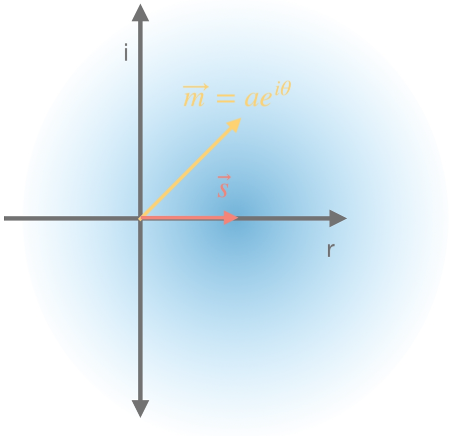
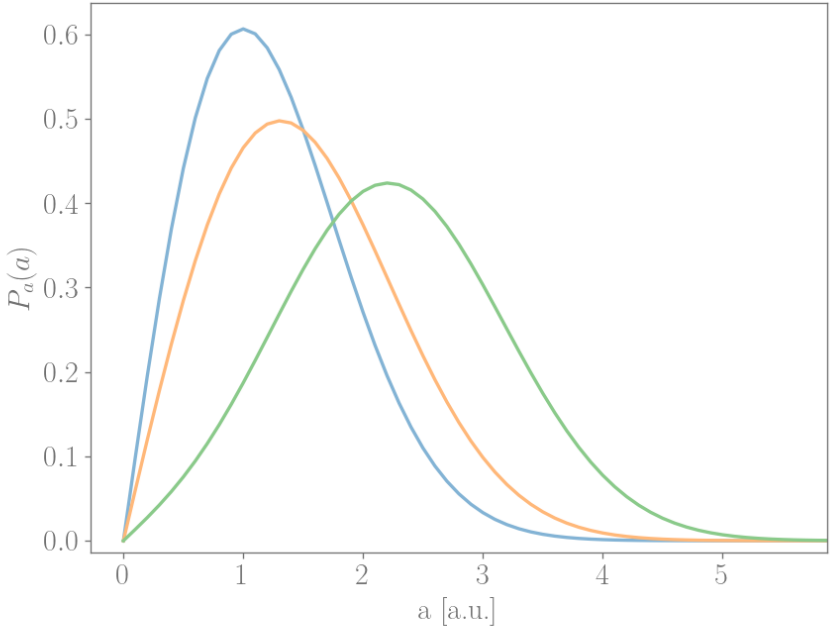
RD event reconstruction and mass sensitivity

Spectral slope model for future improvements
Uncertainty estimation with RICE distribution

KSETA Plenary Workshop 2023, Durbach - 29th March

PhD student: Sara Martinelli





PIERRE AUGER OBSERVATORY

HIRSAP Helmholtz International Research School for Astroparticle Physics

KIT Karlsruher Institut für Technologie

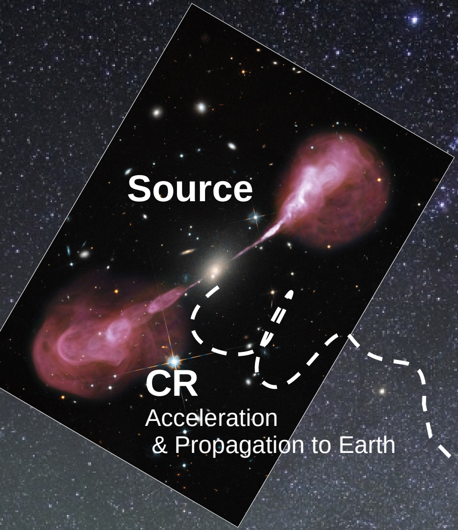
UNSAM

sara.martinelli@kit.edu

Extensive Air Showers

EAS initiated by UHECRs

Ultra-High-Energy Cosmic Rays



CR
Acceleration
& Propagation to Earth

CR
First interaction with
the Earth atmosphere
UHECR E>10^{18}eV

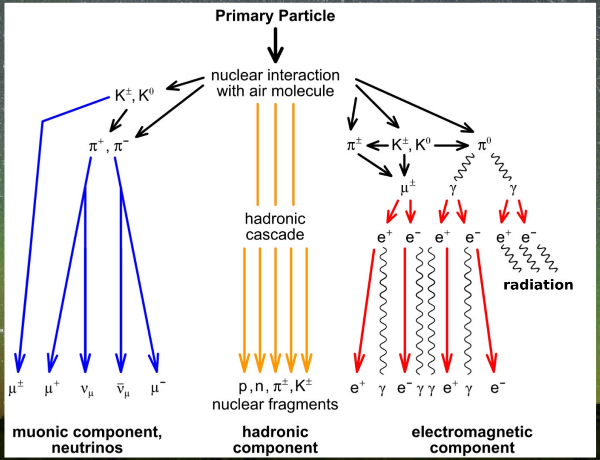
EAS

Development of
particles cascade

Xmax

- **Sources?**
- **Acceleration mechanisms?**
- **Composition?**

This talk: focus on composition and Radio detection of UHECRs at the Pierre Auger Observatory



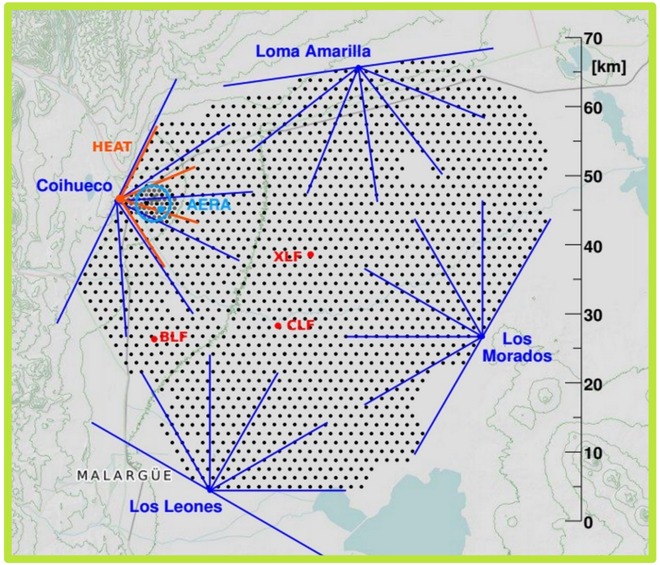
Hybrid detection at Auger

Largest ground-based observatory combining several detection techniques

PIERRE AUGER OBSERVATORY

The Pierre Auger Observatory

Located in Argentina, 3000km^2



Surface Detector (SD)

- Water Cherenkov Detectors (WCDs)
- 1600 stations on 1.5km grid

Fluorescence Detector (FD)

- 24 optical telescopes
- 4 sites
- atm. monitoring

Auger Engineering Radio Array (AERA)

- Radio detectors
- Coincidence with SD
- freq. band: 30-80 MHz
- Vertical showers
- E ~ 10^{17}eV - 10^{19}eV
- 150 stations over 17 km^2

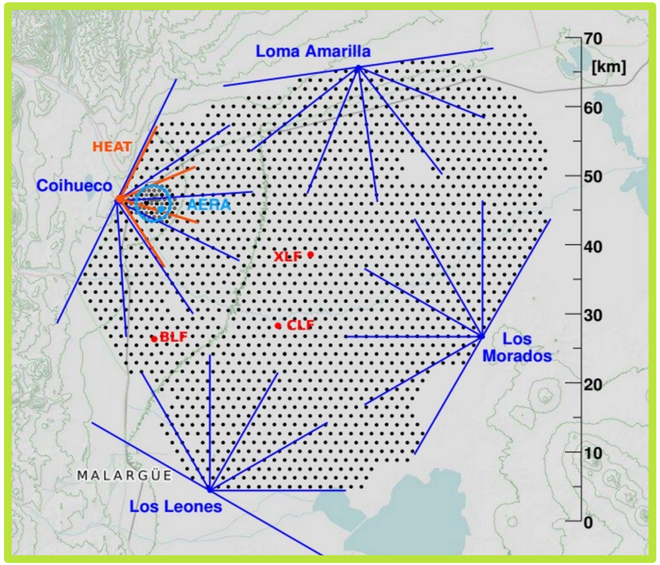
High Elevation Auger Telescopes (HEAT) → more shallow showers

SD-750, SD-433 → lower SD energy threshold

Auger Muon and Infill Ground Array (AMIGA) → buried scintillators

The Pierre Auger Observatory

Located in Argentina, 3000km^2



Fluorescence Detector (FD)

- Sensitive to the mass of the primary particle (X_{max})

- Operation time limited by day light, moon light, clouds …

Surface Detector (SD)

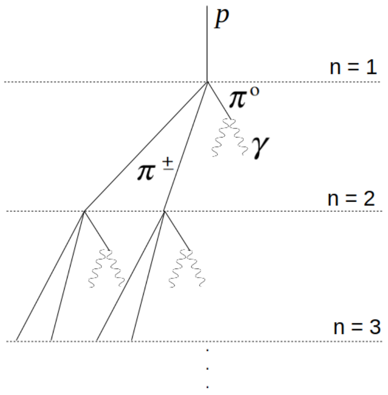
- Not sensitive to the mass

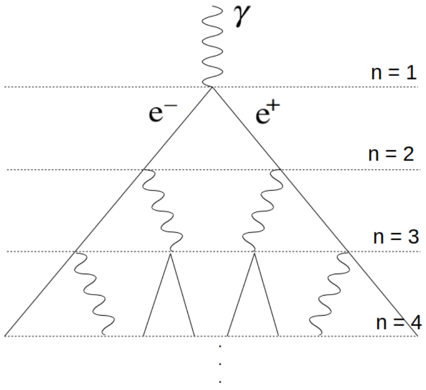
- 100% uptime

⇨ Upgrade to increase UHECR mass sensitivity

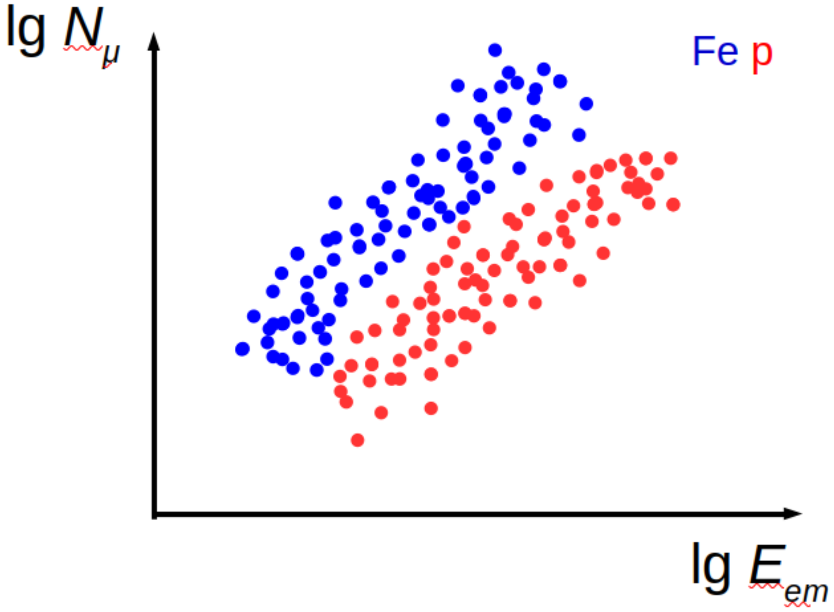
Mass sensitivity

Heitler-Mathews



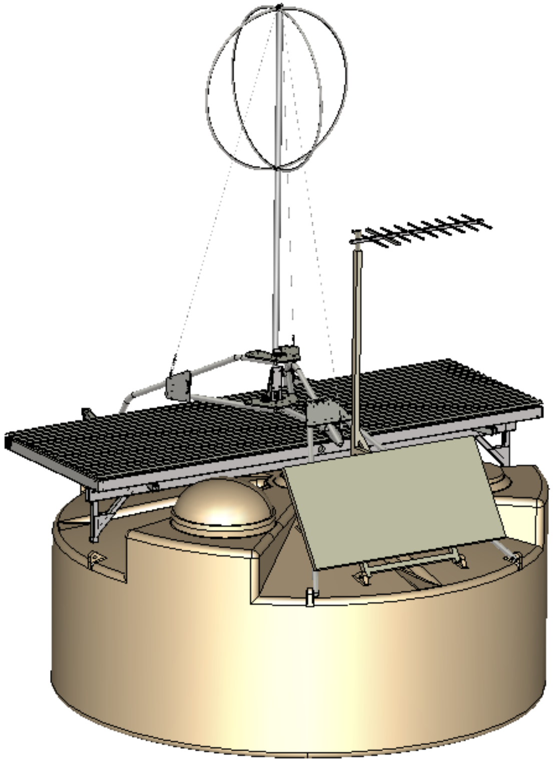


$N_\mu^A \approx A^{0.1} N_\mu^p$ $\quad N_\mu \sim E^{0.9}$ $\quad E_{em} \sim N_e \approx N_e^A \approx N_e^p$



WCD sensitive to muons → Upgrade SD to get sensitive to elm particles

AugerPrime

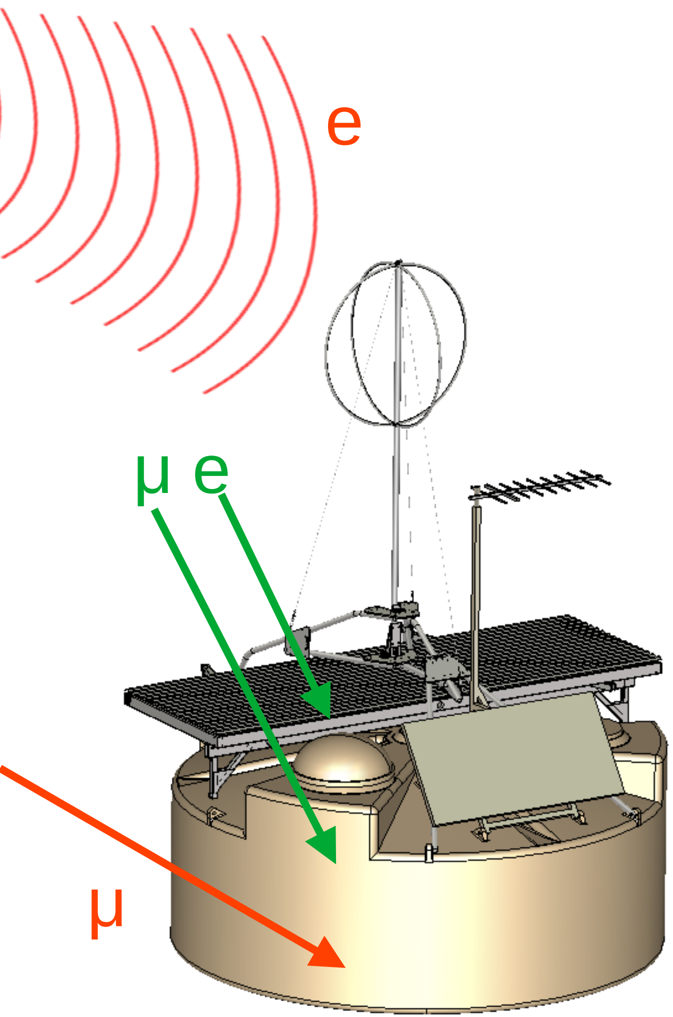


Ongoing upgrade of the Pierre Auger Observatory

Surface Scintillator Detector (SSD)
- 3.84 m^2 plastic scintillator panel on top of 1400 tanks

AugerPrime Radio Detector (RD)
- SALLA located on top of each tank
- Sensitive to radio in 30-80 MHz
- Trigger from WCD
- **Mass deployment started and expected to be over by the end of 2023!**
- RD-EA engineering array (10 antennas) already operative since end of 2021
    - background measurments
    - first showers

AugerPrime



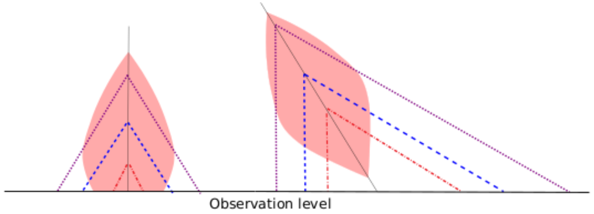
Ongoing upgrade of the Pierre Auger Observatory

SSD+ WCD

SSD loses sensitivity for more inclined air showers
→ vertical air showers ($0° \lesssim \theta \lesssim 60°$)

RD + WCD

Works well with inclined air showers ($65° \lesssim \theta \lesssim 85°$)



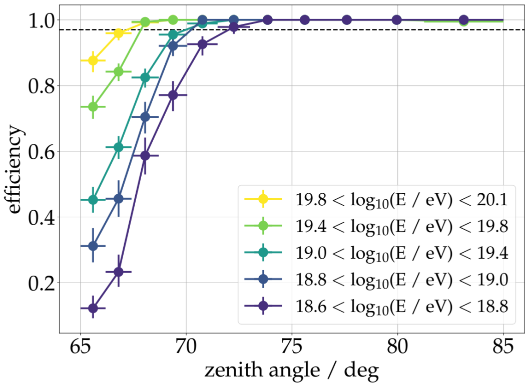
Radio Detector

Complete mass deployment by the end of 2023 ... more air showers data soon!

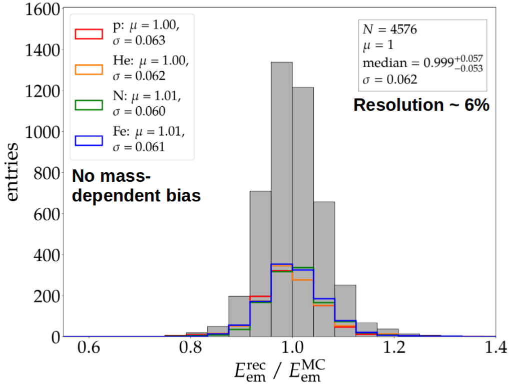
- **Radio reconstruction algorithm implemented in** Offline (official Auger simulation and reconstruction framework)
- Expected performance of RD
- Expected sensitivity to mass composition

→ Ready!

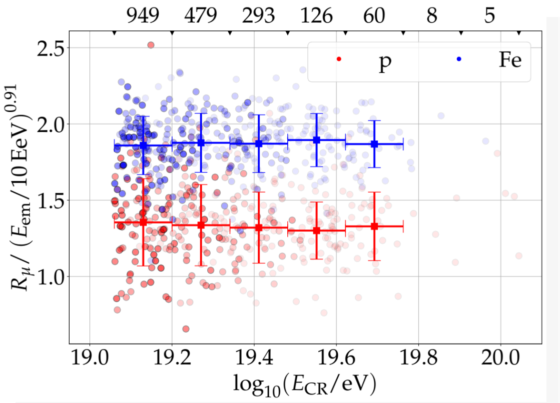
Work done by Felix Schlueter during his PhD



Full efficiency θ ≳ 70°, lg$E$ > 18.8



Energy resolution <10%



Event-by-event p-Fe discrimination ,FOM = 1.6

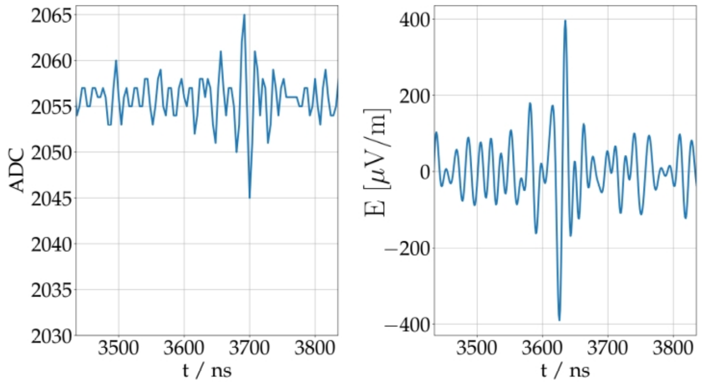
Radio Reconstruction

E-field reconstruction

- Digital to analog conversion, upsampling, Hann window etc.

- Unfolding of the response of the signal-processing chain (LNA, impedance matching, filter amplifiers...)

- **Unfolding of the antenna response** (NEC-2) to get the E-field (EW, NS, N)



Calibrated signals <u>(see Max Buesken's talk on last Monday)</u>

- Decomposition of the E-field in the **shower plane** coordinate system

- Estimation of signal-to-noise ratio (SNR)

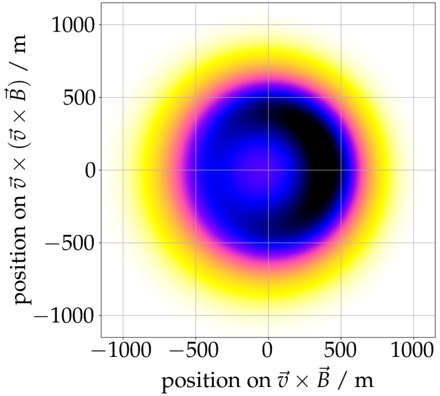
- Estimation of the **energy fluence** f [eV m^{-2}], the energy deposit per unit area

Still room for improvements
(this talk, more in the next slides)

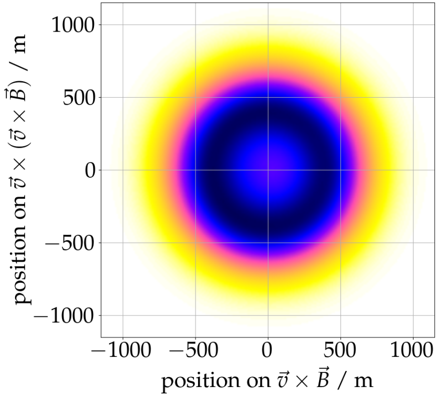
Radio Reconstruction

Geomagnetic energy fluence

- Analytic correction of early-late asymmetry
- Parameterized **subtraction of charge-excess emission**
  → 1-dim LDF

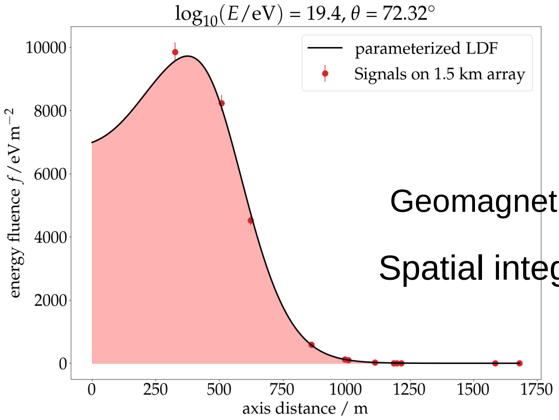


After removing asymmetries



Radiation energy

- **LDF fit** to estimate the geo radiation energy E_{geo} (energy emitted in form of waves)
- Correction on E_{geo}
  compensate for the second-order scaling with the geomagnetic angle and air density at Xmax



Geomagnetic radiation energy E_{geo}

Spatial integral over the energy deposit

Electromagnetic energy

$$S_{geo} \sim E_{em}^2$$

Signal Estimation

SNR quantification over the 3 polarizations

$$\mathrm{SNR} = \left(\frac{|A^{\mathrm{hilb}}_{\mathrm{tot}}|_{\max}}{V^{\mathrm{noise}}_{\mathrm{RMS}}}\right)^2$$

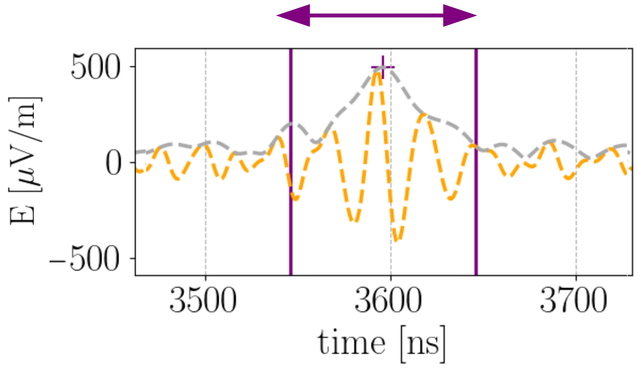
At the moment SNR cut is applied to determine presence of signal in a station

Estimation of the **energy fluence**

$$f = \epsilon_0 c \Delta t \left(\sum_{t_i=t_1}^{t_2} A(t_i)^2 - \frac{t_2 - t_1}{t_4 - t_3}\sum_{t_i=t_3}^{t_4} A(t_i)^2\right)$$

Noise fluence subtraction
[6000-8000]ns

100 ns window around the Hilbert envelope peak

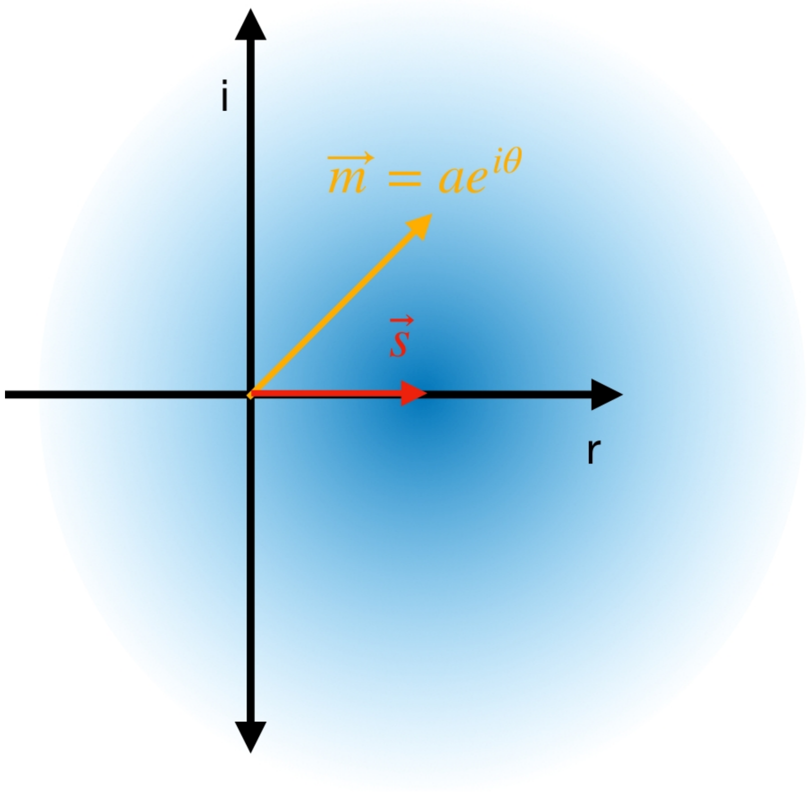


Uncertainties on the **energy fluence** underestimated (see backup slides)

Signal Estimation with Rice

Goal: get rid of the SNR selection and obtain better estimation of the signals and their uncertainties

CAVEAT: radio signal has amplitude and phase!



Our **measurement** can be expressed as:
sum of **constant known phasor s** and a **random phasor sum**

Joint density function for **amplitude and phase:**

$$p_{A\Theta}(a,\theta) = \begin{cases} \dfrac{a}{2\pi\sigma^2} \exp\left[-\dfrac{(a\cos\theta - s)^2 + (a\sin\theta)^2}{2\sigma^2}\right] & -\pi < \theta \leq \pi \text{ and } \\ & a > 0 \\ 0 & \text{otherwise.} \end{cases}$$

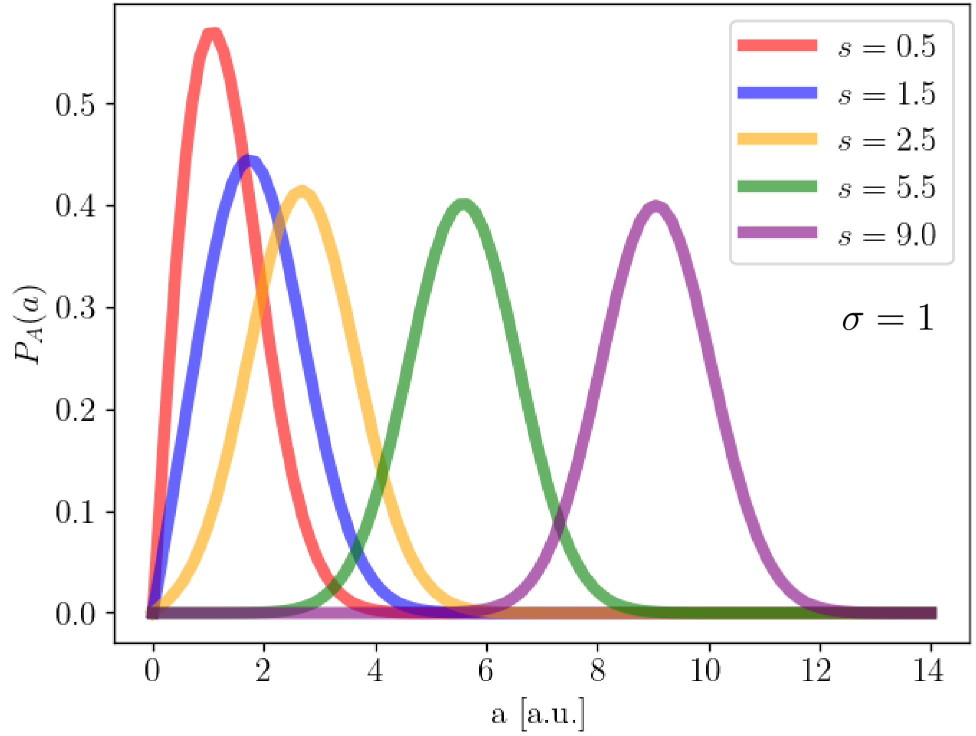
Based on Chapter 2.9 from J. W. Goodman, Statistical Optics (2015)

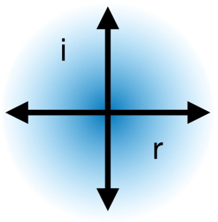
Signal Estimation with Rice

Marginal density function for amplitude only:

$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right)\, I_0\left(\frac{as}{\sigma^2}\right) & \text{a} > 0 \\ 0 & \text{otherwise} \end{cases}$$

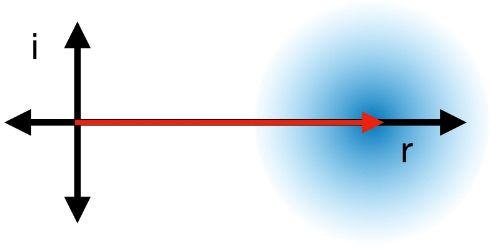
with I_0 modified Bessel function 1st kind of 0-order





No signal

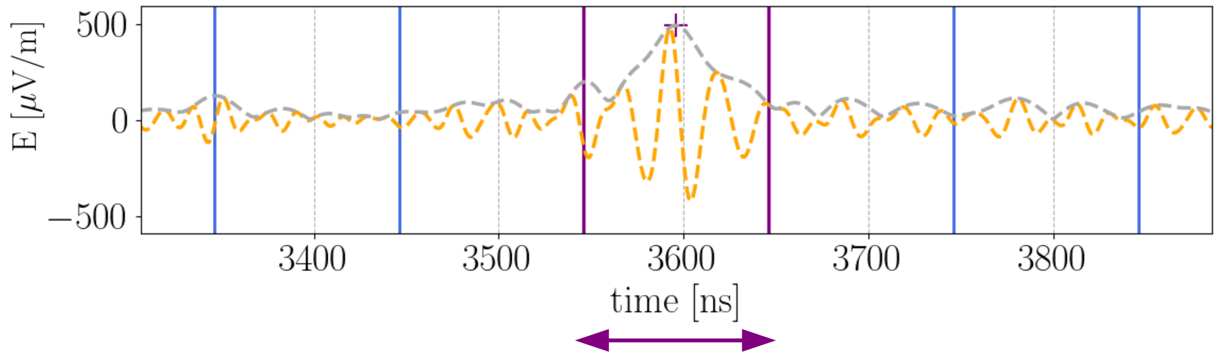
Rayleigh distribution



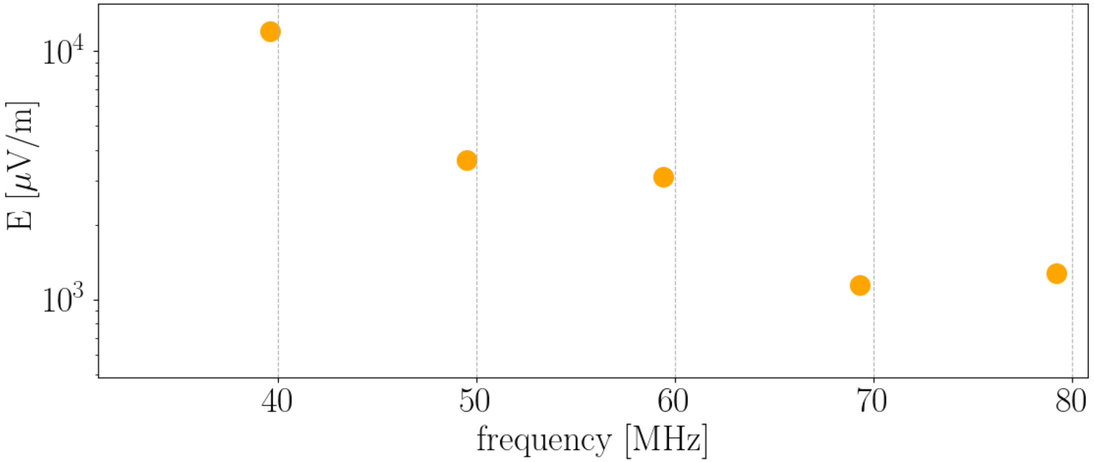
Large signal

Gaussian distribution

Signal Estimation with Rice



100 ns window around the Hilbert envelope peak

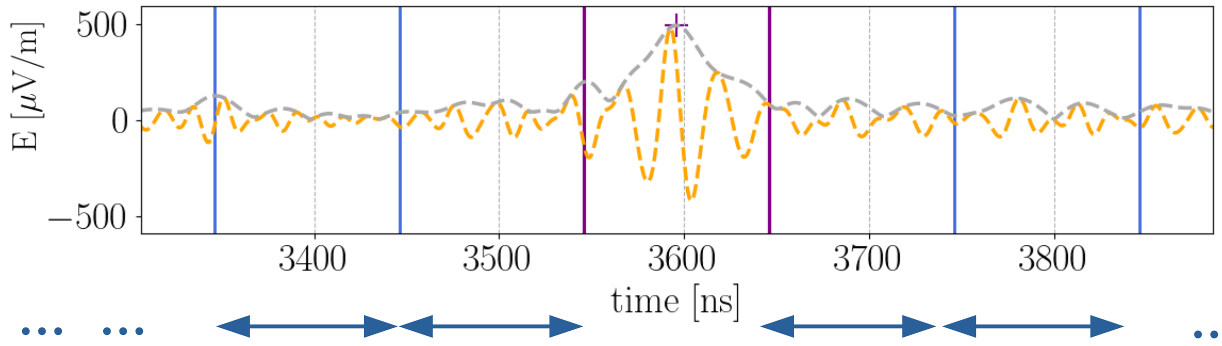


FFT in the signal window

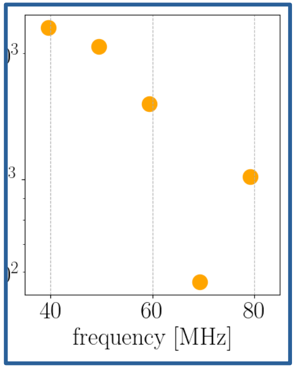
$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right)\, I_0\left(\frac{as}{\sigma^2}\right) & \text{a} > 0 \\ 0 & \text{otherwise} \end{cases}$$

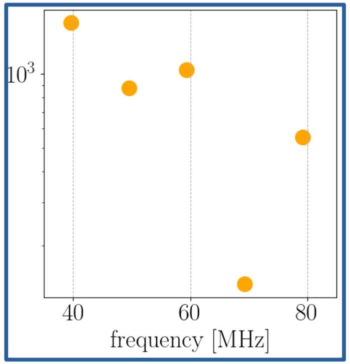
a(f) → our measurement

Signal Estimation with Rice



[80 noise windows 100 ns wide along the 8192 ns trace]()





FFT in the noise windows

$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right) I_0\left(\frac{as}{\sigma^2}\right) & \text{a} > 0 \\ 0 & \text{otherwise} \end{cases}$$

σ(f) → noise level

$$\sigma = \mu / \sqrt{(\pi/2)}$$

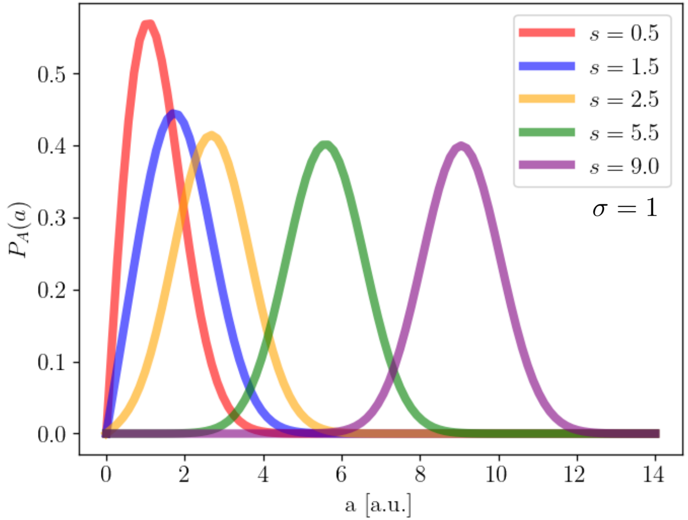
Signal Estimation with Rice

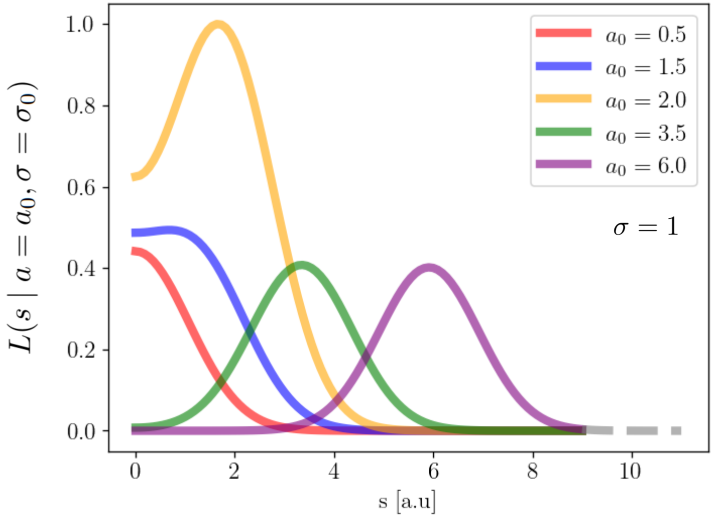
Get estimators of s for each frequency bin using the maximum of the Likelihood function

$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right)\, I_0\left(\frac{as}{\sigma^2}\right) & \text{a} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\left.\frac{dL(s)}{ds}\right|_{s=\hat{s_{\mathrm{ML}}}} = 0$$

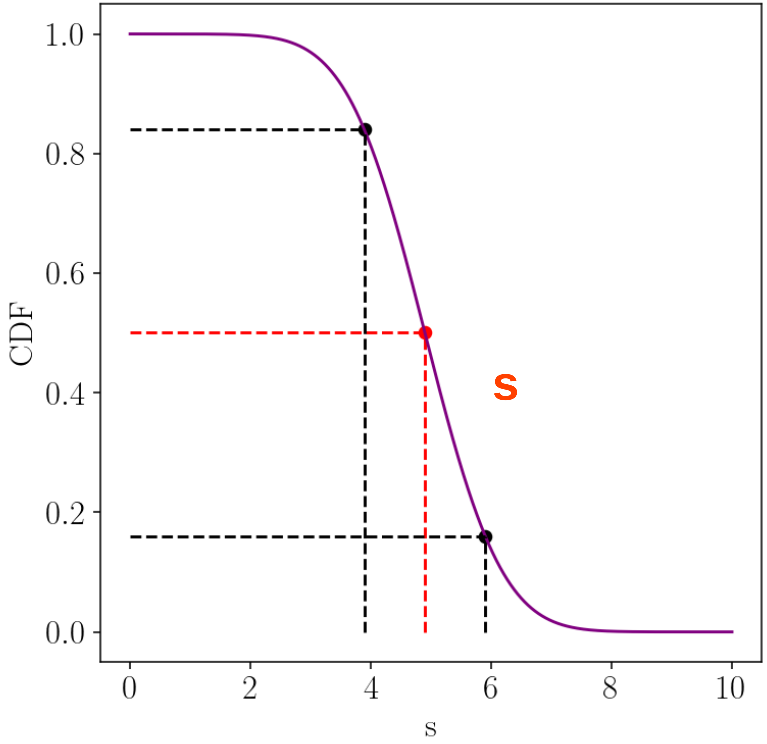
$$L(s \mid a = a_0, \sigma = \sigma_0) = \frac{a}{\sigma^2} \cdot exp\left(-\frac{a^2+s^2}{2\sigma^2}\right) \cdot I_0\left(\frac{as}{\sigma^2}\right), \quad \text{s} \geq 0$$

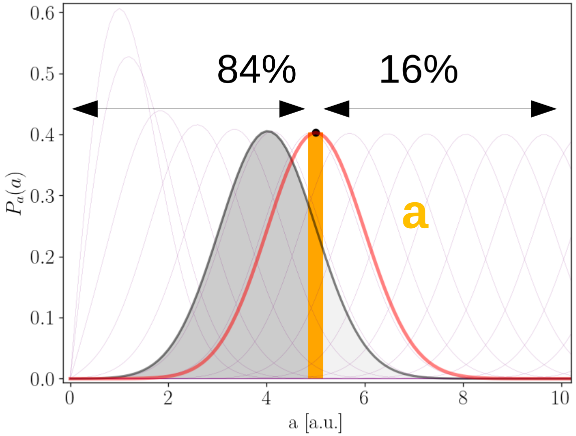




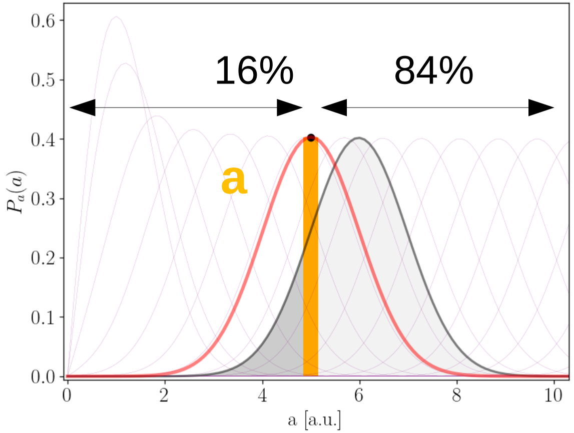
Uncertainty Estimation with Rice

Define CDF intervals to estimate uncertainty





smaller s



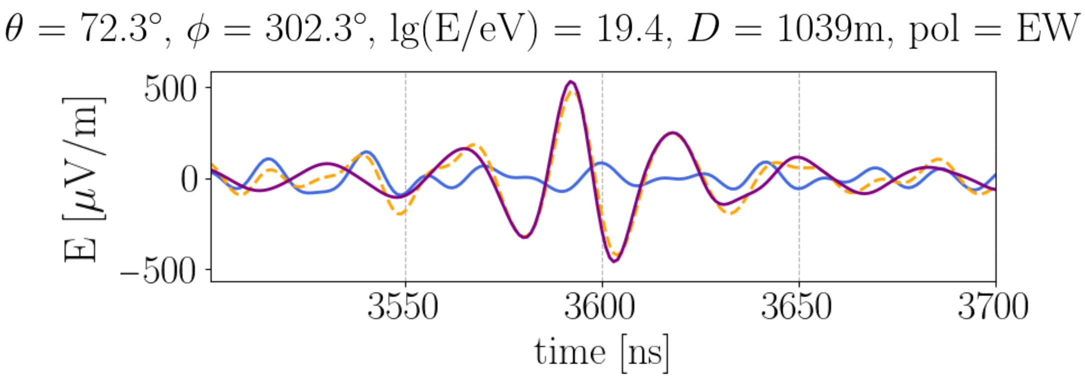
larger s

Bias with Rice (work in progress)

S pure simulated signal

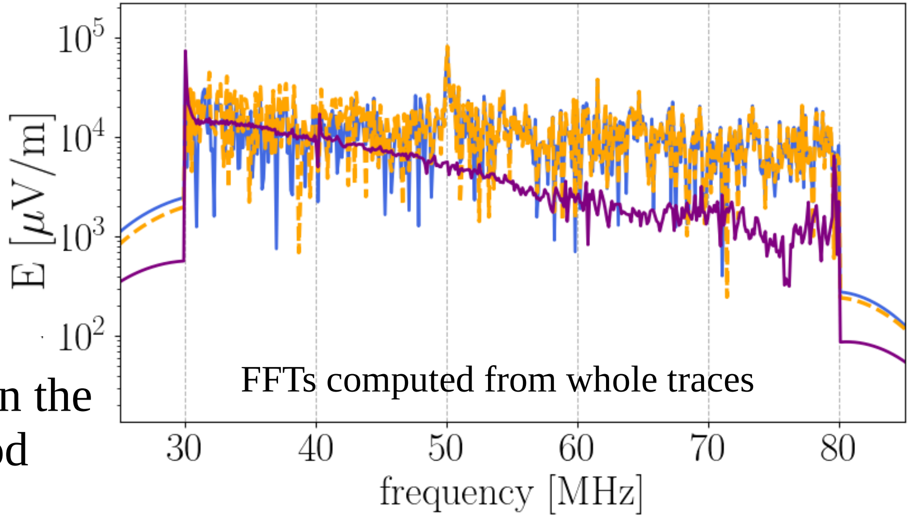
N noise (from background measurements)

S + N measurement



Add to the same simulation different noise and **study the bias obtained for the energy fluence**

→ Compare to the bias of the actual standard estimation

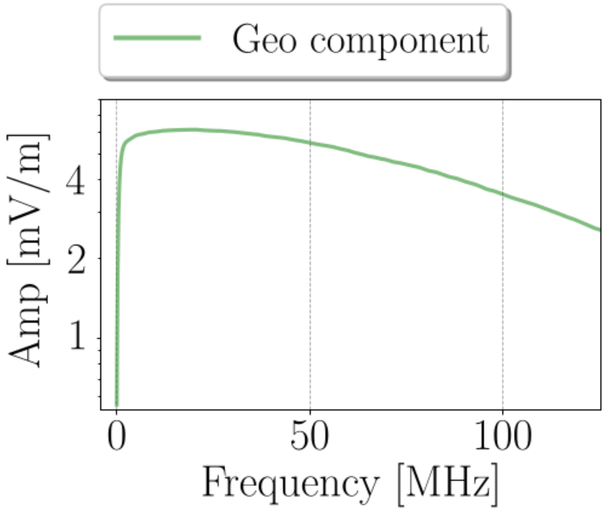


Offline: storing f, s, sigma, uncertainties in the ParameterStorage to have the full likelihood

18

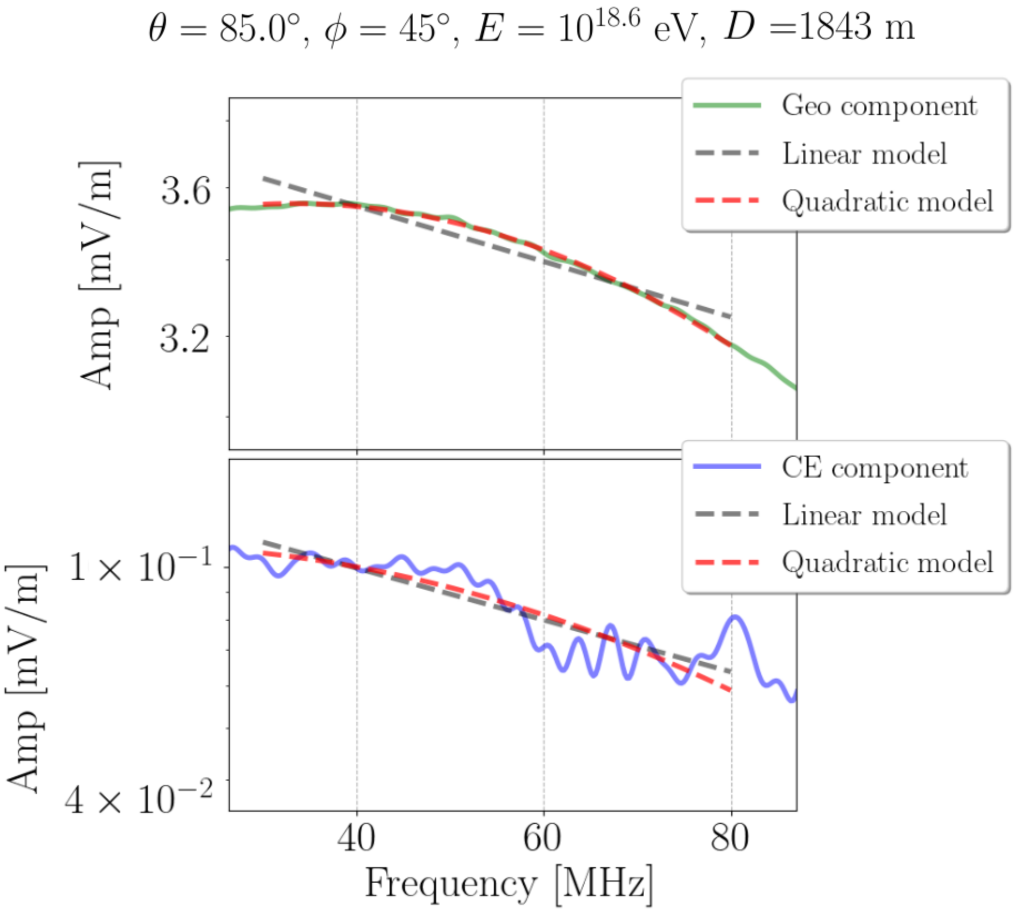
Frequency spectrum

$\theta = 85.0°,\ \phi = 45°,\ E = 10^{18.6}$ eV, D =1843 m



A further improvement would be to use a Signal model describing the shape of the frequency spectrum and fit all points maximizing the likelihood $\mathscr{L} = \sum_i -\log(P(x_i \mid S_i, \sigma)) = -\sum_i \log(\mathrm{Rice}(x_i, S_i, \sigma))$

Frequency spectrum



$$L(f) = A \cdot 10^{\,m_{\mathrm{f}} \cdot (f - f_0)}$$

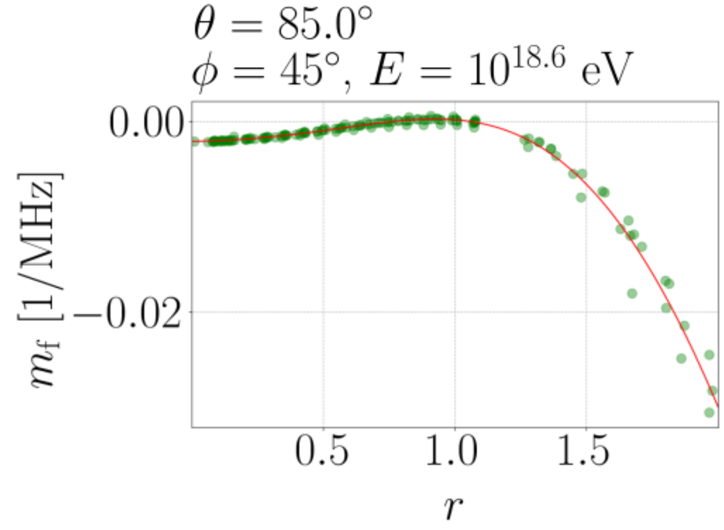
Frequency slope

$$Q(f) = A \cdot 10^{\,m_{\mathrm{f}} \cdot (f - f_0) + m_{\mathrm{f2}} \cdot (f - f_0)^2}$$

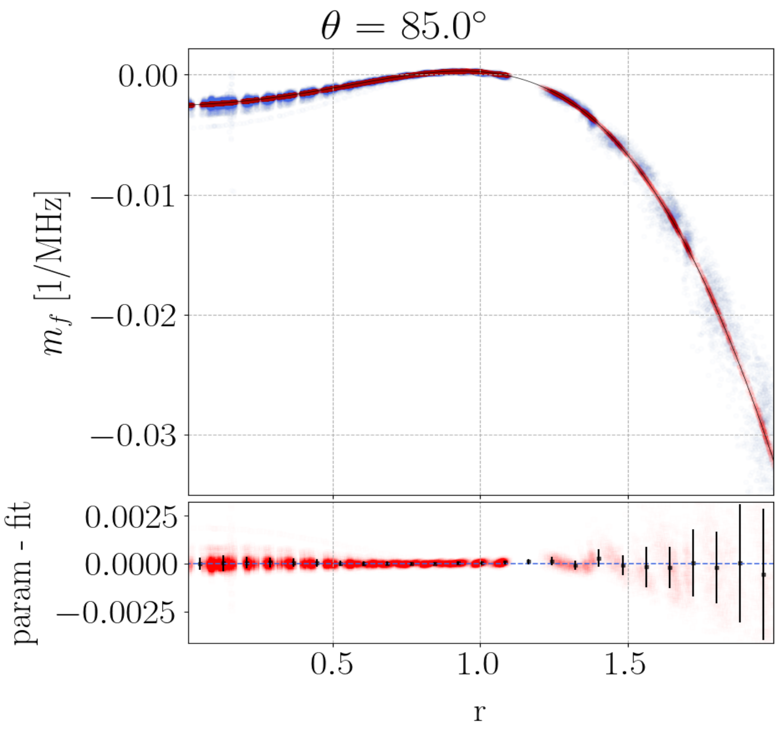
Frequency slope

Lateral distance function

$$f(r) = A_1 \cdot \left[-e^{B\cdot(r-r_0)} + A_2 \cdot e^{-C\cdot(r-r_0)^2}\right]$$



Parameterization as a function of the geometrical distance between core and shower maximum



Can be used to constrain geometry

Conclusions

- Mass deployment of RD completed by the end of 2023

- Radio reconstruction algorithm available in Offline

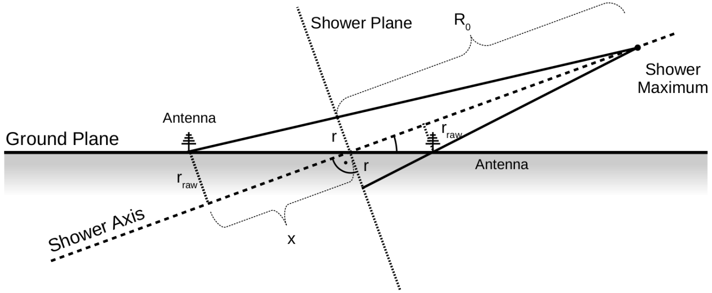
- Ongoing improvements about the signal estimation using the Rice distribution willl be soon implemented in Offline

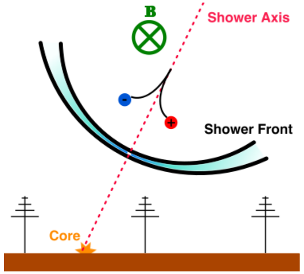
- Parameterization of the spectral content of radio emission in 30-80 MHz available for future improvements (constrain geometry)

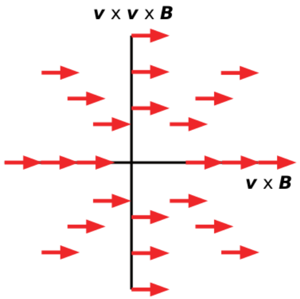
Backup

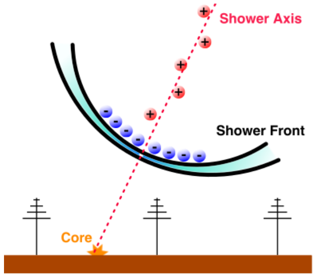
Early-late asymmetry

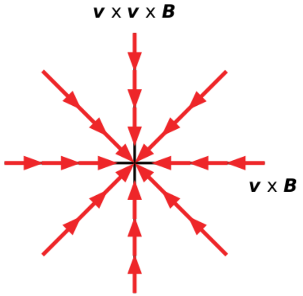
Charge-excess asymmetry

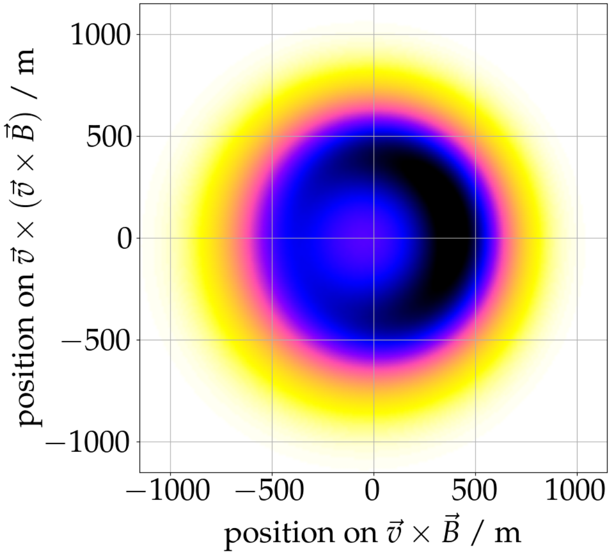


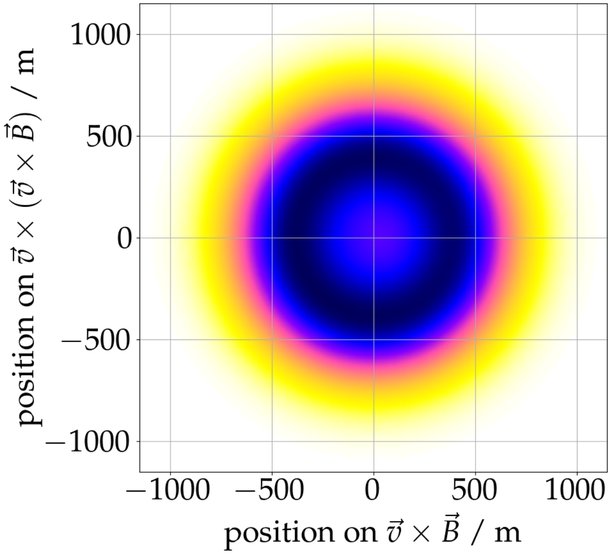












Energy fluence Uncertainty model

$$\sigma_{\mathrm{f}}^2 = \sigma_{\mathrm{Gauss}}^2 + \sigma_{\mathrm{det}}^2 + f_{\mathrm{noise}}^2$$

signal window

$$f = \epsilon_0 c \Delta t \left(\sum_{t_i=t_1}^{t_2} A(t_i)^2 - \frac{t_2 - t_1}{t_4 - t_3} \sum_{t_i=t_3}^{t_4} A(t_i)^2 \right)$$

noise window

$$f_{\mathrm{noise}}^2 = \frac{t_2 - t_1}{t_4 - t_3} \sum_{t_i=t_3}^{t_4} A(t_i)^2$$

fluence determined in the noise window

(introduced for AERA long time ago)

uncertainty due to noise after subtracting it

amplitudes are superposition of signal and white noise Gaussian distributed

sampling 1 ns

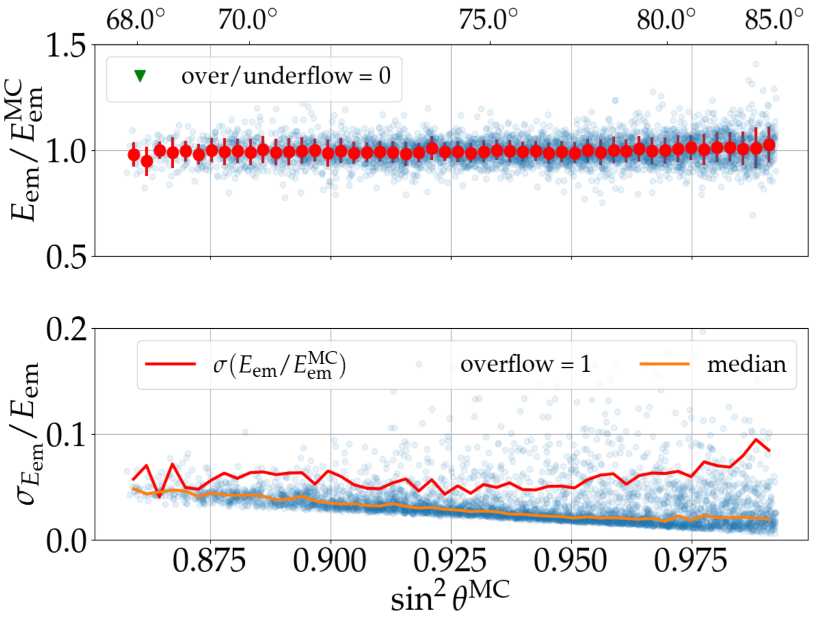
$$\sigma_{\mathrm{Gauss}}^2 = 4|f|\epsilon_0 c \Delta t V_{\mathrm{RMS}}^{\mathrm{noise}^2} + 2(t_2 - t_1)(\epsilon_0 c)^2 \Delta t V_{\mathrm{RMS}}^{\mathrm{noise}^4}$$

$\propto f_{\mathrm{noise}}^2$ *calculated in the noise window*

$$\sigma_{\mathrm{det}} = \sigma_{\mathrm{A}}^{\mathrm{detector}} \cdot 2 \cdot f$$

f~A² *5% uncertainty on the amplitudes* $\sigma_{\mathrm{A}}^{\mathrm{detector}} = 0.05$ *(probably overestimated vs data)*

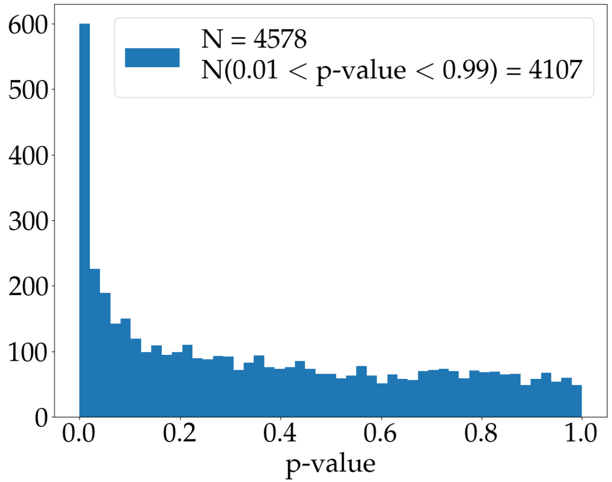
smearing the signal implemented in the detector simulation (antennas variation)



Resolution: how good we reconstruct the energy, std of Eem /EemMC in zenith bins

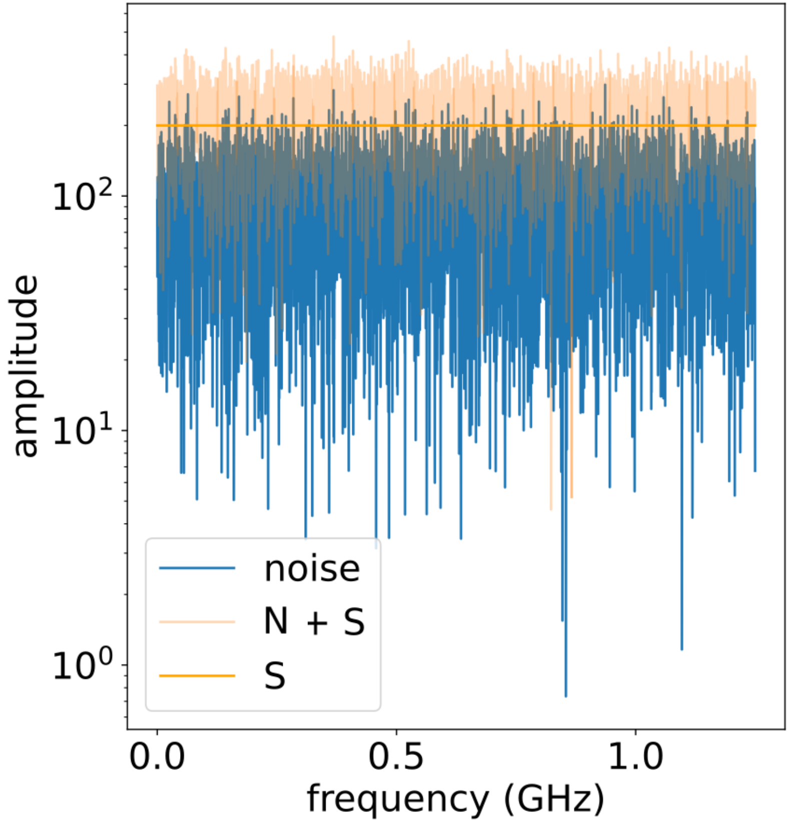
The relative uncertainties do not match the resolution → uncertainties underestimated

Goodness of the LDF fits:
too small p-values → uncertainties probably underestimated

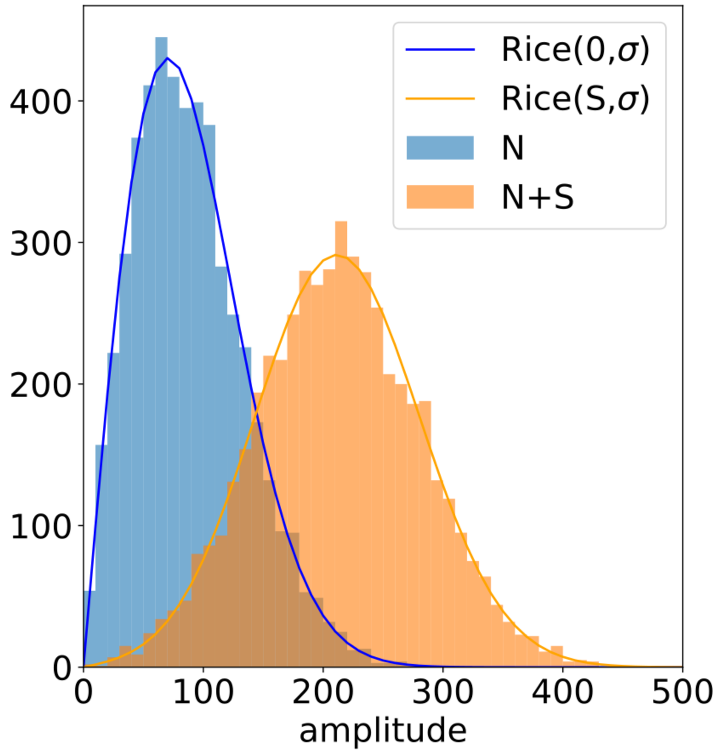


Rice distribution: a trivial example

S(t) = Aδ(t − t 0)
N noise (random generated)



FFTs computed from whole traces



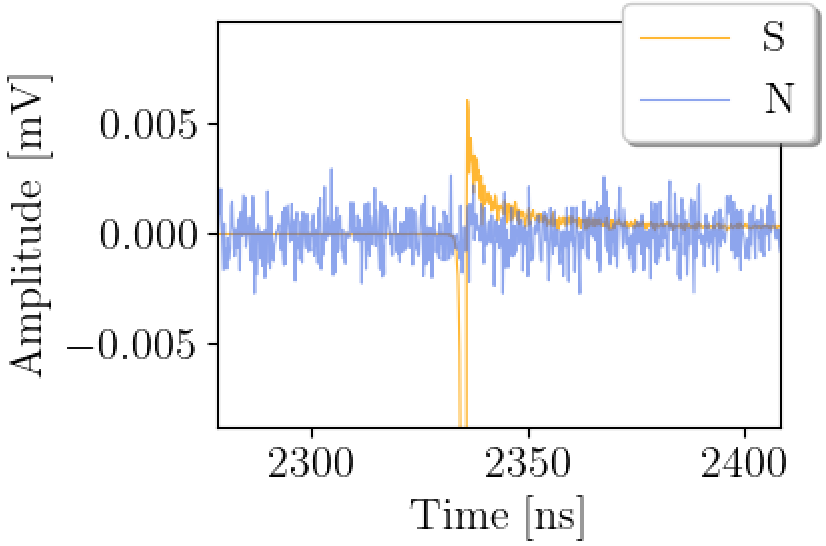
$\sigma = \mu / \sqrt{(\pi/2)}$

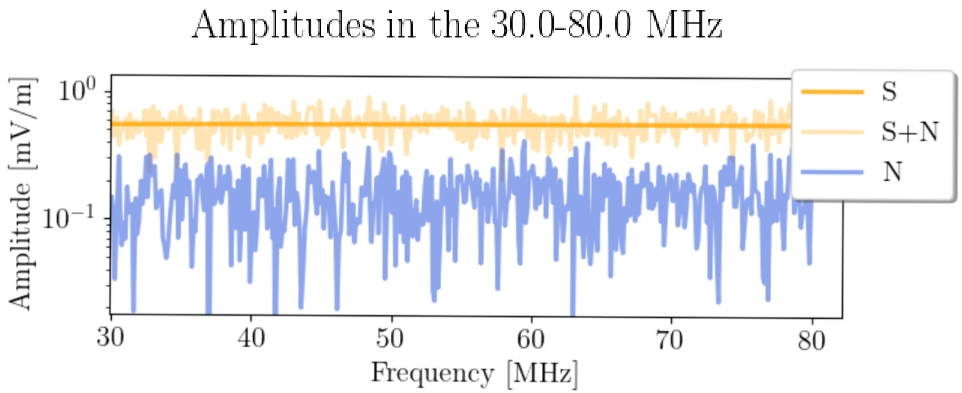
of the noise trace

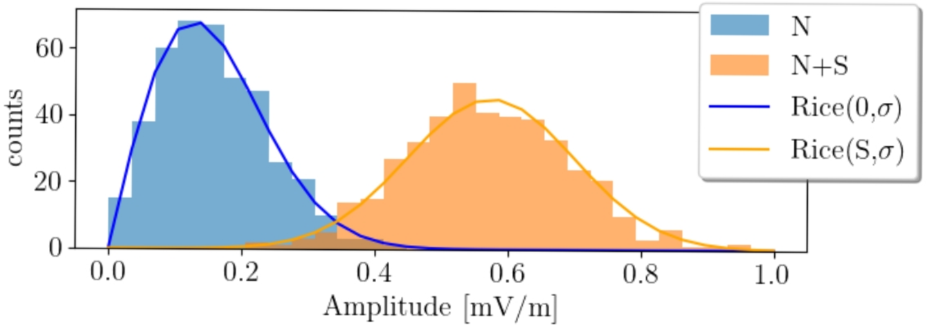
S mean value of the signal trace

Rice distribution: a less trivial example (a)

S simulated signal (geo component)
N noise (random generated)





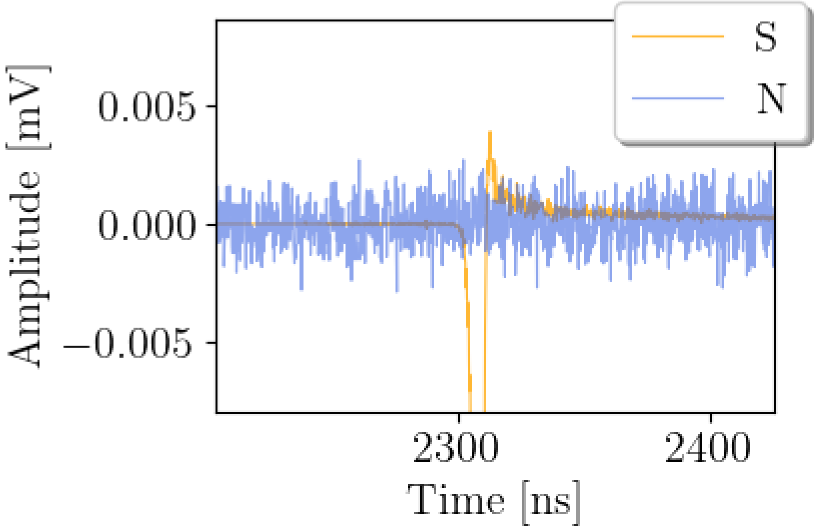


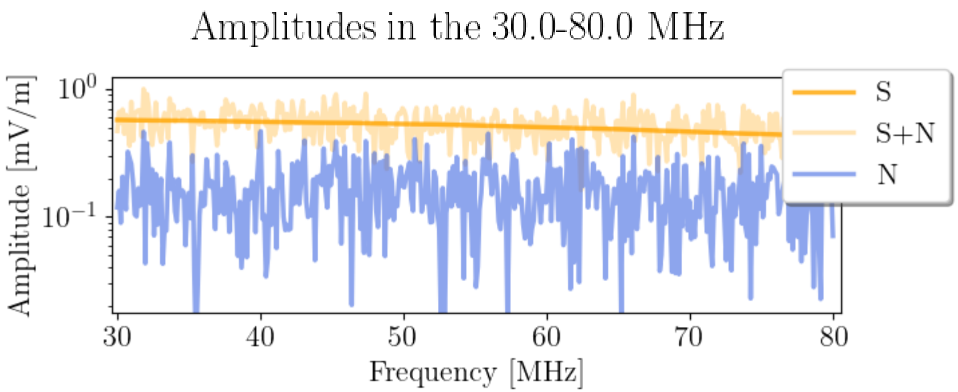
S mean value of the signal trace

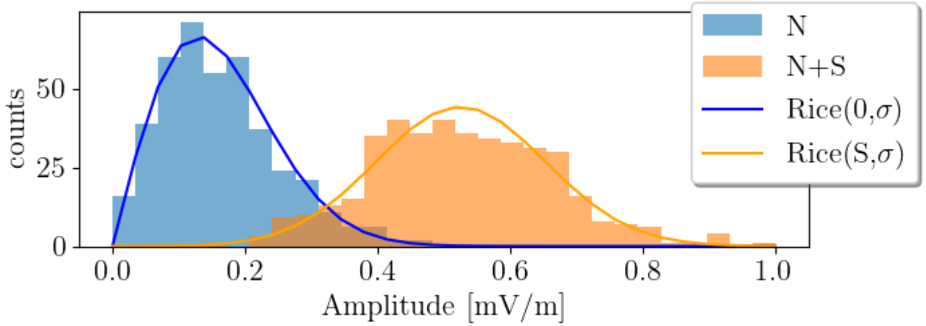
$$\sigma = \mu / \sqrt{(\pi/2)}$$

Rice distribution: a less trivial example (b)

S simulated signal (geo component)
N noise (random generated)







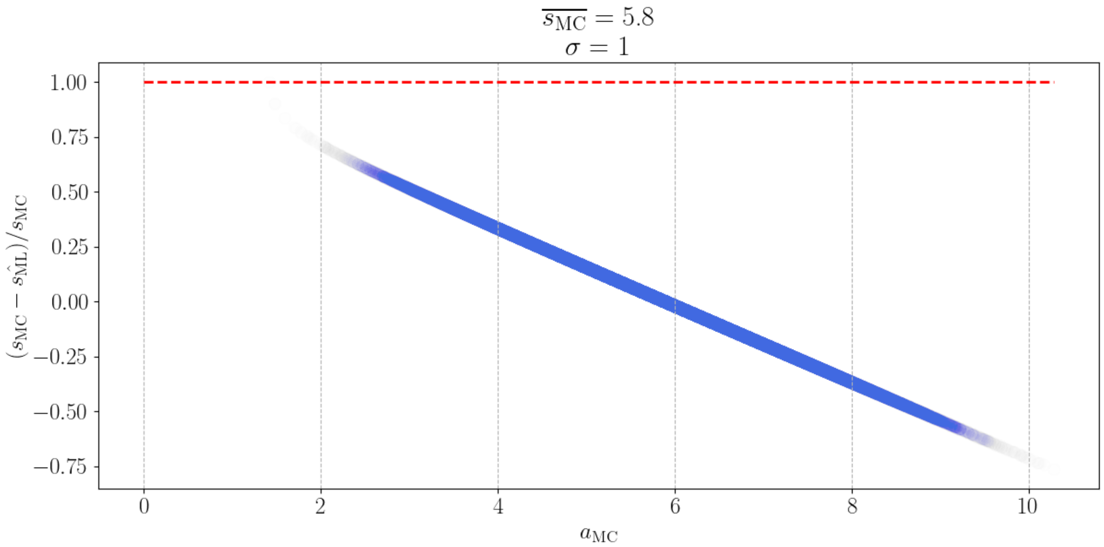
S mean value of the signal trace

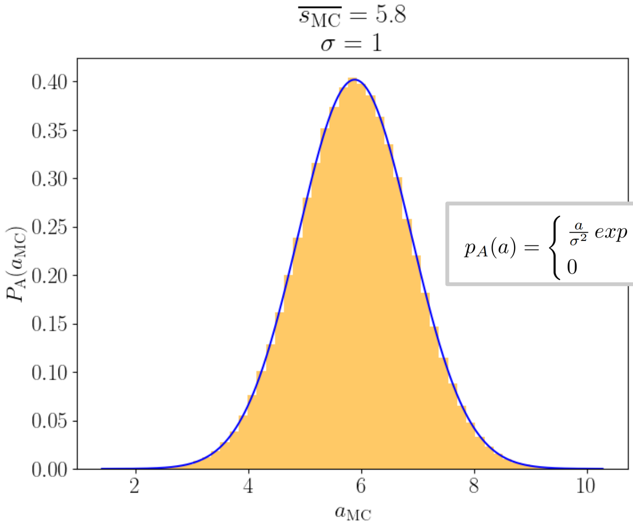
$$\sigma = \mu / \sqrt{(\pi/2)}$$

Slightly steeper frequency spectrum

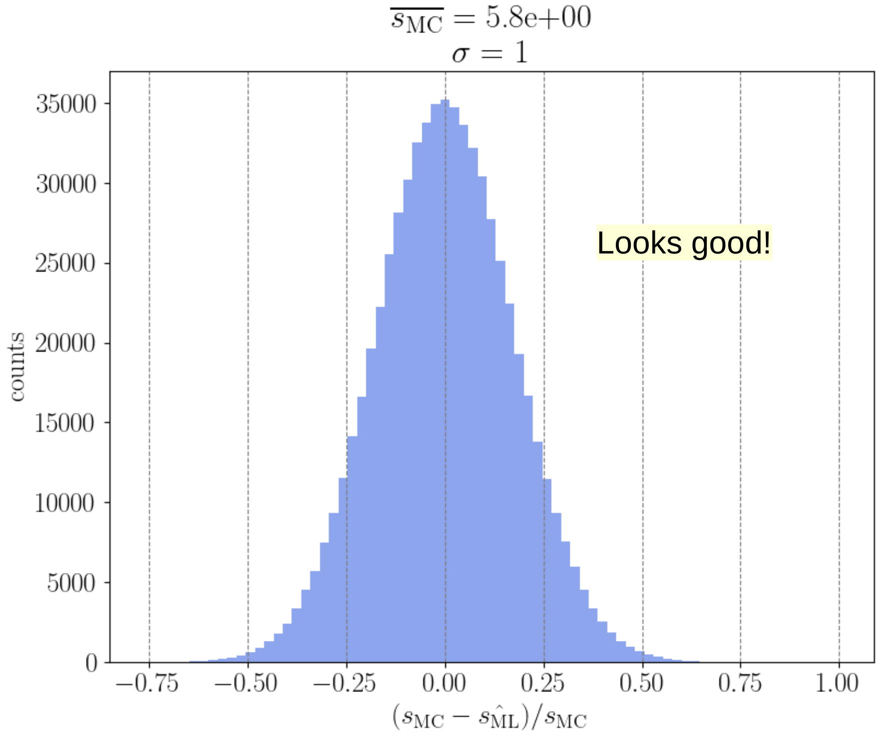
Toy MC

Binned MC signal &ML signal estimator



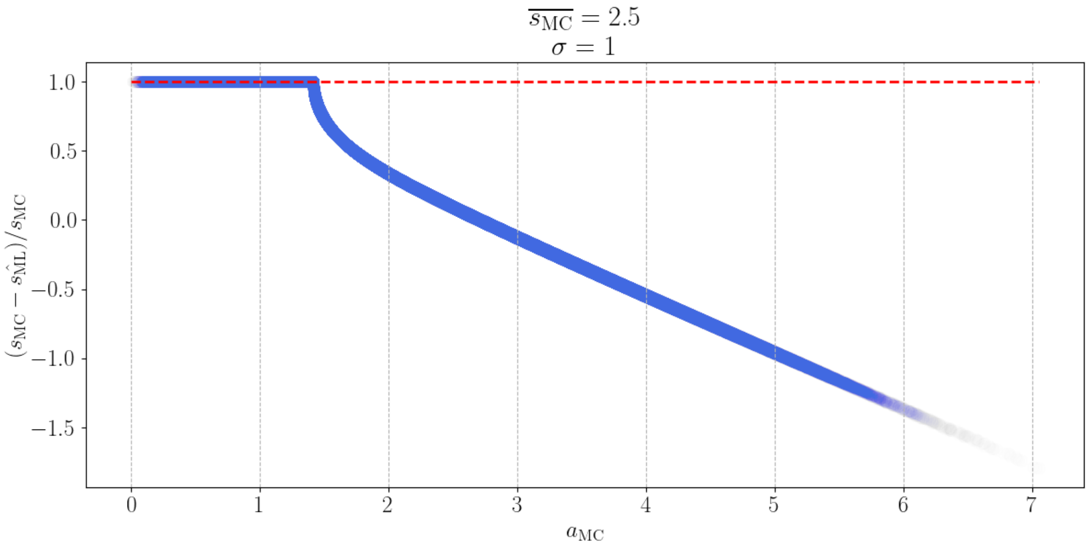


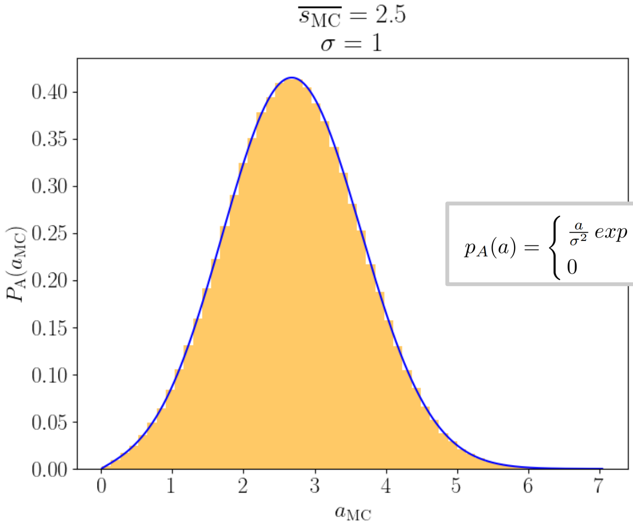
$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right)\, I_0\left(\frac{as}{\sigma^2}\right) & \text{a} > 0 \\ 0 & \text{otherwise} \end{cases}$$



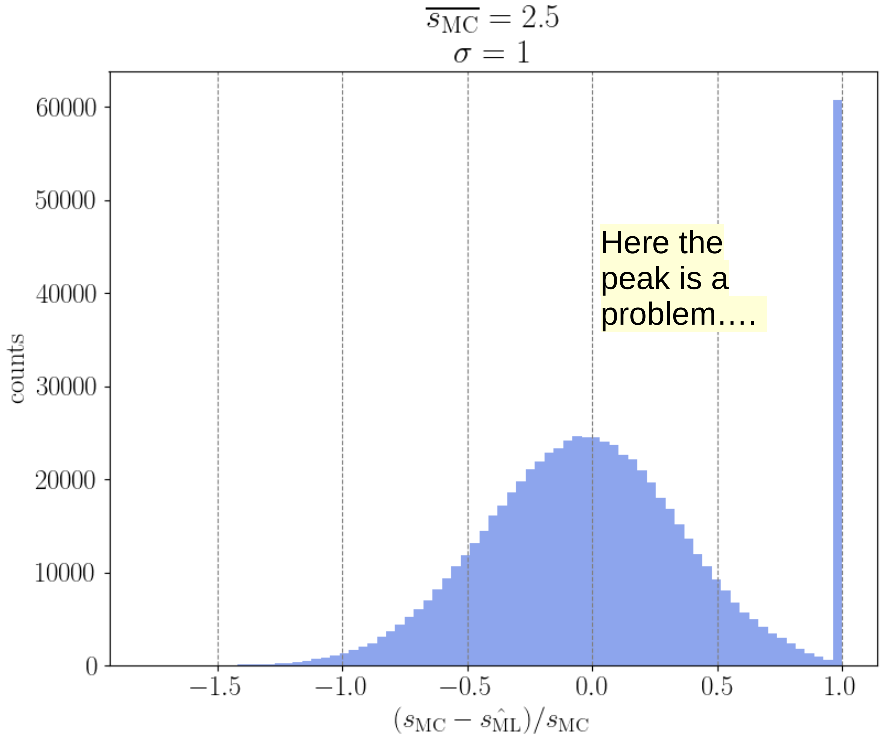
Toy MC

Binned MC signal &ML signal estimator



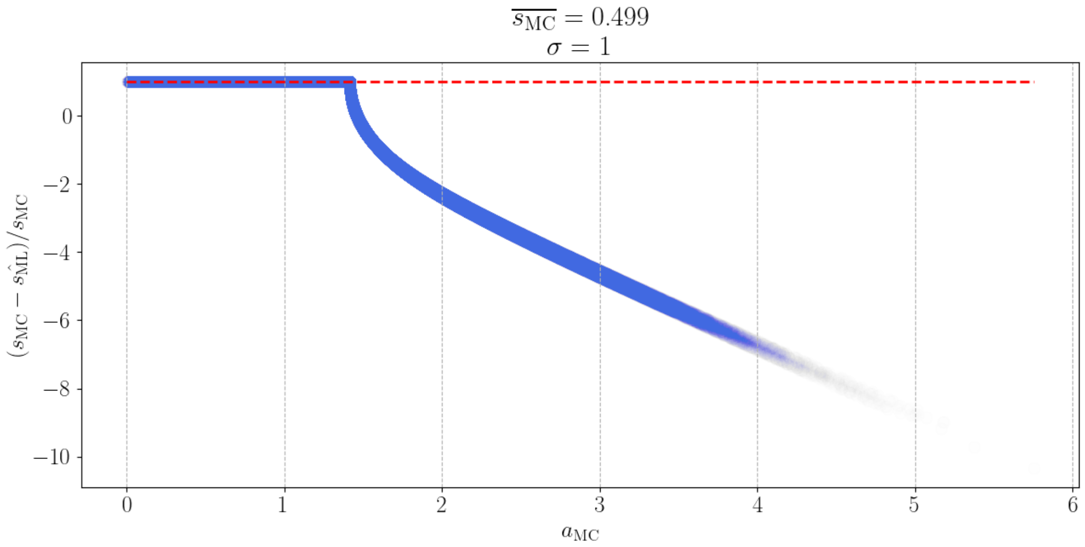


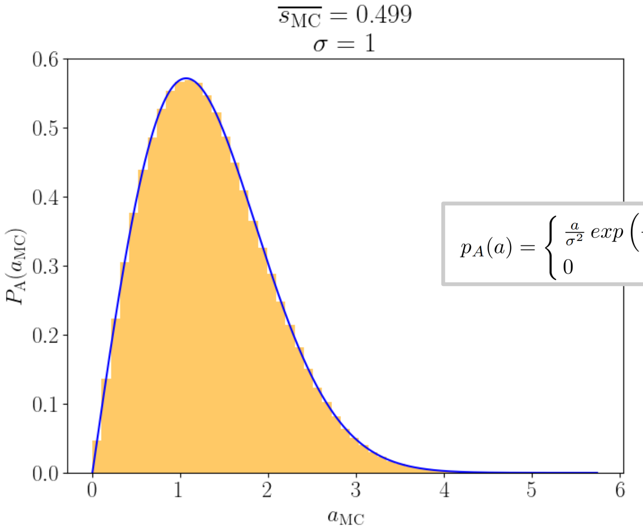
$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right)\, I_0\left(\frac{as}{\sigma^2}\right) & \text{a > 0} \\ 0 & \text{otherwise} \end{cases}$$



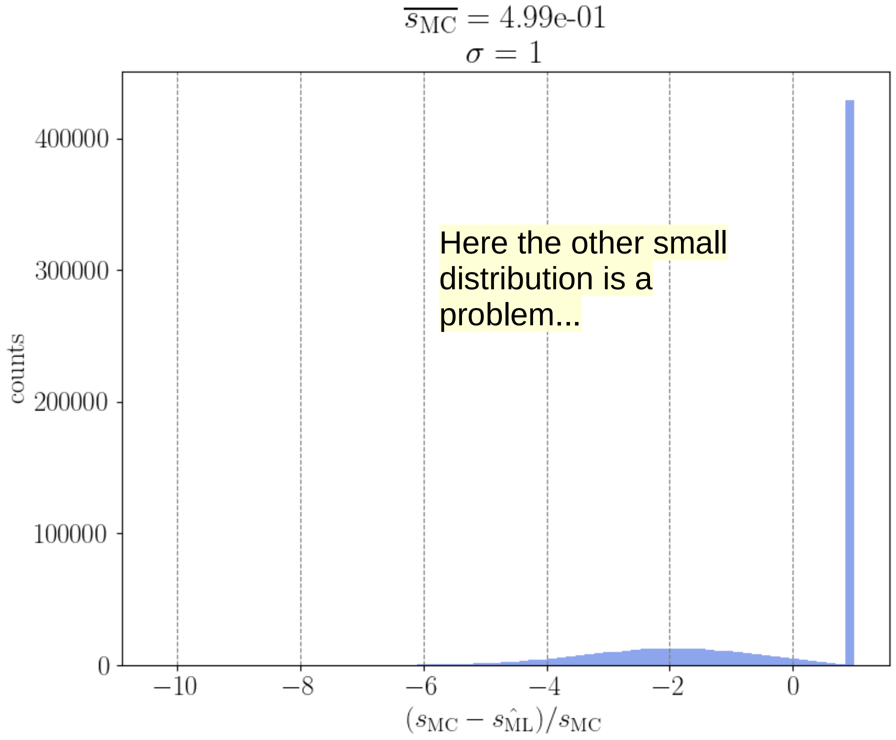
Toy MC

Binned MC signal &ML signal estimator



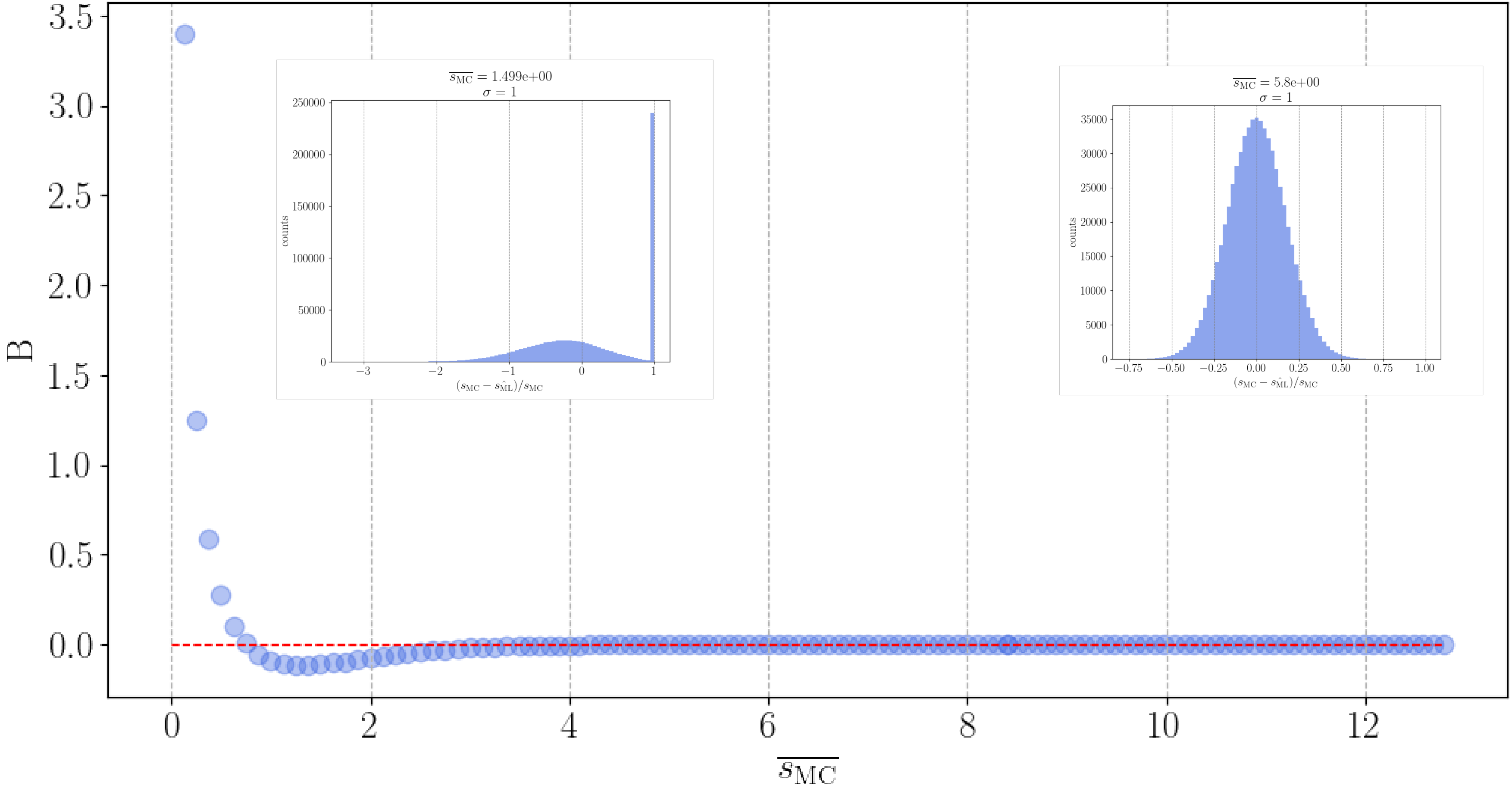


$$p_A(a) = \begin{cases} \frac{a}{\sigma^2}\, exp\left(-\frac{a^2+s^2}{2\sigma^2}\right)\, I_0\left(\frac{as}{\sigma^2}\right) & \text{a} > 0 \\ 0 & \text{otherwise} \end{cases}$$



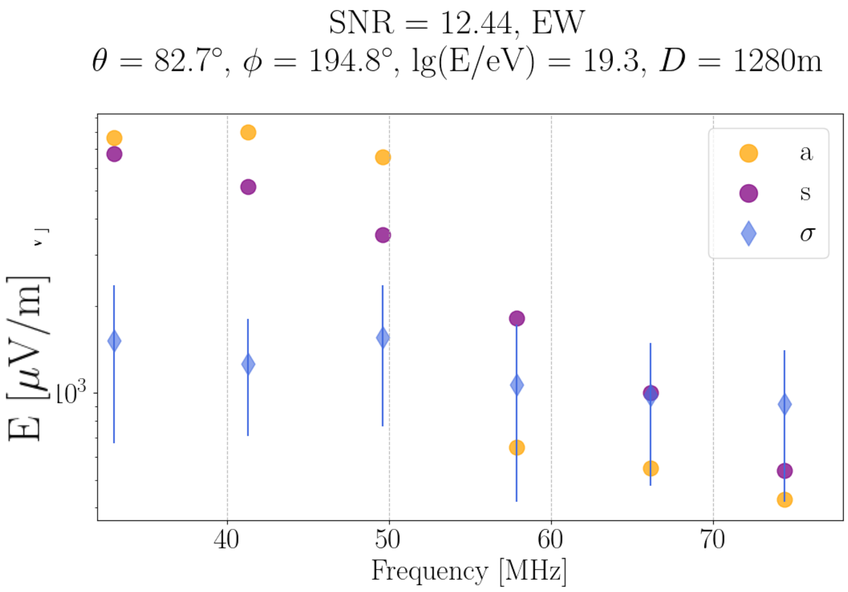
Bias normalized to the mean value of MC bin

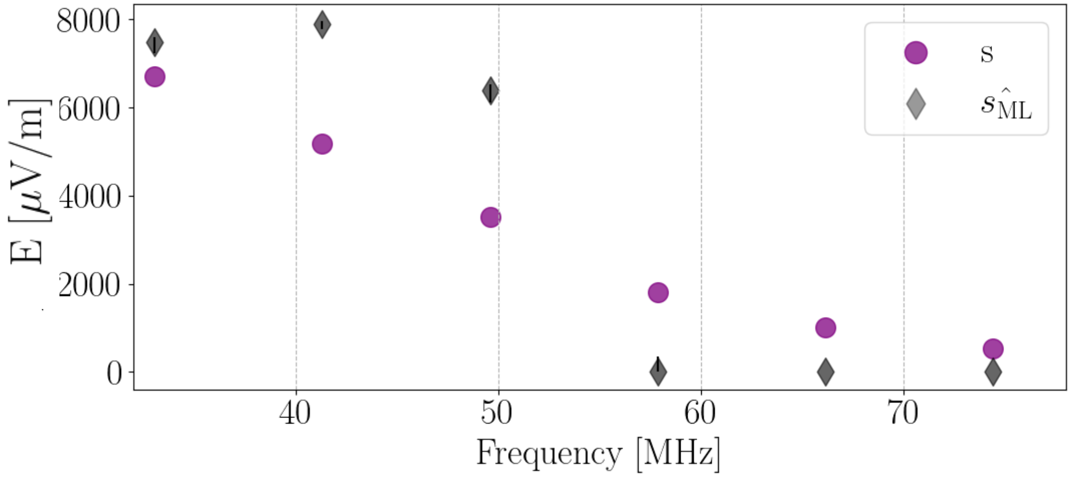
$$B = (\overline{\hat{s_{\mathrm{ML}}}} - \overline{s_{\mathrm{MC}}})/\overline{s_{\mathrm{MC}}}$$
$$\sigma = 1$$



Signal Estimation with Rice

Example of estimators for a given trace, s from pure simulation without noise

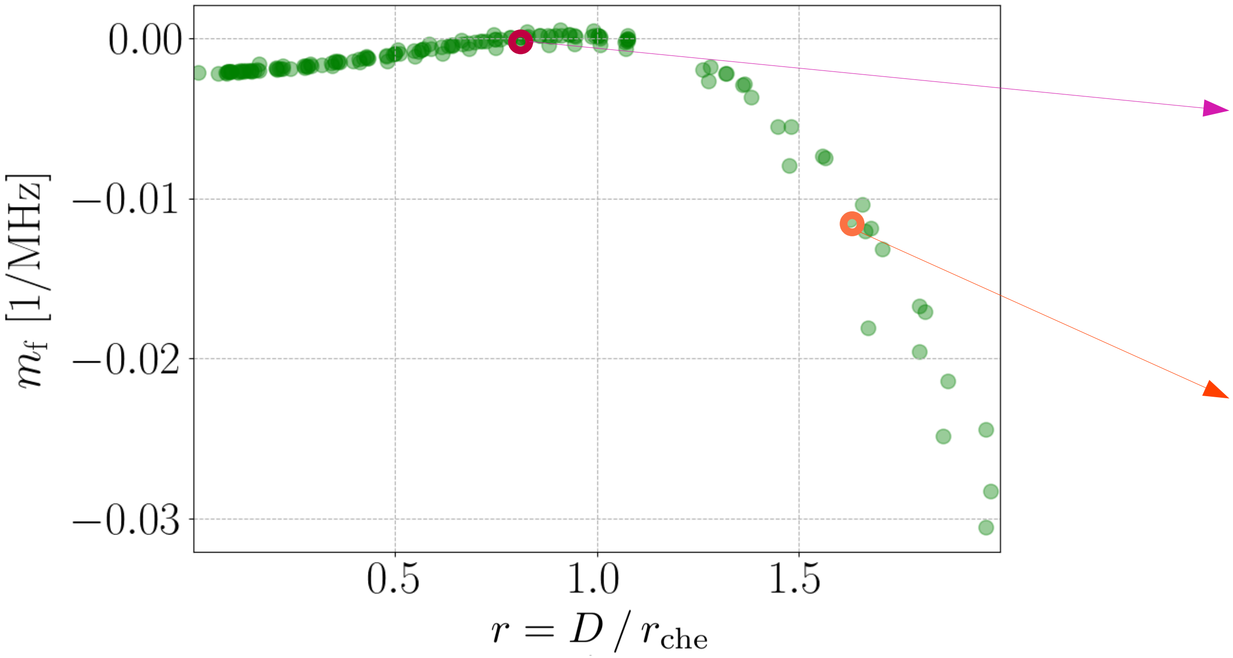


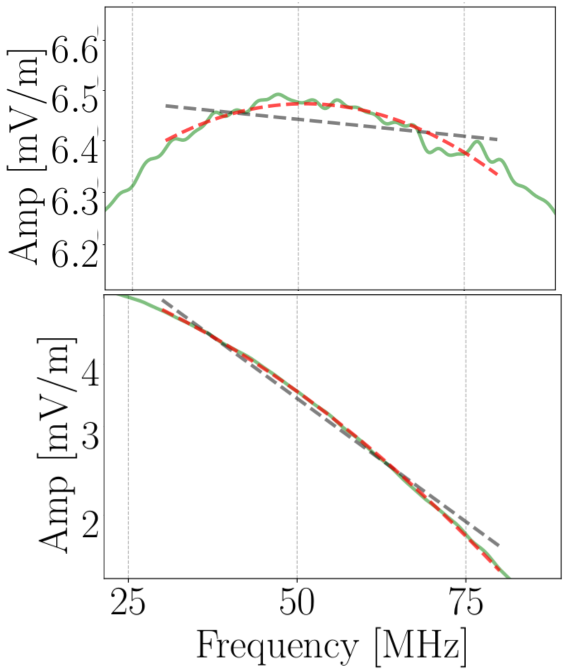


Single-event lateral distribution

Frequency slopes of the Geo component from a simulation having $\theta = 85.0°,\ \phi = 45°,\ E = 10^{18.6}$ eV

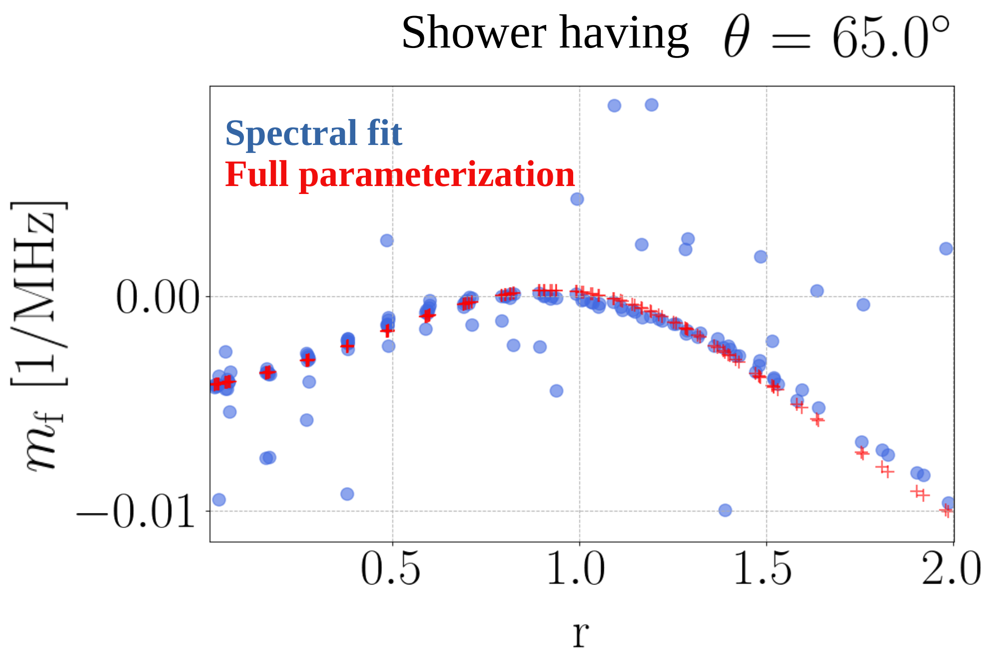
Spectral fits at different r

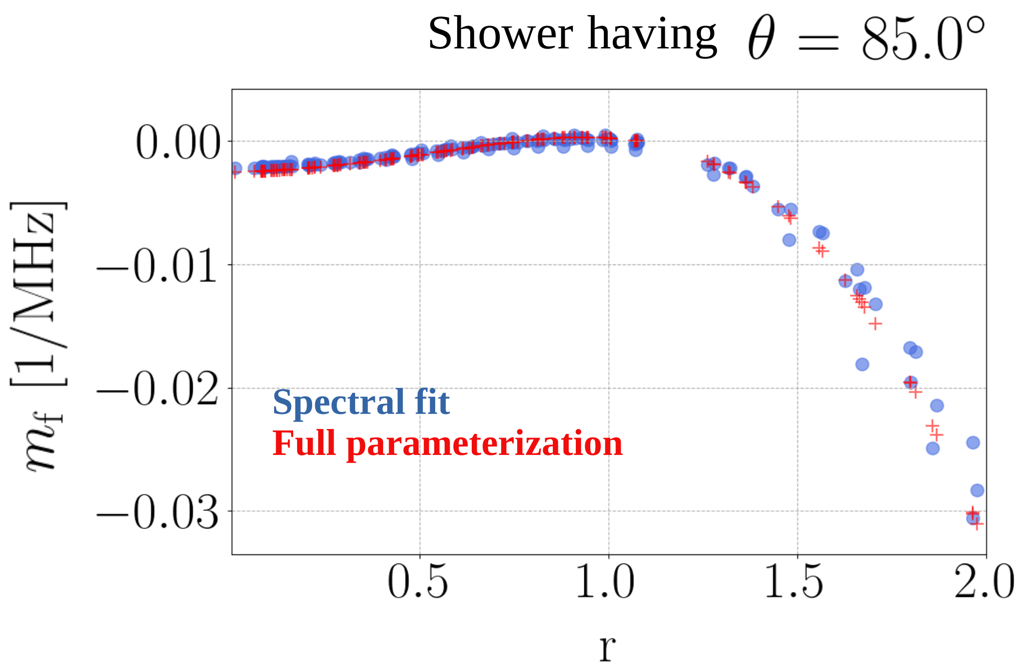




Flattening of the spectrum → shorter pulses and coherence

Slope lateral distribution: fit vs parameterized values





Parameterization: Geomagnetic frequency slope

$$f(r) = A_1 \cdot \left[- e^{B\cdot(r-r_0)} + A_2 \cdot e^{-C\cdot(r-r_0)^2} \right]$$

$$r_0\,(d_{\max}) = \frac{-3.558}{d_{\max}} \cdot \log(d_{\max}) + 1.738$$

$$\frac{B(d_{\max})}{C\,(d_{\max})} = \frac{-7.078}{d_{\max}} \cdot \log(d_{\max}) + 1.625$$

$$A_2(d_{\max}) = \frac{3.468}{d_{\max}} \cdot [\log(0.2335 \cdot d_{\max}) + 0.4805\,\mathrm{km}] + 0.5863$$

$$C\,(d_{\max}) = 0.9985^{-d_{\max}/\mathrm{km}} - \frac{0.2155}{\mathrm{km}} \cdot \log(d_{\max} - 11.69\,\mathrm{km}) + 0.6492$$

$$A_1(d_{\max}) = \left[\frac{-0.3792}{d_{\max}} \cdot \log(0.2008 \cdot d_{\max}) + \frac{4.057 \cdot 10^{-5}}{\mathrm{km}} \cdot d_{\max} + 0.0359 \right] \mathrm{MHz}^{-1}$$

Xmax dependence