FAST Reconstruction:

- A) Top-down Reconstruction
- B) Neural Network Reconstruction

A) Top-down Reconstruction

- Traditional air shower reconstruction techniques use **a bottom-up approach**
- Two-step approach:
  - 1) **Fit for the shower geometry** (zenith, azimuth and core position) using a track o triggered pixel across the camera (only 2 parameters are extracted from each PMT => the total signal and centroid time of each pixel)
  - 2) **Fit a Gaisser-Hillas** profile to the energy deposited as a function of slant depth

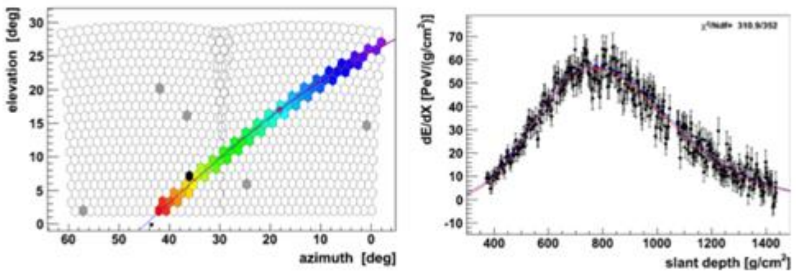

Figure 7.1. Left: Track of triggered pixels across the camera of the Pierre Auger Observatory (Auger) Fluorescence Detector (FD). Right: The Gaisser-Hillas profile fit to the energy deposited as a function of atmospheric depth.

- **The top-down algorithm** attempts to derive information about air shower parameters using **a maximum likelihood technique**

1) Reconstruction Technique:

- A maximum likelihood **comparison between measured data and semi-analytical simulation**
- The reconstruction algorithm uses all available information at the level of individual **PMT traces** in order to simultaneously reconstruct **the shower geometry, Xmax and energy**

Likelihood Function:

- The reconstruction of air shower is specific case of the more generalized problem of estimating a set of $\vec{a}$ **given a set of measured values** $\vec{x}$
  - The parameters $\vec{a}$: **the shower geometry** (θ, ϕ, x, y), **atmospheric depth of shower maximum, Xmax and energy**
  - **Maximazing the likelihood function**: $\mathcal{L}(\vec{x}\,|\vec{a}\,) = \prod_{i}^{N} P(x_i|\vec{a}\,)$
  - The likelihood function **is maximized when the predicted (simulated) signals best match the measured signals** => log-likelihood: $\ln \mathcal{L}(\vec{x}\,|\vec{a}\,) = \sum_{i}^{N} P(x_i|\vec{a}\,)$
    - (Due to the complexity of the likelihood function and its simulation requirements, it is not possible to minimize -lnL analytically -> numerical methods must be employed in the minimisation)

Signal Uncertainly

- **The probability of observing a signal** within an individual time bin is dependent on the **expected signal and backgroung fluctuations**
  - **The night-sky background**
  - The variance of the noise in early time bins of the measured trace can be used as an estimate fot the total backgrounf noise measured by a PMT
  - **The total background:** $\sigma^2_{\text{bkgd}} = n_{\text{bkgd}}\,(1 + V_g) + \sigma^2_{\text{elec}}$
    - n…the mean number of photoelectrons measured from the NSB, V…the gain variance of the PMT, σ…the electronic noise

Event Likelihood Function

- The total event log-likelihood function is given as the **sum of probabilities over all pixels and signal bins**: $\ln \mathcal{L}(\vec{x}\,|\vec{a}\,) = \sum_k^{N_{\text{pix}}} \sum_i^{N_{\text{bins}}} P_k(x_i|\vec{a}\,)$
  - P…the probability of measuring a signal of x photoelectrons in the ith time bin of pixel k

Parameter Uncertainties

- The statistical uncertainties based on the **1σ likelihood contours**
- The value for which the likelihood function must increase by in order to change a parameter **by 1σ depends on the number of parameters (degrees of freedom)** which are being reconstructed simultaneously (two-parameter reconstruction -> **two degrees of freedom (Xmax and energy)**

2) Simulation of the FAST telescope:

- The FAST reconstruction requires a realistic simulation of the detector response to enable a maximum likelihood comparison at the level of individual PMT traces to reconstruct the shower parameters

- FAST-sim: **C++, Auger Offline software framework** (a modified version)

The modules:

- **FASTProfileSimulator**: produces an **analytical Gaisser-Hillas profile** based on the input shower parameters, and subsequentrly calulates the energy deposited in buns of atmospheric depth $f_{GH}(X) = \left(\frac{dE}{dX}\right)_{X_{\max}} \left(\frac{X - X_0}{X_{\max} - X_0}\right)^{\frac{X_{\max}-X_0}{\lambda}} e^{-\frac{X - X_{\max}}{\lambda}}$
- **FASTEventGenerator**: configures **the shower timing, geometry, and coordinate system-related parameters** including the input core position of the simulated shower

```
<!-- A sequence for FAST shower simulation -->
<sequenceFile>
  <enableTiming/>
  <moduleControl>

    <loop numTimes="1">

      <module> FASTProfileSimulatorCG    </module>
      <module> FASTEventGeneratorCG      </module>
      <module> ShowerLightSimulatorCG    </module>
      <module> FASTSimulatorCG           </module>
      <module> FASTEventFileExporterUA   </module>

    </loop>

   </moduleControl>
</sequenceFile>
```

Code 7.1 An example sequence file for the simulation of a single FAST event.

- **ShowerLightSimulator**: **the number of fluorescence photons** at the shower track is calculated using AIRFLY fluorescence model
  - The Cherenkov light contribution is also calculated based on the number of shower electrons above the Cherenkov threshold in air
- **FASTSimulator**: the FAST detector simulation procedure
  - **Propagating photons** through a parametrized molecular and aerosol atmosphere to each FAST telescope, taking into account the wavelength-dependent atmospheric transmission
- **FASTEventFileExporter**: the results of the simulation are written into the **FASTEventFile** format using this module

3) Reconstruction Algorithm:

The modules:

- **FASTEventFileReader**: processes a FASTEventFile and extracts the FASTEvent form the input ROOT file (the input event can either be a simulated event generated from the FAST-sim program or a real measured event)
- **FASTTopDownReconstructor**: is the main processing module of the reconstruction program
- **FASTEventFileExporter**

```
<!-- A sequence for FAST shower reconstruction -->
<sequenceFile>
  <enableTiming/>
  <moduleControl>

    <loop numTimes="unbounded" pushEventToStack="yes">

      <module> FASTEventFileReaderUA       </module>
      <module> FASTTopDownReconstructorUA  </module>
      <module> FASTEventFileExporterUA     </module>

    </loop>

   </moduleControl>
</sequenceFile>
```

Code 7.2 An example sequence file for the reconstruction of FAST events.

Pixel Calibration

- The steps of the reconstruction procedure:
  - 1) The conversion form ADC counts to **photoelectrons in order to compare with the output of the detector simulation**
  - 2) The baseline is subtracted from the trace of each pixel
    - **Fluctuations around the baseline are a combination of both NSB and electronic noise**

Likelihood Search

- Once the measured traces have been calibrated, they can be compared with simulated traces
- **The minimization procedure: the measured trace is compared to simulated traces from many variations of the shower parameters**

Absolute Time Offset

- During the reconstruction process, **the absolute time offset between the data and simulated traces must be determined**
- The timing (and trigger) of each event is **dependent on background fluctuations**
- During the reconstruction procedure**, the best-fit absolute time offset** is determined for each iteration of the minimization

4) Testing the Reconstruction Algorithm on Simulated Events:

- A dataset: 1000 events with fixed energy of $10^{19,5}$ eV, Xmax: EPOS-LHC hadronic model, the arrival directions are sampled from a realistic sinθcosθ distribution, the core positions are sampled uniformly within a 10 km radius of the center of a triangular cell of FAST station

Xmax Reconstruction

- The resolution will significantly degrade as additional shower parameters are added to the reconstruction, and the first of the shower parameters are shifted their true values

The pull distribution: $\mathrm{Pull}(X) = \frac{X_{\mathrm{rec}} - X_{\mathrm{true}}}{\sigma(X_{\mathrm{rec}})}$

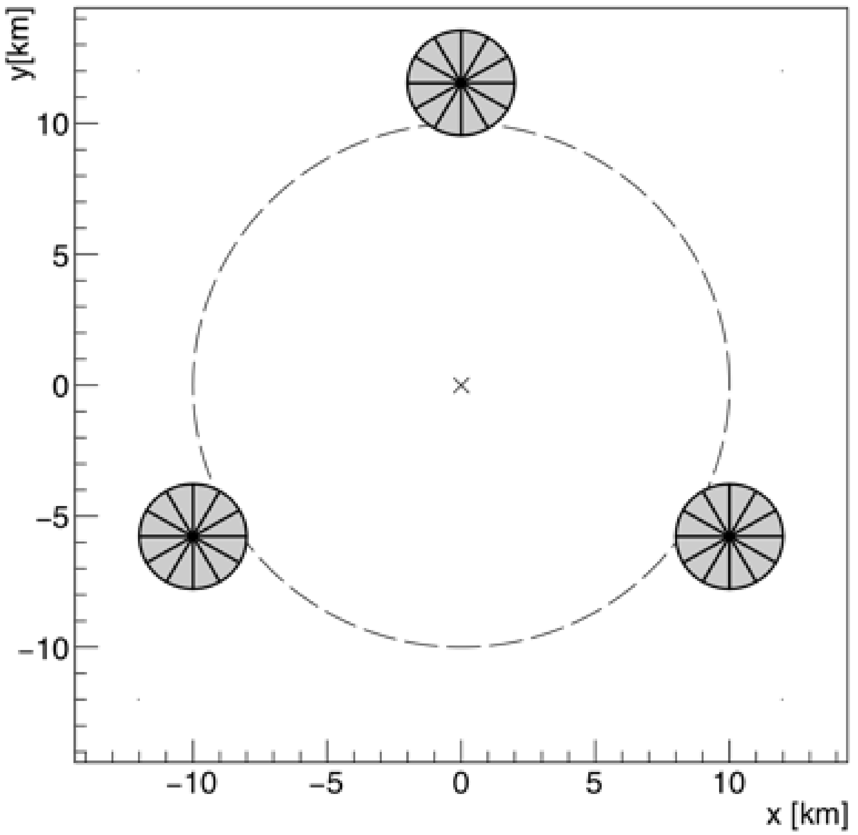

Figure 7.7. The configuration of the simulated FAST stations. The dashed line represents the region within which core positions of simulated showers are sampled uniformly, and the cross represents the centre of the cell.

Geometry Reconstruction

- **The energy and Xmax: the Gaisser-Hillas profile**; other parameters are expected to be more complex due to the require transformation from signal along the shower axis to camera response as a function of time
- **The shower geometry has a direct influence on the camera response as a function of time** since a change in the position or orientation of the shower axis will change the time the signal arrives at the telescope aperture

Hybrid Reconstruction

- **FAST can operate in two distinct modes**
  - 1) **The full reconstruction using a triangular array of FAST station** which are able to measure the shower stereoscopically
  - 2) **An independent reconstruction of the shower geometry using a SD array**
    - FAST telescopes are only responsible for the reconstruction of **Xmax and the shower energy**

- FAST prototype telescopes at TA: 3 FAST telescopes, an azimuthal FoV of 90°, the shower energy is fixed to $10^{19,5}$ eV, Xmax: EPOS-LHC hadronic model, top-down reconstruction model
  - Very little bias, Xmax resolution: 23 g/cm^2, energy resolution: 7%

Full Reconstruction

- Energy: $10^{19,5}$ eV, Xmax: 750 g/cm^2, zenith angle: 30°, Gaisser-Hillas shape parameters: X_0: -121 g/cm^2, λ: 61g/cm^2

Parameter	Simulated Value	First Guess	Reconstructed Value
$X_{\max}$ [g/cm^2]	750	750	751.8±9.6
Energy [EeV]	31.6	31.6	31.2±0.7
Zenith [deg]	30	30	31.2±0.3
Azimuth [deg]	50	50	49.8±0.8
CoreX [m]	500	500	516.0±45.6
CoreY [m]	-500	-500	-515.9±34.4

Table 7.2. Summary of the simulated and reconstructed shower parameters for an example simulated event where the true parameters are passed as the first guesses.

B) Neural Network Reconstruction

- A neutral network has been designed to reconstruct extensive air shower (EAS) properties from measured parameters of the FAST detector. Three parameters are input to the neural network per photomultiplier tube (PMT)
  - *Centroid time* (signal-weighted time average) – provides information about the relative time of arrival of signals at each PMT
  - *Total signal* – provides information about total signal measured from the shower as well as the relative signal between PMTs
  - *Pulse height* – provides additional information about the shape of the signal pulse including asymmetry

2 Machine Learning and Neutral Networks

- During training of the neural network, the weights are varied in order to **minimize the difference between the predicted and true parameters**
- **Each neuron in a given layer is connected to all neurons in the previous layer** (a dense or fully connected layer): $\vec{y} = W \cdot \vec{x} + \vec{b}$ $\quad \vec{y} = f\left(W \cdot \vec{x} + \vec{b}\right)$
  - **x represents the input form the previous neuron**, **W is the weight connecting the neurons**, **b is a bias term**, **y represents the values of the next layer of neurons**, **f represents a given activation function**
- The neutral network must predict continuous numerical values for each of the shower parameters -> a rectified linear unit (ReLU) activation function
- The predicted output parameters of a neutral network are compared to the desired output parameters **using a so-called loss function**
- Training a neural network is essentially **a maximum likelihood problem**

Detector Simulation

- **The three** most significant stations will provide adequate descriptions of the event for **a first guess of the shower parameters**
- **The top-down reconstruction** can then include all information from any additional triggered FAST stations
- The FAST telescopes are simulated using **the FAST-sim detector simulation**
- The total background noise, and the electronic noise were determined using data measured with the FAST prototype telescopes at TA
  - **The electronics noise** is determined by recording background data with **the shutter closed**
  - **The background noise** is dominated by the NSB with **the shutter open**

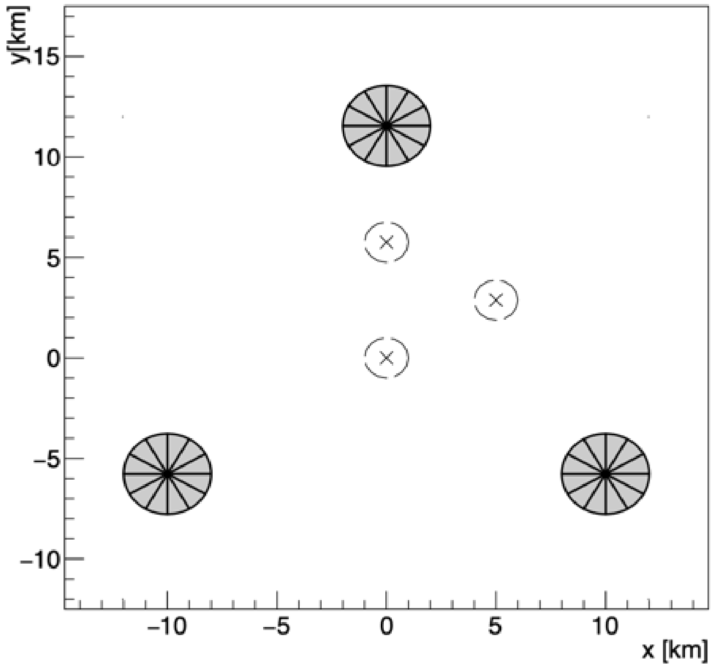

Figure 8.2. The configuration of the simulated FAST stations used for the training and testing of the neural network reconstruction. The dashed lines represent the regions within which core positions of the simulated showers are sampled uniformly, and the crosses represent their centres.

Simulated Training Data Set

- Important to provide uniform sampling of the desired output parameters
- **To determine the range of Xmax to sample, it is useful to consider the predictions of hadronic models for pure proton and iron distribution at the limits of the energy range of the simulations**
  - Pure iron 10^{18} eV (1EeV): 650 g/cm^2, pure protons 10^{20} eV (100EeV): 850 g/cm^2
  - The values of Xmax used to train the neural network will be sampled from a uniform distribution between 500 and 1200 g/cm^2
  - => Xmax ~ U(500, 1200) g/cm^2
- **The shower geometry (θ, ϕ, x, y): uniform sampling of the individual parameters would not provide a uniform sampling of the physical phase space**
  - **The shower arrival direction** (zenith and azimuth) **must be sampled a way that hey populate a hemisphere uniformly** => this cannot be achieve by sampling the zenith angle uniformly, since an infinitesimal change in azimuth depends on the zenith angle -> the solid angle is smaller at the vertical (small zenith angle)

- **The core position is sampled uniformly inside a circle, rather than from x and y independently**
  => a uniform sampling over the entire area of the circle

- **The angle is sampled uniformly:** $\theta \sim U(0, 2\pi)$

  - A circle with center x0, y0; radius R, angle θ, polar distance r

- The probability of sampling a point within a distance r of the centre: $P(r) = \frac{r^2}{R^2}$

$$r = R\sqrt{x}$$

Network Architecture and Training

a) Pre-processing

- The parameters of characterize the pixel pulses: total signal, pulse height, centroid time

- **FAST pixel trace:** each trace -> the start and stop times which maximize the signal-to-noise ratio (SNR) in a given pixel i: $\left(\frac{S}{N}\right)_i = \frac{S_i}{\sigma_i \sqrt{k_{\text{stop}} - k_{\text{start}}}}$
  - Si…the signal measured in the ith pixel, σ_i…standard deviation of the background noise (the minimum pulse width: 300 ns = 3 time bins)

Signal Traces

- **The total signal S:** $S = \sum_{j=k_{\text{start}}}^{k_{\text{stop}}} s_j$
  - The signal measured by these **PMTs have a large dynamic range** (important to consider the effect of extremely large input parameters to the neural network)
- **The normalized total signal:** $\hat{S}_i = \frac{S_i}{S_0}$
  - **S0…the average total signal of the entire training data set**

Pulse Height

- The pulse height: **the maximum signal of all 100 ns bins within the pulse window**
- **The average pulse height** of the entire training data set: $\hat{h} = \frac{h_i}{h_0}$

Arrival Time

- Standard FD reconstruction algorithms utilize **-> the arrival times of the shower signal in each pixel to estimate the shower geometry** (the FAST camera only contains four PMTs per telescope)
- **An estimate of the signal arrival time:** $\bar{t} = \frac{\sum_{j=k_{\mathrm{start}}}^{k_{\mathrm{stop}}} s_j t_j}{\sum_{j=k_{\mathrm{start}}}^{k_{\mathrm{stop}}} s_j}$
    - Sj…the signal in units of photoelectrons in the jtg trace bin, tj…the midpoint of the jth trace bin
- **The normalized centroid time:** $\hat{t_i} = \frac{t_i - t_0}{\sigma_t}$
    - This normalization of identically applied to the training data and to the independent validation and test data sets

b) Pulse Trigger Threshold

- To reduce such noise, only **pulses above some pre-defined significance threshold** should be included in the training and subsequent reconstruction
- **Apply a high significance threshold cut to the pixel pulse** => **this can have the effect of removing many (useful) pulses from the event** => a balance must be found between maintaining good quality signals, and a large enough quantity of signals
- The simulation: 100,000 events; for each of the 144 pixels in an event – calculate its maximum SNR
  - The noise distribution: a mean of 3.5 σ
  - Above 6 σ the distribution purely consists of signal pulses (a significant number of pulses with signal exist below 6 σ, many of which are potentially worth including in the reconstruction
  - The peak of the signal distribution approximately matches the peak of the noise distribution
  - The bulk of these pulses are signals which are well below the background fluctuation => it is impossible to recover any useful information from them

- At approximately 5 σ the noise pulses have significantly fallen in number to the point where it is expected that a pulse is more likely to actually contain signal than not

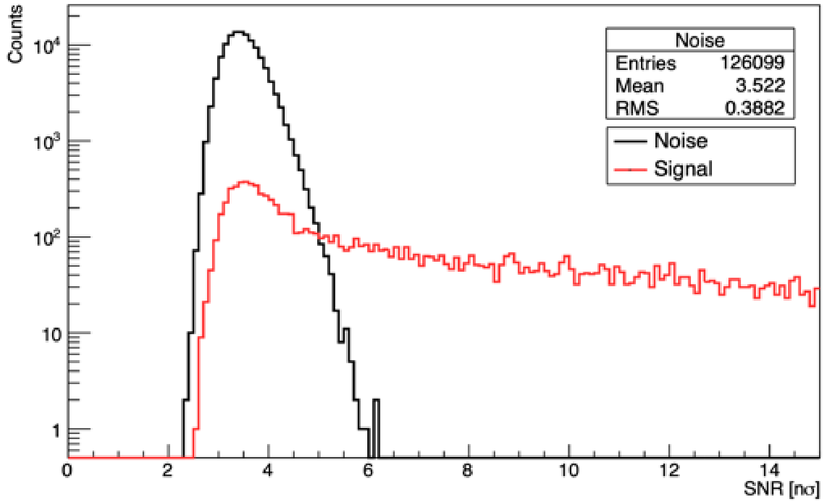

Figure 8.3. The maximum SNR for signal (red) and noise (black) pulses from simulated FAST events. Since only the maximum SNR of each pulse is included, values smaller than approximately 2σ were not obtained in this study.

c) Network Architecture

- A neural network with here hidden layers before being passed to the output layer **which predicts the six shower parameters**

d) Training

- 500,000 events have been simulated in each of **the three core regions**
- The simulated data sets are separated into two samples each: **80% for the actual training, and 20% for validation**
- The mean squared error: $\mathcal{L} = \frac{1}{n}\sum_{i}^{n}(x_i - y_i)^2$
  - Xi and yi…the predicted and true value of arch of the n=6 shower parameters: Xmax, energy, θ, ϕ, x, y
- **The neural networks are trained for up to 1000 epochs**

e) Output Parameters

- **Each of the output shower parameters are normalized** so that the predicted output parameters approximately lie in the range [-1, 1]
- **The transformation**: $x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$
  - x…represents each of the shower parameters, xmin and xmax…the minimum and maximum values of each parameter in the entire training data set
- The reverse normalization: $x_{\text{out}} = x_{\text{min}} + (x_{\text{max}} - x_{\text{min}})\, x_{\text{norm}}.$
  - The core position of each event is normalized using fixed values of xmin = 0 m, and xmax = 12000 m
- The output arrival direction (two parameters: zenith angle [0, pi/2] and azimuth [0,2pi])

$$x = \sin\theta\cos\phi$$
$$y = \sin\theta\sin\phi$$
$$z = \sqrt{1 - x^2 - y^2}$$

  - r = 1; x and y…predicted by the neural network, z…can be calculated

Reconstruction Performance

- A data set: 10000 events; the simulated events are sampled uniformly in Xmax and energy in order to test the reconstruction over the full training range
- **The reconstructed values of each parameter are compared to their true values**
    - The neural network is able to predict the desired values as indicated by the strong correlation between the true and reconstructed shower parameters
- The simulation: 10 EeV to 100 EeV in steps of 10 EeV; the arrival directionas are sampled form a realistic $\sin\theta\cos\theta$ distribution; the core positions are sampled uniformly within the three regions

Xmax Bias and Resolution

- The reconstruction bias: $\Delta X_{max} = X_{max}^{rec} - X_{max}^{true}$
  - The Xmax bias **is smallest for the central core position** with a maximum bias of 5 g/cm^2
- The Xmax resolution**: indicates a strong energy dependence** with resolution of 70g/m2 at 10 EeV -> small 25g/cm2 at the highest energies
  - **The central position has the best resolution, followed by the right position, and finally the upper(?) position**
    - This dependence on core position can be explained by the average distance to the three FAST stations

Energy Bias and Resolution

- The energy reconstruction bias and resolution as a function of energy for the three core positions => There is an energy dependent reconstruction bias (**largest for the right core position**)

Arrival Direction Bias and Resolution

- **The bias is very small:** except for the lowest energy bin of the right core position
- **A small core position dependence in the zenith angle resolution**, with the central position again performing the best and the upper position performing the worst