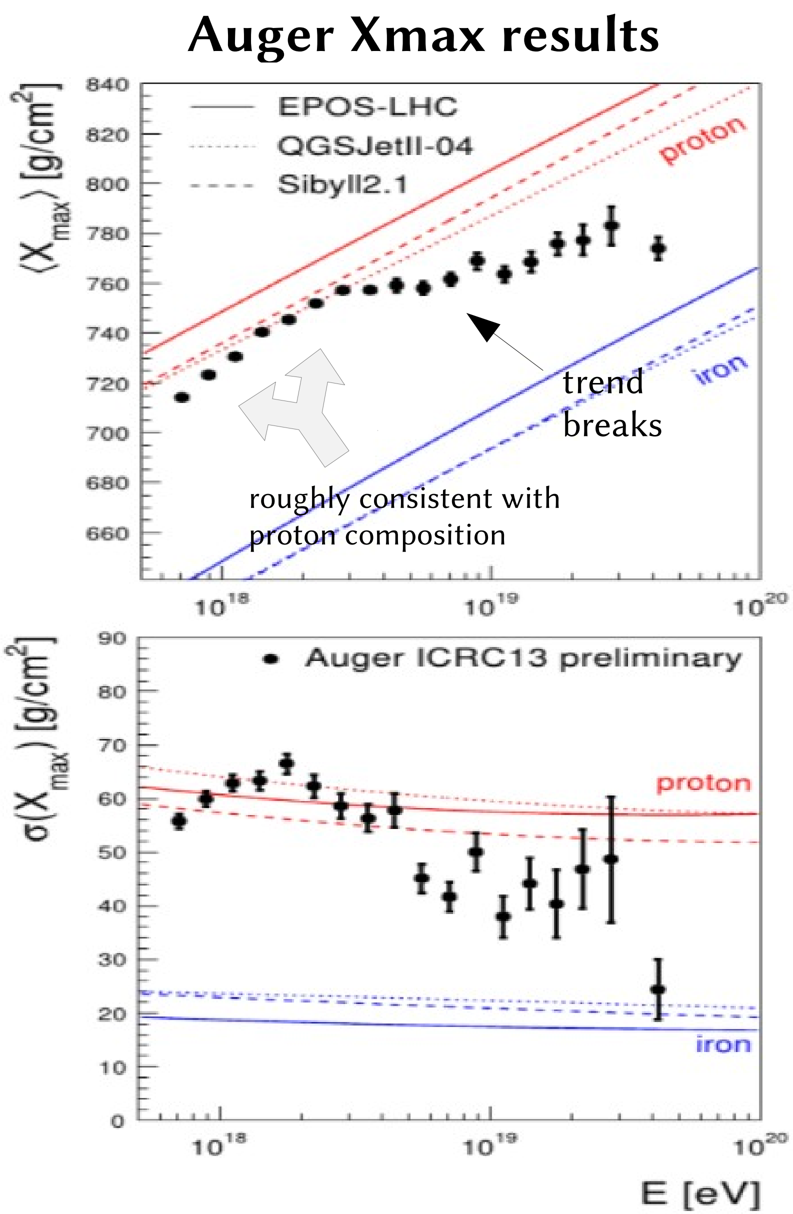
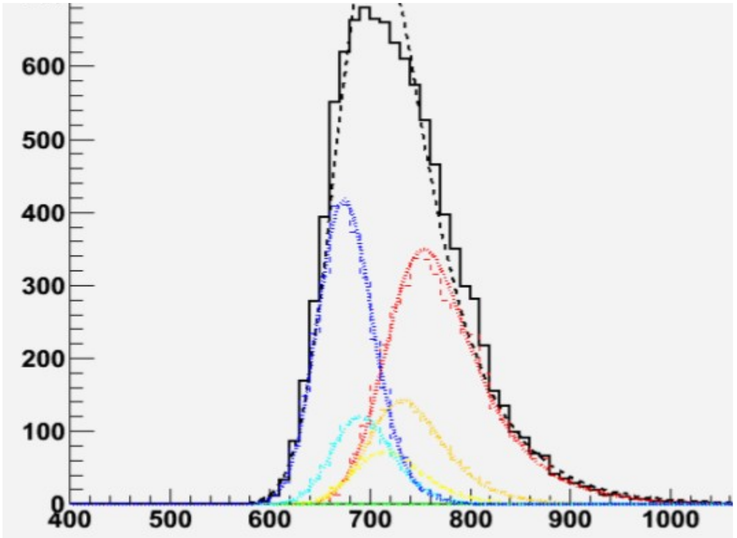
Xmax analysis using skewness

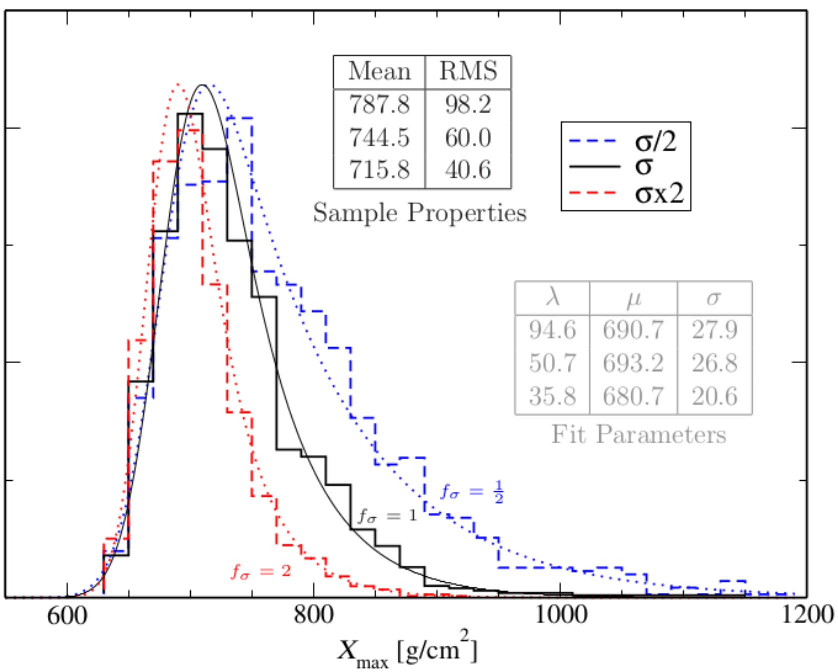
James Stapleton
The Ohio State University
Center for Cosmology and Astro-Particle Physics
October 12th, 2013



- depth of shower maximum is sensitive to UHECR composition
  - Heitler model, nuclear superposition, etc.
- standard approach: compare data/MC Xmax mean and RMS
  - mean Xmax consistent with protons below ~$10^{18.6}$ eV
  - Xmax trend follows log E (as expected for a pure composition)
  - broken trendline usually taken to indicate a change in mass over energy

Xmax data: interpretation





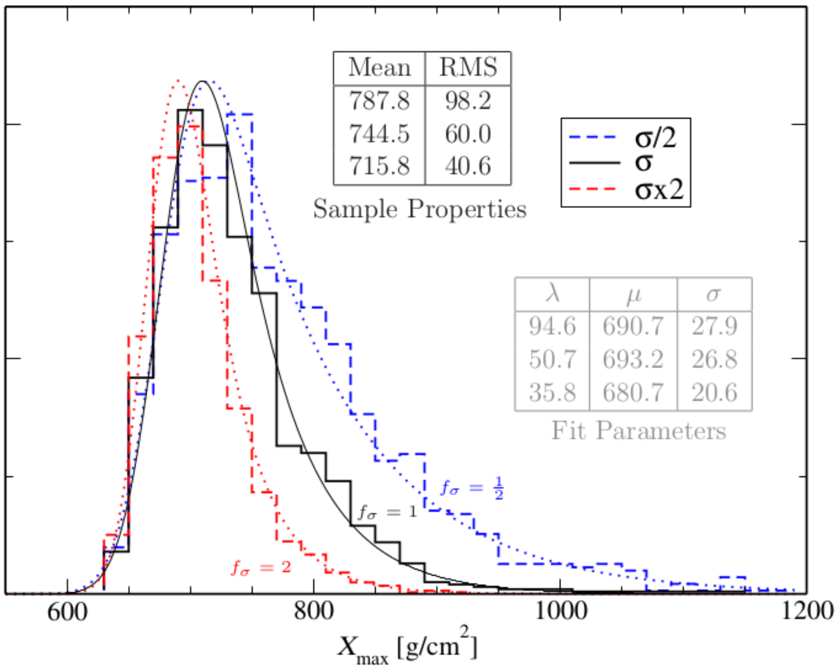
Intrinsic shower-to-shower fluctuations in depth reduce 'mass resolution'

- we only shows **mean** of Xmax distribution
- Xmax distributed about mean with significant variance
- distributions for different masses overlap
- composition mixtures mean <Xmax> is an average over a composite distribution made of multiple, overlapping pure-mass distributions

Hadronic interactions in air shower further complicate the picture

- MC simulations use hadronic phenomenology instead of perturbative QCD
- BUT phenomenology is extrapolated to energies and momenta at which they cannot be directly tested
- result: large systematic uncertainties on simulated Xmax

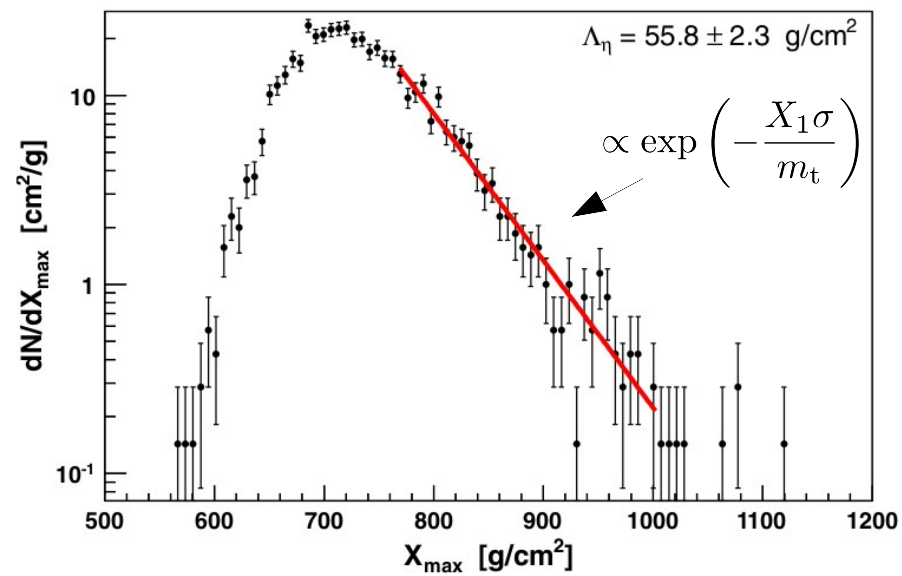
Xmax asymmetry and early interactions



Distribution tail

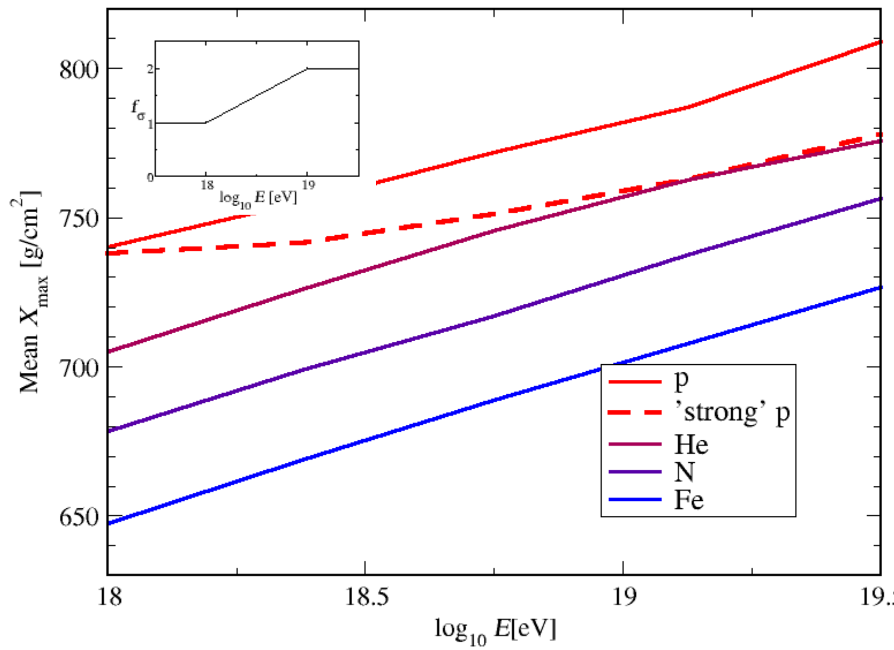
- intrinsic asymmetry in Xmax fluctuations
- due to intrinsic asymmetry in depths of first interaction X_1
- Sensitive to cross-section in X_1!
- Proton-air cross-section measurement
  - use proton-like Xmax data below break energy
  - Auger p-air cross-section measurement

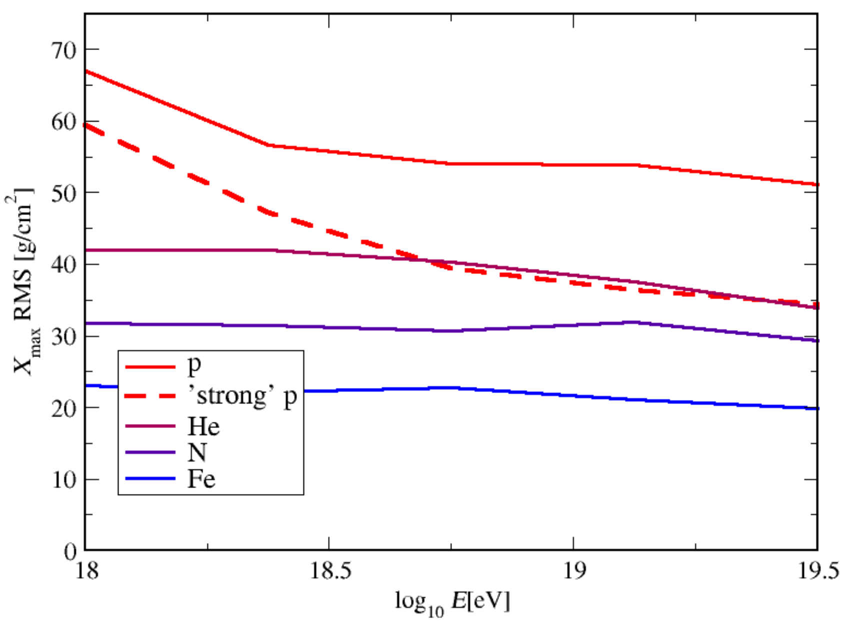
- Note: Xmax **mean** and **width** also depend strongly on early cross-section
- BUT these quantities are used to estimate primary UHECR mass
- **Mass estimates can easily be confounded by cross-section systematics**



Example: artificially scale cross-sections

- CONEX simulation using QGSJet model
- QGSJet predicts total cross-sections for (p, pi, kaon incident on nucleus)
- apply scaling factor f_σ to cross-section
- f_σ changes slowly from 1 to 2 between 10^{18} eV and 10^{19} eV





Could be interpreted as a change in mass!

Why is asymmetry so sensitive to X1?

What can we do about it?

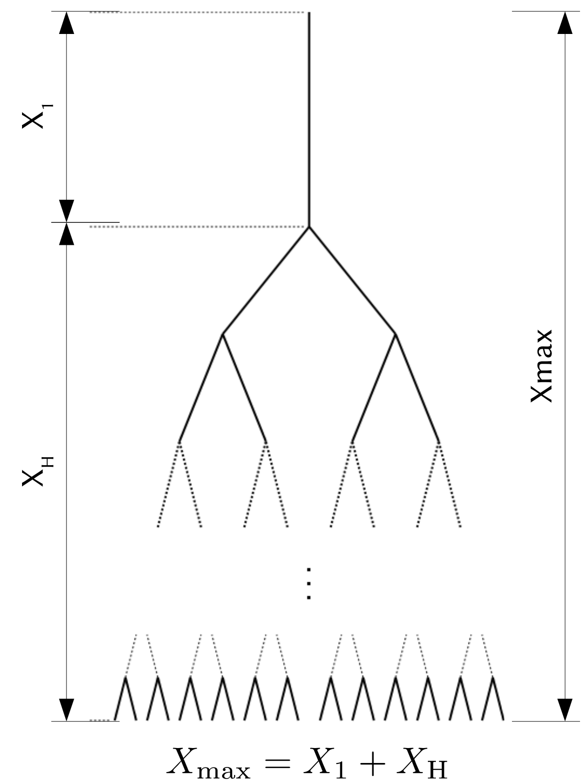
Separate sources of shower-to-shower fluctuations in depth

- all cascade generations have intrinsic fluctuations in depth
- *earlier* fluctuations have greater influence than later fluctuations

Early interactions

- X_1: exponentially distributed ~ exp($-X_1$/lambda)
  - *when measured in COLUMN DENSITY* [g/cm^2]
- lambda inversely proportional to UHECR-air cross-section

Easiest extension of cascade picture: treat X_1 separately from remaining distance to Xmax



Three-parameter distribution of Xmax

Combine distributions of X_1 and X_H

- we know the distribution of X_1
- we know the distribution of X_H is 'significantly wide'
  - use a normal distribution:
    - ***N*** (Xmax; mean=eta, std.dev.=tau)
- convolution gives the distribution of Xmax = X_1 + X_H
- $f_3(X_{\mathrm{max}}; \lambda, \eta, \tau) = \int dX_1' \frac{1}{\lambda} e^{-\frac{X_1'}{\lambda}} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2}\left(\frac{X_{\mathrm{max}} - X_1' - \eta}{\tau}\right)^2}$

f_3 shape parameters

- lambda: mean X_1 (interaction length of UHECR in atmosphere)
- eta: mean X_H
- tau: standard deviation of X_H

Benefits

- f_3 provides a good fit to simulated Xmax and real data
- provides parametric treatment of uncertain cross-sections
- tau absorbs (symmetric) Xmax error systematics
- statistical moments of f_3 can be expressed as functions of shape parameters!

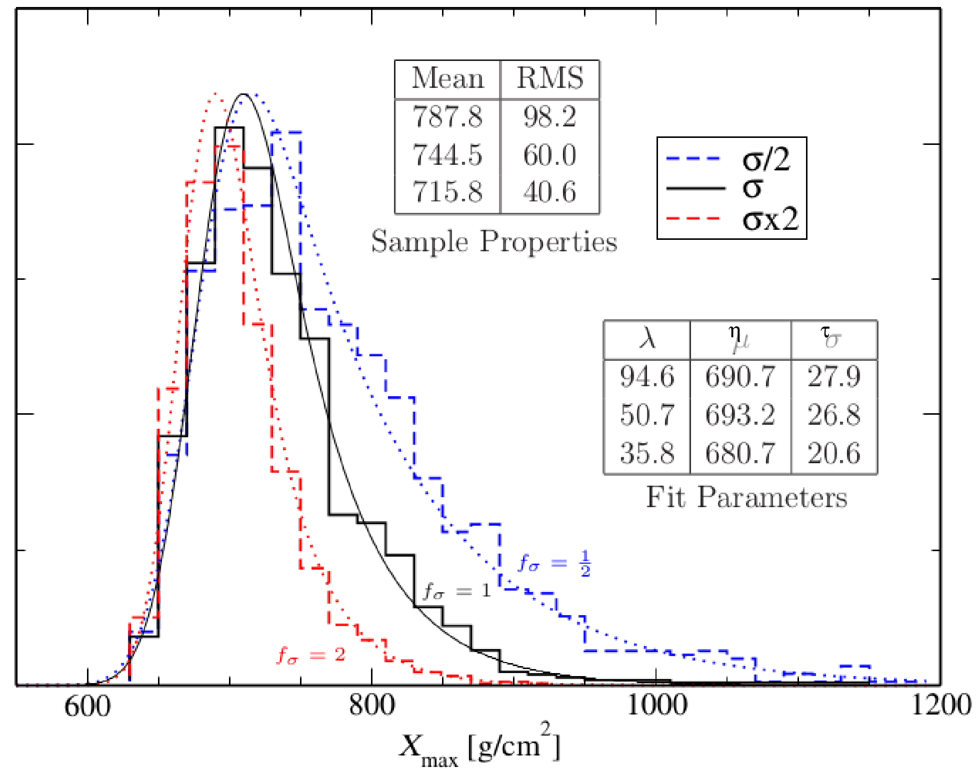
We can separately parameterize cascade development at highest energies (most uncertainty) and lower-energies (better understood)

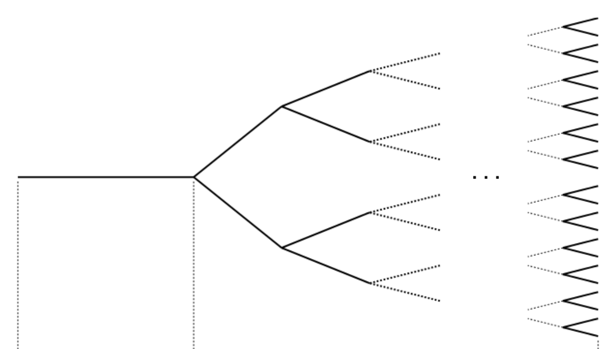
Earlier cascade interactions

- fewer branches/particles
- depth fluctuations have greater overall effect on depth of Xmax
- occur with the highest energies
- are the most vulnerable to systematics in hadronic phenomenology

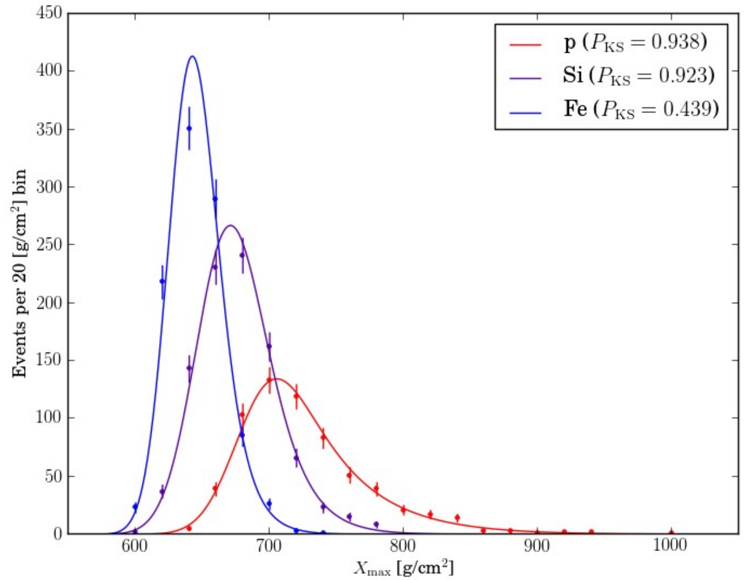
Later cascade interactions

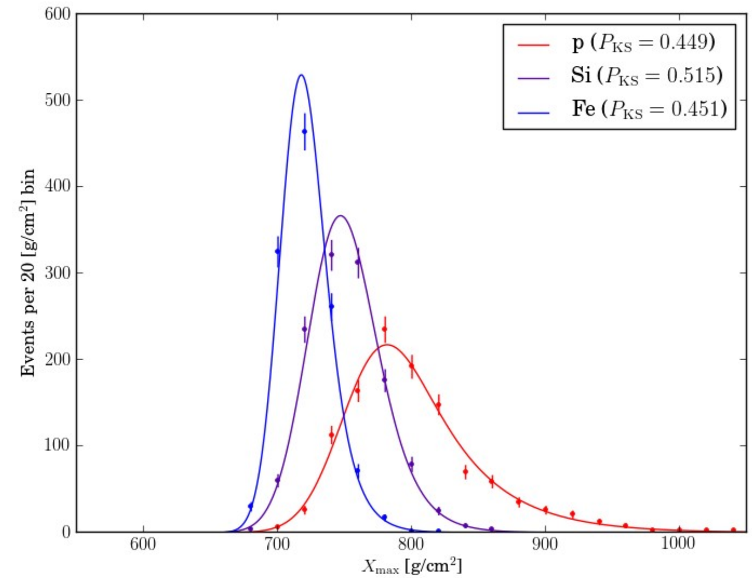
- involve MANY branches/particles
- are *most* cascade interactions
- more valid use of Heitler model
- better phenomenological predictions
- asymmetric distance fluctuations are 'averaged out' more efficiently



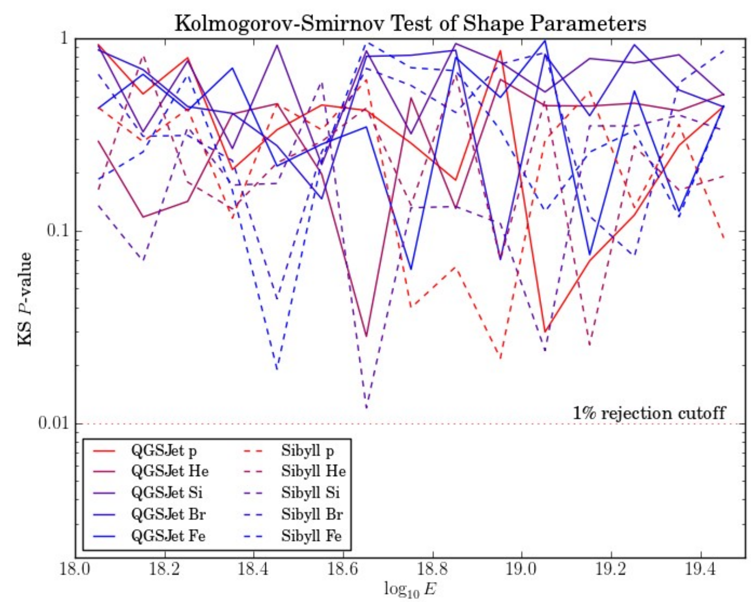


Reproducing the Xmax distribution





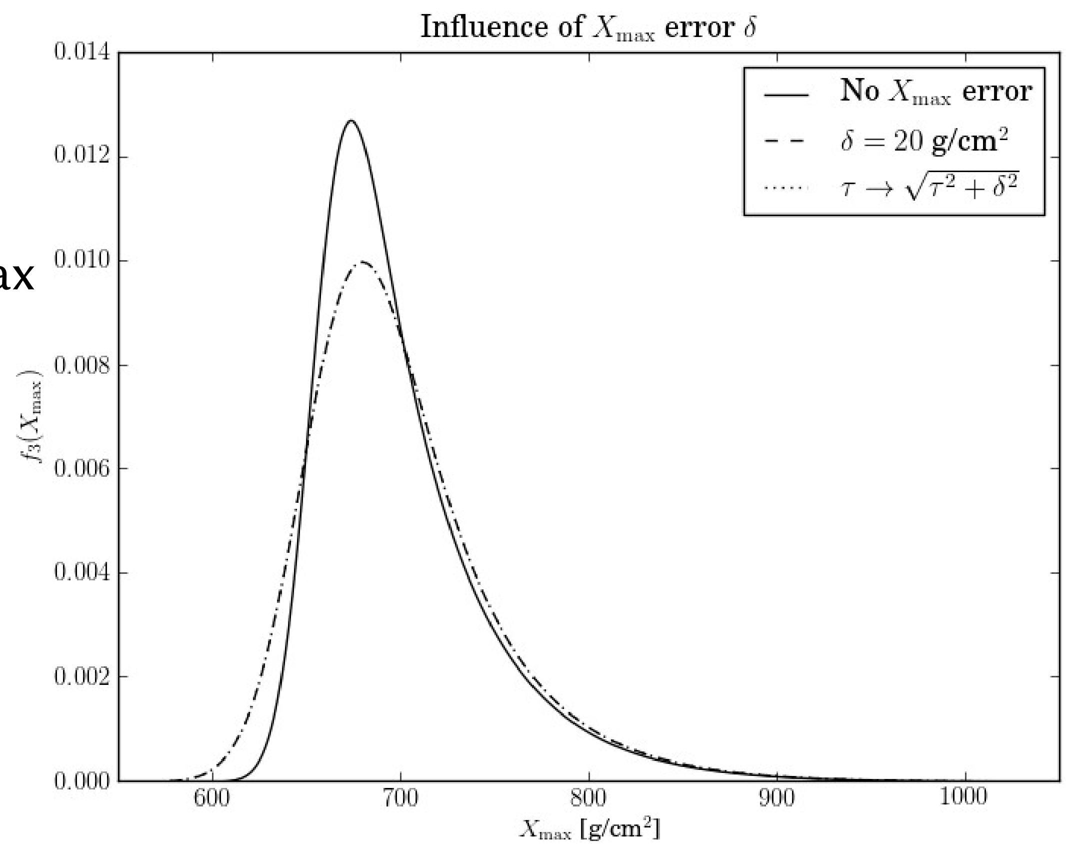
- f_3 describes Xmax well for:
  - energies $10^{18} - 10^{19.5}$
  - $A = (1, 4, 14, 35, 56)$
- f_3 previously discussed for this reason
  - GAP 2009-078, 2010-105, 2010-108, 2011-041, 2011-064, 2012-030, ...?



Xmax measurement error

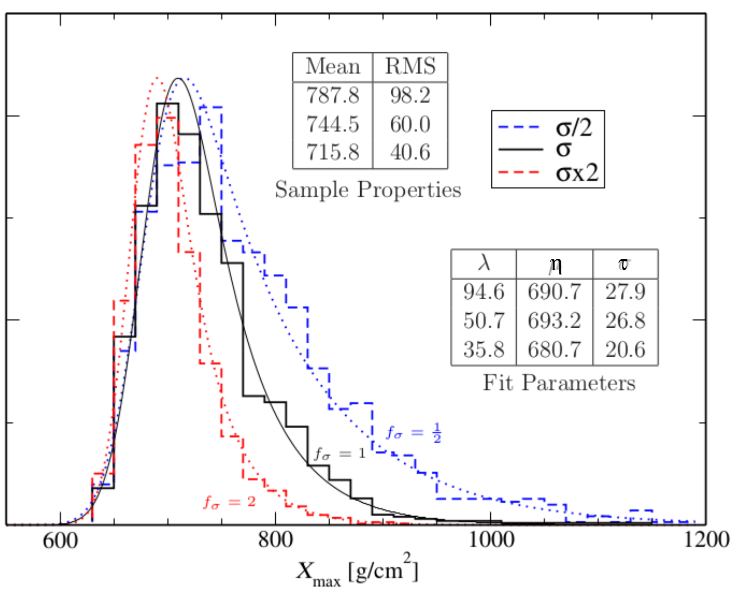
- another neat property: tau absorbs Xmax measurement error
  - (for Gaussian models of Xmax error)
- convolution already contains one Gaussian distribution
  - Gaussians combine
- no integral needed to 'smear' Xmax distribution for error

$$f_3(X_{\mathrm{max}};\lambda,\eta,\tau)\otimes G(\Delta X_{\mathrm{max}};0,\delta) = f_3(X_{\mathrm{max}};\lambda,\eta,\sqrt{\tau^2+\delta^2})$$



Interpretation of shape parameters

- statistical moments as simple functions of shape parameters
  - invert relationships
  - calculate shape parameters as simple functions of statistical moments
  - forms make their physical meaning clear
- lambda is really just a measure of skewness
- eta: 'X_1-corrected' measure of mean Xmax
- tau: 'X_1 corrected' measure of variance
- these 'corrections' make the mean and variance more resistant to cross-section systematics



$$f_3(X_{\max};\lambda,\eta,\tau) = \int dX_1{}' \, \frac{1}{\lambda} e^{-\frac{X_1{}'}{\lambda}} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2}\left(\frac{X_{\max}-X_1{}'-\eta}{\tau}\right)^2}$$

$$\langle X_{\max}\rangle = \lambda + \eta \qquad \lambda = \left(\frac{\theta}{2}\right)^{\frac{1}{3}}$$

$$\sigma^2 = \lambda^2 + \tau^2 \quad \Longleftrightarrow \quad \eta = \langle X_{\max}\rangle - \lambda$$

$$\theta = 2\lambda^3 \qquad \tau^2 = \sigma^2 - \lambda^2$$

mean, standard deviation, and 'unnormalized' skewness (defined as expectation values)

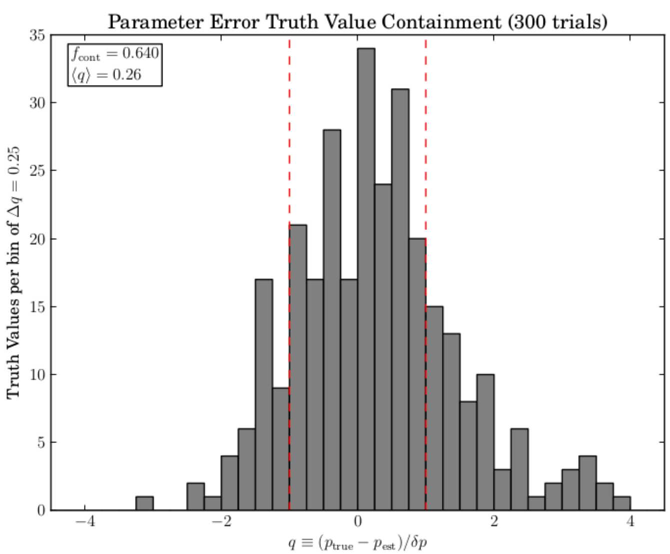
$$\bar{x} \equiv \langle x \rangle$$

$$\sigma \equiv \left\langle (x-\bar{x})^2 \right\rangle$$

$$\theta \equiv \left\langle (x-\bar{x})^3 \right\rangle$$

Parameter <--> statistic relationship greatly facilitates data analysis

- mean, variance, and skewness have statistical estimators
  - e.g. unbiased estimator of population variance: $\sigma_e^2 = \frac{\sum x_i^2 - \left(\sum x_i\right)^2/n}{n-1}$
- parameters get statistical estimators!
  - no messy curve fitting
  - minimal bias
  - estimators can weight data to account for non-uniform exposure
  - get error estimates via resampling methods (jackknife/bootstrap)



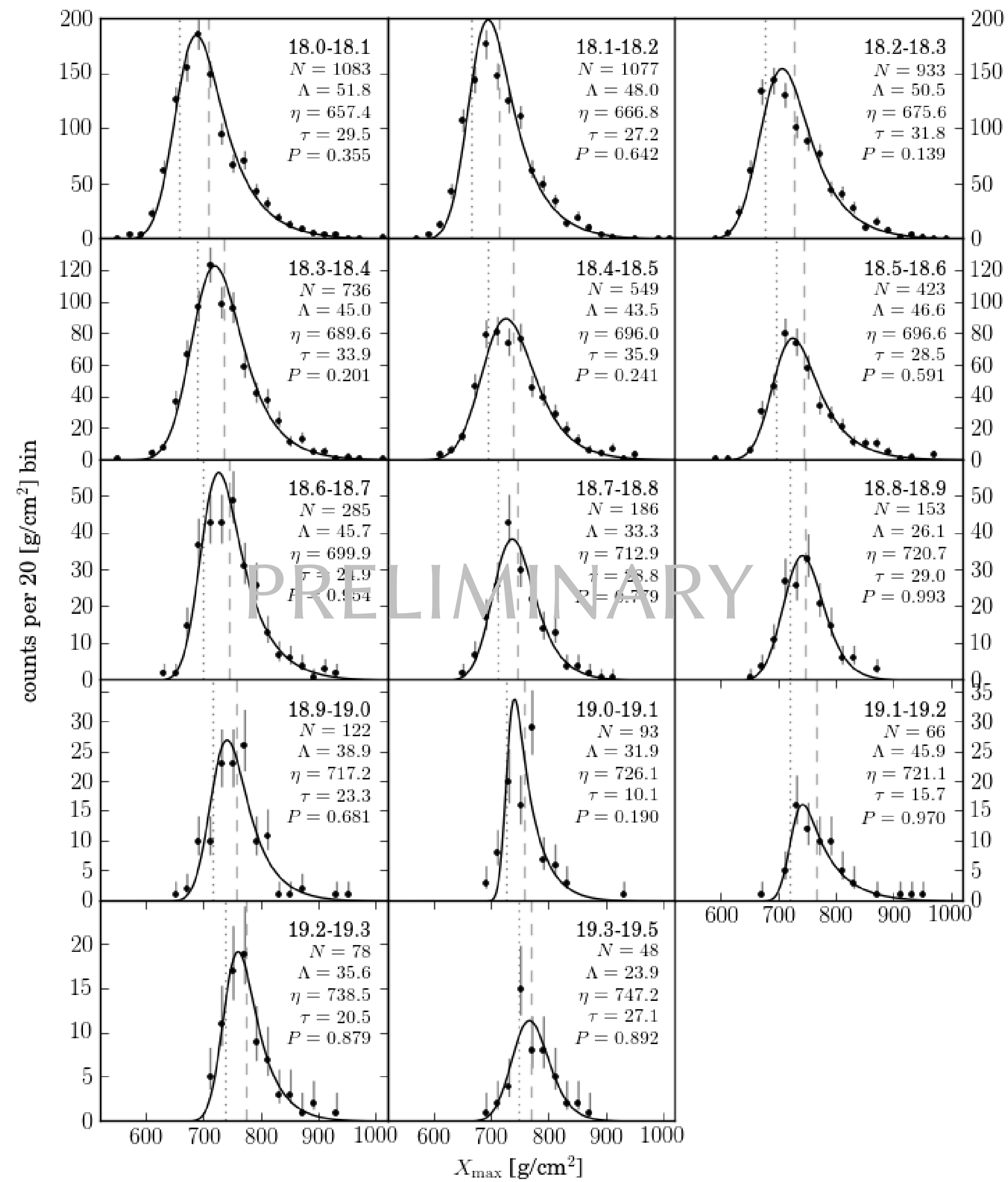
lambda

Example: test estimators for bias

- choose 'truth' parameters
  - (lambda, eta, tau) = (45, 650, 25) g/cm2
- loop over trials:
  - sample from truth distribution to Monte Carlo a 'fake' data set
  - estimate (lambda, eta, tau) from fake data set
  - estimate errors on (lambda, eta, tau)
  - compare estimated value/error bar to truth value: $q = \frac{p_{\rm truth} - p_{\rm est}}{\delta p_{\rm est}}$

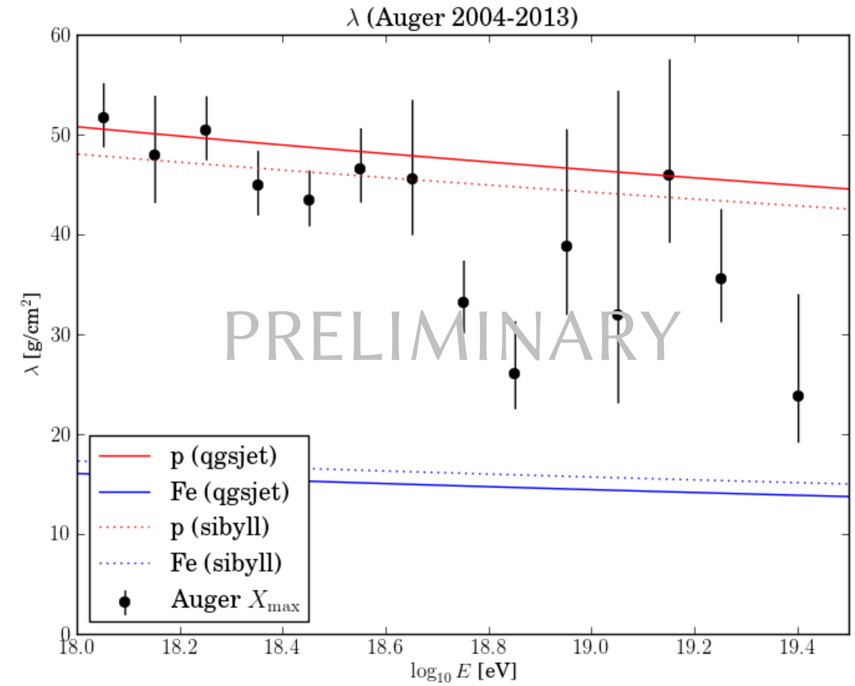
Xmax data analysis

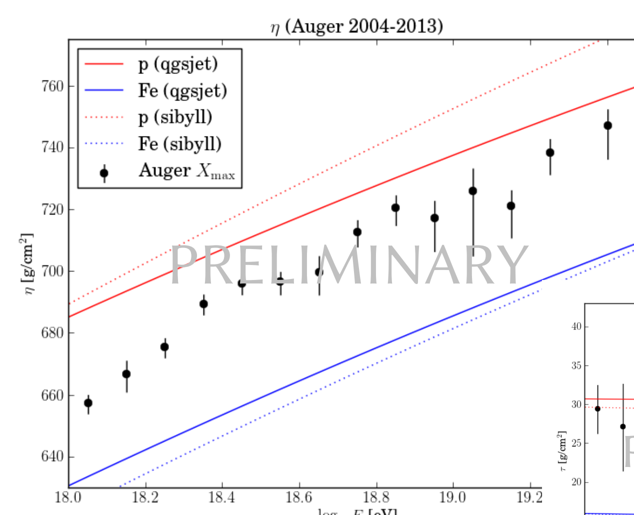
- Auger Xmax data
  - Observer v9r1
  - 2004 – Jan 2013
  - data selection/anti-bias cuts follow 2010 Xmax PRL
  - no fitting, so no goodness-of-fit measure
  - Kolmogorov-Smirnov test
    - *P*-values indicate good fits to data at all energies

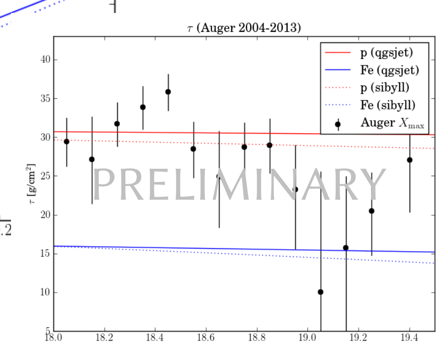


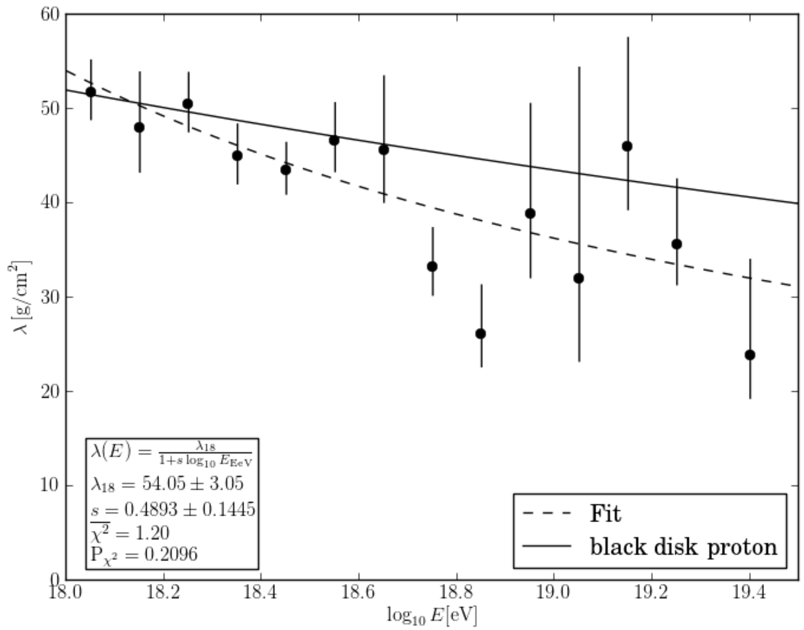
Xmax data analysis

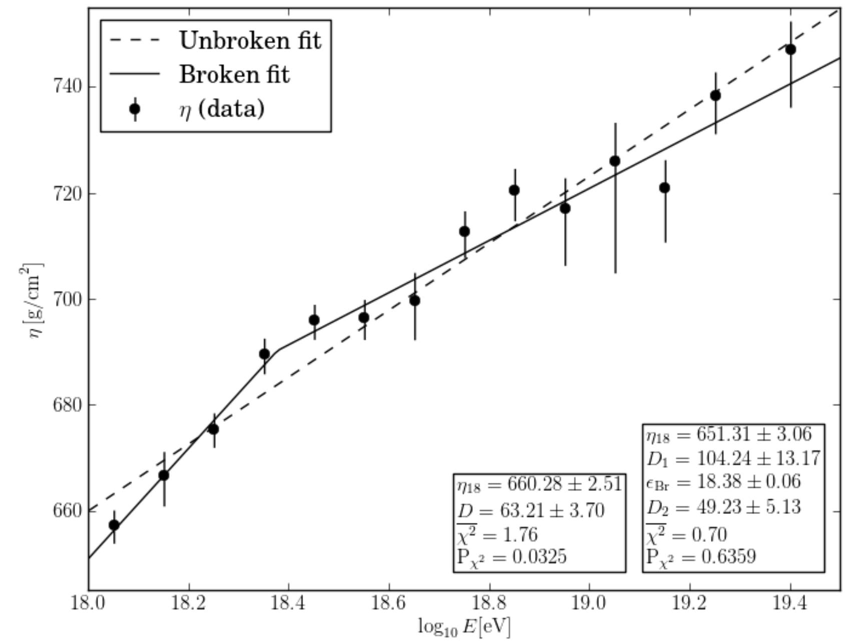
- lambda: consistent with proton-air interaction length below 18.7
  - above 18.7: break from trend?
- eta: slope appears to break between 18.4 and 18.7
  - also somewhat consistent with an unbroken linear trend
- tau yields no clear information
  - absorbs Xmax error systematics
- more data would help
- Xmax efficiency/acceptance study needed
  - current Xmax anti-bias data cuts attempt to unbias ***mean Xmax only***











- lower-energy lambda is consistent with simple cross-section model
  - Block-Halzen 'black disk' proton (2012)
- also consistent with a single trend

- eta gives us an elongation rate
- broken trend in $\log E$ is a better fit

Xmax data analysis

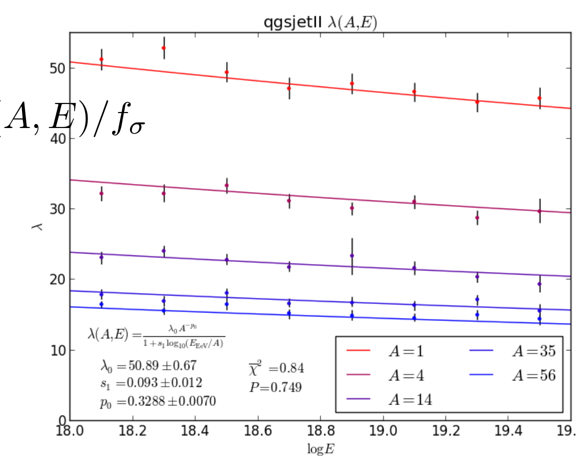
- what should 'standard Xmax analysis' really be?
- current analyses focus on precision measurement of Xmax mean, variance
- we are effectively promoting the use of:
  - 'X_1 corrected' mean
  - 'X_1 corrected' variance
  - skewness

- continued collection of longitudinal profile data
  - Auger recently/currently releasing data with more statistics, better control of systematics
  - future projects (like JEM-EUSO) will provide more longitudinal profile data
- with additional statistics, analyses with higher moments will become viable
- **we should at least add skewness to standard Xmax data analysis**

- Next: facilitate adoption
  - Function has been used by others within collaboration
  - Numerical implementation can be difficult
- Various numerical tools/software
  - Code for fast/accurate evaluation
    - $f_3(x;\lambda \rightarrow 0,\eta,\tau) \rightarrow G(x;\eta,\tau)$
    - Delta function in integral
  - Fitting binned X_{max}
    - Least-square fits have systematic problems with tail/width
    - Log-likelihood fits yield results which closely match calculated parameters
  - Fast random sampling of f_3
  - Parameter error estimation via resampling methods
  - f_3 integral/CDF for easy application of the Kolmogorov-Smirnov test

- Provide Monte-Carlo trained composite distributions for mass mixture scenarios
  - Fit f_3 to Monte Carlo predictions
  - Scan over UHECR primaries with different mass, energy
  - Insert cross-section scaling parameter: $\lambda \mapsto \lambda/f_\sigma$
- Provide parameters as a function of $\ln(A)$, $\log_{10}(E)$

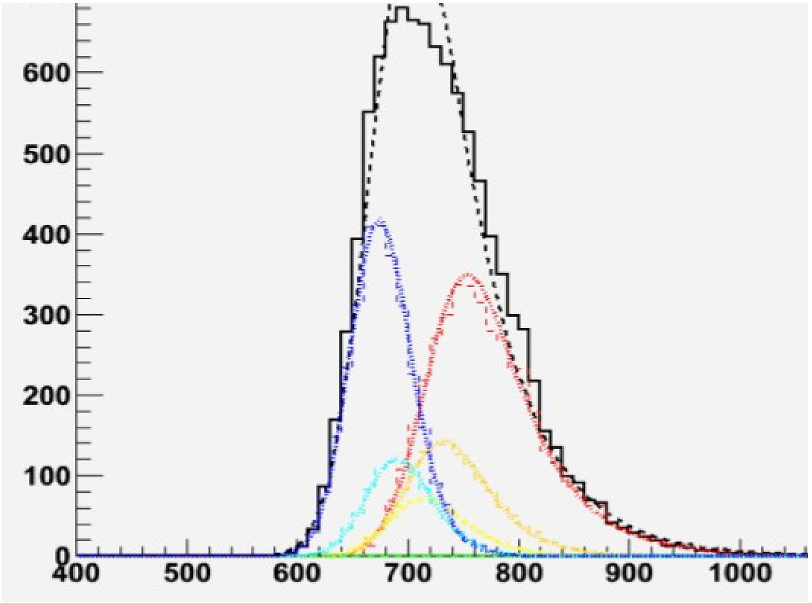
$$\lambda_s(f_\sigma, A, E) = \lambda(A, E)/f_\sigma$$
$$\eta(A, E)$$
$$\tau(A, E)$$



- Provide $f_3(X_{max};A,E,f_\sigma)$
  - (And facilities to re-train parameters using your favorite Monte Carlo simulations)

Mixed-mass formulation

- Most immediate problem:
  - f_3 shape parameter analysis outlined so far is **ONLY VALID WHEN APPLIED TO PURE-COMPOSITION XMAX DISTRIBUTIONS**
  - shape parameters of composite distribution **LOSE PHYSICAL MEANING**



- build composite distribution from superposition of underlying (pure-mass) distributions:

$$f_{\mathrm{Tot}}(X_{\mathrm{max}}) = \sum_A c_A\, f_3(X_{\mathrm{max}}; \lambda_A, \eta_A, \tau_A)$$

$$= \int d\alpha\ P(\alpha)\, f_3\left(X_{\mathrm{max}}; \lambda(\alpha), \eta(\alpha), \tau(\alpha)\right) \qquad \alpha \equiv \ln(A)$$

- Utilize MC-trained parameters for different masses
  - lambda ≈ polynomial in logA, logE
  - eta = polynomial in logA, logE
  - tau = polynomial in logA, logE
- Keep cross-section scaling factor f_σ

Next, compute total moments from Descriptive Parameters and mass distribution P(α)

Statistical moments of **composite** distributions	⇔	Statistical moments of **underlying** distributions

$$\langle X\rangle_{\mathrm{T}} = \int d\alpha\, P(\alpha)\langle X\rangle_\alpha$$

$$\sigma^2_{X\,\mathrm{T}} = \int d\alpha\, P(\alpha)\left(\sigma^2_{X\,\alpha} + (\langle X\rangle_\alpha - \langle X\rangle_{\mathrm{T}})^2\right)$$

$$\theta_{X\,\mathrm{T}} = \int d\alpha\, P(\alpha)\left(\theta_{X\,\alpha} + 3\,(\langle X\rangle_\alpha - \langle X\rangle_{\mathrm{T}})\,\sigma^2_{X\,\alpha} + (\langle X\rangle_\alpha - \langle X\rangle_{\mathrm{T}})^3\right)$$

Result: if we knew the mass distribution P(α), we could calculate the **total** Xmax moments (which we can already observe, of course…)

$$\langle X\rangle_{\mathrm{T}} = \int d\alpha\, P(\alpha)[\lambda(\alpha) + \eta(\alpha)]$$

$$\sigma^2_{X\,\mathrm{T}} = \int d\alpha\, P(\alpha)[\lambda^2(\alpha) + \eta^2(\alpha)]$$

$$\theta_{X\,\mathrm{T}} = \int d\alpha\, P(\alpha)[2\lambda^3(\alpha) + 3\,(\lambda(\alpha) + \eta(\alpha) - \langle X\rangle_{\mathrm{T}})\,\sigma^2_{X\,\mathrm{T}} + (\lambda(\alpha) + \eta(\alpha) - \langle X\rangle_{\mathrm{T}})^3]$$

- Moments of superposed Xmax distribution can be calculated from moments of underlying Xmax distribution
- Moments of underlying distributions can be written as polynomial functions of $\alpha = \ln(A)$
- under $P(\alpha)d\alpha$ integral, α polynomial is converted to linear combinations of $P(\alpha)$ distribution moments!
- Example: mean of composite Xmax distribution:

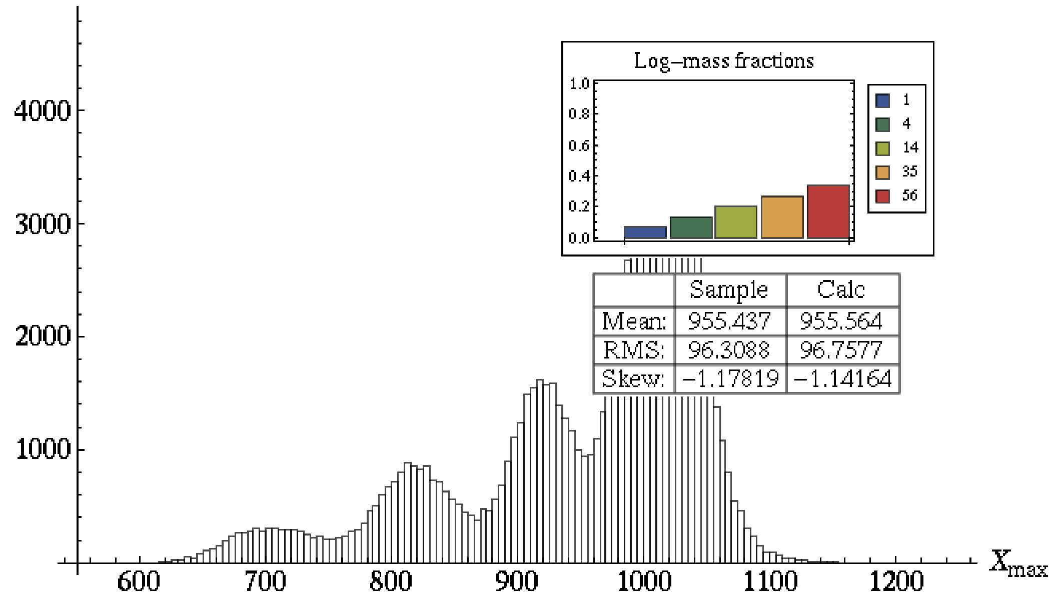
$$\bar{X}_{\mathrm{T}} = \lambda_0 \left(1 - p_0\langle\alpha\rangle + \frac{p_0^2}{2}\right)\left(1 - \frac{p_0^2}{2}\sigma_\alpha^2\right)$$
$$+(D_{01} + D_{11}\bar{\epsilon})\langle\alpha\rangle + D_{02}\langle\alpha\rangle^2 + D_{02}\sigma_\alpha^2 + D_{k0}\bar{\epsilon}^k$$

- Linear transformation!

$$\begin{bmatrix}\langle\alpha\rangle \\ \sigma_\alpha \\ \theta_\alpha\end{bmatrix} = \begin{bmatrix} & \tilde{Q} & \end{bmatrix}\left(\begin{bmatrix}\langle X_{\max}\rangle \\ \sigma_{X_{\max}} \\ \theta_{X_{\max}}\end{bmatrix} - \begin{bmatrix}r_1 \\ r_2 \\ r_3\end{bmatrix}\right)$$

  - Q_{ij} and r_j are functions of f_σ, polynomial constants from MC-training, and powers of $\log E$

- **Compute moments of ln(A) while retaining the ability to semi-analytically scale highest-energy cross-sections**



	Sample	Calc
Mean:	955.437	955.564
RMS:	96.3088	96.7577
Skew:	−1.17819	−1.14164

Conclusion

- f_3 distribution
  - well-motivated
  - describes Xmax well
  - already widely recognized
  - facilitates real-world data analysis
- three-moment analysis is a natural extension to two-moment analysis
  - especially as more data are collected!
- parameter/moment relation can be useful in many ways

- future work
  - anti-bias cuts which target Xmax RMS, skewness
  - full extension to composition mixtures

also, small Python module to aid evaluation:
`http://physics.ohio-state.edu/~jcs/downloads/2013-07-01/f3_eval.tar.bz2`

Log-normal distribution

- Another problem: X_H adds significant skewness to Xmax (for medium, high mass showers)
- BUT f_3 uses a normal distribution for X_H!
- log-normal distribution is better-motivated for cascades
  - unfortunately, log-normal moments are more complex functions of shape parameters

Second interaction(s)

- Single shower: average second interaction depth $\langle X_2 \rangle$ proportional to Xmax?
- Gamma distribution
  - Single/multiple particle species
- Average third interaction? Nth interaction? General description? Independent vs. exchangeable variables?