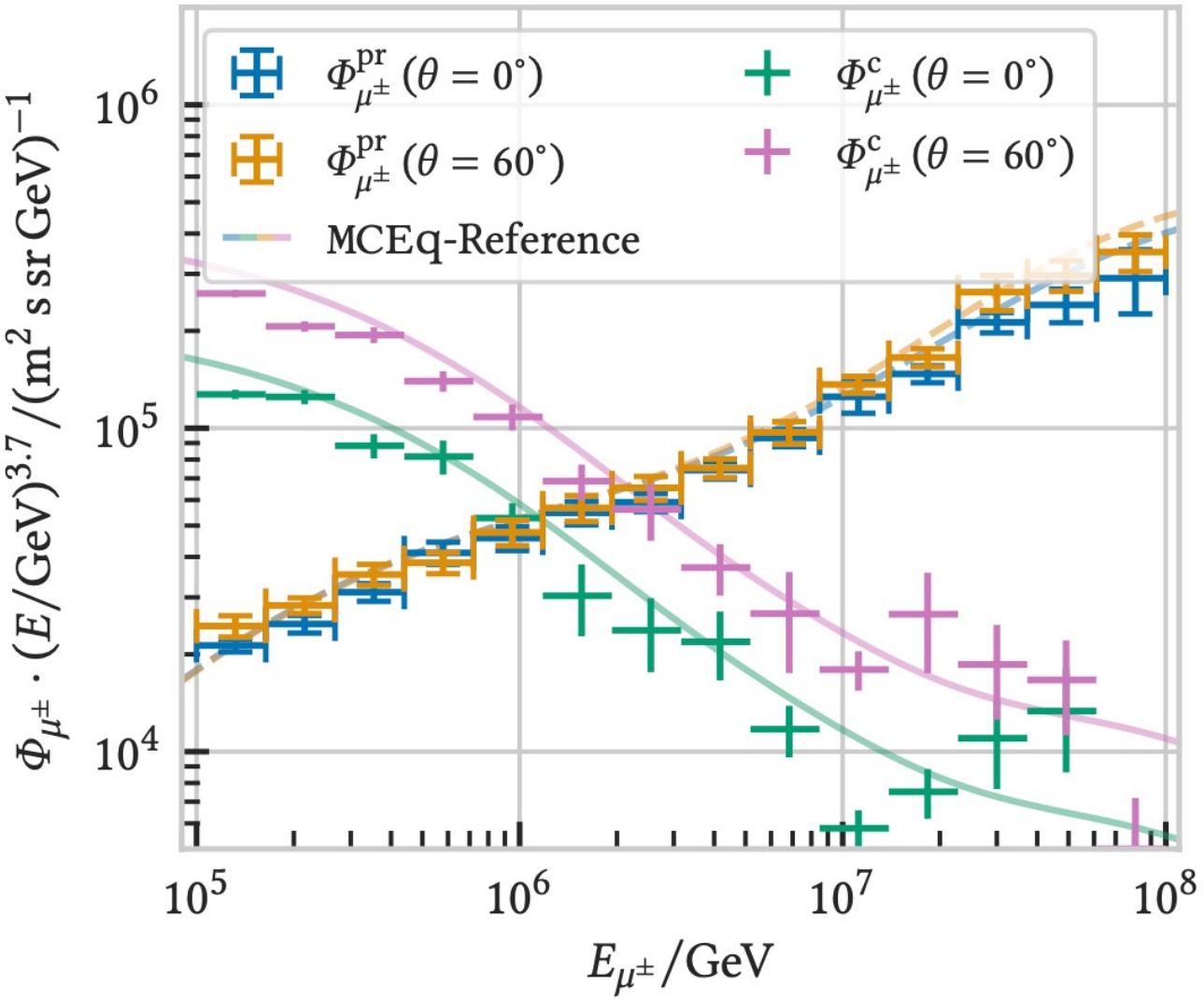
- Muon bundle
- Multiple muon tracks
- ➢ Most energetic muon: “leading muon”

Analysis Overview



Simulation: CORSIKA 7 + PANAMA

MCEq: tool to numerically solve the cascade equations that describe the evolution of particle densities as they propagate through a gaseous, dense medium
https://github.com/mceq-project/MCEq



- Good agreement for inclusive flux

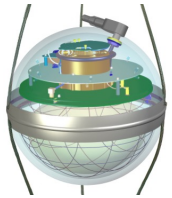
Python package developed – PANAMA

- Execute CORSIKA 7 (multi core)
- Read DAT files → pandas DataFrames
- Parse EHIST option
- Calculate primary weightings

https://github.com/The-Ludwig/PANAMA
Neste & Gutjahr et al., EPJ C, 85, 759 (2025)

pascal.gutjahr@adelaide.edu.au

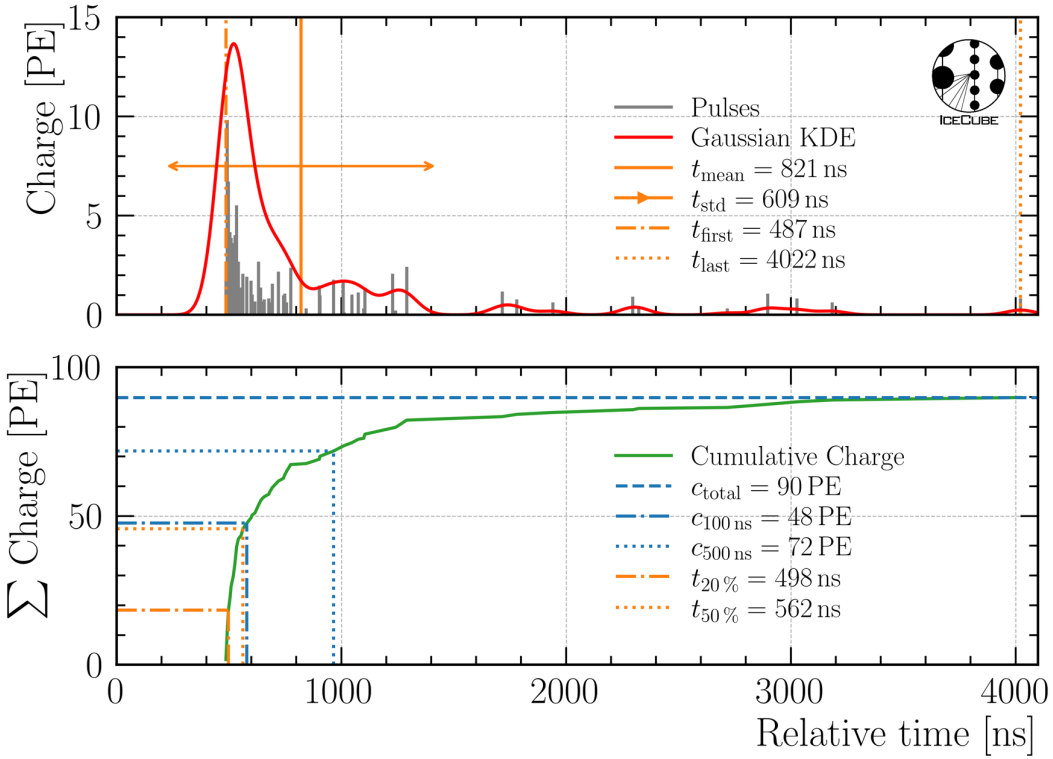
DNN Reconstructions

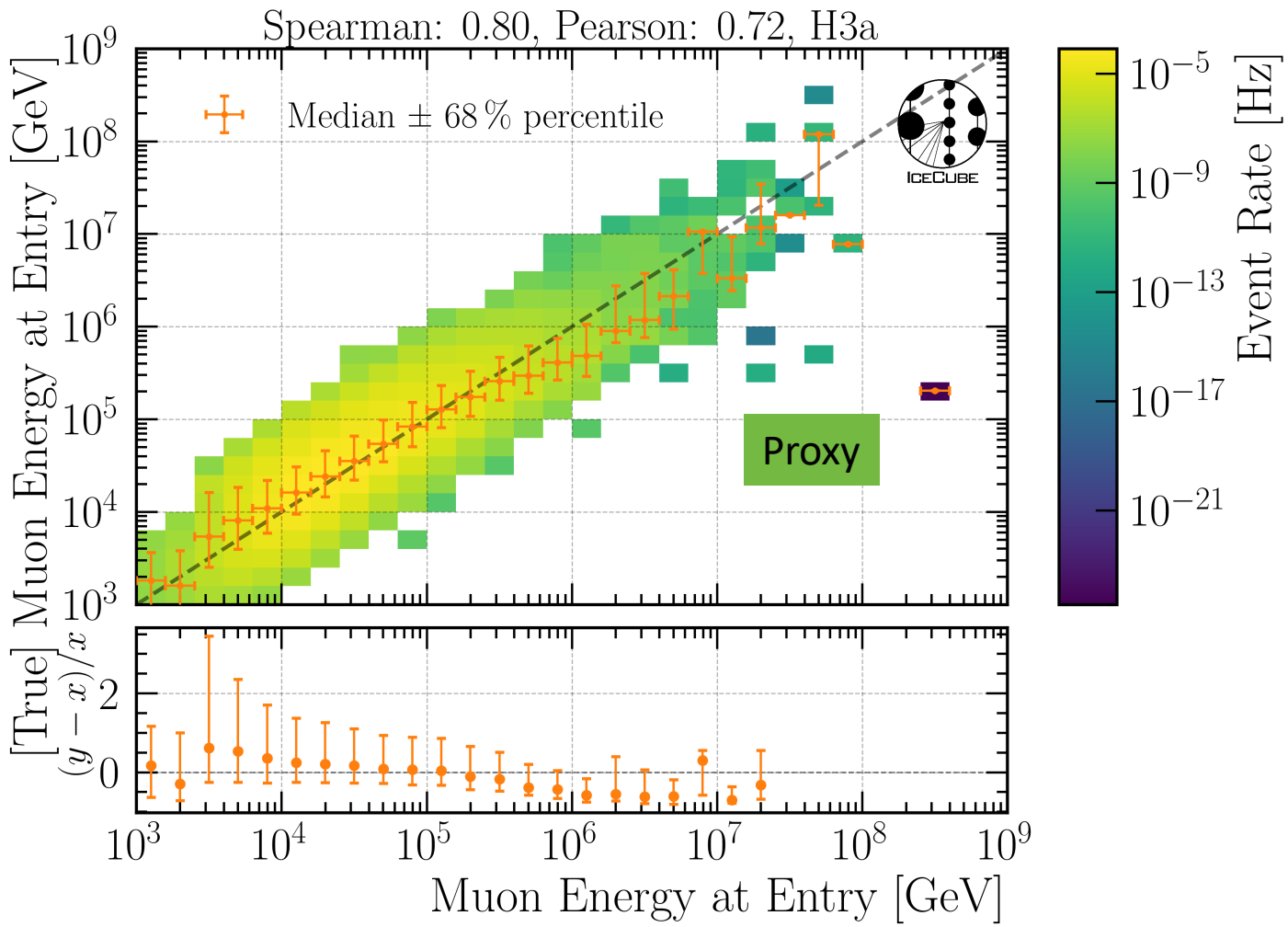


- Targets: energy, direction, track geometry
- Each DOM records a pulse series
  - Input features: charge + time per DOM
- Trained on several simulation sets

➢ 4 final networks for the analysis

➢ 16 quantities + uncertainties



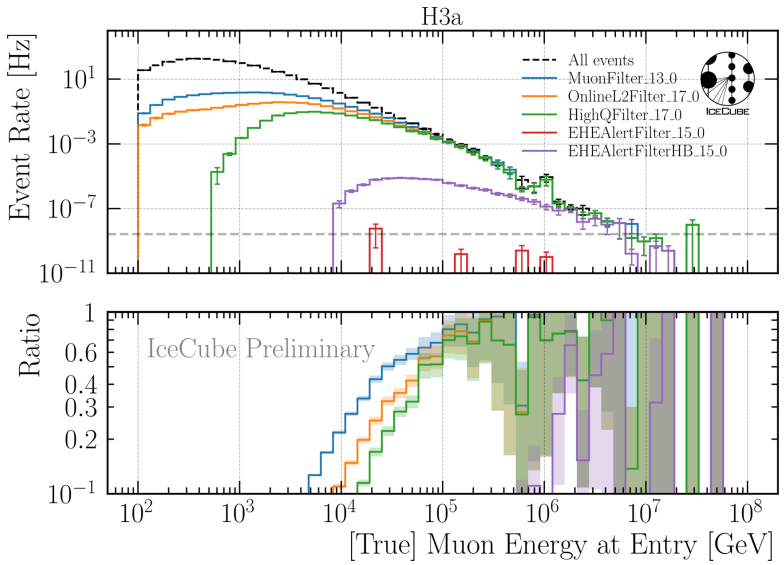


➢ 1. Leading-muon energy reconstruction

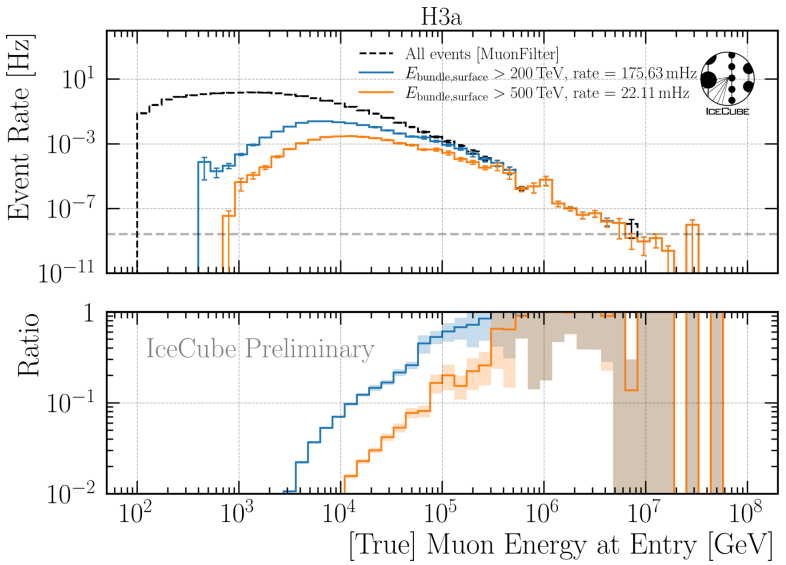
Selection

Adelaide University

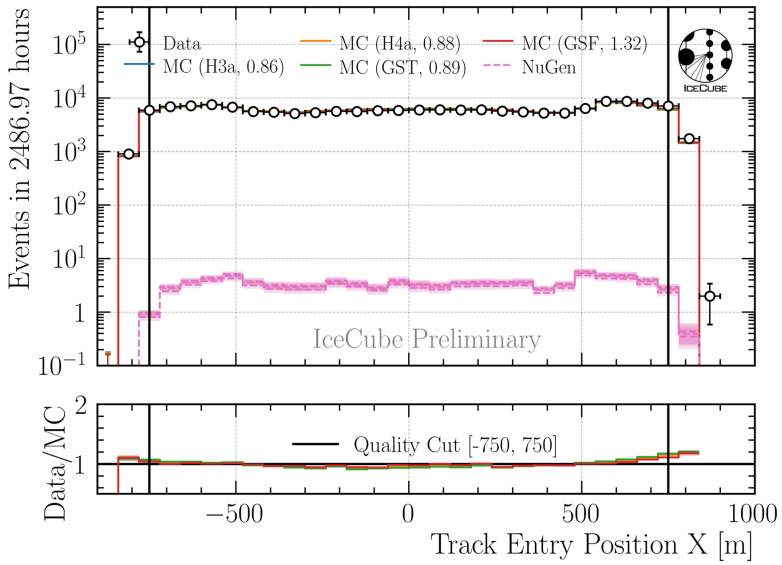
Level 3: muon filter



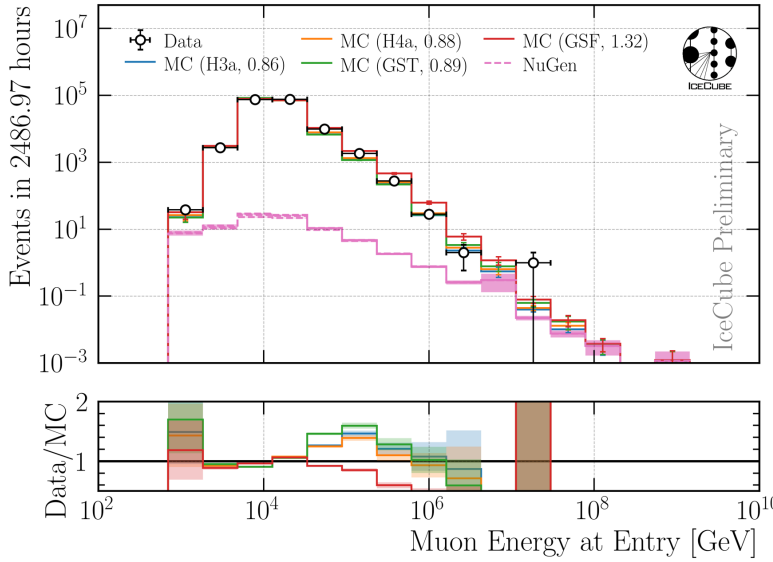
Level 4: bundle energy at surface > 500 TeV



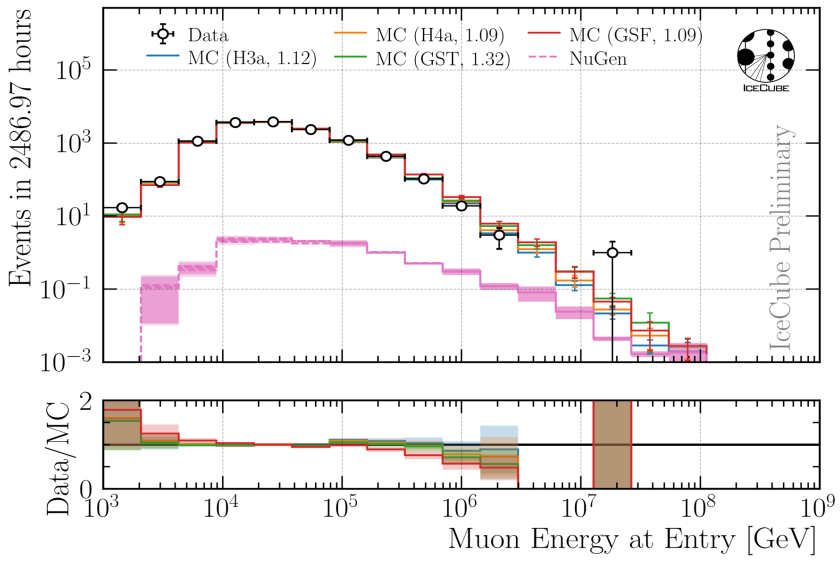
Level 5: 23 data/MC quality cuts



Final level: leadingness > 40 %



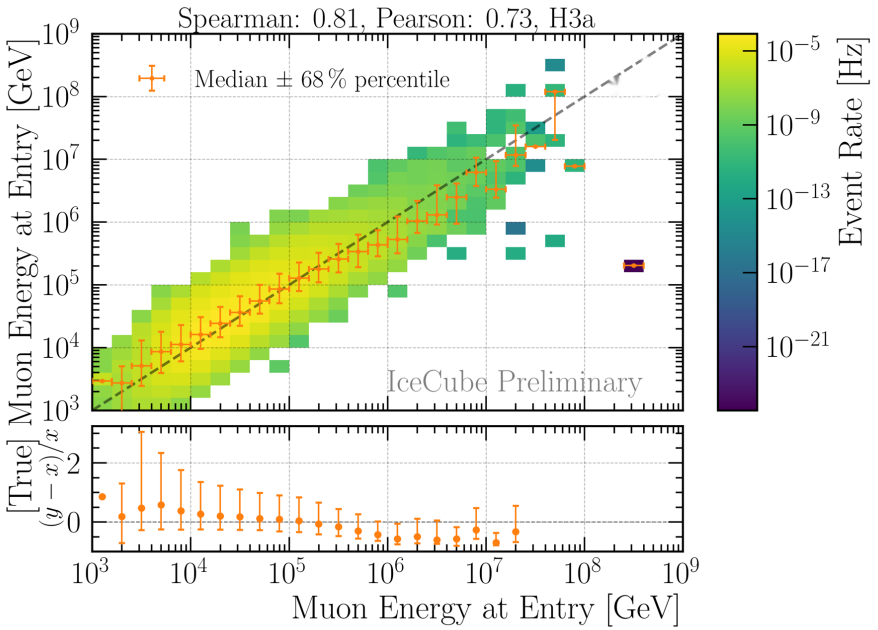
$$L = \frac{\text{Leading Energy at Entry}}{\text{Bundle Energy at Entry}} > 40\ \%$$



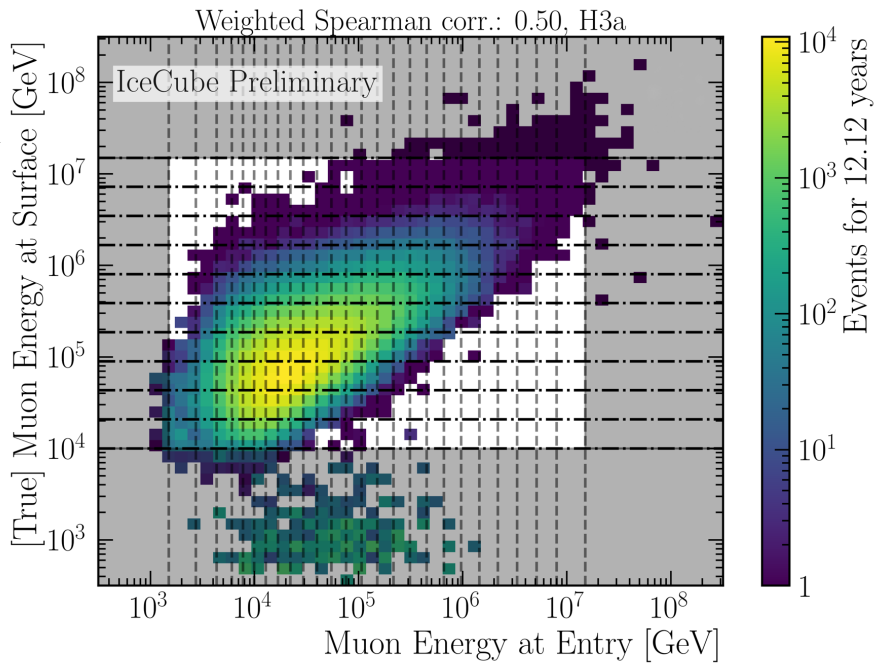
*Numbers in brackets give the scaling factor applied to match the total number of data events.

Unfolding Procedure

$$g(y) = \int_{E_0}^{E_1} A(E_\mu, y) f(E_\mu)\, \mathrm{d}E_\mu + b(y)$$



Use reconstructed leading muon energy at entry to unfold individual muon energy at surface



2. Maximum likelihod method:

$$\mathcal{L}(\vec{g}|\vec{f}) = \prod_{j=1}^{M} \frac{\lambda_j^{g_j}}{g_j!} \exp(-\lambda_j)$$

$$= \prod_{j=1}^{M} \frac{(A\vec{f})_j^{g_j}}{g_j!} \exp(-(A\vec{f})_j)$$

3. Thikonov regularization:

$$t(\vec{f}) = -\frac{1}{2}(C\vec{f})^T(\tau 1)^{-1}(C\vec{f})$$

4. Optical ice systematics:

$$\mathbf{A}' = \mathbf{w}_j^\top \cdot \mathbf{A}$$

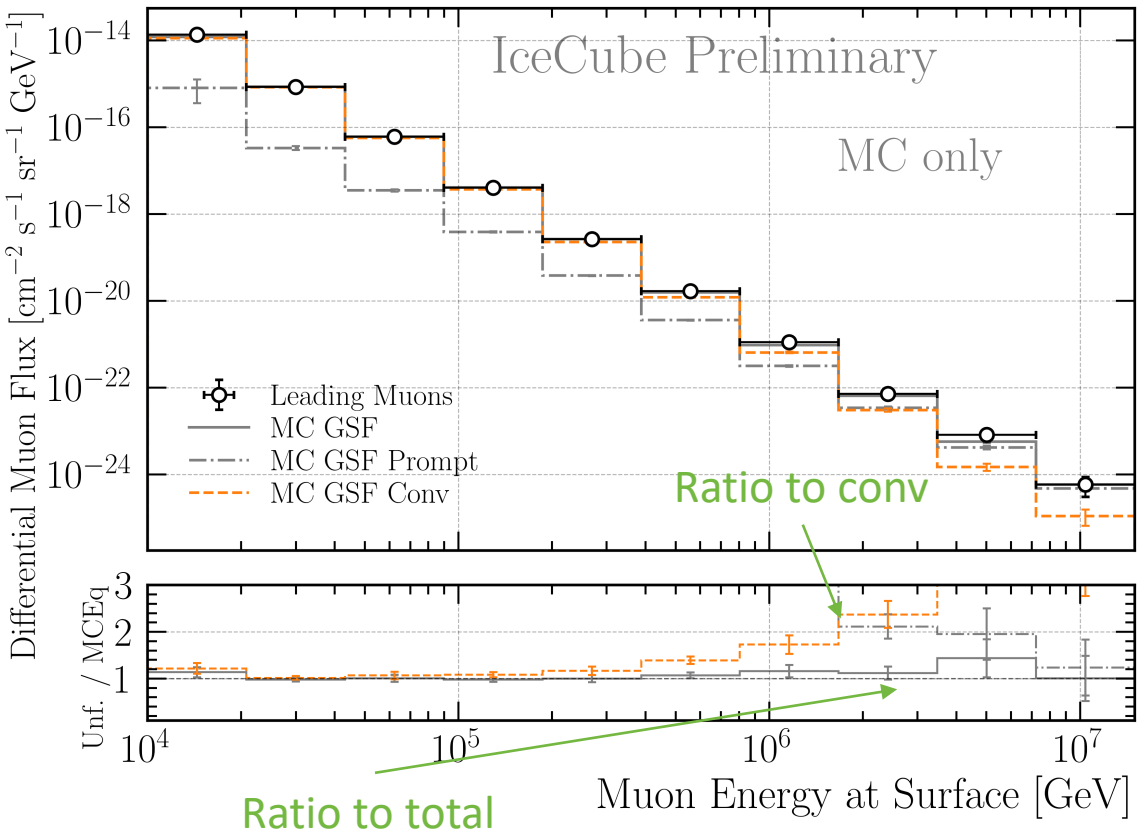
5. Maximize $\log(\mathcal{L}(\vec{g}|\vec{f})) + t(\vec{f})$ with respect to $\vec{f}$

GitHub **funfolding** by M. Börner

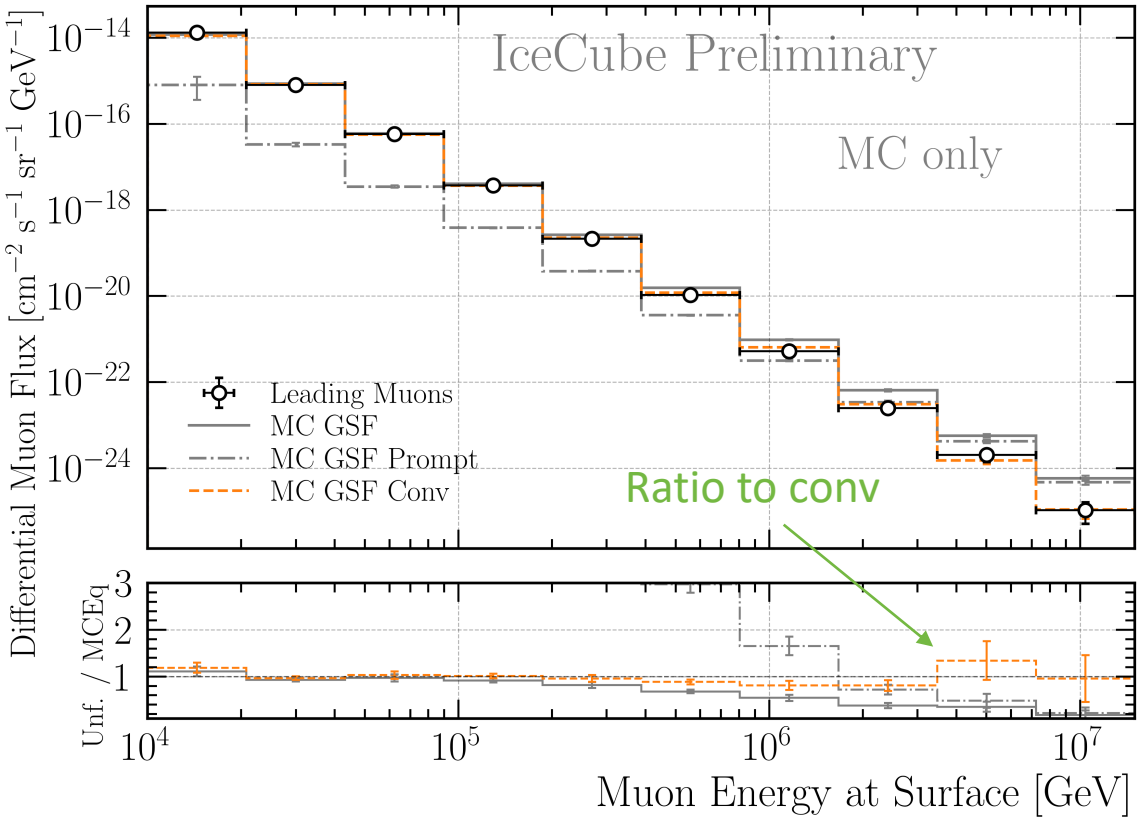
6. Perform 4 unfoldings: one per primary model (H3a, H4a, GST, GSF)
→ Average = Final result

Sensitive to Prompt?

sample has prompt & conv



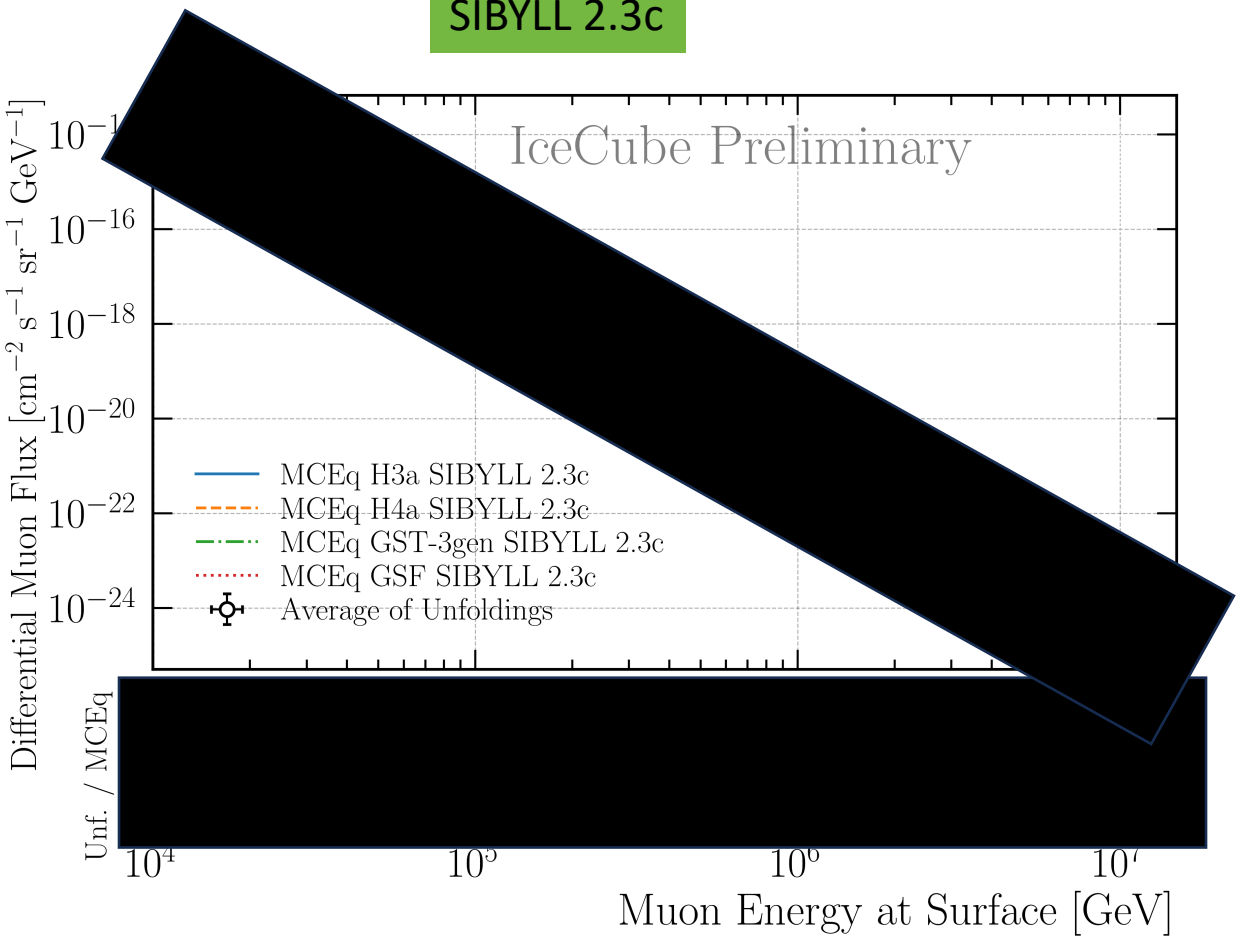
sample has only conv



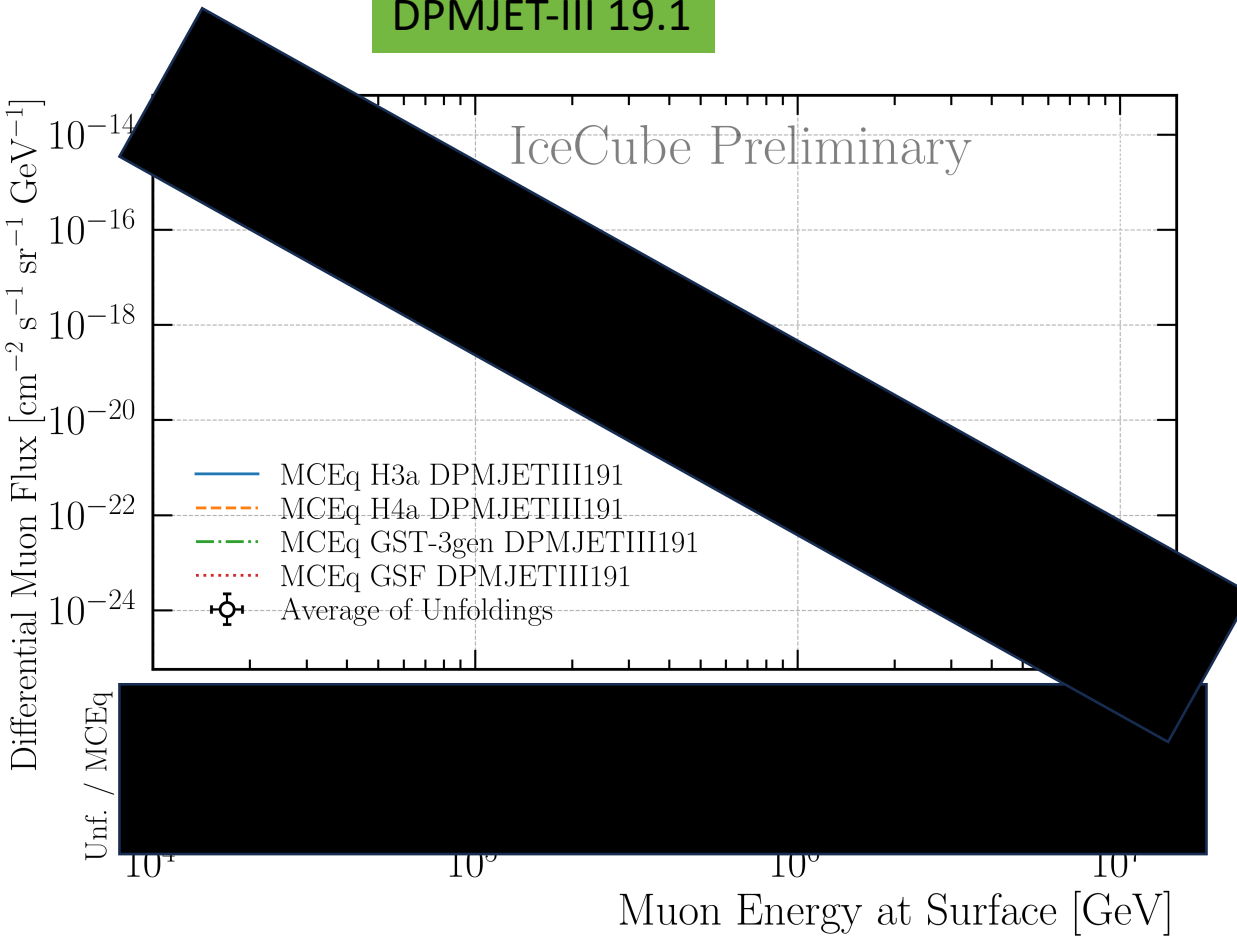
➢ Unfolding is sensitive to the prompt component

Compare Unfolded Flux with MCEq Predictions

SIBYLL 2.3c



DPMJET-III 19.1



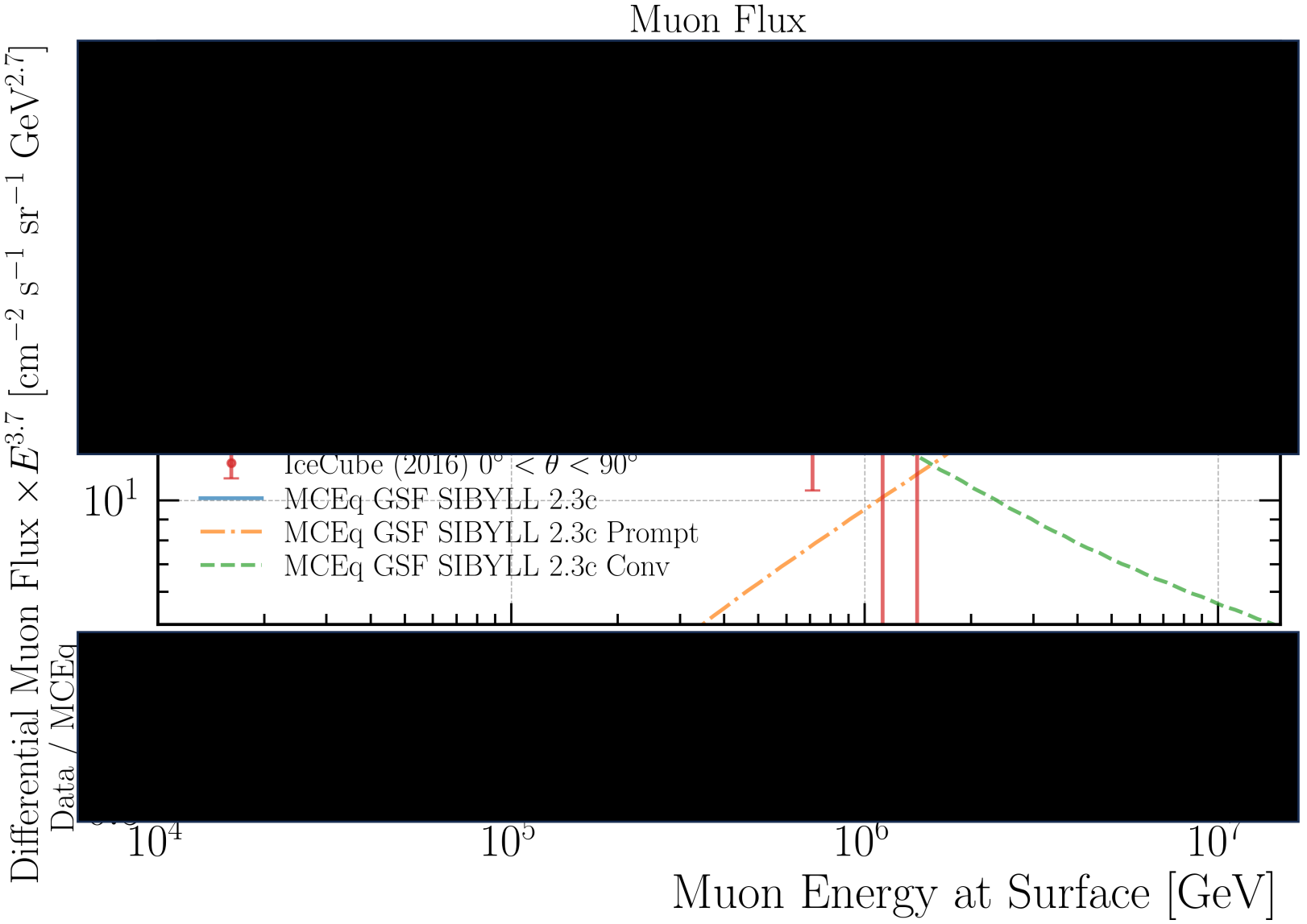
- Best agreement found for GSF with SIBYLL 2.3c

- For DPMJET: shapes fit to unfolded flux, but normalization offset (H3a/H4a/GST)
- GSF shows shape mismatch

Compare to Previous IceCube Result

- This work:
  zenith 0°-78°
- IceCube[1]:
  1. zenith 0°-90° (correlated unc.)
  2. zenith 0°-60° (no unc.)

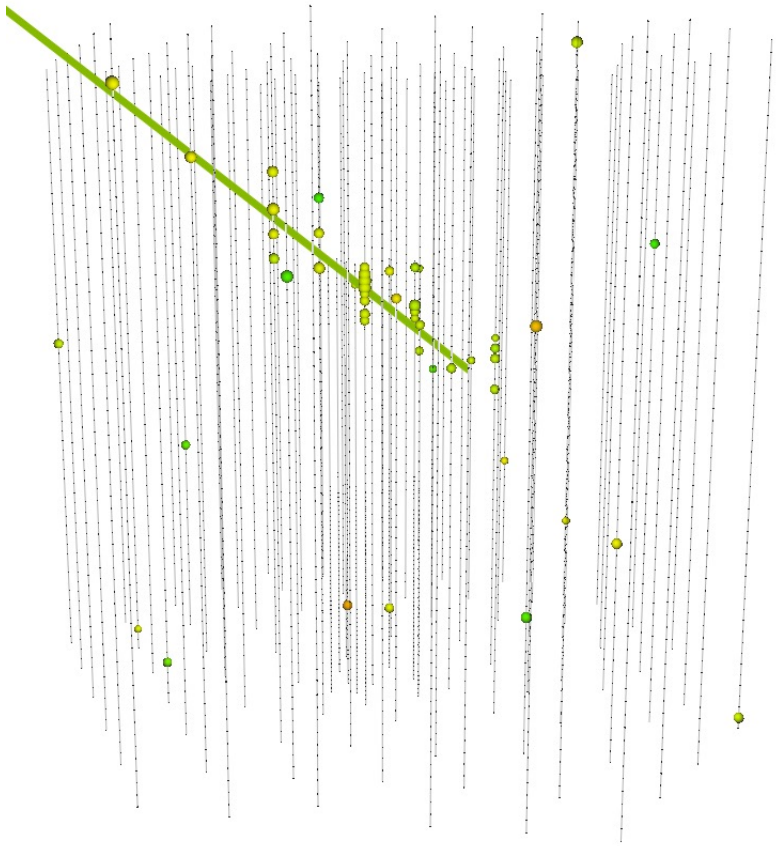
- ➢ Measurements are compatible
- ➢ Uncertainties significantly reduced
- ➢ Extended energy range by 1 order of magnitude

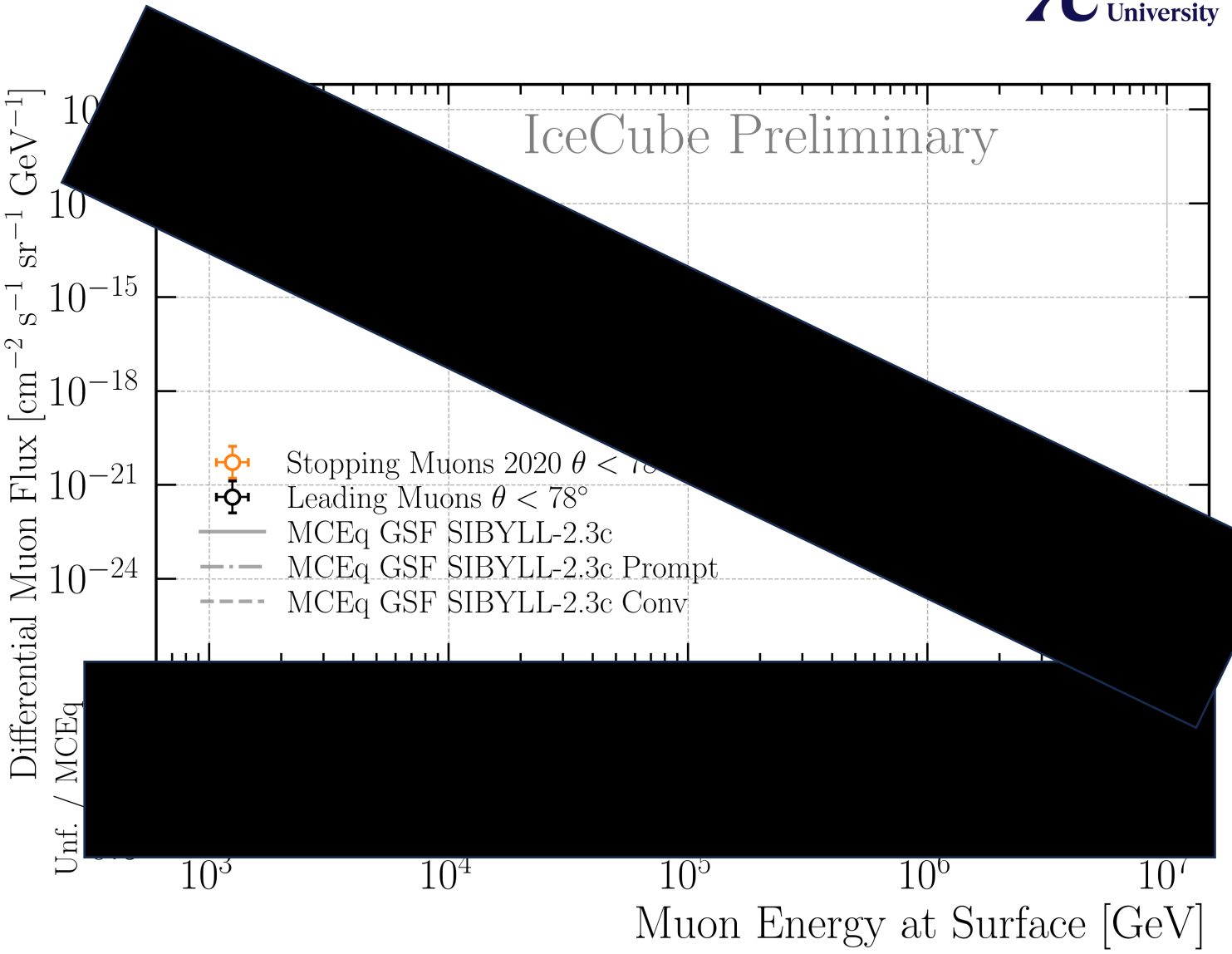


*Unfolded flux is scaled by the logarithmic bin center, which is an approximation

[1]Astropart.Phys. 78 (2016) 1-27

Include Stopping Muons





Stopping muons

- Decay in the detector volume
- Sensitive to low-energy range from 600 GeV to 10 TeV (conventional component)

➢ Unfolded year 2020 of IceCube data → slight tension around 10 TeV

Fit Muon Flux Components

- Muon flux per bin *i*:

$$\mu_i(N_{\text{prompt}}, N_{\text{conv}}) = N_{\text{conv}} \cdot \Phi^{\text{baseline}}_{\text{conv},i} + N_{\text{prompt}} \cdot \Phi^{\text{baseline}}_{\text{prompt},i}$$

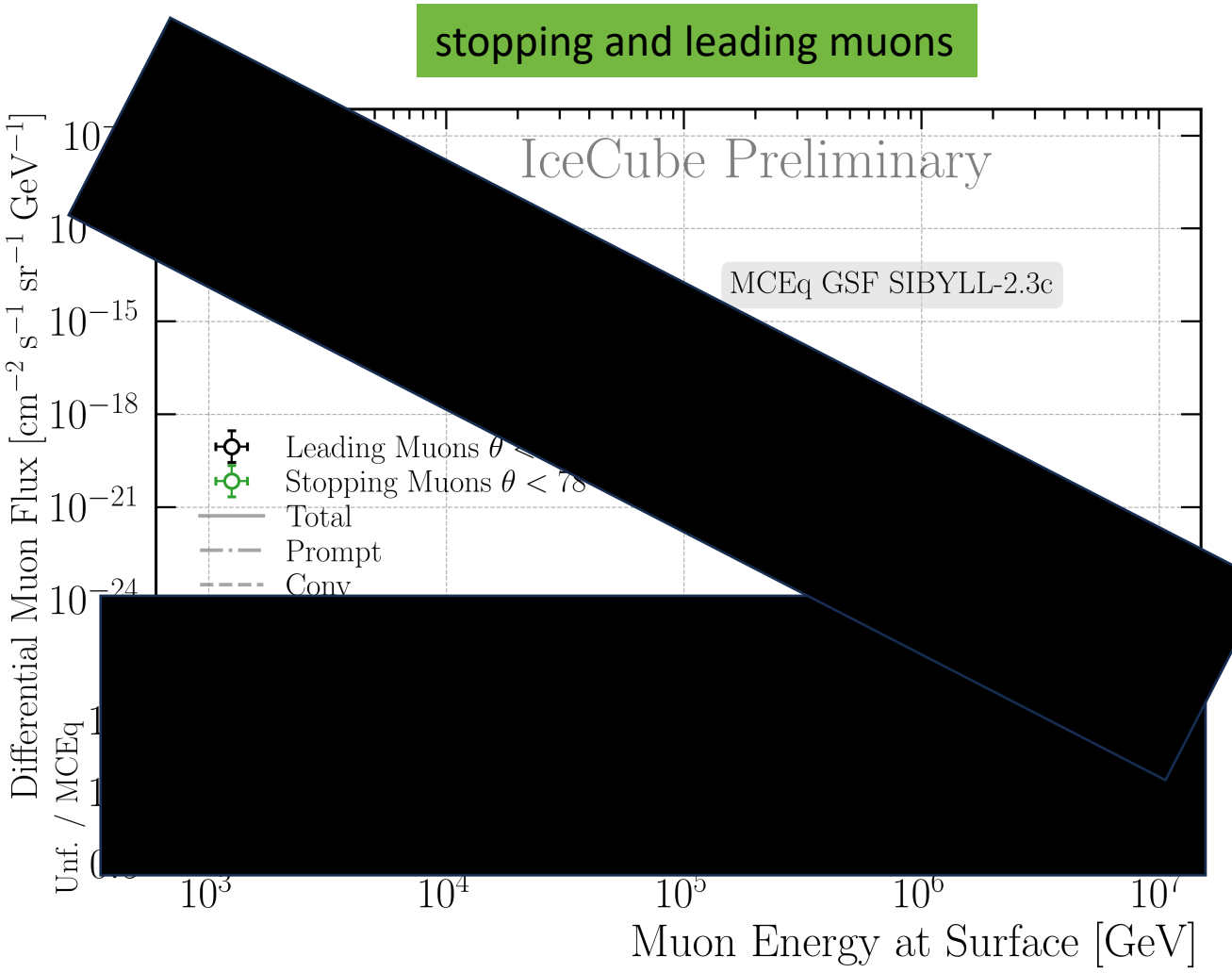
- Minimize Chi2 over all bins:

$$\chi^2(N_{\text{prompt}}, N_{\text{conv}}) = \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

- Fit normalization for prompt and conventional for a given model prediction
  - Here: MCEq with SIBYLL 2.3c weighted to GSF

- Several model predictions can be tested

➢ Measure prompt muon flux
➢ Fine-tune hadronic interaction models



Conclusion & Outlook

Muon puzzle

- Still unsolved with post-LHC models
- Auger 2024–26: deeper X_{max} and more far-forward hadronic energy

LHC

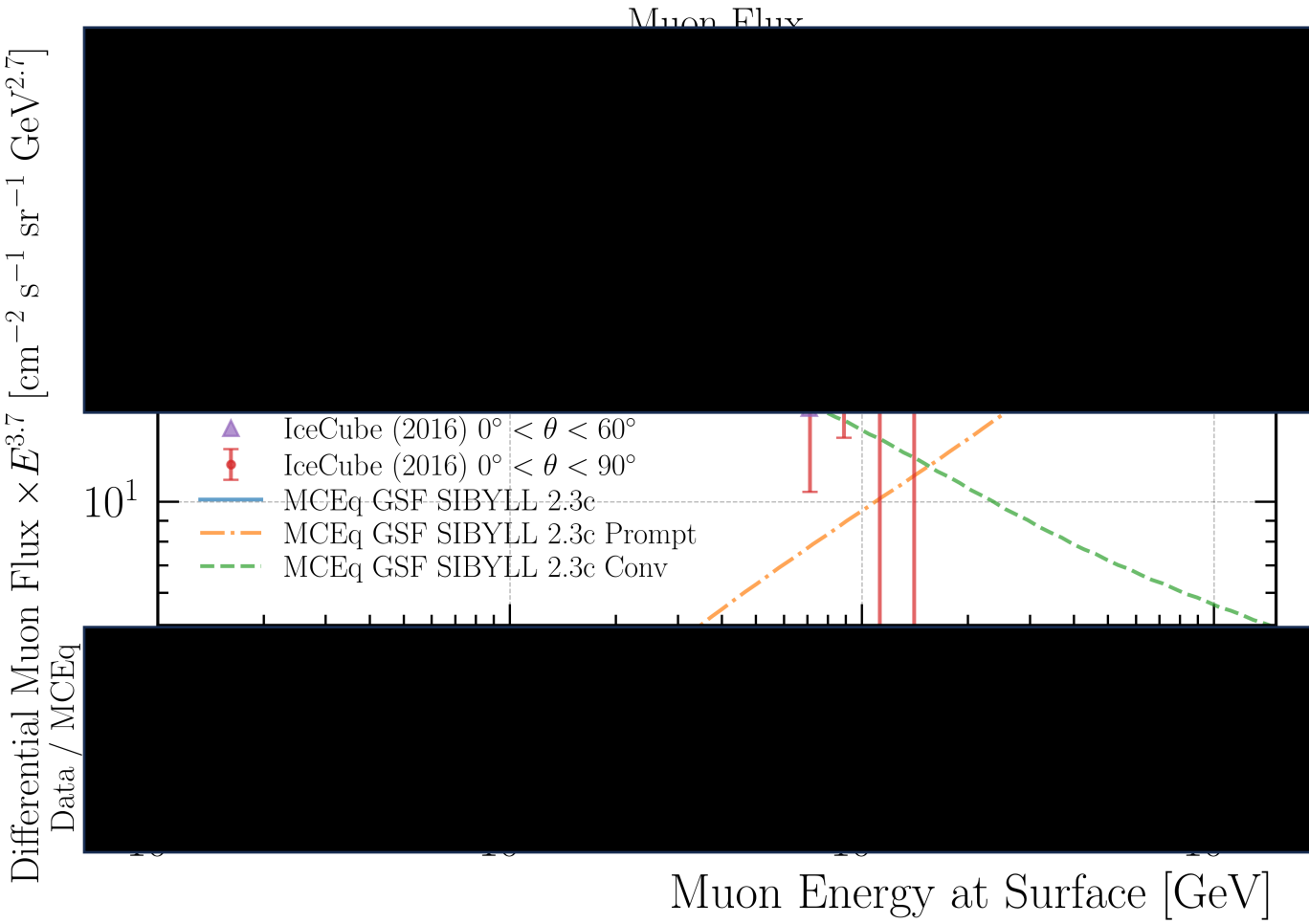
- First p–O cross section (ATLAS); FASER: π/K larger than expected

Atmospheric muons with IceCube

- Spectrum ~600 GeV to ~10 PeV (leading muons: 12 years of data)
- Best agreement: GSF + SIBYLL 2.3c; prompt needed above ~1 PeV

Outlook

- Compare with post-LHC models
- Run 3: K/π (LHCb, FASER), LHCb p–O
- Global model tuning



Thank you for your attention!

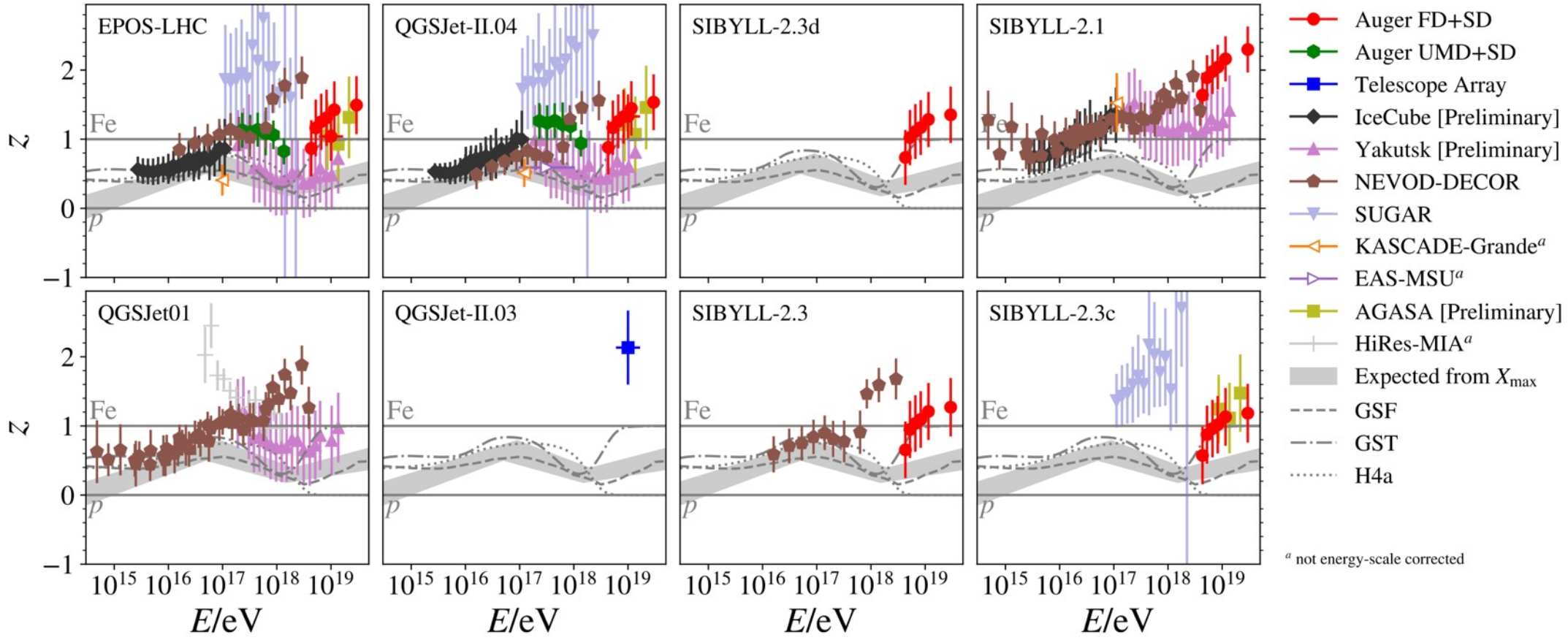
pascal.gutjahr@adelaide.edu.au

Backup

The Muon Puzzle

”muon number”

$$z = \frac{\ln\langle N_\mu\rangle - \ln\langle N_\mu\rangle_p}{\ln\langle N_\mu\rangle_{\mathrm{Fe}} - \ln\langle N_\mu\rangle_p}$$

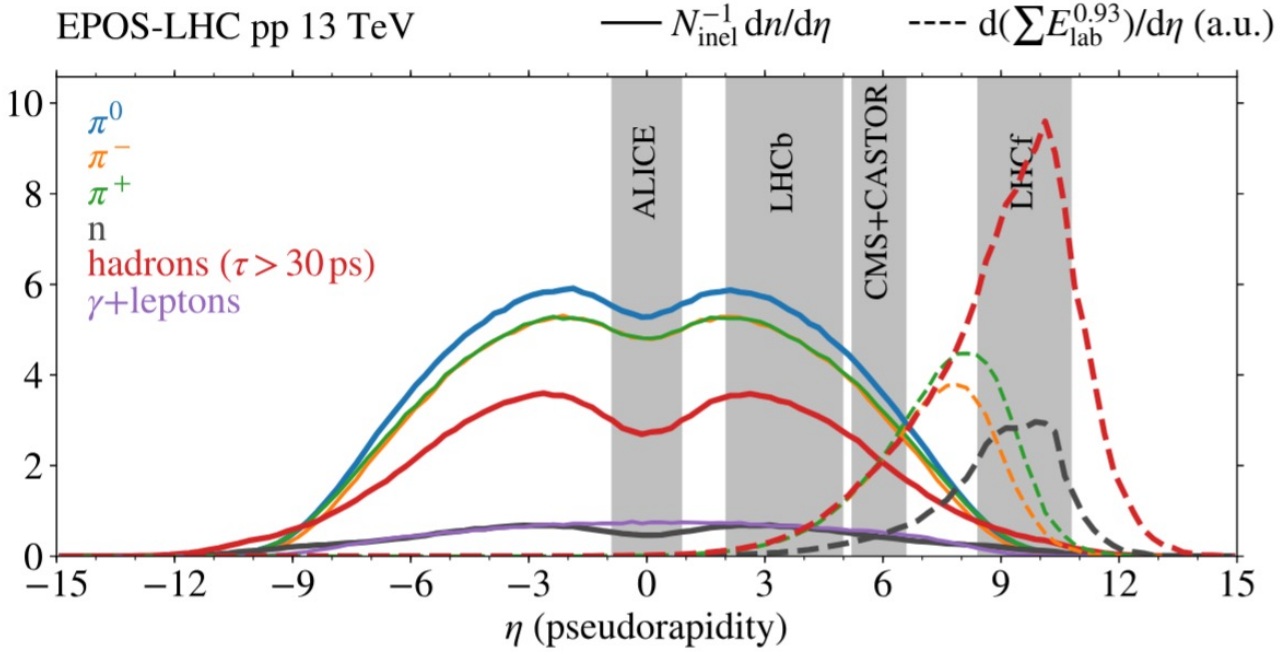


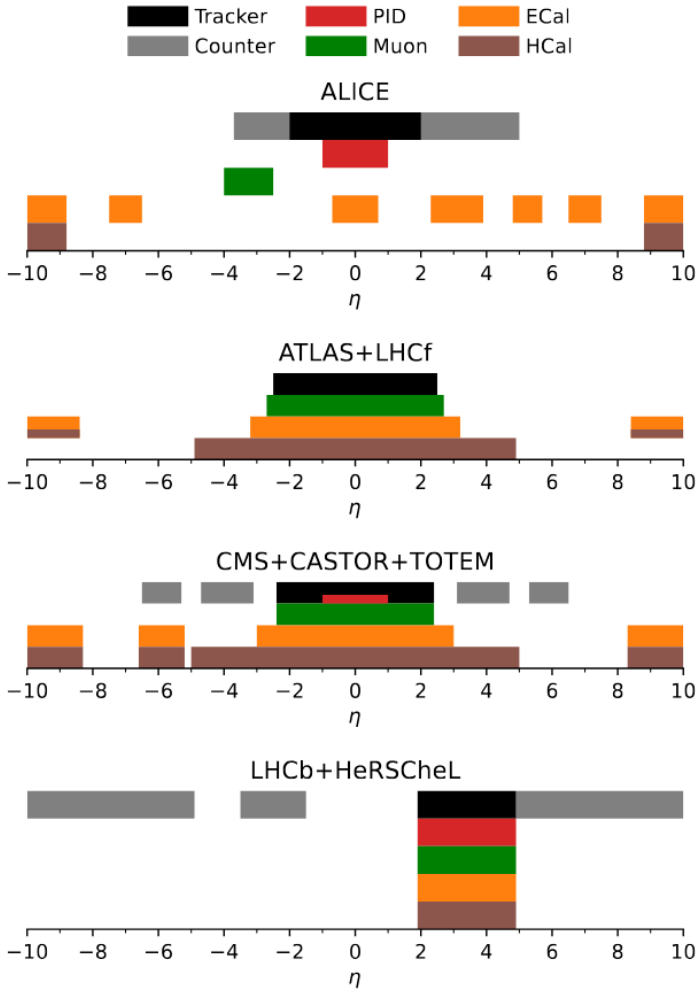
Arxiv: 2108.08341

Adelaide University

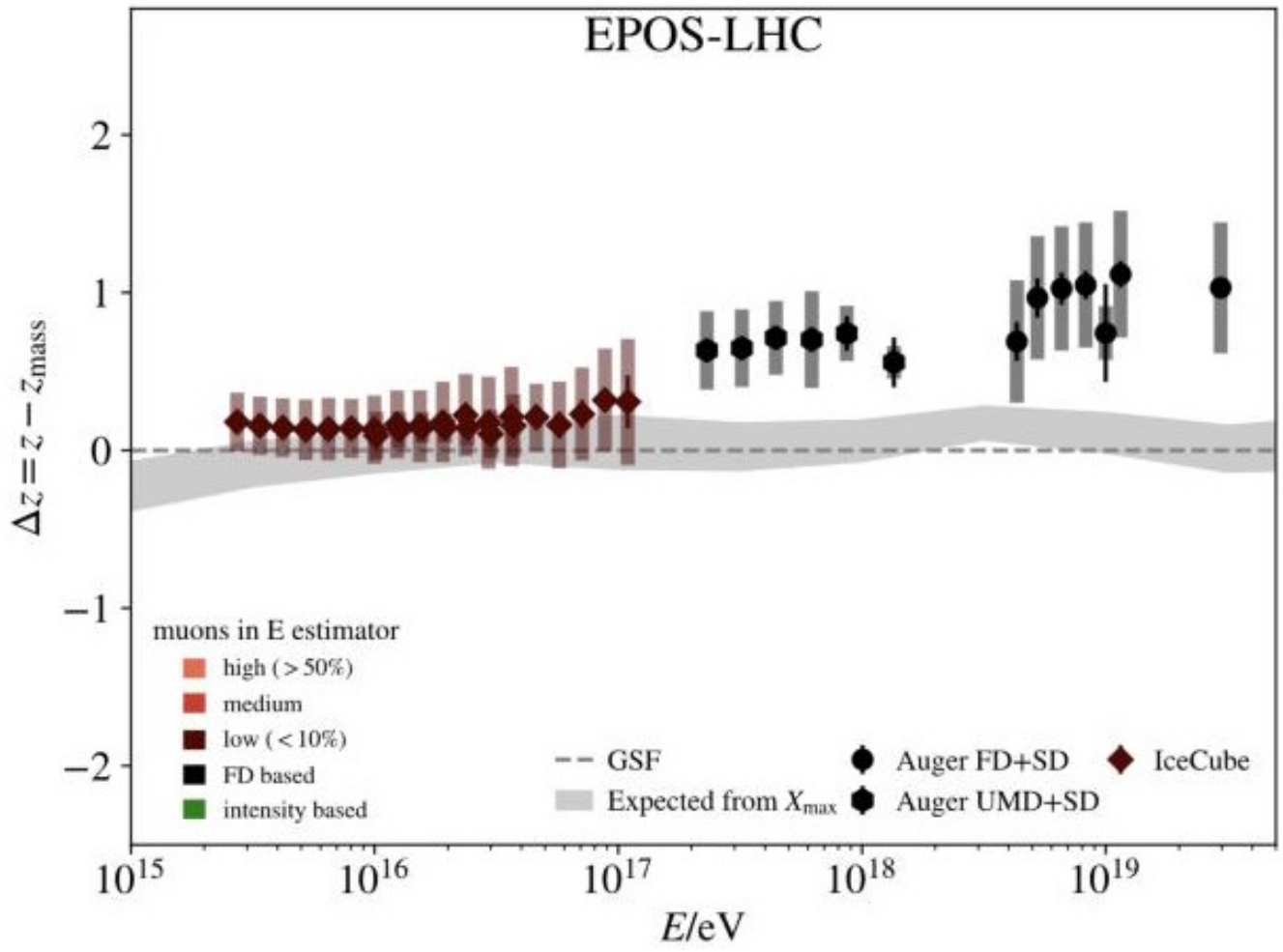
Which Collisions Produce the Muons?

"Muon production weight"
how many muons would be produced in the shower by secondaries of this collision





Muon Content vs Energy: 2026 Compilation



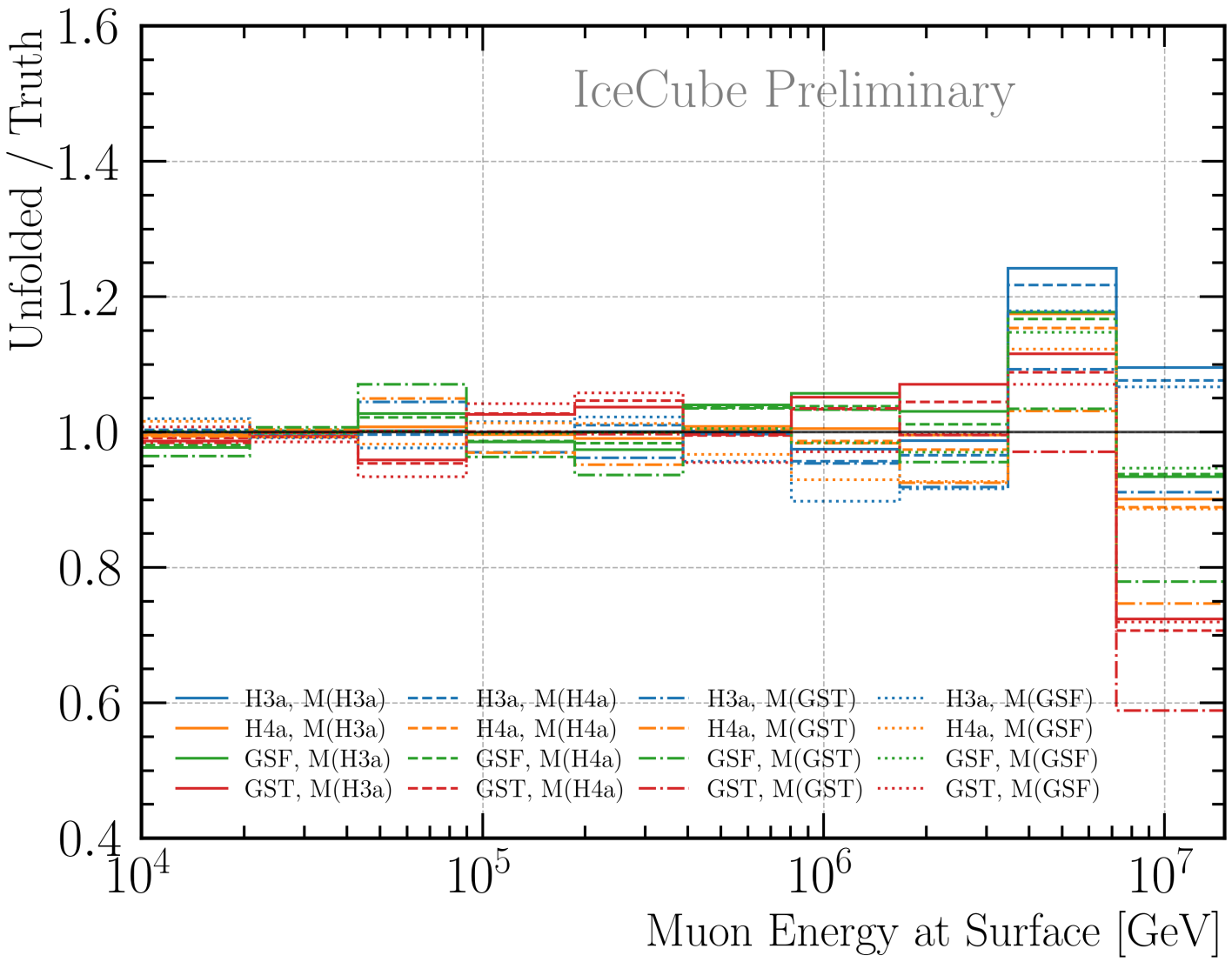
J. Albrecht et al., Nature Rev. Phys. 8 (2026) 98, arXiv:2508.21796 [figure: Fig. 2, EPOS-LHC]

Injection Recovery Test

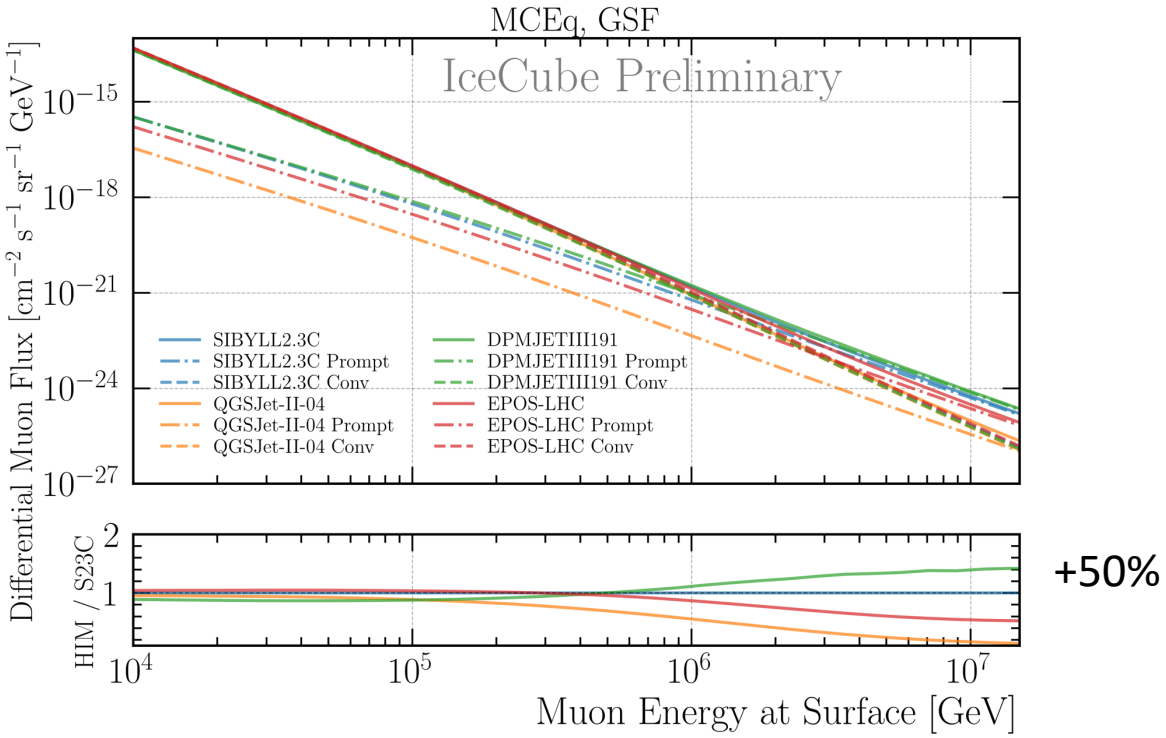
- Train unfolding on H3a, H4a, GSF, GST
- Unfold 4 MC weighted datasets for the 4 primary models
- Divide each unfolding by its true MC distribution
- 16 unfoldings in total

➢ Ratio of 1 expected
➢ Offset to 1 → bias

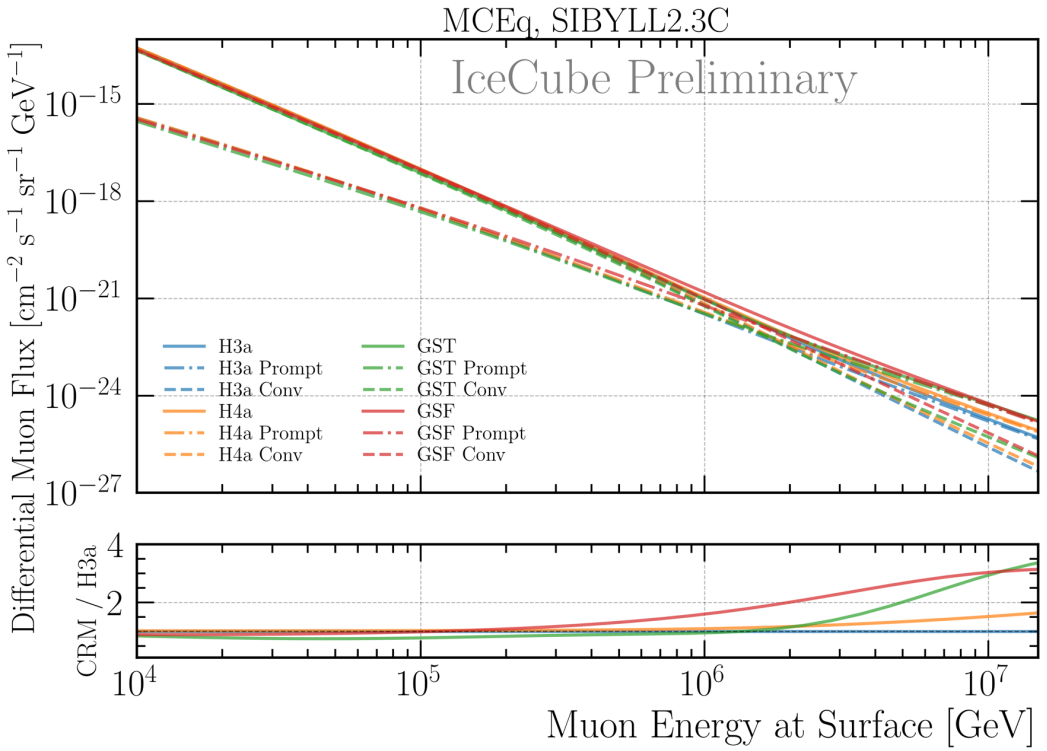
➢ Maximum distance to true distribution
➢ Add these biases as an uncertainty to the unfolding result



Impact of Hadronic Interaction Models?



+50%



+250%

- Only SIBYLL and DPMJET include charm
- Constant offset < 100 TeV → irrelevant for unfolding
- Up to 50% difference at highest energies
- Difference between primary models about 250%
- Unfolding recovers this spectral index change within uncertainties

Unfolding Procedure

- Unfold dataset 4 times, once per primary model (H3a, H4a, GST, GSF)
- Uncertainty per unfolding k:

$$\sigma_k = \sqrt{\sigma_{\mathrm{unf},k}^2 + \left(b_{\mathrm{unf,bias}}\,\hat{\phi}_k\right)^2 + \left(\delta_{A_{\mathrm{eff}}}\,\hat{\phi}_k\right)^2}$$

- Average:

$$\bar{\Phi}_{\mathrm{unf}} = \frac{1}{4}\sum_{k=1}^{4}\hat{\phi}_k$$

- Uncertainty on average:

$$\sigma_{\mathrm{corr}} = \frac{1}{4}\sum_{k=1}^{4}\sigma_k$$

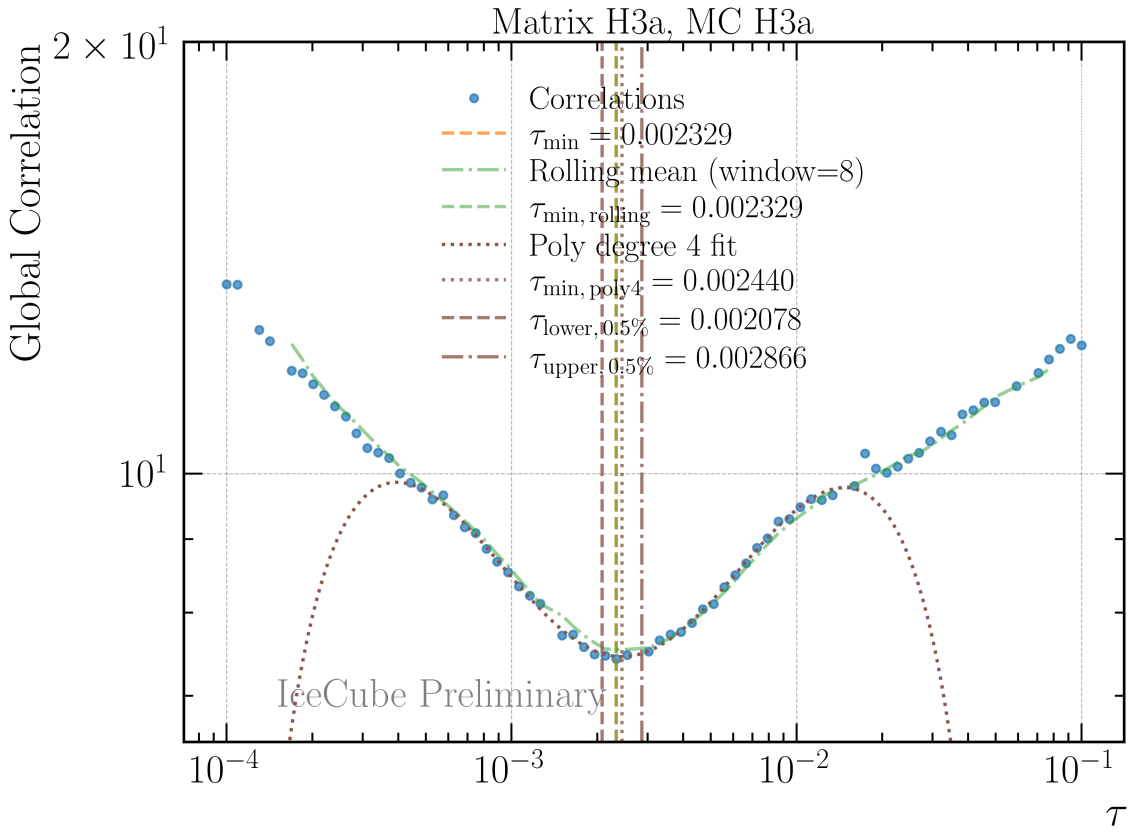
- Deviation between unfoldings:

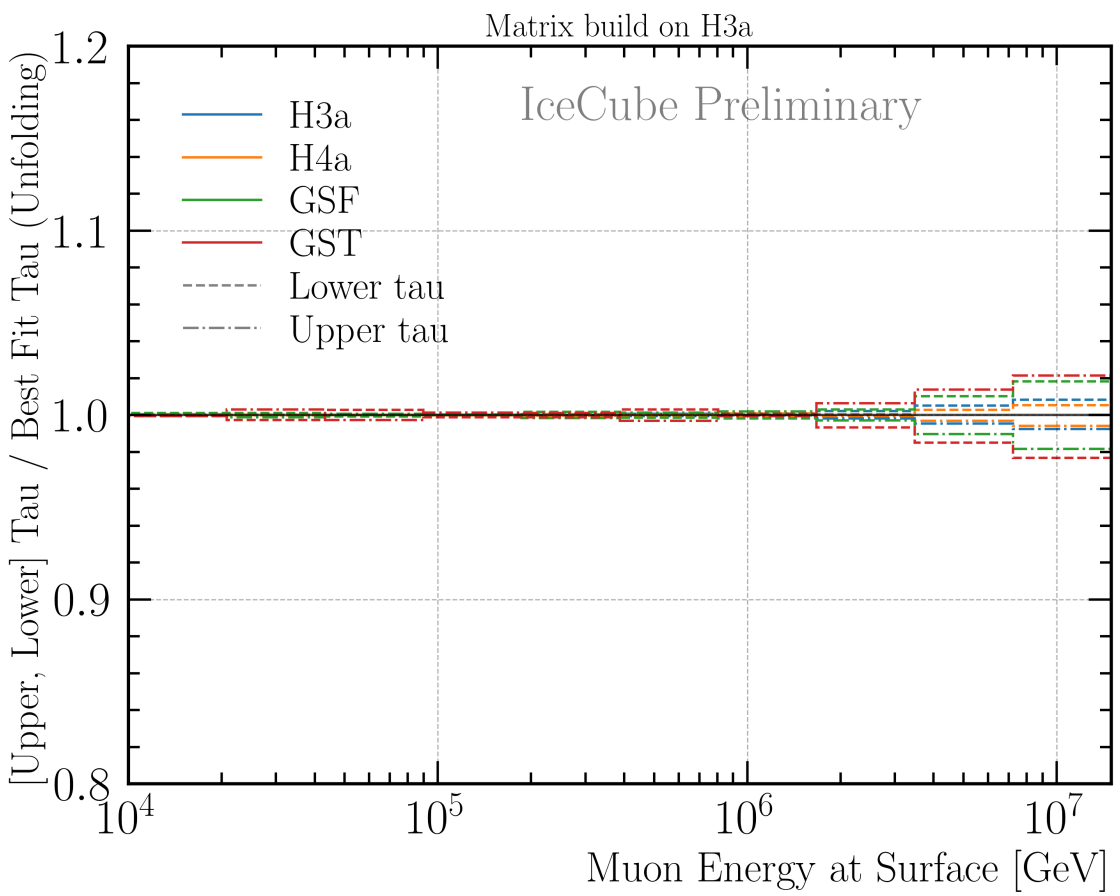
$$\sigma_{\mathrm{spread}} = \sqrt{\frac{1}{4-1}\sum_{k=1}^{4}\left(\hat{\phi}_k - \bar{\Phi}_{\mathrm{unf}}\right)^2}$$

- Total uncertainty:

$$\sigma_{\mathrm{tot}} = \sqrt{\sigma_{\mathrm{corr}}^2 + \sigma_{\mathrm{spread}}^2}$$

Check Impact of Regularization Variations





Coincident Primary Background Estimation

