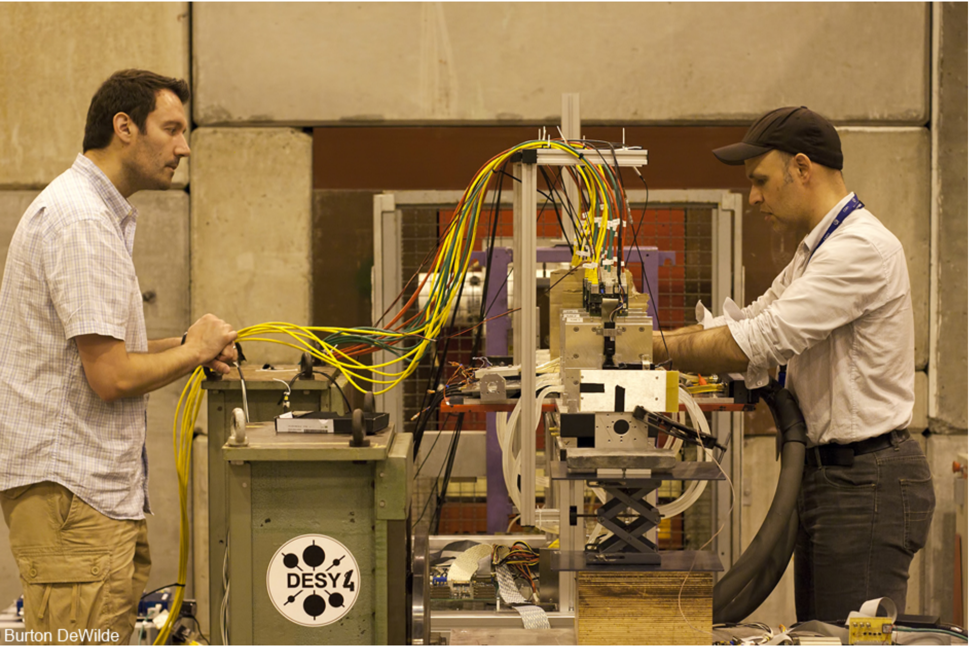
# Statistics and track reconstruction.

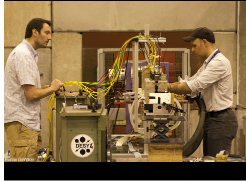
Håvard Gjersdal
hgjersdal@gmail.com

18. januar 2015




2015-01-18

# DESY-2015



- For my thesis, I worked on test beam characterization of ATLAS 3D pixels
- Analysis of the performance of ATLAS 3D pixels
- Straight line track reconstruction and alignment
- Development of new methods for improving reconstruction

# Test beam reconstruction

Raw data

↓

Set of hits

↓

Noise filtered set of hits

↓

Clusters

↓

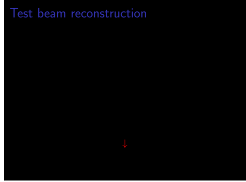
Measurements

↓

Tracks







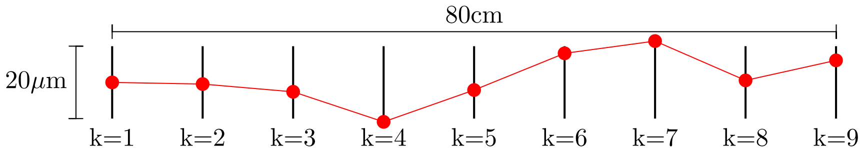

- Reconstruction of test beam data consists of several steps.
- This talk will focus on the final red arrow, going from measurements to a mathematical description of particle tracks.
- Some theoretical background for track fitting will be given. Mainly discuss algorithms based on the Kalman filter, not GBL. Straight lines, not B-field. A lot of it still applies.
- I will also discuss alignment and estimation of material and resolutions.

# Track fitting

## Goal

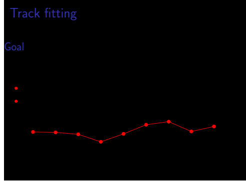
Get best possible estimate of particle positions and angles in device under test planes

- Least squares estimation of parameters
- Straight lines through air, scatter in planes







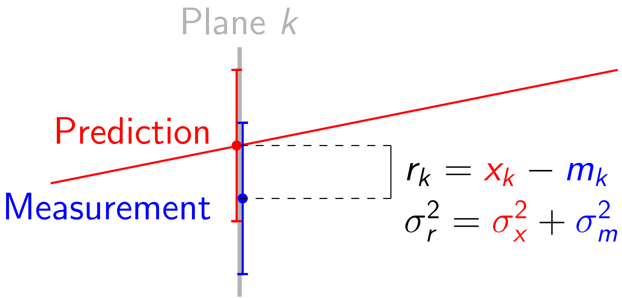




- Standard deviation calculated from Highland formula, from thickness of plane and energy of the particle.
- Assumed to be Gaussian, but in fact it is not. Not a problem at high energies, but the tails that extend further than for a Gaussian might be a problem at lower energies.

2015-01-18

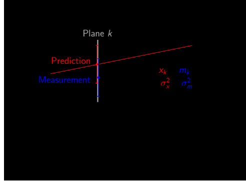
DESY-2015

# Background

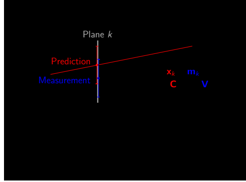


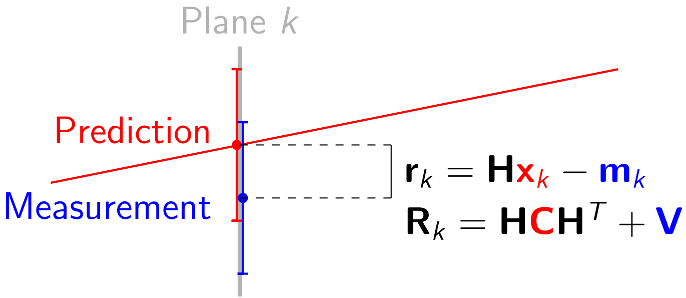

$$r_k = x_k - m_k$$
$$\sigma_r^2 = \sigma_x^2 + \sigma_m^2$$

Prediction $(x, \sigma_x^2)$
Measurement $(m, \sigma_m^2)$
Residuals $(r, \sigma_r^2)$



- Residuals are the difference between model prediction and the measurement.
- Uncertainties are available for the prediction and the measurement.
- Pred and meas uncertainties are used to estimate the uncertainties in the residuals.

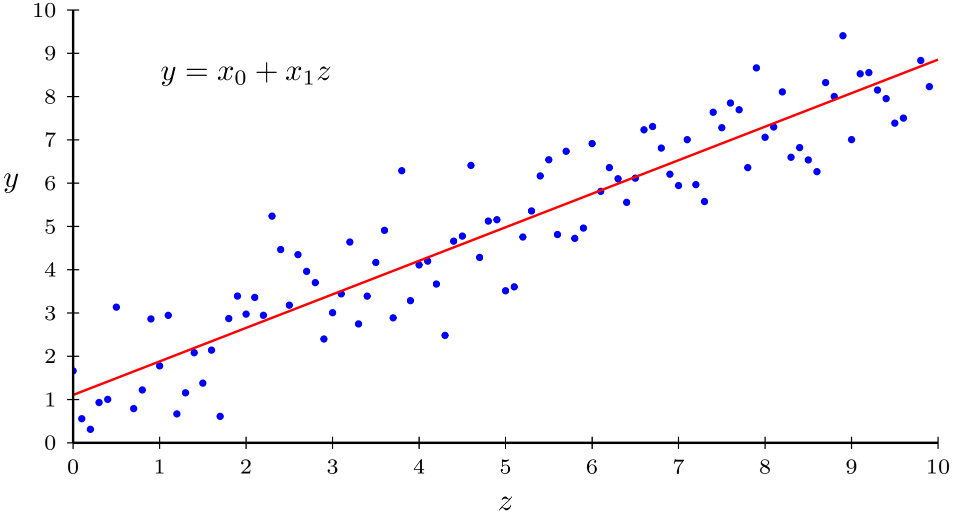





Prediction $(\mathbf{x}, \mathbf{C})$
Measurement $(\mathbf{m}, \mathbf{V})$
Residuals $(\mathbf{r}, \mathbf{R})$

- Measurements are generally one or two dimesional.
- Predictions are four or five dimensional
- Uncertainties are described by covariance matrices.
- The matrix **H** projects from prediction parameter space to measurement space.

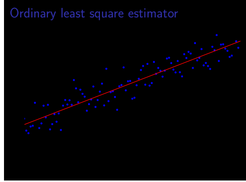
# Ordinary least square estimator




Find the $\boldsymbol{x}$ that minimizes $\sum r^2$








- Residuals are the vertical distance between red line and blue dots.
- OLS minimizes the sum of squared residuals.

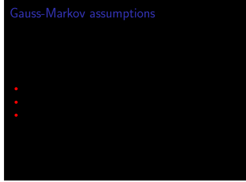
# Gauss-Markov assumptions

The ordinary least squares estimator is the optimal linear estimator if

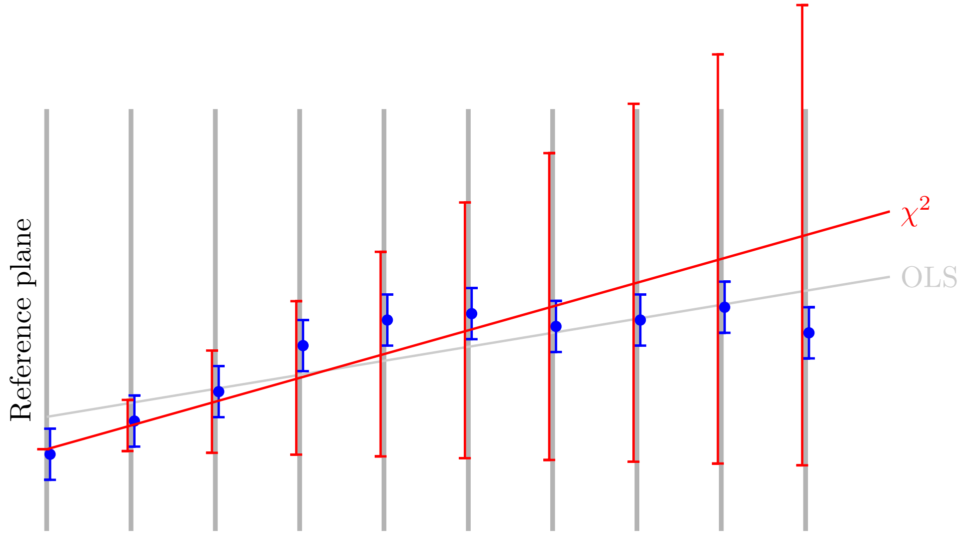
- Measurements are unbiased.
- The uncertainties of all the measurements are the same.
- The measurement errors are uncorrelated.





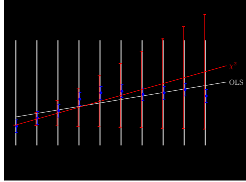



- Optimal means unbiased, minimum variance estimator.




Find the $\boldsymbol{x}$ that minimizes $\chi^2 = \sum (r/\sigma_r)^2 = \sum \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}$





- Particle tracks are broken lines, and the uncertainties of the residuals grow with distance from reference plane due to multiple scattering.
- Blue error bars are measurement uncertainties, red error bars are residual uncertainties.
- In this case OLS does not give good estimates at all.
- A $\chi^2$ fit uses information about residual uncertainties, not only measurement uncertainties.
- Can easily deal with measurements with different uncertainties.

# Gauss-Markov assumptions revised

$\chi^2$-minimization is the optimal linear estimator if

- Measurements are unbiased.
- The uncertainties of all residuals are correctly estimated.
- The measurement errors are uncorrelated.





Gauss-Markov assumptions revised

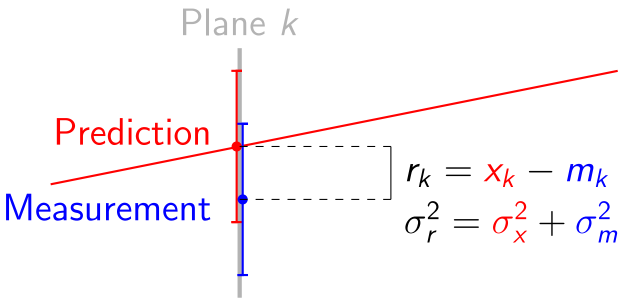
- The uncertainties of all residuals are correctly estimated.

- Unbiased: Alignment must be right. The measurements must be where you think they are.
- Residual uncertainties: One must know the amount of scattering that occurs, and the resolution of the measurements accurately.

# Test statistics

- Test statistics are used to study how well the data fits with the model.
- A test of implementation, track model, and system description.

# $\chi^2$ minimization



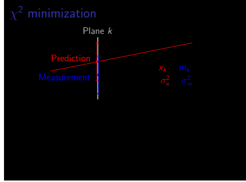

$$r_k/\sigma_r \sim \text{normal distribution}$$

$$\sum^n (r_k/\sigma_r)^2 \sim \chi^2 \text{ with } n \text{ degrees of freedom}$$

$$\sum^n \left(\mathbf{r}_k^T \mathbf{R}_k^{-1} \mathbf{r}_k^T\right)^2 \sim \chi^2 \text{ with } n \times \dim(\mathbf{r}) \text{ degrees of freedom}$$

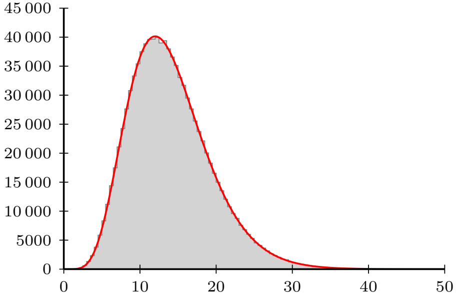


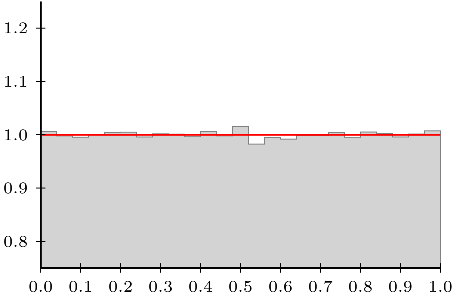




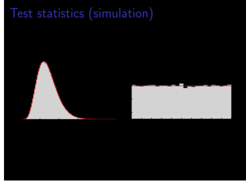

- Track fitting is $\chi^2$ minimization, minimization of the sum of squared normalized residuals.
- If the GM-assumtions hold, and the residuals are Gaussian, the fitted $\chi^2$ should follow a $\chi^2$ distribution.
- ndof $=$ number of measurements ($\times$ 2) - the number of fitted parameters.

# Test statistics (simulation)



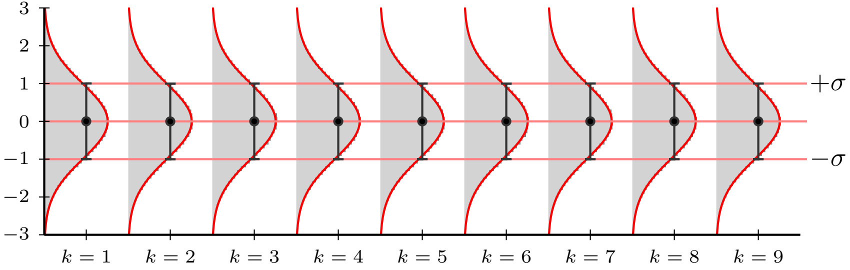


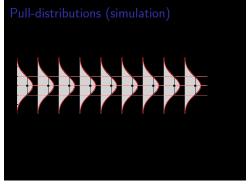



- p-values are the integral of the true $\chi^2$ distribution from the observed $\chi^2$ to $+\infty$.
- If the data is truly $\chi^2$ distributed, the p-values should follow a uniform distribution. Easier to see deviations from model with this visualization.
- Large deviations from the true $\chi^2$-distribution can indicate problems with the track fits, and suboptimal performance.
- Non-Gaussian residual uncertainties will also make the fits deviate from true $\chi^2$, but are not necessarily a problem.

# Pull-distributions (simulation)






└─Pull-distributions (simulation)

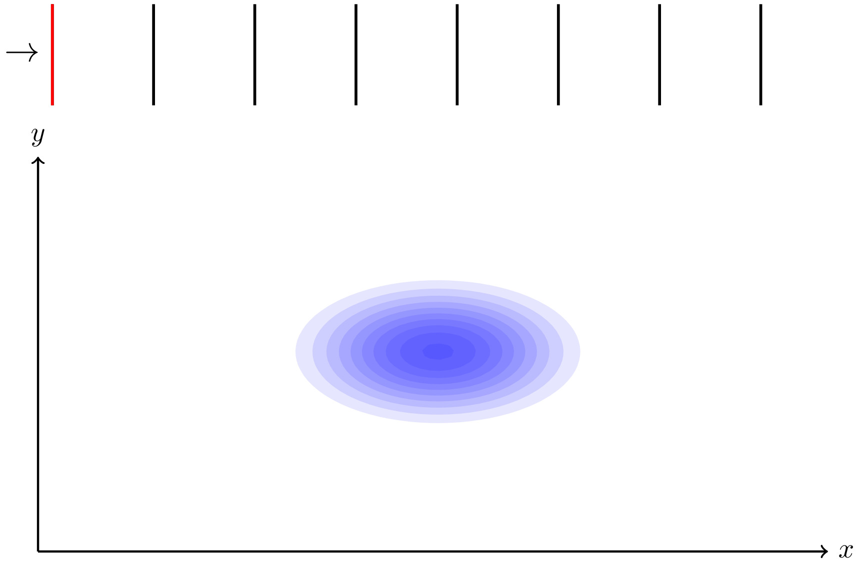
- Pull distributions are simply the normalized residual distributions in the different planes.
- A good tool for looking for problems with specific detector planes.

# Track fitting with the Kalman filter

- Recursive formulation of $\chi^2$-minimization.
- Fast compared to global $\chi^2$ minimization.
- Can be used to build other algorithms for:
  - Track finding
  - Outlier rejection
  - Dealing with non-Gaussian energy loss.

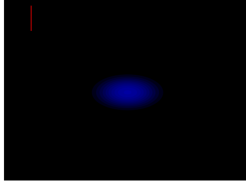



- Mathematics of the Kalman filter are discussed many places.
- Instead of going into much mathematical detail, I will go through a KF track fit, looking at the evolution of the track estimates.

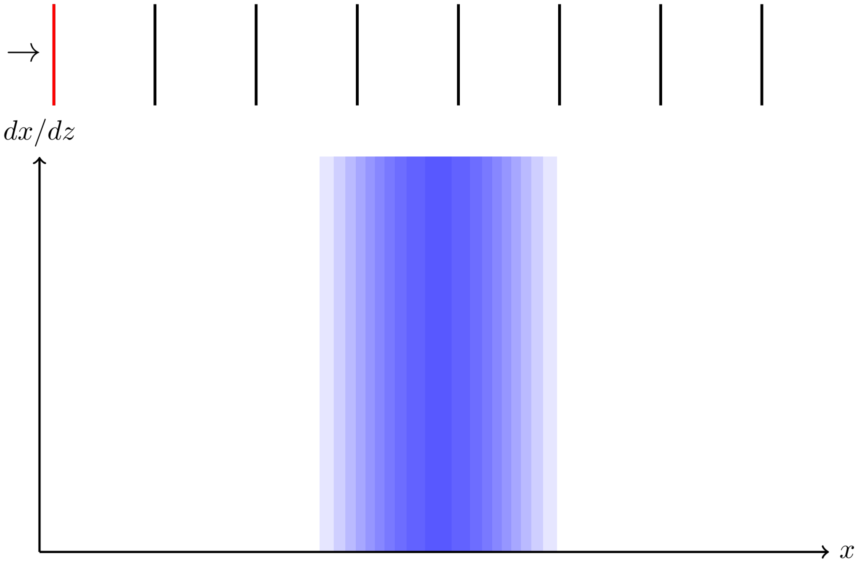


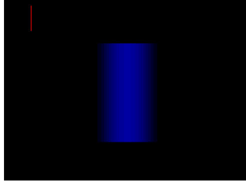

$$\mathcal{N}(\mathbf{x}';\mathbf{x},\mathbf{C}) = \frac{1}{2\pi\sqrt{|\mathbf{C}|}} e^{-\frac{1}{2}(\mathbf{x}'-\mathbf{x})^T\mathbf{C}^{-1}(\mathbf{x}'-\mathbf{x})}$$





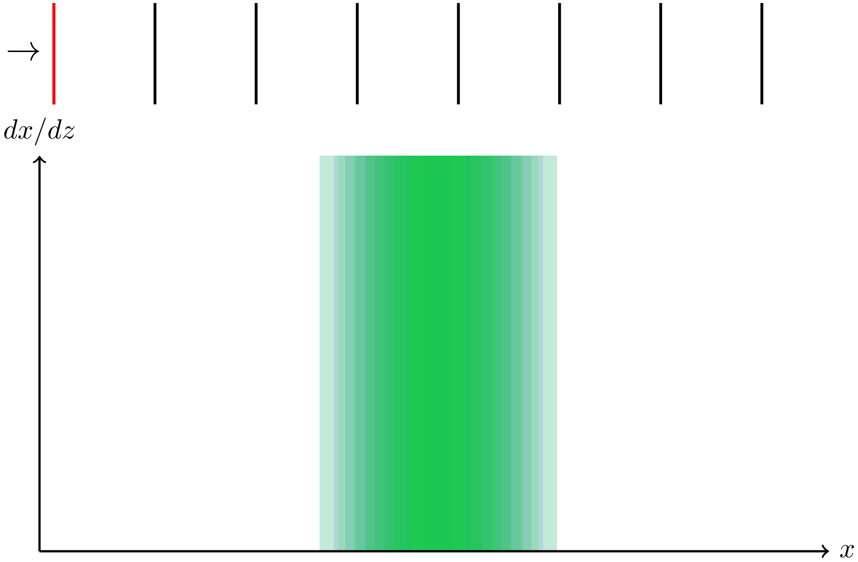
- Here we see the measurement as a two dimensional probability distribution in measurement space/a detector plane.



$$\mathcal{N}(\mathbf{x}';\mathbf{x},\mathbf{C}) \sim \cancel{\frac{1}{2\pi\sqrt{|\mathbf{C}|}}}\, e^{-\frac{1}{2}(\mathbf{x}'-\mathbf{x})^T\mathbf{C}^{-1}(\mathbf{x}'-\mathbf{x})}$$

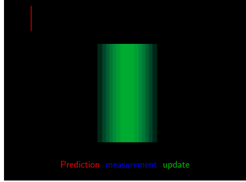
- A more interesting way of viewing the measurement in parameter space, $x$ vs $dx/dz$, position versus angle.
- The measurement has no information about the angle of the track, the variance in the angle direction is $\infty$.



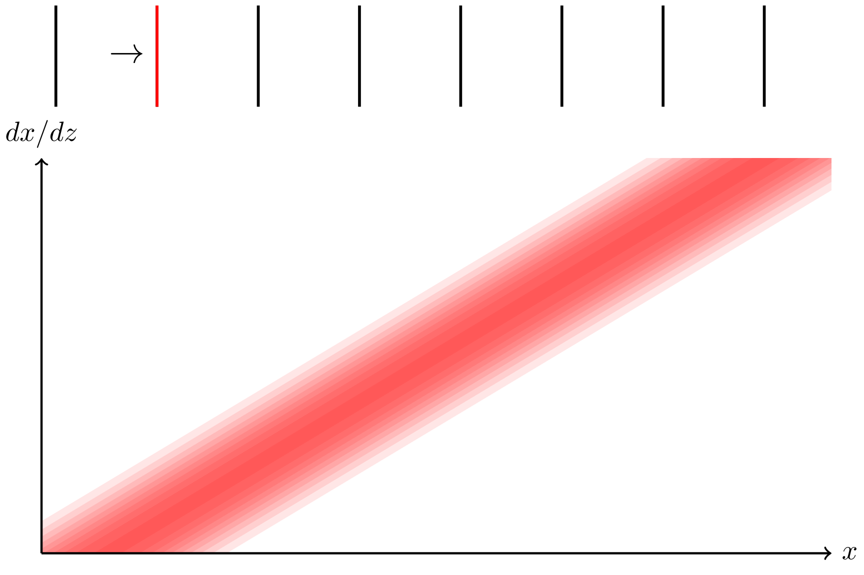

Prediction, measurement, update

2015-01-18

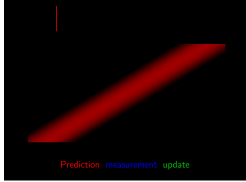
DESY-2015



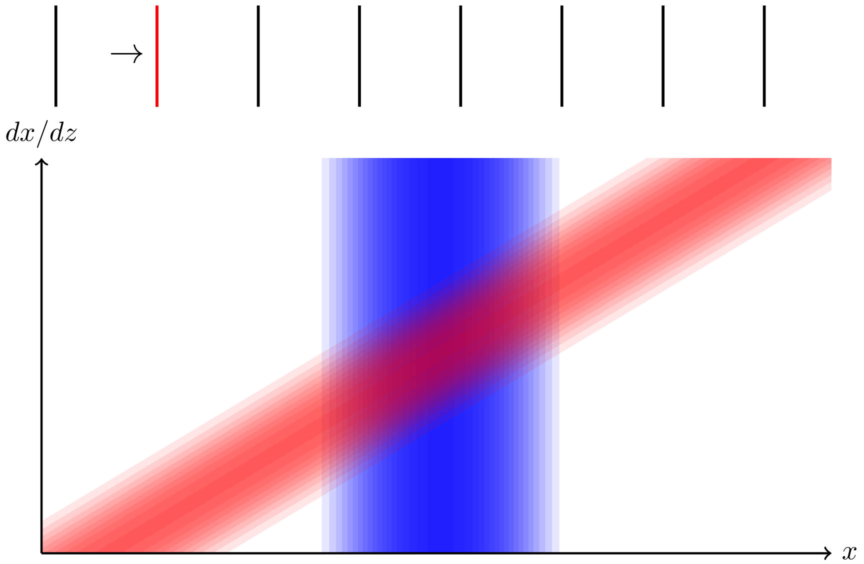

- The first step in the Kalman filter is to copy the information of the measurement in the first plane.




Prediction, measurement, update



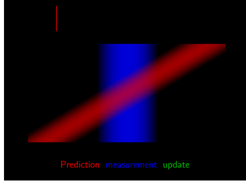

- The next step is to make a prediction in the second plane.
- The resulting prediction has $\infty$ uncertainties in both angle and position
- BUT, it has information about the correlation between the two.
- If the angle points in the positive direction, the measurement will be to the right compared to the first measurement, and vice versa,



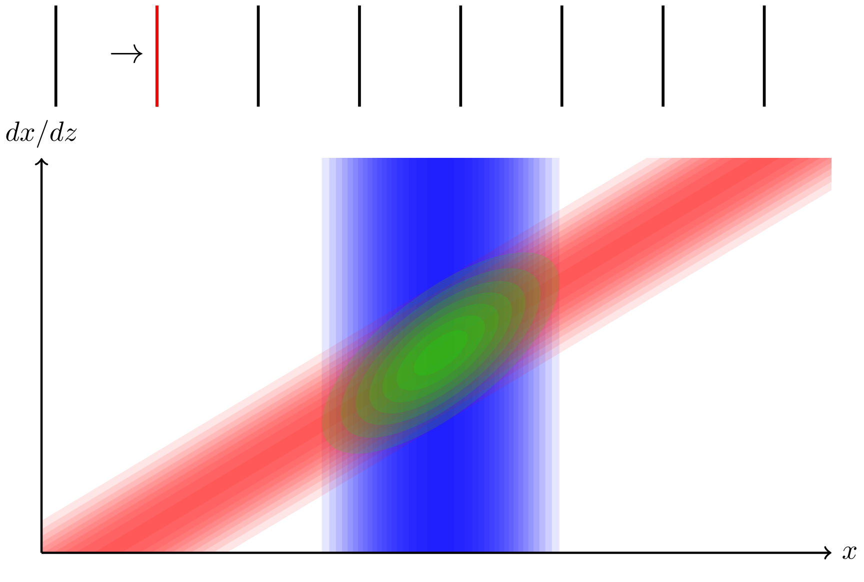

Prediction, measurement, update



DESY-2015

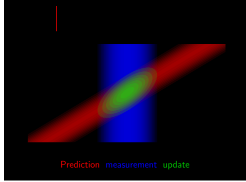



- The measurement is then combined with the measurement in the plane



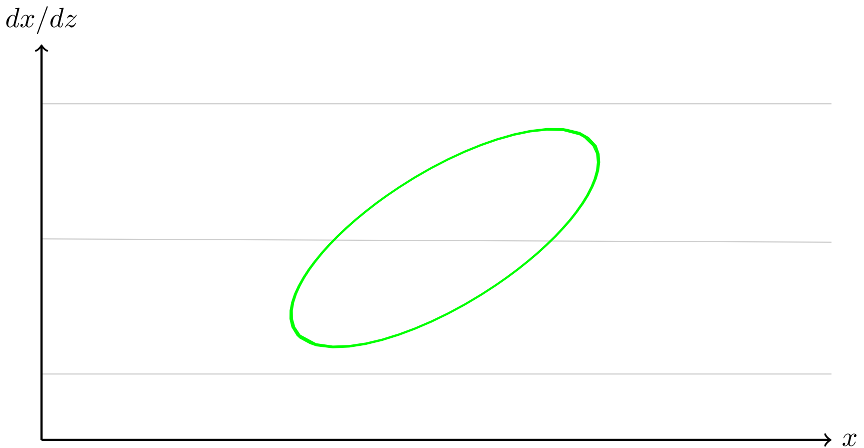

Prediction, measurement, update






- Leading to a measurement with a finite covariance matrix.

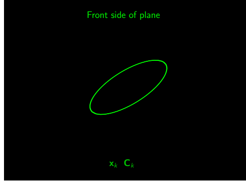
## Front side of plane




$\mathbf{x}_k$, $\mathbf{C}_k$

2015-01-18

DESY-2015



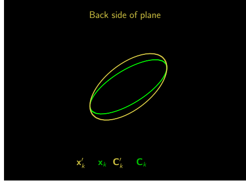

- A closer look at propagation:

# Back side of plane


$$\mathbf{x}'_k = \mathbf{x}_k,\ \mathbf{C}'_k = \mathbf{C}_k + \mathbf{Q}_k$$

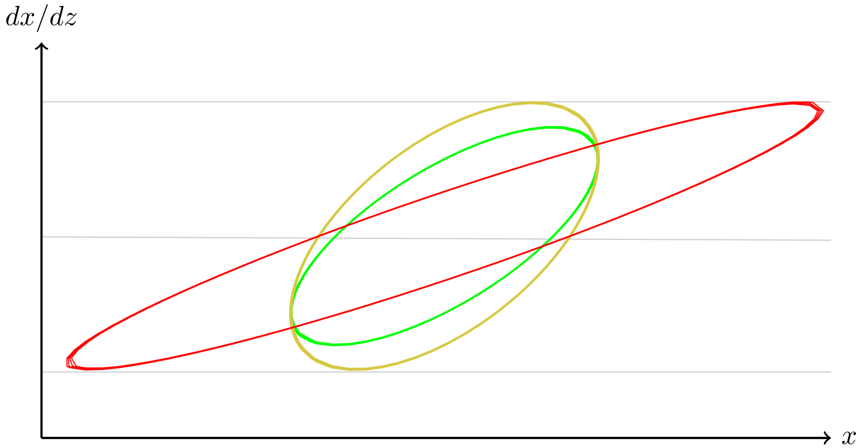
2015-01-18

DESY-2015




- First the estimate is propagated to the back of the plane.
- Scattering can occur. Uncertainties in angle increase.
- Uncertainties in position does not increase (if plane is thin.)

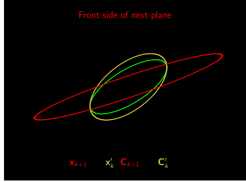
## Front side of next plane



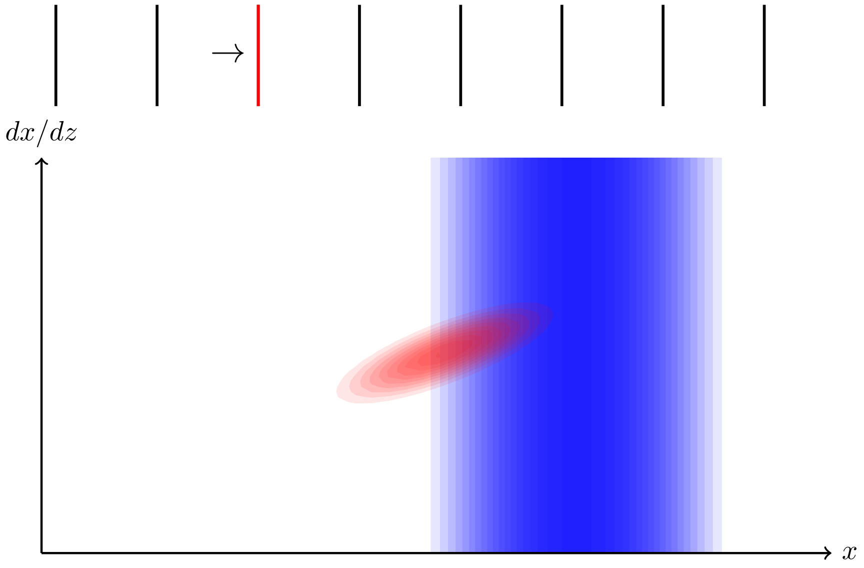

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}'_k,\ \mathbf{C}_{k+1} = \mathbf{F}\mathbf{C}'_k\mathbf{F}^T$$





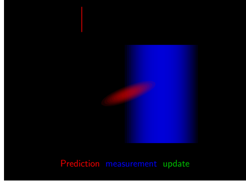


- Straight line propagation through air.
- Angle uncertainties are the same
- Due to initial uncertainty in angle, the uncertainty in position will increase.
- The longer the propagation length, the larger the position uncertainties will be.

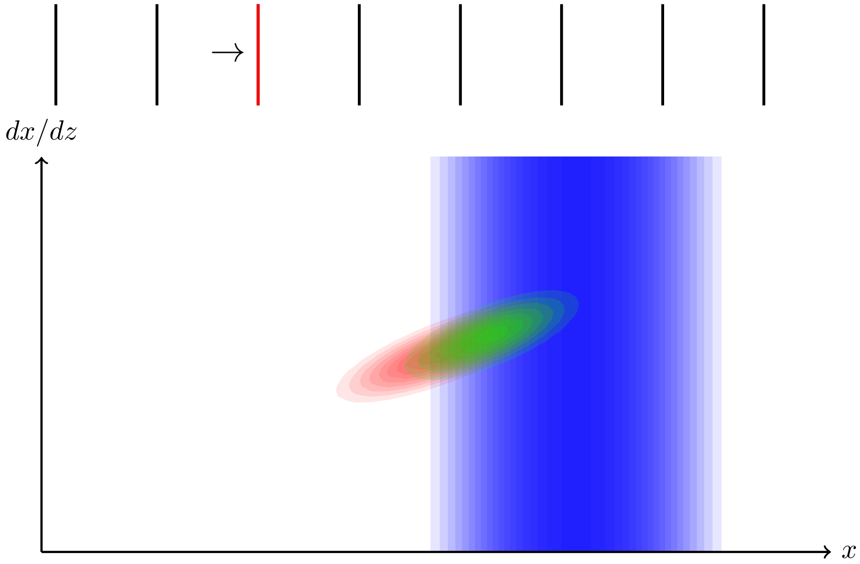



Prediction, measurement, update






- In the third plane, prediction has finite covariance matrix, measurement does not.

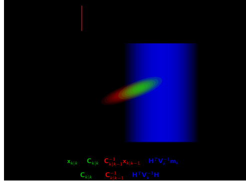



$$\mathbf{x}_{k|k} = \mathbf{C}_{k|k}\left(\mathbf{C}_{k|k-1}^{-1}\mathbf{x}_{k|k-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{m}_k\right)$$

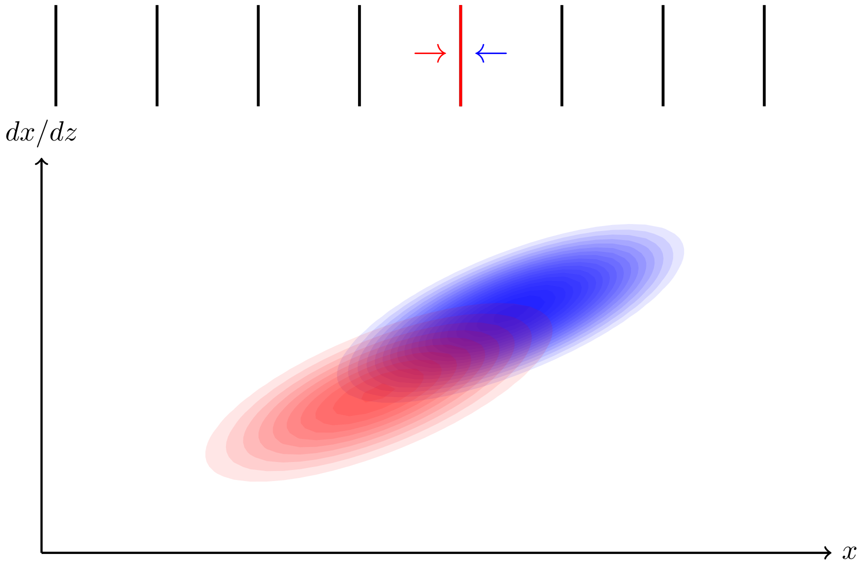
$$\mathbf{C}_{k|k} = \left(\mathbf{C}_{k|k-1}^{-1} + \mathbf{H}^T\mathbf{V}_k^{-1}\mathbf{H}\right)^{-1}$$



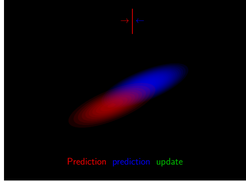




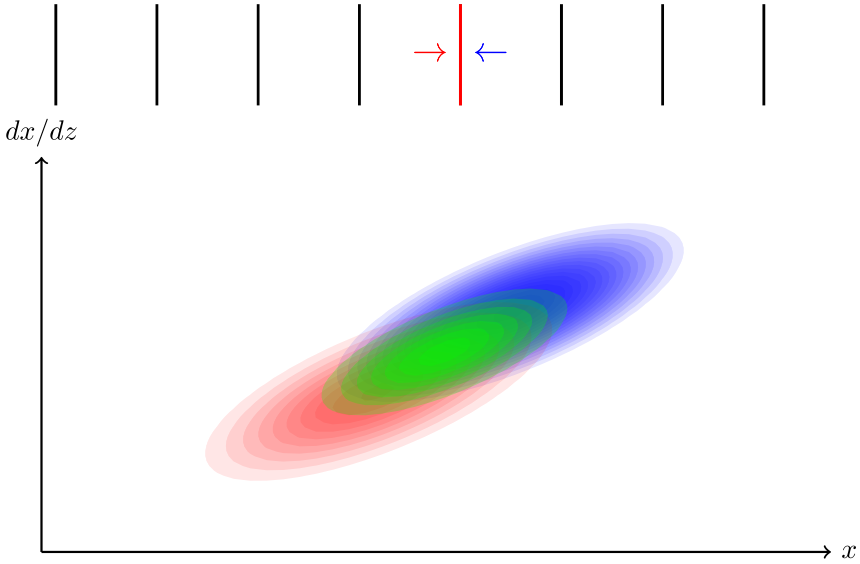
- Combining the measurement in the plane with the prediction looks similar to this in every following plane.
- The maths in the bottom shows the actual algorithm, combining information from the prediction and the measurement.




Prediction, prediction, update




- The Kalman filter estimate is only optimal when all measurements are included, meaning in the last measurement plane.
- To obtain optimal estimate for a plane in the middle of the detector system , information from a Kalman filter running in the forward direction is combined with information from a KF running in the BW direction.

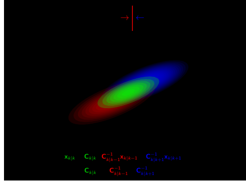



$$\mathbf{x}_{k|k} = \mathbf{C}_{k|k}\left(\mathbf{C}^{-1}_{k|k-1}\mathbf{x}_{k|k-1} + \mathbf{C}^{-1}_{k|k+1}\mathbf{x}_{k|k+1}\right)$$

$$\mathbf{C}_{k|k} = \left(\mathbf{C}^{-1}_{k|k-1} + \mathbf{C}^{-1}_{k|k+1}\right)^{-1}$$





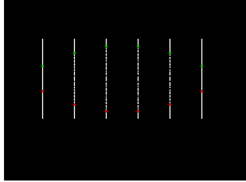


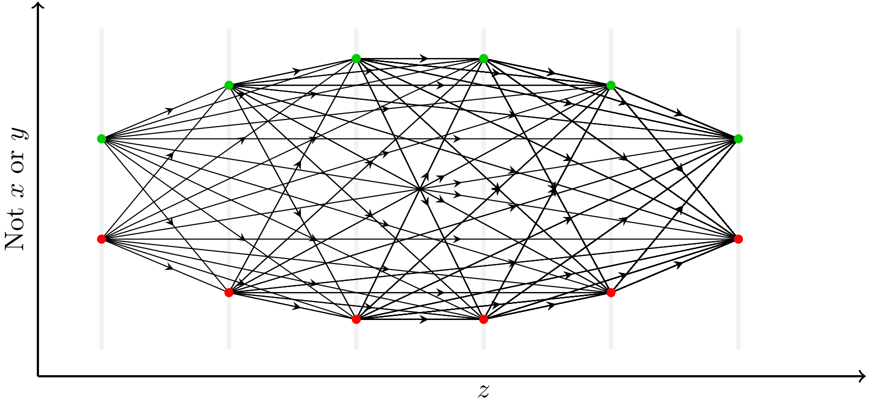
- The result, called the smoothed prediction, contain information from all measurements, except in the current plane.
- The maths shown here are equivalent to the algorithm combining a prediction with an update.
- This is simply the linear algebra equivalent of a weighted average, where the weights are the inverted covariance matrices.

# Track finding

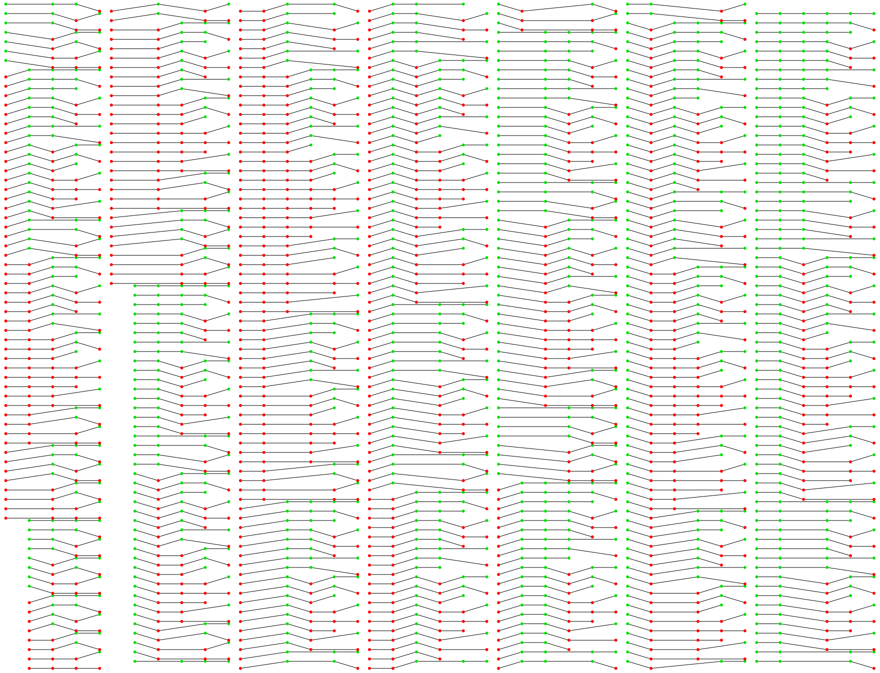


- KF expects a set of meauremets, one or zero per plane, to be fitted.
- Real data contains noise hits, hits belonging to other tracks, detection inefficiencies.
- Identifying sets of measurements created by the same particle is called track finding, and is pattern recognition.



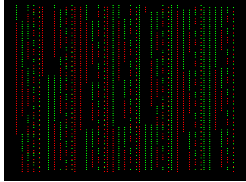



- Graph theory can be used to describe the problem of track finding. It can be seen as a directed acyclic graph.
- Nodes are measurement, in the example two per plane.
- All nodes are connected to all other nodes corresponding to measurements further downstream with an edge (a line).
- A path is a set of connected edges. Any such path can be a track.
- In the graph, the horizontal position is $z$, vertical position is not connected to position, but a measurement index.



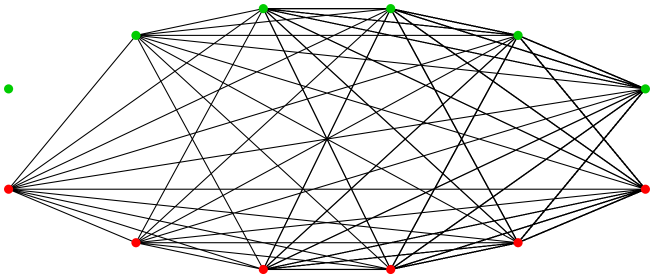
2015-01-18

# DESY-2015

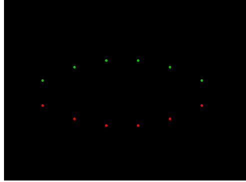


- The graph hides the complexity of the problem.
- Here you see all >400 paths containing 4, 5 or 6 measurements.
- If there was 10 measurements per plane, there would be approx 1M such paths.

A node in the first plane is chosen as a starting point

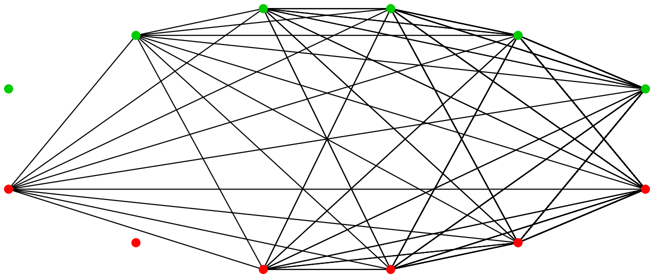




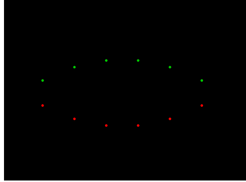


- The combinatorial Kalman filter is a depth-first search for paths in this graph that look like tracks.
- The search starts in a node in the first plane, and is performed by following edges.
- Following an edge means a KF prediction.

# Discarding an edge excludes sub-graph

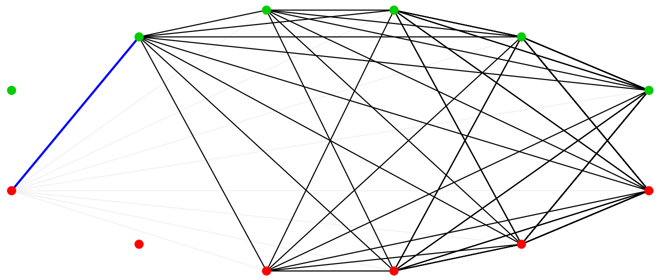






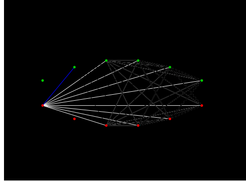
- Excluding an edge from the search excludes an entire subgraoh, greatly reducing complexity.

# Initial edges are followed based on angle cuts



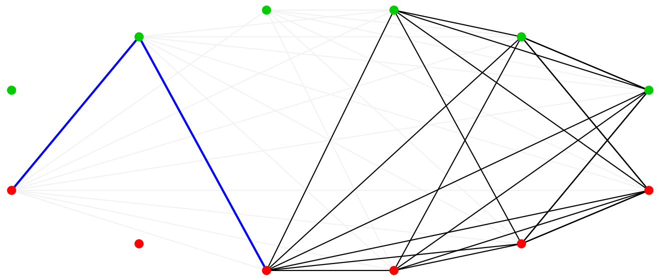
2015-01-18

DESY-2015

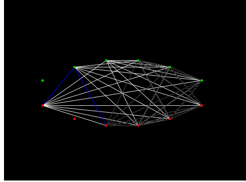


- Short edges are followed first.
- The first edges to be followed are excluded or included based on the angle between the nodes it connects.

The following edges are followed based on $\chi^2$ cuts

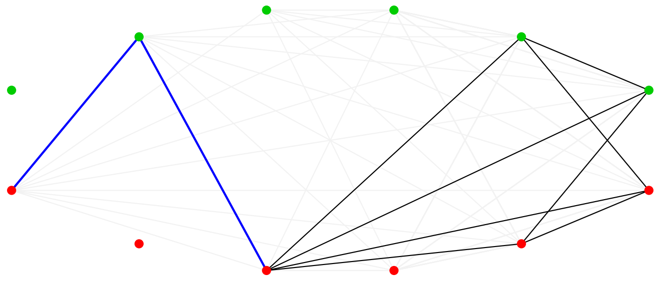




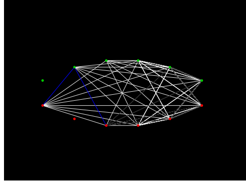


- With 2 nodes, $\chi^2$ increment are used. Obtained by KF updates.

# Follow short edges depth first

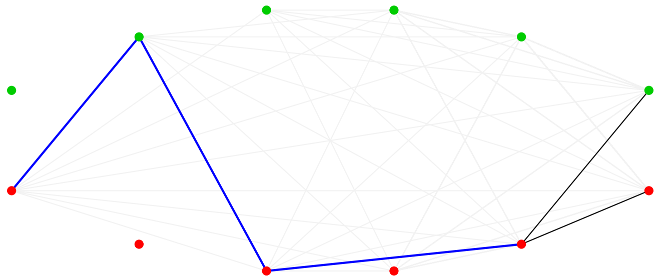






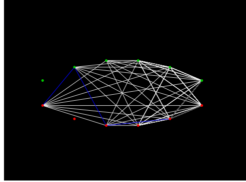
- Edges are only followed if all shorter edges have been excluded.

# Follow short edges depth first

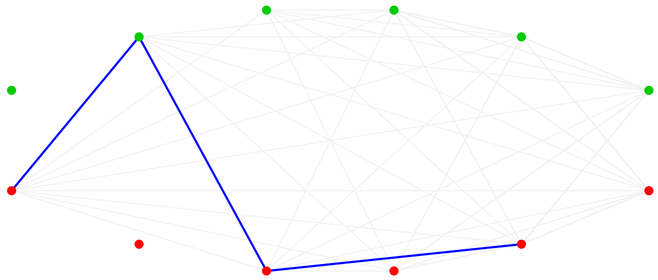


2015-01-18

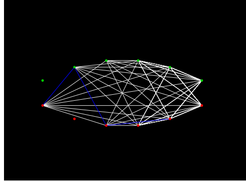
DESY-2015



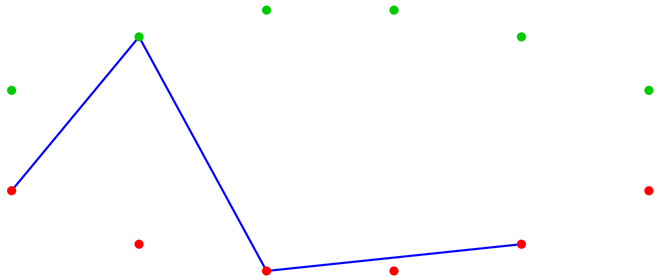
## When the path can no longer grow, it is rejected or accepted



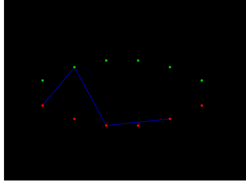
2015-01-18 DESY-2015



## If accepted, the track is stored for analysis

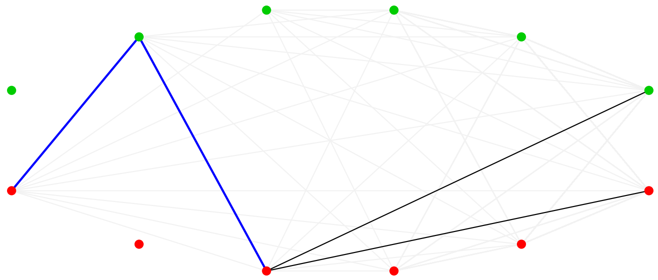




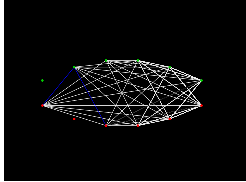


- Before track is accepted, it must pass a global $\chi^2$ cut in addition to the $\chi^2$ increment cut.
- If the cut in $\chi^2$ increment is equal to the global cut, every good candidate will be found.
- A more aggressive cut in $\chi^2$ increment can speed the search up greatly, though.

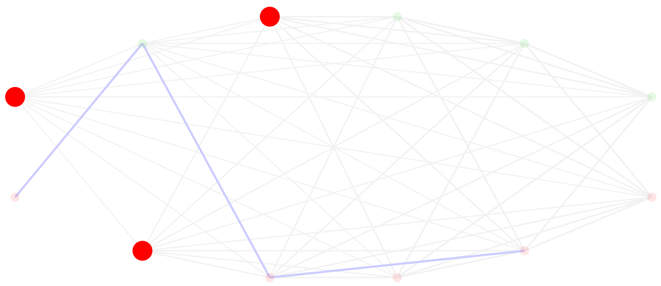
# If rejected, edges skipping more planes are considered



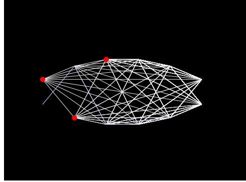




Searches are started in all nodes that can lead to a track

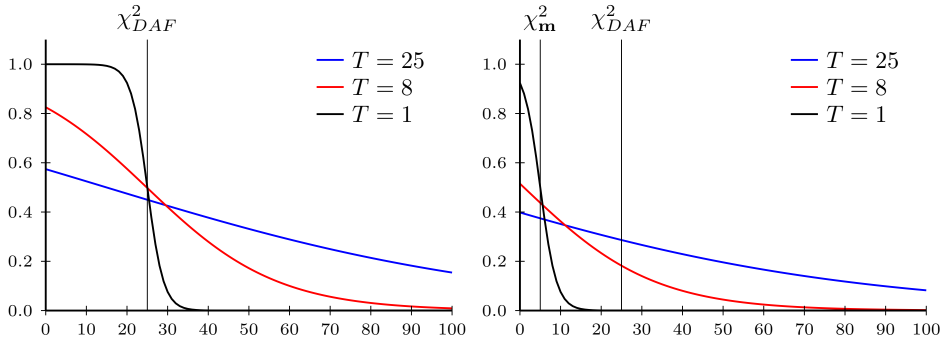






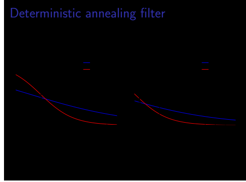
- There can be several tracks per event, search cannot stop after finding a single good one.
- No point in starting searches in planes where the paths cannot become long enough.

# Deterministic annealing filter





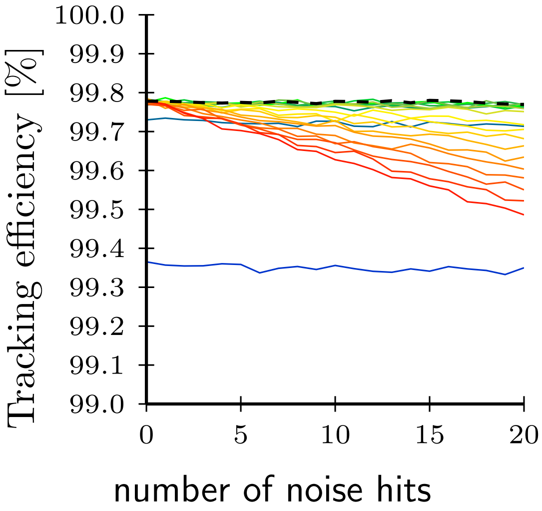




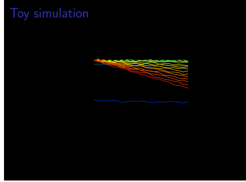

- Weakness of CKF: only information from forward running KF available when deciding on including measurement
- The DAF is an algorithm that can improve a track candidate from CKF or any other track finder.
- Uses smoothed predictions, meaning all available information, when deciding on including a measurement.
- The DAF considers all the measurements in all the planes by assigning weights to them.
- DAF can include measurements that were excluded by the track finder, and exclude measurements included by the track finder. Not just an outlier rejection.

# Toy simulation



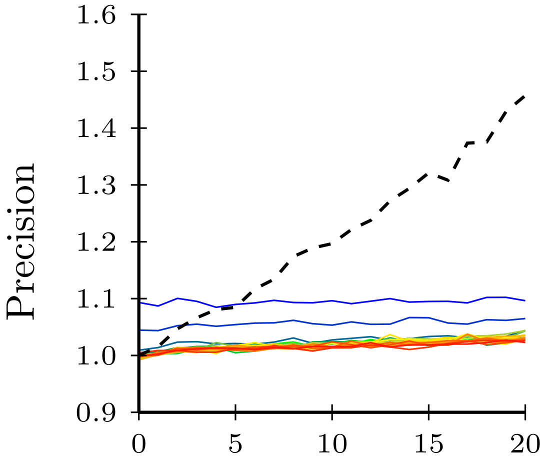




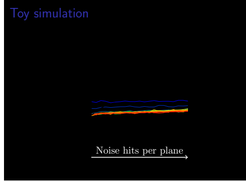

- Simulation with 6 planes, 120GeV pions, 95% detection efficiency, resolution of 4.3µm.
- Only tracks with more than 4 measurements are included.
- Binomial probability of having 4 or more hits is 99.8%.
- Dashed line is CKF: near perfect, if properly configured.
- Colored lines are CKF followed by DAF with different configurations: Near perfect if properly configured
- Noise distributed in a 0.5cm$\times$ 0.5cm area. A noise density of 20 here, means 80 noise hits per cm$^2$. MUCH higher than what I have seen in data.

# Toy simulation



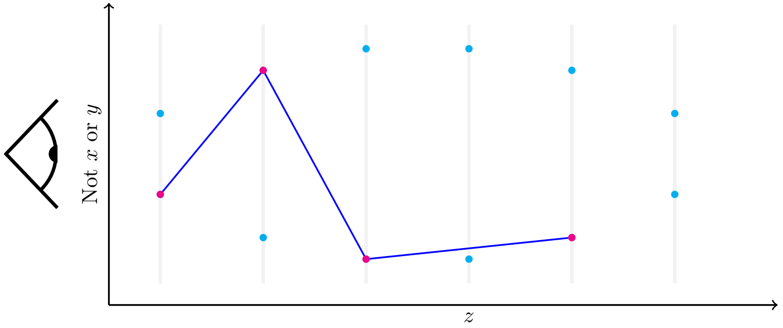

number of noise hits





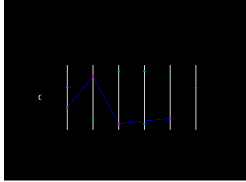

- Precision is defined as a normalized determinant of the empirical covariance matrix.
- DAF beats CKF alone in precision with high noise densities.
- CKF + DAF has near perfect track finding efficiency with near perfect precision.
- Each point on each line is the result of simulating, track finding and fitting of 1M tracks. Takes a few minutes to generate. The methods are quite fast.

The cluster track finder takes advantage of the collimated beam.



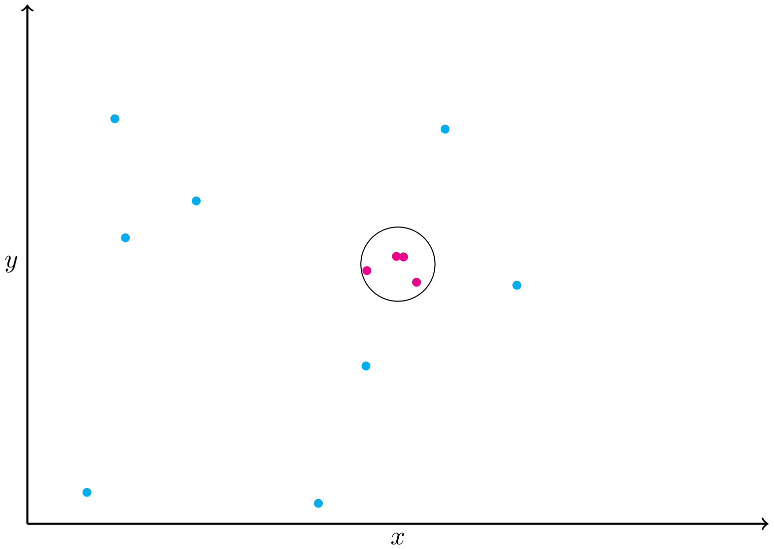


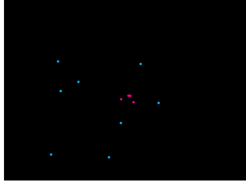


- A simpler to understand and implement technique for track finding.

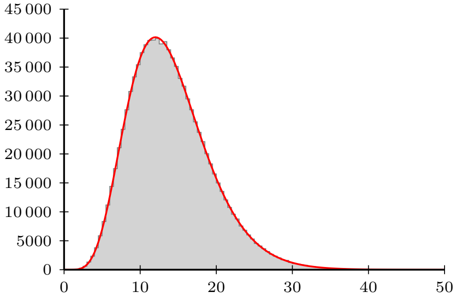
# The cluster track finder takes advantage of the collimated beam.

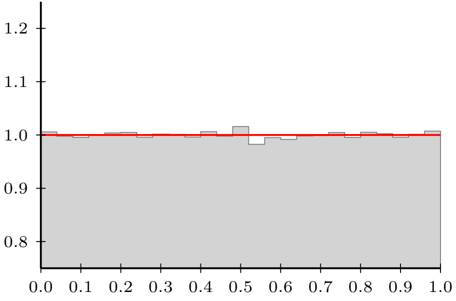


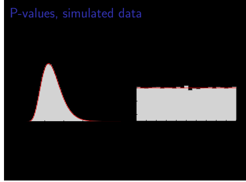
- All measurements are propagated into the first plane by the nominal beam angle.
- Track finding is a simple cluster finder.
- This approach was used for APIX test beam data.
- Works well when combined with DAF and with low noise densities.
- Not as good or as fast as CKF+DAF, though.

# P-values, simulated data



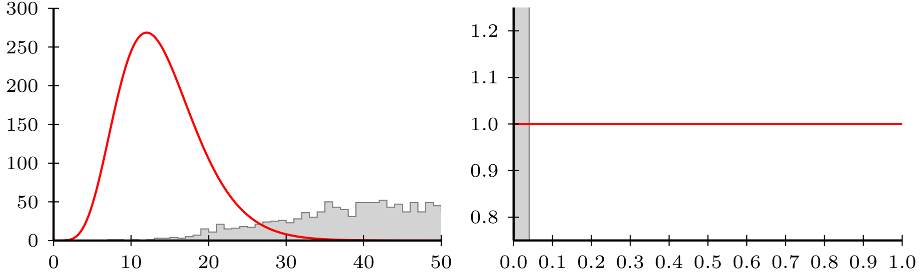





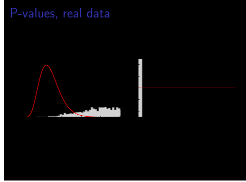




- These are the same chi2 values I showed earlier.

# P-values, real data








- Before the tracker detector is properly configured, test statistics can look more like this.
- Several parameters are needed to describe the tracking detector: Alignment, resolution, material.
- Especially in a test beam experiment, these parameters are hard to know exactly a priori.

# Detector alignment

- Obtaining a geometry description of the detector system
- Measurements during mounting not enough.
- The sum of independent random numbers following $\chi^2$-distributions is also a $\chi^2$ distribution

$$\chi^2(n_{doftot} = n_{track} \times n_{dof}) = \sum^{n_{track}} \chi^2(n_{dof})$$

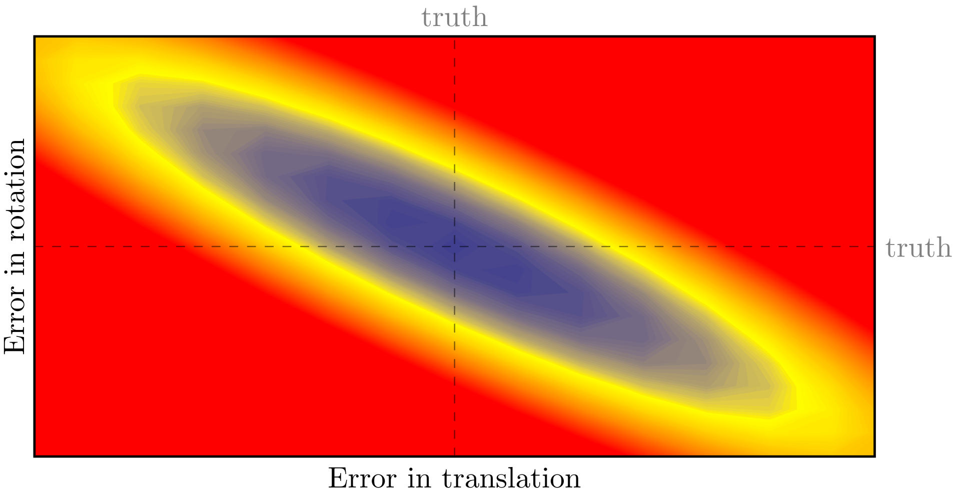




Detector alignment

- Resolution of the EUDET telescope is a few µm. Hard to get this kind of precision from measurements during mounting.
- Numerical methods for alignment are needed.
- The geometry of the planes are described by three translations, and three rotations.
- With a very collimated beam, like in CERN NA, there is very little sensitivity for translations in the $z$-direction.
- There are some advantages to using scale factors instead of real rotations in terms of numerical stability and simplicity. If sensors are very large, and rotated so that propagation differences become significant, real rotations would be better.

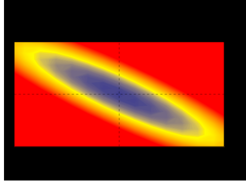
Track sample $\chi^2$ as a function of alignment parameters




Alignment: Find alignment parameters that minimize track sample $\chi^2$.

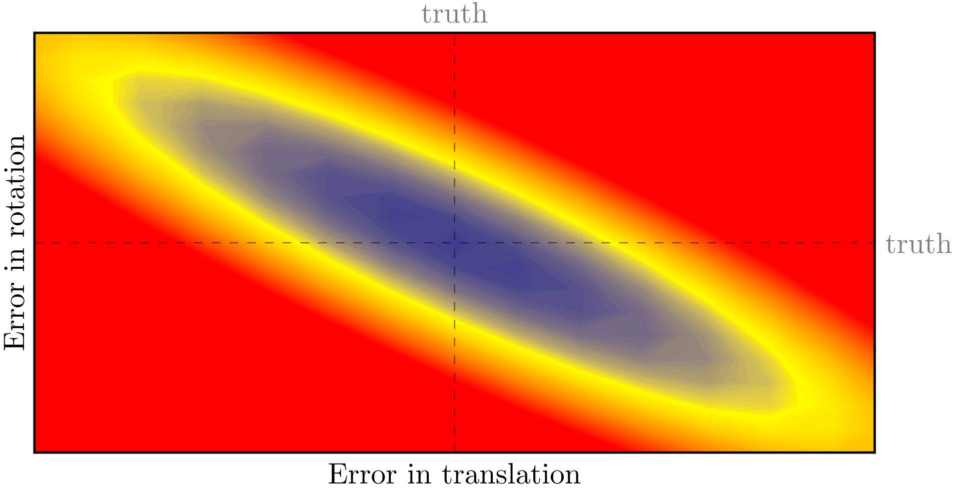






- $\chi^2$ as a function of alignment parameters of a plane.
- Red color means high $\chi^2$, blue means low.
- If geometry of all other planes are correct, the minimum $\chi^2$ is at the true position of the plane.

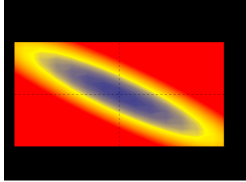
## Track sample $\chi^2$ as a function of alignment parameters




Alignment: Find alignment parameters that minimize track sample $\chi^2$.

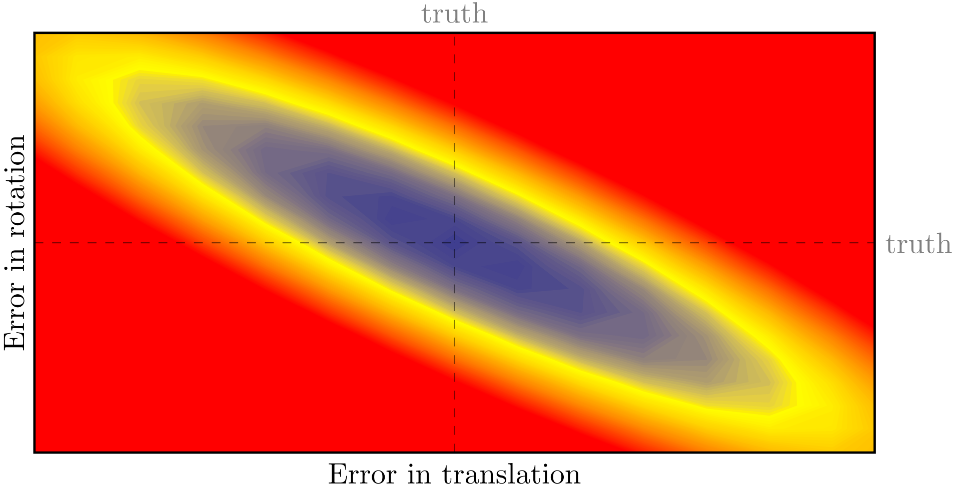






- If one of the other planes were moved out og their position, the position and value of the minimum in this plot would change.
- It is not possible to first align one plane, then the next. Both planes must be movable at the same time to find the true minimum.
- With 9 detector planes and 5 alignment parameters per plane, this is a search for a minimum in a 45 dimensional parameter space.
- Not all these parameters can be free in the search! If all parameters are free, the global coordinate system can potentially become deformed.

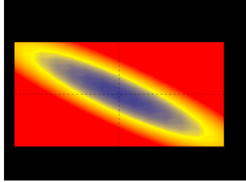
## Track sample $\chi^2$ as a function of alignment parameters




Alignment: Find alignment parameters that minimize track sample $\chi^2$.

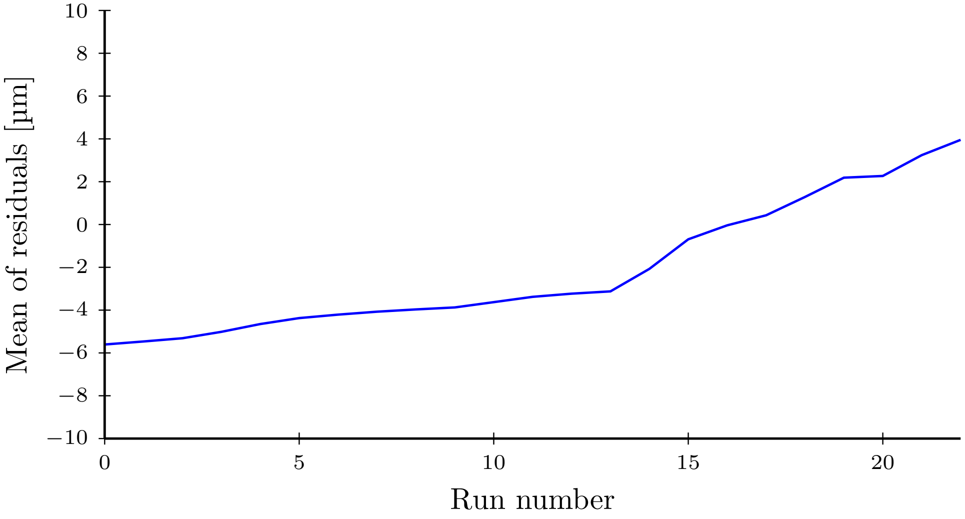






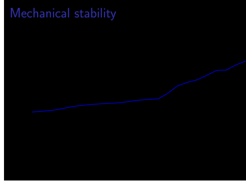
- Multidimensional minimization is generally very slow.
- With a very fast track fitter, it is possible to do alignment with iterative searches available from f.eks GSL or root Minuit. Still slow, an hour or so for 1M tracks.
- Luckily there exists a program called Millepede, that can solve alignment problems without iteration. Very fast.

# Mechanical stability



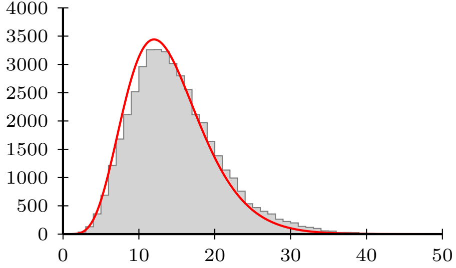


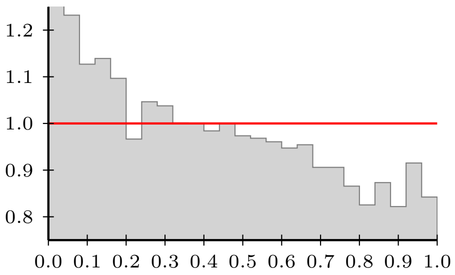


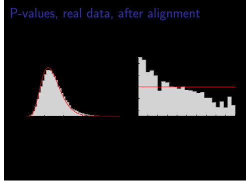

- With a tracking resolution of a few µm, mechanical stability of the setup becomes very important.
- For high resolution studies it can be a good idea to figure out data ranges where the setup is stable.

# P-values, real data, after alignment






└P-values, real data, after alignment

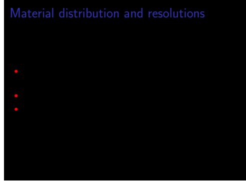
- Much better, but not perfect.

# Material distribution and resolutions

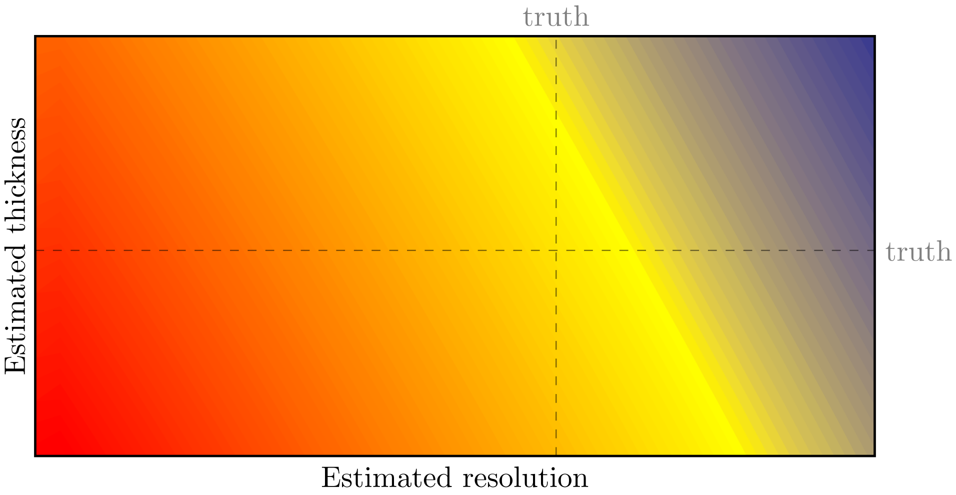
- Affect uncertainties of measurements, predictions and updates
- Not known at time of data taking
- Estimation from data is new!





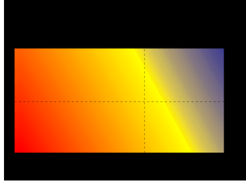


# Average $\chi^2$ as a function of thickness and resolution







- No minimum in *chi*$^2$ around true parameter values.
- Residuals divided by infinite uncertainties $\to$ 0.
- Infinitely large resolutions and thicknesses minimize $\chi^2$.

$$\sum \left( \left\langle \frac{d_i}{\sigma_{d_i}} \right\rangle^2 + \left[ \left\langle \left( \frac{d_i}{\sigma_{df_i}} \right)^2 \right\rangle - 1 \right] \right)^2$$

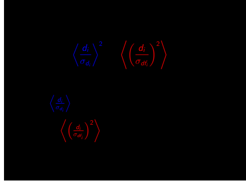
$d_0, d_1, \ldots d_n$ are all residuals and state parameter differences

The mean, $\left\langle \frac{d_i}{\sigma_{d_i}} \right\rangle$ should be 0 from alignment

The variance, $\left\langle \left( \frac{d_i}{\sigma_{df_i}} \right)^2 \right\rangle$ should be 1 if uncertainties are correct

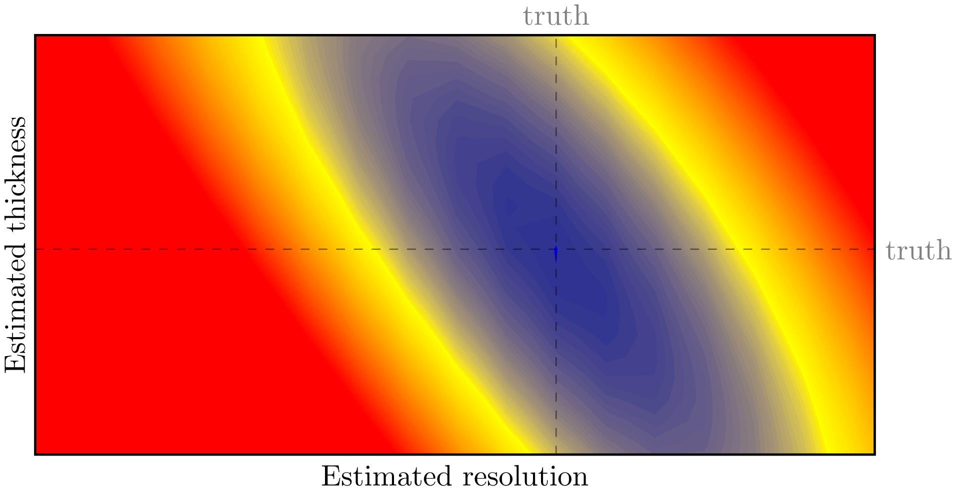




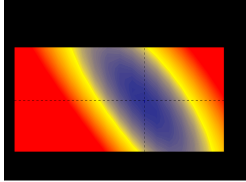


- Estimation of material and resolution can be done with multidimensional minimization.
- The trick is finding a weight function with a minimum at the true parameter values.
- Several such weigh functions were tested (paper: Gjersdal, Frühwirth, Nadler, Strandlie).

## SDR3 as a function of thickness and resolution

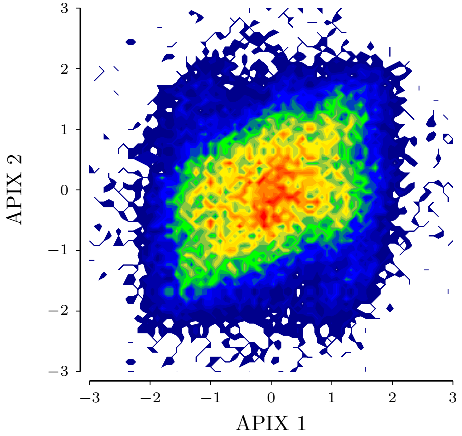


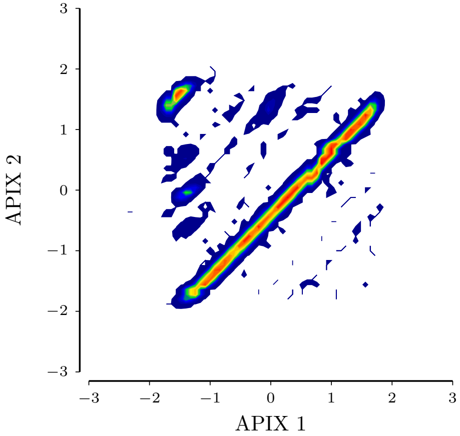


- Finding such a minimum requires a fast track fitter, as general mulidimensional minimizers are needed. (No millepede).
- Refitting a track sample when track finding is solved is an embarassingly parallel task.
- A simple test beam geometry can easily fit in GPU memory, and one can then fit thousands of tracks in parallel. Possible to solve the problem in seconds in the straight line case.

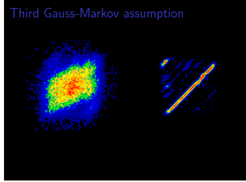
# Third Gauss-Markov assumption



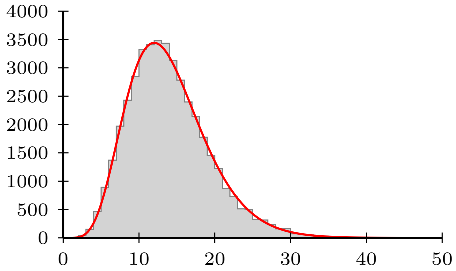


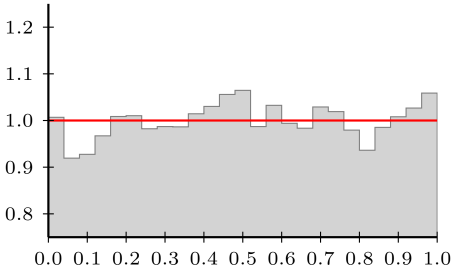



- If measurement errors are large compared to telescope resolution, and to a large degree determined by the position of the particle within the cell, errors can be correlated between measurement planes.
- The example is here showing correlations between residual errors in two neighboring APIX planes.
- In this case, including the APIX planes in the track fit decreases the precision of the track fitter in the long pixel direction.

# P-values, real data, material and resolution estimated

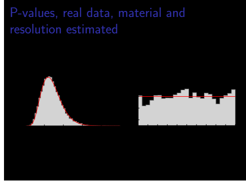




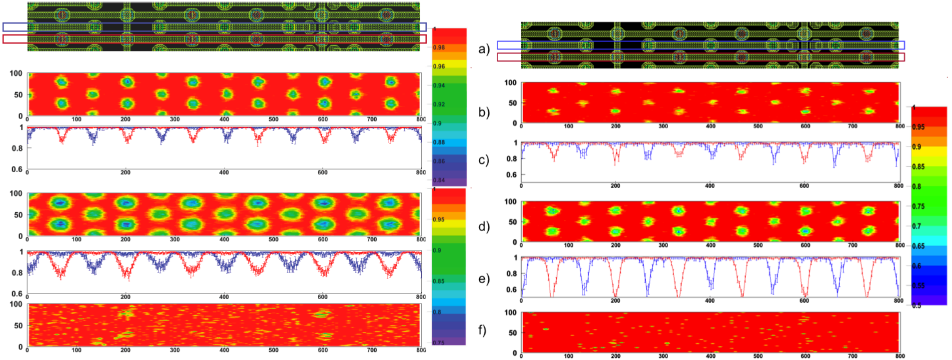



└P-values, real data, material and resolution estimated




- After alignment and mat/res estimation, near perfect recreation of the true $\chi^2$ distribution.
- Despite correlated errors and non Gaussian APIX response.
- Correct test statistics indicates that: Track fits are optimal, track finder/outlier rejector have optimal running conditions, estimated track uncertainties are correct.

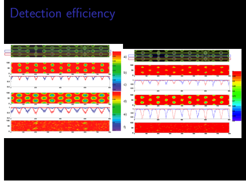
# Detection efficiency



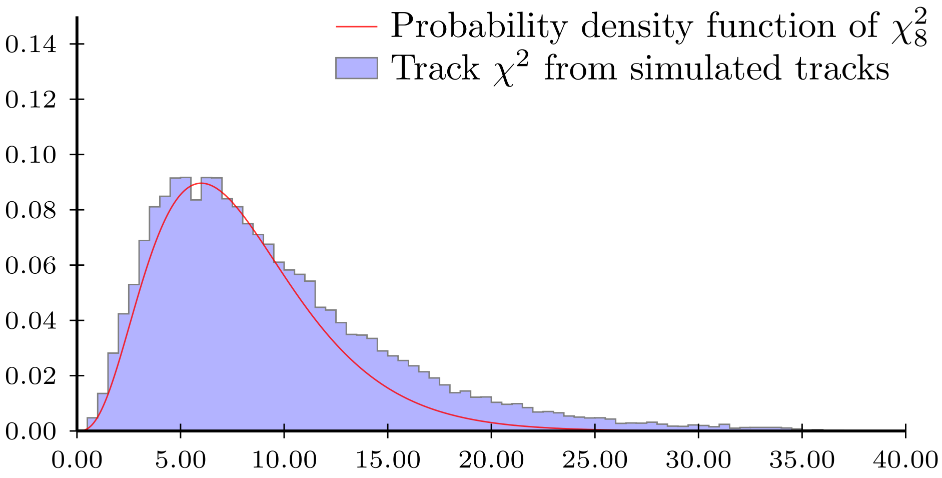

2015-01-18

DESY-2015

└Detection efficiency

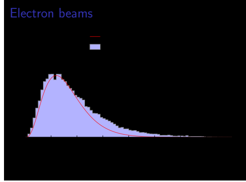


# Electron beams








└Electron beams

- Electrons lose energy mainly through Bremsstrahlung.
- This is a highly non-Gaussian process, where the electrons can lose a large fraction of its energy in a single interaction.
- Electron data fitted with standard KF not truly $\chi^2$-distributed.
- Specialized electron track fitters based on KF: Gaussian Sum Filter (GSF) and Dynamic Noise Adjustment (DNA).
- The weight function presented here for material and resolution estimation will not work on electrons. Should be possible to come up with an electron track weight function.

# Summary

## More detail:

*Straight line track reconstruction for the ATLAS IBL testbeam with the EUDET telescope*. Gjersdal, Røhne, Strandlie. ATL-INDET-PUB-2014-003.

*Optimizing track reconstruction by simultaneous estimation of material and resolutions*. Gjersdal, Frühwirth, Nadler, Strandlie. JINST, 8, 2013.

*Test beam track reconstruction and analysis of ATLAS 3D pixel detectors*. Håvard Gjersdal. PhD thesis, UiO, 2014.

## Code:

https://github.com/hgjersdal/eigen-track-fitter

Thanks for listening!





Summary