. In order to centralize the fitting process an iterative refitting algorithm servs as a wrapper. Its function is to do several iterations for each fit varying the fitting limits and number of bins until it finds the best solution, based on a minimization of the 1-Chi2/Ndf value. (Chi2-Ndf)/sqrt(2*)

Fitting limits are set as function of the datasheet standard deviation. They are defined symmetrically around the mean value for the Gaussian and Gaussian X Linear approach. For the Landau X Gaussian implementation, they are defined asymmetrically by increasing or reducing at the dataset's minimum and maximum value correspondingly. The convention I am following is:

Symmetric limits: `upper/lower_limit = mean +/- (k/2)*strdv with 3 < k < 9`

Asymmetric limits: `lower_limit = min + ((k-5)/10)*strdv with 3 < k < 9`
`upper_limit = max - ((k-5)/2)*strdv with 3 < k < 9`

The above approach will generate 5 different cases.

For each limit variation, I perform 7 different bin number variations. In discrete datasets, bin numbers variations are defined by estimating a reasonable maximum and minimum number of bins for each case. Maximum number of bins is defined by calculating the dataset resolution (see PS1) and using it to divide the fitting interval, as this has been defined above. The minimum number of bins is defined as the fitting interval divided by the standard deviation. Intermediate values are then established using the scheme described in the attached table. For non-discrete variable datasets, such a definition becomes problematic, so I revert to varying the number of bins around the sqrt(n_elemts) value.

<table>
<tr><th>Dataset Type</th><th>Statistics Case</th><th colspan="4">Bin number array (7 cases)</th></tr>
<tr><td rowspan="7">Discrete Datasets</td><td rowspan="2">$\frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} < \sqrt{N_{elements}} < N_{bins\_max}$</td><td>Lower 3 bin number variations</td><td>Optimum Bin number</td><td colspan="2">Higher 3 bin number variations</td></tr>
<tr><td>$\left\lceil \left\lceil\sqrt{N_{elements}}\right\rceil - n \times \left\lceil \frac{\sqrt{N_{elements}} - \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma}}{3} \right\rceil \right\rceil$
with 1 < n < 3</td><td>$\left\lceil\sqrt{N_{elements}}\right\rceil$</td><td colspan="2">$\left\lfloor \left\lceil\sqrt{N_{elements}}\right\rceil + n \times \left\lfloor \frac{N_{bins\_max} - \sqrt{N_{elements}}}{3} \right\rfloor \right\rfloor$
with 1 < n < 3 **</td></tr>
<tr><td rowspan="2">$\sqrt{N_{elements}} \leq \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} < N_{bins\_max}$</td><td>Lowest bin number</td><td colspan="3">Rest of the bin number array</td></tr>
<tr><td rowspan="2">$\left\lfloor\sqrt{N_{elements}}\right\rfloor$</td><td rowspan="2" colspan="3">$\left\lfloor\sqrt{N_{elements}}\right\rfloor + n \times \left\lceil \left|N_{bins\_max} - \sqrt{N_{elements}}\right| / 6 \right\rceil$
with 1 < n < 6 **</td></tr>
<tr><td>$\sqrt{N_{elements}} \leq N_{bins\_max} < \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma}$</td></tr>
<tr><td>$N_{bins\_max} \leq \sqrt{N_{elements}}$</td><td colspan="4">$n \times \left\lfloor N_{bins\_max}/7 \right\rfloor$
with 1 < n < 7</td></tr>
<tr><td rowspan="3">Non – Discrete Datasets</td><td rowspan="2">$\frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} < \sqrt{N_{elements}}$</td><td>Lower 3 bin number variations</td><td>Optimum Bin number</td><td colspan="2">Higher 3 bin number variations</td></tr>
<tr><td>$\left\lceil \left\lceil\sqrt{N_{elements}}\right\rceil - n \times \left\lceil \frac{\sqrt{N_{elements}} - \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma}}{3} \right\rceil \right\rceil$
with 1 < n < 3</td><td>$\left\lceil\sqrt{N_{elements}}\right\rceil$</td><td colspan="2">$\left\lfloor \left\lceil\sqrt{N_{elements}}\right\rceil + n \times \left\lfloor \frac{\frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma}}{3} \right\rfloor \right\rfloor$
with 1 < n < 3</td></tr>
<tr><td>$\sqrt{N_{elements}} \leq \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma}$</td><td>$\left\lceil \left\lceil \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} \right\rceil - n \times \left\lceil \frac{\frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} - \sqrt{N_{elements}}}{3} \right\rceil \right\rceil$
with 1 < n < 3</td><td>$\left\lceil \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} \right\rceil$</td><td colspan="2">$\left\lfloor \frac{\left|\lim\limits_{fit} High - \lim\limits_{fit} Low\right|}{\sigma} + n \times \left\lfloor \frac{\sqrt{N_{elements}}}{3} \right\rfloor \right\rfloor$
with 1 < n < 3</td></tr>
</table>

- **An additional divider of the number of bins is applied at the higher part of the binning array, by a factor of 5 or power of 10 in the case where the resolution is too fine, quantified by bin_max > 5*sqrt(n_elements) or bin_max > 10*sqrt(n_elements) respectively.
- *** A reduction of the number of bins is applied in the case where the number of elements is smaller than the number of bins. The number of bins is then set equal to the number of entries in the dataset.
- Binning scheme is identical for all types of fits (I know it preferably shouldn't be)

All of the above are valid for uniformly binned datasets. In a non-uniformly binned dataset (such as FFT, where the resolution varies with respect to frequency), the assumption is made that enough statistics exists to have at least one hit per resolution bin. In such a case, a rebinding is applied using consecutive different values as centers for neighboring bins. Rebinding can then be applied on the dataset by increasing each bin limits proportionally to its width (decreasing here would make no sense) and recalculating bin contents in the new configuration.

PS1 on resolution calculations:
The absolute difference of each element of the data set with every other element is computed and a new dataset of differences is generated. Extreme values (varying more than 4 orders of magnitude from the mean of these new dataset) are excluded to account for floating point rounding errors (typically present because of instrument controller bit count differences). Finally, the minimum of the differences dataset is considered as the resolution of the discretized variable.

n-on-p and p-on-n production
23 x 6” wafers n-type 320 µm thick wafers (available at CNM)
25 x 6” wafers p-type 300 µm thick wafers (available at CNM)
Not thicker sensors, nor thinner
In general TimePix4 matrices
Include some TimePix3 matrices

This will be an AIDAinova run
Designed for highest voltage operation (extreme edges. Even mm with several guard rings)
No active edge
Paraleine coating because one has to take care of the shielding between edge and pads. Edges are going to be to high voltages and chip sort of grounded
Around 500 -600V operation voltage, after that we see breakdown
200µm thick sensors with 450µm edge goes to 2kV (edge is 2,5 times the thickness of the sensor)
We will collect both electrons and holes and there should be no different for the ASIC, the limiting factor in terms of timing would be the sensors

How many wafers of each type to produce?
Yield is expect to be 90%, test be including temporary metal at CNM
Expected more than 10 good sensors per wafer (6” wafers)
8 wafers per batch is the minimum, one wafer will for sure break in processing because of thinness
Wait for a first design of sensors

Using SiGe wafers for timing applications, 1kOhm wafers, we should also include them????

Next meeting will be in 2-3 weeks’ time, the week of the 15th of October (Thursday) at 16:00