NOISE and CORRELATION

R. Schieder

1. Physics Institute, University of Cologne

Table of content

Introductory Remarks

The following is a compendium of the results of some investigations, which I wrote down in the past in order to better understand the properties of spectrometers I have been working with. In particular, Acousto-Optical-Spectrometers (AOS) were in the main focus, and it seemed necessary to learn about the properties of AOS as well as of other spectrometer types in more detail. It was also essential to deal with problems arising when dealing with spectrometer data afterwards. This led to more detailed discussions of errors and the characteristics of noise, in particular of correlation in data streams and their influence on experimental errors in general. The main focus was certainly on the noise as it comes out of the spectrometers, and it turned out, that it can be quite tricky to understand the details in full depth. In addition, some of the tasks were also focused on the methods of laboratory spectroscopy which are also an issue for a detailed noise discussion. Therefore, my feeling is, it might be useful for others to have the following articles available. Some articles repeat the arguments of previous ones or of presentations in literature in order to make it easier to follow the arguments.

I. About Correlation

The problem to provide of meaningful estimates of statistical errors in experimental data is one of the most neglected tasks in all experimental sciences although it belongs to any experimental result that an error is also provided in order to establish some credibility for the presented results. In a lot of cases, this is may not be of particular concern, because the signal to noise ratio of the data involved may be high enough not to worry about statistics but, on the other hand, the question whether there are systematic errors or not is still something to take care of. The situation is very particular in radio-astronomical spectroscopy, where the line signals observed are typically very weak and the background is huge and rather noisy. But also in experimental laboratory physics, the signal to noise ratio is frequently of major concern since the easy experiments are mostly already done and the more difficult ones are usually hard to carry out because of marginal signal to noise. In all such cases a very detailed analysis of the statistical performance of the instrumentation should be undertaken, and for this the nature of the noise should be well understood. In the special case of radio-astronomical instrumentation the use of Allan-variance measurements has proven to be extremely valuable. By means of the Allan-variance plot it is easy to distinguish between white noise and low frequency noise or drift noise by investigating the slope of the plot. Drift noise means that on shorter time scales, the fluctuations in the output of the instrument are strongly correlated so that their impact on the experimental results can be greatly reduced just by measuring differences taken within short time intervals. This makes the Allan-variance a very good tool to study the impact of correlation in experimental data.

When considering correlation between two different data sets $y_1(x)$ and $y_2(x)$, one uses the definition of the first order correlation function:

$$G_{12}^{(1)}(\tau) = < y_1(x) \cdot y_2^*(x+\tau) >_x$$

The superscript "*" stands for the conjugate complex value of eventually complex data. In most cases it can be neglected. The subscript "$_x$" for the expectation "<>" means that one has to average over the argument "x", it could be the time, a distance or else depending on the problem under consideration.
When exchanging the argument τ by $-\tau$, the correlation function changes like:

$$G_{12}^{(1)}(-\tau) = < y_1(x) \cdot y_2^*(x-\tau) >_x = < y_1(x'+\tau) \cdot y_2^*(x') >_{x'} = < y_2(x') \cdot y_1^*(x'+\tau) >_{x'}^* = G_{21}^{(1)^*}(\tau)$$

while using $x' = x-\tau$. If we are discussing the output of a simple measuring instrument like a Volt-meter for example, the data y are generally real valued and we can drop the conjugate complex indicator. In this case we have

$$G_{12}^{(1)}(\tau) = < y_1(x) \cdot y_2(x+\tau) >_x$$

Mostly, one wants to investigate the "autocorrelation function":

$$G_{11}^{(1)} = < y_1(x) \cdot y_1(x+\tau) >_x$$

It provides important information about the noise characteristics of any system.
When changing the argument to its negative value then we have:

$$G_{12}^{(1)}(-\tau) = G_{21}^{(1)}(\tau) \text{ and } G_{11}^{(1)}(-\tau) = G_{11}^{(1)}(\tau)$$

It indicates that the auto-correlation function of real valued data is always symmetric with respect to the argument.

One should distinguish between noisy data, which are mostly real valued and amplitudes of optical fields for example, which are typically complex. If we have a data set taken as a sequence in time "*t*" from a standard instrument, the correlation function is real valued and symmetric in time. A Fourier-Transform-Spectrometer (FTS) provides the first order auto-correlation function of the incident electromagnetic field as a function of optical delay $\tau = c \cdot \Delta x$ with Δx the optical path-length difference between the two arms of the spectrometer. In principal, we therefore should get now a complex auto-correlation function $G_{11}^{(1)}(\tau)$, which one should determine for both, negative and positive values. But, since we have $G_{11}^{(1)}(-\tau) = G_{11}^{(1)}(\tau)^*$, we still would need only the data for non-negative values of τ. However, since we are measuring the power of the interferometer output, we get only real data, which are representing the so called "visibility function" of the spectrometer. In principle, one can also measure the "Sin-" and "Cos-" visibility just by using two outputs of the two-beam interferometer, which is equivalent to the measurement of the complex correlation function. It might be essential to determine it for both negative and positive values of τ in order to take care of the inherent phase-shifts within the two-beam interferometer.

How to describe correlation of experimental data

Correlation between different data means that the statistical behavior of the data is partially identical. At full correlation all statistical fluctuations are the same for all data, and without correlation the statistical fluctuations are completely independent from each other. The usual way to describe correlation is done by the definition of the first order correlation function $G_{12}^{(1)}(\tau)$ (see above). Data sets of arbitrary length are certainly not available so that the true expectation <> cannot be determined. Instead, one has to deal with finite data sets, and one needs modified procedures to derive knowledge of the true correlation function. Frequently, as mentioned above, the auto-correlation function is considered, which is calculated while both data sets $\{y_1\}$ and $\{y_2\}$ are identical.

$$G_{11}^{(1)}(\tau) = < y_1(t) \cdot y_1(t+\tau) >_t = < y_1(t+\tau) \cdot y_1(t) >_t$$

The value of $G_{11}^{(1)}$ at $\tau = 0$ is then:

$$G_{11}^{(1)}(0) = < y_1^2(t) >_t$$

At extremely long delay τ all statistical correlation should disappear for any signal with random behavior. We can separate the mean value of $y_i(t)$ and its fluctuations $dy_i(t)$:

$$y_i(t) = < y_i(t) >_t + dy_i(t)$$

and we get

$$< y_1(t) \cdot y_2(t+\tau) >_t = < (< y_1(t) >_t + dy_1(t)) \cdot (< y_2(t+\tau) >_t + dy_2(t+\tau)) >_t =$$
$$= < y_1(t) >_t \cdot < y_2(t+\tau) >_t + < y_1(t) >_t \cdot < dy_2(t+\tau) >_t + < dy_1(t) >_t \cdot < y_2(t+\tau) >_t + < dy_1(t) \cdot dy_2(t+\tau) >_t =$$
$$= < y_1(t) >_t \cdot < y_2(t) >_t + < dy_1(t) \cdot dy_2(t+\tau) >_t$$

since $< dy_1(t+\tau) >_t = < dy_1(t) >_t = < dy_2(t) >_t = 0$.

If τ is supposed to be very large, the fluctuations are not correlated, i.e. the product of both is zero on average, and we have then:

$$< dy_1(t) \cdot dy_2(t+\tau) >_t = < dy_1(t) >_t \cdot < dy_2(t+\tau) >_t = 0$$

Thus we have:

$$G_{12}^{(1)}(\infty) = < y_1(t) >_t \cdot < y_2(t) >_t \quad \text{and} \quad G_{11}^{(1)}(\infty) = < y_1(t) >_t^2$$

The variance of the statistical distribution of one data set is then described by

$$\sigma_1^2 = < y_1^2(t) >_t - < y_1(t) >_t^2 = G_{11}^{(1)}(0) - G_{11}^{(1)}(\infty)$$

We can simplify things, when removing the mean value from the data. Then we rewrite:

$$G_{12}^{(1)}(\tau) = < (y_1(t) - < y_1(t) >_t) \cdot (y_2(t+\tau) - < y_2(t+\tau) >_t) >_t = < dy_1(t) \cdot dy_2(t+\tau) >_t$$

and

$$G_{11}^{(1)}(\tau) = < (y_1(t) - < y_1(t) >_t) \cdot (y_1(t+\tau) - < y_1(t+\tau) >_t) >_t = < dy_1(t) \cdot dy_1(t+\tau) >_t$$

The variance of a data set is then given by

$$\sigma_i^2 = < dy_i(t) \cdot dy_i(t) >_t = G_{ii}^{(1)}(0), \quad i = 1,2$$

The correlation function is an important quantity in systems like Michelson Interferometers, Array-Antennas, etc.. The correlated signal from two telescopes contains all the available (!) information about the spatial distribution of the observed sources.

The cross-correlation between data sets is not often used, but sometimes it is practical to determine the statistical similarity of data sets of two real time spectrometers. For example, it might be necessary to determine the relative frequency calibration. If both spectrometers have taken data simultaneously, and if the measured signal was just pure white noise (i.e. it is identical for both spectrometers), then the cross-correlation will become maximum if the frequency shift between the two spectra is zero. If the frequency separation between the frequency pixels of the spectrometers is not identical, then the correlation becomes strongly degraded but will not disappear. It might be helpful to make the frequency separation of the pixels of both spectrometers identical as much as possible. This can be done with a proper resampling procedure (see below). Since the simultaneous sampling with both spectrometers guarantees that all noise fluctuations are identical, one can expect that the cross-correlation becomes large. Such procedures might be useful for the relative calibration of Digital Fourier Transform Spectrometers, Auto-Correlators, Filter-Banks or Acousto-Optical Spectrometers, which might not have an extremely well known frequency calibration.

White Noise

Before we start to investigate the various possible noise characteristics we should first determine how we understand what is called “white noise”. If all data points of a set of data are statistically independent from each other, we consider the data set as representing white noise. There exists no possibility to describe a data stream representing white noise analytically! It can only be characterized by a correlation function like a Dirac Delta-function $\delta(\tau)$:

$$G_{11}^{(1)}(\tau) \;=\; \Gamma_0 \cdot \delta(\tau)$$

Γ_0 describes the power, better the square of the amplitude of the white noise fluctuations. The word “white” reflects the assumption that the spectral power density, i.e. the power found at any frequency is constant and identical at all frequencies. The Fourier-transform of a Delta-function function shows this behavior. The function diverges at $\tau = 0$, which surely cannot happen in practice. Therefore, in reality we have to consider “band-limited” white noise, which says that the noise fluctuations do not have frequency components at arbitrary high frequencies. The Delta-function is then understood as a “spiky” function of finite width and with finite maximum, which is determined by the maximum frequency visible in the noise. It may be described for example by a simple box of very small width δ, a Sinc- or a Sinc^2-function like:

$$\delta(\tau) \to \frac{\sin(2\pi \cdot \tau / \tau_0)}{\pi \cdot \tau} \quad \text{or} \quad \delta(\tau) \to \frac{\sin^2(2\pi \cdot \tau / \tau_0)}{2\tau_0 \cdot (\pi \cdot \tau / \tau_0)^2} \quad \text{with } \tau_0 \text{ very small.}$$

Its value at $\tau = 0$ is now equal to $2/\tau_0$, while the width $\Delta\tau$ is equal to $\tau_0/2$ (according to the “Rayleigh-criterion”). The second expression has the advantage that it is non-negative everywhere, so that it avoids anti-correlation. Negative values of the correlation function represent “anti-correlation”. It can be illustrated by considering an accidental movement of the content of one data point in one direction; the value of the neighboured data point would preferably move in the opposite direction. (In case of normal or positive correlation the opposite would be more likely.) The expressions keep the value of the correlation function finite at $\tau=0$, but remains narrow with a small time constant τ_0. Still, the integral over τ is equal to unity. Other replacements for the Dirac Delta-function like a Gaussian or a Lorentzian are also possible.

$$\delta(\tau) \to \frac{1}{\tau_0} \cdot \exp\{-\pi \cdot (\tau / \tau_0)^2\} \;, \quad \delta(\tau) \to \frac{\tau_0/\pi}{\tau^2 + \tau_0^2} \text{ with very small } \tau_0$$

Frequently, a normalized correlation function is used, which is defined by:

$$g_{12}^{(1)}(\tau) \;=\; \frac{< y_1(t) \cdot y_2(t+\tau) >_t}{\sqrt{< y_1^2(t) >_t \cdot < y_2^2(t) >_t}} \quad \text{and} \quad g_{11}^{(1)}(\tau) \;=\; \frac{< y_1(t) \cdot y_1(t+\tau) >_t}{< y_1^2(t) >_t}$$

Using this definition, the value of $g_{11}^{(1)}$ becomes automatically “1” at zero. The value of $g_{12}^{(1)}(0)$ on the other hand stands for the relative amount of correlation between the two data sets. When considering noise only it is again reasonable to subtract the mean from all data which modifies the expressions to

$$g_{12}^{(1)}(\tau) \;=\; \frac{< dy_1(t) \cdot dy_2(t+\tau) >_t}{\sqrt{< dy_1^2(t) >_t \cdot < dy_2^2(t) >_t}} \quad \text{and} \quad g_{11}^{(1)}(\tau) \;=\; \frac{< dy_1(t) \cdot dy_1(t+\tau) >_t}{< dy_1^2(t) >_t}$$

(In the following we mostly drop the subscript argument “t” to simplify things.)

In reality, it is common that data adjacent in time are partly correlated due to inevitable time constants of the system. In this case we can determine the auto-correlation function by means of one simple parameter τ_s, which influences the data like

$$y(t) = \frac{1}{\tau_s} \cdot \int_{-\infty}^{t} x(t') \cdot \exp\{-(t-t')/\tau_s\} \cdot dt'$$

The input *x(t),* which we assume as white noise, is the signal without the influence of the filter. We can easily calculate the output correlation function with:

$$G(\tau) = < y(t) \cdot y(t+\tau) > = < \frac{1}{\tau_s^2} \cdot \int_{-\infty}^{t} \int_{-\infty}^{t+\tau} x(t') \cdot x(t'') \cdot \exp\{-(t-t')/\tau_s\} \cdot \exp\{-(t+\tau-t'')/\tau_s\} \cdot dt' \cdot dt'' > =$$

$$= \frac{1}{\tau_s^2} \cdot \int_{-\infty}^{t} \int_{-\infty}^{t+\tau} < x(t') \cdot x(t'') > \cdot \exp\{-(t-t')/\tau_s\} \cdot \exp\{-(t+\tau-t'')/\tau_s\} \cdot dt' \cdot dt'' =$$

$$= \frac{1}{\tau_s^2} \cdot \int_{-\infty}^{t} \int_{-\infty}^{t+\tau} \Gamma_0 \cdot \delta(t'-t'') \cdot \exp\{-(t-t')/\tau_s\} \cdot \exp\{-(t+\tau-t'')/\tau_s\} \cdot dt' \cdot dt'' = \frac{\Gamma_0}{2 \cdot \tau_s} \cdot \exp\{-\frac{|\tau|}{\tau_s}\}$$

The normalized correlation function becomes: $g(\tau) = \exp\{-|\tau|/\tau_s\}$.

If the data are sampled at times $\tau = \tau_k = k \cdot \tau_0$ then we can rewrite:

$$g_k = g(\tau_k) = \exp\{-|\tau_k|/\tau_s\} = \exp\{-|k| \cdot \tau_0/\tau_s\} \quad \text{with} \quad k = 0, \pm 1, \pm 2, \ldots \qquad [1]$$

Thus we have found the correlation at the presence of time constants, which are usually inevitable

It should be mentioned that higher order correlations are also defined in literature like for example (see e.g. in Loudon[1]):

$$G_{1234} = < y_1(t+\tau_1) \cdot y_2(t+\tau_2) \cdot y_3(t+\tau_3) \cdot y_4(t+\tau_4) >$$

This function is useful when describing the interference of many field amplitudes (in this case 4) in optical systems like a multi mirror telescope for example. It then describes the effect of the combination of all different optical fields in dependence of the different delay time τ_i. If they are all made equal, one "sees" the ideal image which is equivalent to the image of a large perfect telescope, which has openings in the telescope aperture at the places where the real telescopes are located. This is a rather difficult undertaking, since one needs optical delay lines for each telescope in order to compensate for the different path-lengths τ_i simultaneously. In reality one usually combines only pairs of telescopes and calculates the correlation of each pair individually. Higher order correlation is then neglected.

"Real" data sets

During a real experiment the data do not appear as continuous functions, but instead, they are a finite data set of discrete data points. Thus the above definitions need a refinement to deal with such data. We assume that we have a data set of *N* data, all taken at different time, say $t = n \cdot \tau_0$ (n=1,2,…). The expectations <> are now replaced by averages of a sum, e.g. the correlation between two data sets "1" and "2" should be calculated now by:

$$G_{12}^{(1)}(k) = \frac{1}{N-k} \sum_{n=1}^{N-k} y_1(n) \cdot y_2(n+k),$$

The arguments "*n*" refer to the different times the data have been collected. They can also mean the pixel number of the output of a spectrometer for example.

In most cases it is necessary to subtract the mean of the data with:

$$dy_i(n) = y_i(n) - \frac{1}{N} \cdot \sum_{m=1}^{N} y_i(m), \quad i = 1,2$$

$$G_{12}^{(1)} = \frac{1}{N-k} \cdot \sum_{n=1}^{N-k} dy_1(n) \cdot dy_2(n+k) \quad \text{and} \quad G_{11}^{(1)} = \frac{1}{N-k} \cdot \sum_{n=1}^{N-k} dy_1(n) \cdot dy_1(n+k)$$

[1] Rodney Loudon, The quantum theory of light, Chapter 5; Clarendon Press Oxford (1973)

Similar we can write:

$$g_{12}^{(1)}(k) = \frac{\frac{1}{N-k}\cdot\sum_{n=1}^{N-k} dy_1(n)\cdot dy_2(n+k)}{\sqrt{\frac{1}{N-k}\sum_{n=1}^{N-k} dy_1^2(n)\cdot\frac{1}{N-k}\sum_{n=k+1}^{N} dy_2^2(n)}} \quad \text{and} \quad g_{11}^{(1)}(k) = \frac{\frac{1}{N-k}\cdot\sum_{n=1}^{N-k} dy_1(n)\cdot dy_1(n+k)}{\sqrt{\frac{1}{N-k}\sum_{n=1}^{N-k} dy_1^2(n)\cdot\frac{1}{N-k}\sum_{n=k+1}^{N} dy_1^2(n)}} \qquad [2]$$

For the variance of a data set we need to subtract the mean of the data, and we have therefore:

$$\sigma_y^2 = < y^2 > - < y >^2 \rightarrow \frac{1}{N}\cdot\sum_{n=1}^{N} y^2(n) - \left[\frac{1}{N}\cdot\sum_{n=1}^{N} y(n)\right]^2 = \frac{1}{N}\sum_{n=1}^{N} dy^2(n)$$

One should keep in mind that the expectation <> is not the same as the mean of a finite sum, because one will never be able to determine a fully reliable average from a finite sample. This has consequences for the error estimates. Using the sums given above it should now not be too complicated to determine a reasonable estimate (or expectation) of the correlation within a discrete set of data.

Error estimate of correlated data

If there are contributions to the noise, which are caused by slow processes, i.e. “drifts”, it is clear that this should have some effect on the error estimates of the experimental data. Slow drifts are causing correlation between data when taken within short time intervals. Frequently, one simple question is to be considered: How is the error of a finite data-set with *N* data samples estimated correctly? For this we start with the calculation of the expectation value of the variance according to the usual definition:

$$\sigma_{true}^2 = < y^2 > - < y >^2 = < [y - < y >]^2 >$$

For a finite data set with *N* data we do not know the exact value of the expectation values, therefore the expression converts into an expectation of an experimental variance

$$\sigma_{\exp}^2 = < [y_n - \frac{1}{N}\cdot\sum_{m=1}^{N} y_m]^2 > .$$

This is not the “true” variance! With the help of the expectation value of each of the y_n we can modify this.

$$\sigma_{\exp}^2 = < [dy_n - \frac{1}{N}\cdot\sum_{m=1}^{N} dy_m]^2 > = < dy_n^2 > - < [\frac{1}{N}\cdot\sum_{m=1}^{N} dy_m]^2 >$$

with $dy_n = y_n - < y_n >$

$< dy_n^2 > = < dy^2 > = \sigma_{true}^2$ stands for the real error of the data, which we don’t know, since we use a not exact expression for the evaluation of $<y_n^2>$. σ_{true}^2 is the same for all *N* data y_n, therefore we can write:

$$\sigma_{\exp}^2 = < dy_n^2 > - < [\frac{1}{N}\cdot\sum_{m=1}^{N} dy_m]^2 > = < dy^2 > - < [\frac{1}{N}\cdot\sum_{m=1}^{N} dy_m]^2 > =$$

$$= < dy^2 > - \frac{1}{N^2}\cdot\sum_{n=1}^{N}\sum_{m=1}^{N} < dy_n \cdot dy_m >$$

The expectation $<dy^2>$ is the true variance σ_{true}^2 of the data, which we want to find. There are *N* terms in the double sum with *n* = *m*, which contribute $N\cdot<dy_n^2>$. Thus we can continue:

$$\sigma_{\exp}^2 = (1 - \frac{1}{N})\cdot\sigma_{true}^2 - \frac{1}{N^2}\cdot\sum_{k=1}^{N-1}\left\{\sum_{n=1}^{N-k} < dy_n \cdot dy_{n+k} > + \sum_{n=k+1}^{N} < dy_n \cdot dy_{n-k} >\right\}$$

The expectation values $<dy_n \cdot dy_{n+k}>$ and $<dy_n \cdot dy_{n-k}>$ depend solely on the separation *k* of the samples and not on *n*. Therefore we can write:

$$< dy_n \cdot dy_{n+k} > = < dy_n \cdot dy_{n-k} > = < dy_n^2 > \cdot g_k$$

g_k is the first order auto-correlation function of the data set:

$$g_k = \frac{< dy_n \cdot dy_{n+k} >}{\sqrt{< dy_n^2 > \cdot < dy_n^2 >}} = \frac{< dy_n \cdot dy_{n+k} >}{< dy_n^2 >} .$$

The expectation should be independent on the suffix *n*. (By the way, we have shortened the writing of $g_{11}^{(1)}(k)$ to g_k to simplify things.)

In total we get now:

$$\sigma^2_{\exp} = \frac{\sigma^2_{true}}{N} \cdot \left\{ N - 1 - 2 \cdot \sum_{k=1}^{N-1} (1 - \frac{k}{N}) \cdot g_k \right\}.$$

From this it follows that a correct estimate of the true variance of the data is given by:

$$\sigma^2_{true} = \frac{N}{N-1} \cdot \frac{\sigma^2_{\exp}}{1 - \frac{2}{N-1} \sum_{k=1}^{N-1} (1 - \frac{k}{N}) \cdot g_k} = \frac{\sum_{n=1}^{N} [y_n - \frac{1}{N} \cdot \sum_{m=1}^{N} y_m]^2}{N - 1 - 2 \cdot \sum_{k=1}^{N-1} (1 - \frac{k}{N}) \cdot g_k} \qquad [3]$$

It describes how the correlation influences the standard deviation of the data. Certainly, we do not have precise knowledge about the values g_k of the correlation function because of insufficient information, but it is probably good enough to use values found by the finite sums as given above (Eq.[2]). If one deals with the output of a single pixel instrument, it might be sufficient in some cases to determine consider the time constant of the system so that the values of g_k can be calculated using Eq.[1] above. It is a bit surprising that the influence of the correlation seems to reduce with N large, i.e. if N is much larger than the value of k, where g_k is approaching zero. But it becomes understandable, if one remembers that for small N one does not deal with a significant sample of data, so that the fluctuations are not fully covered. For sufficiently large N, or if there is no correlation between the data (white noise), the formula leads to the standard expression

$$\sigma^2_{true} \xrightarrow{N\,large} \frac{1}{N\text{-}1} \cdot \sum_{n=1}^{N} \left[y_n - \frac{1}{N} \cdot \sum_{m=1}^{N} y_m \right]^2, \qquad [4]$$

since a few terms in the sum in the denominator don't count much compared to a large N. The fairly complicated derivation here has the advantage that we do not need any particular assumption about the statistical distribution of the data. We just use the definition of the variance as given by the standard definition. (Whether or not this definition is reasonable in all cases might be subject for further discussion.)

The correlation plays only a significant role as long as N is rather small. This is a similar statement that one needs to observe much longer than the typical correlation time of the system determines. Nevertheless, for a correct estimate of the variance, instead of Eq.[4], one should use Eq.[3]. In case the values of the correlation function are well known, this expression provides a much better estimate of the true variance of the statistical distribution of the data. For example, if all data are fully correlated so that the values of g_k are equal to "1" for all k, both, the denominator and the numerator in Eq.[3] become zero. It means that the variance becomes undefined! This is quite "natural" since all data are identical in this case, so that a reasonable error estimate must be impossible. When neglecting the correlation correction, the variance according to Eq.[4] is clearly useless, since the calculated variance becomes now zero. This is particularly important, when estimating the radiometric noise seen on the baseline of a spectrum as observed with a real time spectrometer of a radiometer system while using only a few pixels of the spectrum for the evaluation of the baseline *rms* for example.

The error of the mean

More interesting is the error of the mean of several data points, since it is usually provided together with the result of an experiment. This is a more serious affair, because we need to determine the expected standard deviation of the mean of data y_n with:

$$z = \frac{1}{N} \cdot \sum_{n=1}^{N} y_n$$

with N the number of available data. With the usual definition of the variance

$$\sigma^2_N = < z^2 > - < z >^2$$

we can now determine, how the error of the new data develops.

$$\sigma^2_N = < \left[\frac{1}{N} \cdot \sum_{n=1}^{N} y_n \right]^2 > - < \frac{1}{N} \cdot \sum_{n=1}^{N} y_n >^2 =$$

$$= < \left[\frac{1}{N} \cdot \sum_{n=1}^{N} dy_n \right]^2 > = \frac{1}{N^2} \cdot \sum_{n=1}^{N} \sum_{m=1}^{N} < dy_n \cdot dy_m > = \frac{\sigma^2_y}{N^2} \cdot \sum_{n=1}^{N} \sum_{m=1}^{N} g_{n-m}$$

Since $\sigma_y^2 = g_{nn} = \sigma_{true}^2$, we have now finally:

$$\sigma_N^2 = \frac{\sigma_y^2}{N}\cdot\left\{1+2\cdot\sum_{k=1}^{N-1}(1-\tfrac{k}{N})\cdot g_k\right\} = \frac{\sigma_{true}^2}{N_{eff}}$$

with $$N_{eff} = \frac{N}{1+2\cdot\sum_{k=1}^{N-1}(1-\tfrac{k}{N})\cdot g_k} \qquad [5]$$

N_{eff} is the "effective" number of pixels which contribute to the statistics of the mean. Due the correlation one has less information available than one expects, because each data point contains a fraction of the noise information of the neighboring points. It should be noted that the value of σ_{true}^2 is determined using Eq.[3], as found before. In summary, we have now:

$$\sigma_N^2 = \frac{\sum_{n=1}^{N}\left[y_n-\frac{1}{N}\cdot\sum_{m=1}^{N}y_m\right]^2}{N-1-2\cdot\sum_{k=1}^{N-1}\left(1-\frac{k}{N}\right)\cdot g_k}\cdot\frac{1+2\cdot\sum_{k=1}^{N-1}\left(1-\frac{k}{N}\right)\cdot g_k}{N}.$$ Thus we have:

$$\boldsymbol{\sigma_{true}^2} = \frac{\sum_{n=1}^{N}\left[y_n-\frac{1}{N}\cdot\sum_{m=1}^{N}y_m\right]^2}{N-N/N_{eff})} \quad \text{and} \quad \boldsymbol{\sigma_N^2} = \frac{\sum_{n=1}^{N}\left[y_n-\frac{1}{N}\cdot\sum_{m=1}^{N}y_m\right]^2}{N\cdot(N_{eff}-1)} \qquad [6]$$

These results are definitely different from the usual formulas:

$$\sigma_{true}^2 = \frac{\sum_{n=1}^{N}\left[y_n-\frac{1}{N}\sum_{m=1}^{N}y_m\right]^2}{N-1} \quad \text{and} \quad \sigma_N^2 = \frac{\sum_{n=1}^{N}\left[y_n-\frac{1}{N}\sum_{m=1}^{N}y_m\right]^2}{N\cdot(N-1)}$$

Clearly, the effective number of available data N_{eff} is practically always smaller than the number of data N. If we have all N data fully correlated, N_{eff} becomes unity, and the denominator and the sum in the numerator both become zero. Both variances become therefore undefined. The influence of correlation persists even at large N. This is important when considering the accuracy of averages in general. It should be noted that the simple connection between the variance of the individual data points and the error of the mean is not valid anymore (Instead of $\sigma_N^2 = \sigma_{exp}^2/N$ we have now $\sigma_N^2 = \sigma_{true}^2/N_{eff}$). The result here verifies the trivial expectation that we cannot expect a mean with small error, if the data are strongly correlated.

Error of the correlation function

In an experimental situation, it is nearly impossible to derive true results for expectation values since we are always confronted with data sets of finite size. The typical assumption is that the data are affected by white noise only. In this scenario error estimates become rather simple and are usually given in many publications. But in general this might not be correct, since in many cases data are correlated with the consequence that one has to deal with non-white noise! It is therefore necessary to introduce other expressions, which can be used instead. We assume that we have data equidistant in time (or frequency or else), which are sorted by the index n. In total we have N data y_n. For the first order correlation function we have then:

$$g_k = \frac{< dy_n \cdot dy_{n+k} >}{\sqrt{< dy_n^2 > \cdot < dy_{n+k}^2 >}}$$

The index k could mean that the data have been collected with equidistant time intervals τ_0 so that $\tau = \tau_k$ with $\tau_k = k\cdot\tau_0$. It also may describe the channel numbers of a real time spectrometer, which are read out at the same time. k is then the difference between channel numbers.

The definition above is a bit theoretical because we do not know the true expectation of $dy_n \cdot dy_{n+k}$ and the expectation of y_n as well. Therefore one must find a useful approach for finite data sets in order to obtain a reasonable estimate. It is normal practice to replace the expectation by the mean. Thus we write:

$$g_k = \frac{\frac{1}{N-k-1}\cdot\sum_{n=1}^{N-k}\delta y_n\cdot\delta y_{n+k}}{\sqrt{\frac{1}{N-k-1}\cdot\sum_{n=1}^{N-k}\delta y_n^2\cdot\frac{1}{N-k-1}\cdot\sum_{n=k+1}^{N}\delta y_{n+k}^2}}$$

with $\delta y_n = y_n - \frac{1}{N-k}\cdot\sum_{m=1}^{N-k}y_m$ and $\delta y_{n+k} = y_{n+k} - \frac{1}{N-k}\cdot\sum_{m=k+1}^{N}y_m$

(Note that the δy_n are not identical with the dy_n.) The expectation values are here estimated as the means over an eventually large number of data. This definition assumes that all data y_n of the data set follow the same statistics. When using the above definition while dealing with spectrometer data, the mean value $1/(N\text{-}k)\cdot\Sigma y_m$ should be corrected by subtraction of unwanted, but systematic structures in the data set, which are obviously not of statistical origin. It is also important to note that the value of this correlation function is "1" for k=0 by definition.

Good estimates of the values of the correlation function are usually not available with only one data set of limited size. In order to get some feeling about the accuracy of the g_k we can try to evaluate the expected standard deviation of the correlation function with the assumption that we consider only the regimes with $<g_k> = 0$, i.e. for k very large. The data δy_n and δy_{n+k} should be now completely uncorrelated, since the correlation should disappear at large time or pixel separation. For an estimate we consider the value of

$$\sigma_g^2 = < g_k^2 > = \frac{\frac{1}{(N-k-1)^2}\cdot\sum_{n=1}^{N-k}\sum_{m=1}^{N-k} < \delta y_n \cdot \delta y_{n+k} \cdot \delta y_m \cdot \delta y_{m+k} >}{(\sigma^2)^2} \approx \frac{\frac{1}{(N-k-1)^2}\cdot\sum_{n=1}^{N-k}\sum_{m=1}^{N-k} < dy_n \cdot dy_{n+k} \cdot dy_m \cdot dy_{m+k} >}{(\sigma^2)^2}$$

(The error of the variance σ^2 is here neglected.) The values of δy_n and δy_{n+k} and of δy_m and δy_{m+k} are completely uncorrelated, since we assume k very large, so that the terms $<\delta y_n \cdot \delta y_{n+k} \cdot \delta y_m \cdot \delta y_{m+k}>$ can only be non-zero if n=m. (By the way, this is a good example of a higher order correlation function as was mentioned above.) In this case we have two statistically independent products $<\delta y_n^2 \cdot \delta y_{n+k}^2> = <\delta y_n^2> \cdot <\delta y_{n+k}^2> = (\sigma^2)^2$. Thus we find:

$$\sigma_g^2(k >> 1) = \frac{N-k}{(N-k-1)^2} \approx \frac{1}{N-k}$$

This means that the number of data N must be very large in order to obtain an acceptable error margin. For example, the standard deviation is 10% of the maximum (=1) for a data set with 100 data points. This is unacceptable in most cases. Therefore, if possible, one should use several statistically independent data sets in order to find a reasonable error estimate while averaging the values of g_k.

The role of correlation for a telescope

An optical telescope with one large aperture can be understood as a collector of many correlated signal amplitudes. If we subdivide the telescope surface area in N small pieces with signal amplitudes a_m, then we can write for the total signal:

$$S \propto < \left| \sum_m a_m \right|^2 > = \sum_m \sum_n < a_m \cdot a_n^* > = \sum_m \sum_n G_{mn}$$

G_{mn} is the mutual and not normalized correlation function between the sub-areas of the optics. We describe the partial beams by their amplitude and phase:

$$a_m = |a_m| \cdot e^{i\cdot\varphi_m}$$

and we can write: $G_{mn} = <|a_m| \cdot |a_n| \cdot e^{i(\varphi_m - \varphi_n)} > = <|a_m| \cdot |a_n|> \cdot < e^{i(\varphi_m - \varphi_n)} >$

We consider the averages over amplitude and phase as independent because: (i) the amplitudes are determined only by the illumination (from a point-like source and for a telescope of relatively small size) while (ii) the phases are determined by the optical path-lengths within the telescope optics. (If atmospheric influence is to be taken into account, one could consider it as part of the optics of the telescope.) Therefore, correlation between amplitudes and phases should not occur.

If we introduce now the relative correlation g_{mn} between the pieces we write:

$$g_{mn} = \frac{< a_m \cdot a_n^* >}{\sqrt{< |a_m|^2 > \cdot < |a_n|^2 >}} = \frac{< |a_m| \cdot |a_n| >}{\sqrt{< |a_m|^2 > \cdot < |a_n|^2 >}} \cdot < e^{i(\varphi_m - \varphi_n)} >$$

Both may be time-dependent so that $|a_m| = |a_m(t)|$ and $\varphi_m = \varphi_m(t)$. The phases φ_m include also the atmospheric fluctuations which could be dependent on position in the telescope aperture, but, for simplicity reasons, we consider here only a smaller telescope or a telescope with adaptive optics. In this case the time-variability of all amplitudes $|a_m|$ is identical. The same is true for the phases, if we consider an optically stable telescope system. If there are no imaging errors or statistical surface variations involved, the phases are all identical. In this case we have:

$$G_{mn} = < |a_m| \cdot |a_n| >$$

and we get the usual formula:

$$S \propto < \left| \sum_m a_m \right|^2 > = \sum_m \sum_n < a_m \cdot a_n^* > = \sum_m \sum_n |a_m| \cdot |a_n| = \left(\sum_m |a_m| \right)^2$$

This is the performance of a perfect telescope. All amplitudes add coherently and the intensity is therefore the square of the sum of the amplitudes. If the amplitudes are all identical ($a_m = a$), then we can rewrite

$$S \propto \left(\sum_m |a_m| \right)^2 = \left(\sum_m |a| \right)^2 = N^2 \cdot |a|^2$$

On the other hand, in case there is no correlation between the amplitudes a_n, then we have

$$G_{mn} = < |a_m| \cdot |a_n| > = < |a_m|^2 > \cdot \delta_{nm},$$

Again, if the amplitudes $|a_m| = |a|$ are all identical, then we have:

$$S \propto \left| \sum_m a_m \right|^2 = \sum_m \sum_n G_{mn} = |a|^2 \cdot \left(\sum_{m,n} \delta_{nm} \right) = N \cdot |a|^2$$

The incoherent superposition of the sub-amplitudes becomes only 1/N times the signal of a fully coherent superposition. This would be the result of a telescope surface with completely random surface variations (neglecting imaging errors). A realistic telescope will lie somewhere in between (see below: “Ruze formula”). At the same time the spatial resolution is partly lost, the received power is smeared over a larger image.

The Influence of the Surface Roughness of Mirrors: An application

A particular role plays partial correlation due to the perturbation of the reflected field amplitudes from a rough surface. We consider here only the influence of the varying phase differences occurring during the reflection from different parts of the mirror. Coarse variations we consider as imaging errors, which we don’t pursue here. The statistical imaging errors are in close connection to the quality of the mirror, since a rough surface causes scattering of the light and therefore reflection losses of the radiation. Similar, surface roughness on lenses has identical consequences. We assume that all phase variations are of purely random nature and we consider a statistical distribution of the surface errors t with:

$$W(t) \cdot dt = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{\left[-\frac{t^2}{2 \cdot \sigma^2}\right]} \cdot dt$$

“t” stands for the deviation of the real from the ideal surface. The variance, or better the standard deviation $\sqrt{\sigma^2}$, defines the *rms* value of the surface roughness. Since we investigate the result of the surface roughness on the imaging, the contributing phase error becomes twice the surface error *t*.

We want to know, how the consequences look like in the far-field pattern of the radiation. The degradation is immediately visible in the maximum intensity of the signal. We assume an amplitude distribution *A* of the incoming radiation like

$$A(x_m) = a_m$$

at the m-th location within the aperture of the mirror. Then we have in the center of the image for large distances, z → ∞ (or in the focus of the telescope), an amplitude like

$$A(0,0,z) \propto \frac{e^{i \cdot k \cdot z}}{z} \cdot \sum a_m$$

The sum extends over the complete surface of the mirror. Together with the surface errors this expression changes into

$$A(0,0,z) \propto \frac{e^{i \cdot k \cdot z}}{z} \cdot \sum a_m \cdot e^{2 \cdot i \cdot k \cdot t_m} = \frac{e^{i \cdot k \cdot z}}{z} \cdot \sum b_m$$

The amplitude distribution a_m becomes modified by the phase errors of the surface. For the center intensity we have then:

$$I(0,0,z) = C \cdot < |A(0,0,z)|^2 > = C \cdot < \sum_m \sum_n b_m \cdot b_n^* > = C \cdot \sum_m \sum_n < b_m \cdot b_n^* > = C \cdot \sum_m \sum_n G_{mn}$$

The constant C replaces all unimportant factors, which stand for the further details of the imaging. The G_{mn} are the values of the first order correlation function between the amplitudes at the locations *m* and *n*. If we consider not just single dish optics, but a system of a telescope ensemble, i.e. an interferometer antenna, then we might consider the amplitudes a_m and a_n just as single valued amplitudes for each antenna (assuming that each of the antennas has an ideal surface). The above formula is then the basis for the operation of array antennas like the VLA, SMA, ALMA, etc.

When interchanging the indices *m* and *n,* one obtains the conjugate complex value of g_{nm}, so that we can write now:

$$I(0,0,z) \;=\; C \cdot \left\{ \sum_m G_{mm} + 2 \cdot Re\left[\sum_m \sum_{n<m} G_{mn} \right] \right\}$$

The second part in the brackets explains that the total intensity is not just equal to the sum of the individual intensities (ΣG_{mm}). The outcome depends totally on the mutual correlation between the different amplitudes. For a point-like source the signal correlation is complete as long as the path lengths for all antennas are identical. For extended sources this is certainly not the case. In addition, the atmospheric fluctuations play a very important role, because they destroy the correlation between the amplitudes at the different positions.

When determining all values of the correlation function G_{mn} as a function of the phase difference between the various antennas a complete image of the sky within the main beam of one single dish can be constructed. This requires that the distance between the telescopes is varied, which in parts is already achieved when taking advantage of the earth rotation. The values of the correlation function are most easily found when using heterodyne receivers. With optical systems like the VLT it is more difficult.

Let us describe the amplitudes as

$$b_m \;=\; \lceil a_m \rceil \cdot e^{2 \cdot i \cdot k \cdot t_m}$$

An eventual phase of the a_m we may include in the t_m.They stand for the optical path length at locations *m*. We have now for the value of the correlation function:

$$G_{mn} \;=\; |a_m| \cdot |a_n| \cdot e^{2i \cdot k \cdot (t_m - t_n)}$$

and the intensity becomes now:

$$I(0,0,z) \;=\; C \cdot \left\{ \sum_m G_{mm} \;+\; 2 \cdot \sum_m \sum_{n<m} G_{mn}^{Max} . \cos[2 \cdot k \cdot (t_m - t_n)] \right\}, \quad G_{mn}^{Max} \;=\; |a_m| \cdot |a_n|$$

This expression becomes maximum for all path length differences equal to zero. Different positions on sky require new differences in the setting of the path length corrections between the telescopes. Thus, one can construct a full map of the brightness distribution on sky from a single series of measurements. By the way, this is more or less identical with the situation of a single dish telescope, where one gets a complete image in the focal plane of the optics.

Let us come back to the situation with a single dish antenna. We had:

$$I(0,0,z) \;=\; C \cdot < \lceil A(0,0,z) \rceil^2 > \;=\; C \cdot \langle \sum \sum \lceil a_m \rceil \cdot \lceil a_n \rceil \cdot e^{2 \cdot i \cdot k \cdot (t_m - t_n)} \rangle_t$$

We have to average over all path lengths *t* according to the above given probability distribution. We consider a fully random distribution of all t_m. Thus we can also assume that the differences t_m-t_n are also randomly distributed according to a Gaussian probability function. When building the average it is now important that the distribution of the illumination does not influence the statistics. Thus, when picking a certain position m with path length t_m, the statistics of the differences to this position is again the same as that of another position *n* so that the averaging can be done independently on the illumination function. This assumption is true for a Gaussian illumination function for example. Then we can rewrite

$$I(0,0,z) \;=\; C \cdot \langle \sum \sum \lceil a_m \rceil \cdot \lceil a_n \rceil \cdot e^{2 \cdot i \cdot k \cdot (t_m - t_n)} \rangle_t \;=\; C \cdot \sum \sum \lceil a_m \rceil \cdot \lceil a_n \rceil \cdot \langle e^{2 \cdot i \cdot k \cdot (t_m - t_n)} \rangle_t \;=$$

$$\Rightarrow \; C \cdot \sum \sum \lceil a_m \rceil \cdot \lceil a_n \rceil \cdot \int_{-\infty}^{\infty} d(\delta t) \cdot e^{2 \cdot i \cdot k \cdot \delta t} \cdot w(\delta t)$$

δt stands for the differences t_m-t_n. We consider the variations of a_m locally as zero, i.e. the illumination should be a slowly varying function across the aperture of the telescope like one would have with a plane wave radiation field.

Because of the doubled contribution of the surface errors the probability function of the differences t_m-t_n is given by:

$$W(\delta t)\cdot d\delta t \;=\; \frac{1}{\sqrt{4\cdot\pi\cdot\sigma^2}}\cdot e^{-\frac{\delta t^2}{4\cdot\sigma^2}}\cdot dt$$

Assuming that the $|a_n|$ are all equal we need to evaluate now the average over all path length differences.

$$S \;=\; \int_{-\infty}^{\infty} e^{2i\cdot k\cdot\delta t}\cdot W(\delta t)\cdot d(\delta t) \;=\; \frac{1}{\sqrt{4\cdot\pi\cdot\sigma^2}}\cdot\int_{-\infty}^{\infty}[\cos(2\cdot k\cdot\delta t)+i\cdot\sin(2\cdot k\cdot\delta t)]\cdot e^{\frac{\delta t^2}{4\cdot\sigma^2}}\cdot d(\delta t)$$

The integral over the Sin-function is zero, because it is an asymmetrical function. Thus we find finally:

$$S \;=\; \frac{1}{\sqrt{4\cdot\pi\cdot\sigma^2}}\cdot\int_{-\infty}^{\infty}\cos(2\cdot k\cdot\delta t)\cdot e^{-\frac{\delta t^2}{4\cdot\sigma^2}}\cdot d(\delta t) \;=\; e^{-4\cdot k^2\cdot\sigma^2} \;=\; \exp\left[-\left(\tfrac{4\cdot\pi\cdot\sigma}{\lambda}\right)^2\right]$$

while using $k = 2\cdot\pi/\lambda$. This formula is well known in the literature as “Ruze-formula”.

We therefore get for the expectation of the intensity at large distance *z*:

$$I(0,0,z) \;=\; C\cdot\sum\sum|a_m|\cdot|a_n|\cdot exp\left[-\left(\frac{4\pi\cdot\sigma}{\lambda}\right)^2\right] \;=\; I(0,0,z)_{Max}\cdot exp\left[-\left(\frac{4\pi\cdot\sigma}{\lambda}\right)^2\right]$$

The Ruze-factor represents therefore a direct measure of the losses, which are caused by surface roughness of the mirror. The formula is particularly important for radio-telescopes, because it determines the uppermost frequency, where the telescope can be used efficiently. At the same time it should also be a good tool to identify the usefulness of optical mirrors as laser mirrors for example. But, contrary to the general opinion that the accuracy of a mirror surface must not be better than $\lambda/20$ in order to avoid losses due to the surface quality, the Ruze-formula predicts that such mirror provides only 67% efficiency! The missing 33% become scattered into all directions, as is usually well visible, at least in parts. But, apparently it needs a surface quality of $\lambda/100$ for an efficiency of 98.4%! This is not quite in agreement with the experience with optical mirrors, and one can conclude that the simple picture is not fully valid for such optics. It might be possible that the micro-roughness of such mirrors is much better, so that unfavorable losses are not that high in reality.

The Auto-Correlation Function of various noise spectra

As is obvious from the discussion before, it is not only white noise which is of importance for the statistics of experimental data. An example we have seen above, if a time constant or some pre-integration is involved (see Eq.[1]). In general we have to consider all kinds of noise spectra, which have strong dependence on the frequency one is observing at. Noise with a stronger relevance of low frequency components we call “drift-noise”. Such noise is usually described by its spectral power distribution $S(f)$. In this chapter we consider the “auto-correlation function” $G(\tau)$ of such noise. The noise power spectrum is the Fourier transform of the first order auto-correlation function:

$$S(f) \;=\; \int_{-\infty}^{\infty} G(\tau)\cdot\exp(-2\pi i f\tau)\cdot d\tau \;=\; 2\cdot\int_{0}^{\infty} G(\tau)\cdot\cos(2\pi f\tau)\cdot d\tau \quad \text{and}$$

$$G(\tau) \;=\; \int_{-\infty}^{\infty} S(f)\cdot\exp(2\pi i f\tau)\cdot df \;=\; 2\cdot\int_{0}^{\infty} S(f)\cdot\cos(2\pi f\tau)\cdot df\,,$$

The simplified definition for non-negative frequencies is useful for our discussion, because the signal output of most instruments is assumed to be real. Consequently, the correlation function $G(\tau)$ has to be real valued and symmetric in time ($G(-\tau) = G(\tau)$). From this we have also that $S(-f) = S(f)$. (Note that we have simplified: $G(\tau) = G_{11}{}^{(1)}(\tau)$.)

It should be understood that we talk here about the “noise power spectrum” and not about the plain Fourier transform of a sequence of data. The noise power spectrum is found as the sum of the square of the Sin- and the Cos-Fourier transforms of the data. But it is much more adequate to determine the frequency distribution of the noise power by calculating the Fourier transform of the auto-correlation function of the data. Nevertheless, the relation between the power spectrum and the spectrum of the data can be found by means of the complex Fourier transform

$$A(f) \;=\; \int_{-\infty}^{\infty} a(t)\cdot\exp\{-2\pi i f t\}\cdot dt \quad \text{and} \quad a(t) \;=\; \int_{-\infty}^{\infty} A(f)\cdot\exp\{2\pi i f t\}\cdot df$$

a(t) describes the data stream as a function of time and *A(f)* is the Fourier-transform of *a(t)*. The power spectrum is then given by:

$$S(f) = |A(f)|^2$$

Only if the data stream *a(t)* is symmetric, i.e. if *a(-t) = a(t)*, then the simplified expression above applies. At the same time *A(f)* becomes real valued. In general the auto-correlation function $G(\tau)$ is real, so that we are allowed to use just the non-negative frequency range. (We prefer to use the description in frequencies *f* instead of ω, because it avoids all additional and annoying factors in front of the integrals.)

The noise power spectrum is constant for all frequencies in case of white noise, but it reaches zero for large frequencies f in case of drift noise. Often a power law like $1/f^{\alpha}$ is considered, and, in particular, the consequences of $1/f$ -noise ($\alpha = 1$: flicker noise) are widely discussed. In any case, *S(f)* should approach zero at a finite frequency, since any existing instrumentation has definitely a nonzero time constant. Thus, an upper cut-off frequency f_h for *S(f)* can be defined, at least for white noise ($\rightarrow$ band limited white noise).

For the following we make a few additional assumptions, which are at least "reasonable" in most cases.

1.) If the signal output *a(t)* of an instrument is measured continuously, the measured data points are all well confined within a statistical distribution with existing mean and variance. This is an essential assumption, because any instrumentation which does not obey this requirement would never provide reasonable data from the experiment. We can therefore evaluate the variance σ^2 as a finite and non-negative quantity:

$$\sigma^2 = <[a(t) - <a(t)>]^2> = <a^2(t)> - <a(t)>^2 = G(0) - G(\infty) \geq 0$$

On the other hand we have from the definition of the Fourier Transform from above:

$$G(0) = 2 \cdot \int_0^{\infty} S(f) \cdot df$$

G(0) is proportional to the total power contained in the noise spectrum. The power must be finite, because an instrument generating infinite power does not exist! In consequence, *G(∞)* is also finite (see also below). Thus we assume for example, that the signal output does not drift in one direction indefinitely, but fluctuates about an average value with a (eventually) rather long time constant.

2.) *G(τ)* decreases monotonically for all $\tau \geq 0$.
For two independent measurements of the signal, taken at different times, the result of the second measurement becomes statistically more and more independent on the first with increasing time delay between the two measurements. The signals become completely uncorrelated in case the samples are taken with very large time separation. This behaviour is typical for low frequency noise, which should be called here "drift noise with random statistics". For any scientific instrumentation this is how the drifts usually appear. (By the way, this definition rules out any interference in the sense that an increase of correlation with time is not allowed, as it would be possible with periodic perturbations for example.)

3.) *G(τ)* is finite for all τ.
This is a natural requirement when following the assumptions 1.) and 2.). It prohibits the occurrence of a singularity, in particular a singularity at $\tau = 0$ as a Dirac Delta function would imply for example. Therefore the assumption of white noise means in reality, that *G(τ)* behaves mathematically like a Delta-function in case the integration interval about zero is not too small. This is e.g. valid for band-limited white noise.

4.) *G(τ)* approaches a minimum value "sufficiently fast" for large τ.
If we neglect any non-zero average signal, we can assume that $G(\tau \rightarrow \infty) = 0$. It describes the situation of the noise alone, above which the signal is finally seen.

5.) *G(τ)* is an analytical function for all τ.
Thus, *G(τ)* can be expanded into a series expansion about any τ on the real axis. This assumption is essential for the validity of the Fourier transforms given above for all times τ and frequencies f respectively. In consequence there are no sudden jumps of *G(τ)* at any τ, but instead it is a rather smooth function. This is nothing unusual in a real experiment, but it does **not** really include all the properties of a digitized signal for example.

One consequence from the assumptions before is that $G(\tau) \geq 0$ for any τ (there is no anti-correlation). For the validity of the Fourier transform formulas from above it must be assumed (and is also reasonable) that:

$$\int_0^{\infty} G(\tau) \cdot d\tau = \int_0^{\infty} |G(\tau)| \cdot d\tau \quad \text{exists (or better, is finite).}$$

Thus we have for the value of the Fourier transform *S(f)* at zero frequency:

$S(0) = 2 \cdot \int_0^\infty G(\tau) \cdot d\tau$ is finite.

The expression "$G(\tau)$ approaches zero sufficiently fast for large τ" means, that the integral over $G(\tau)$ has to stay finite. Therefore we can conclude, that $G(\tau)$ must go to zero faster than $1/\tau$ for sufficiently large τ.

If the noise is analyzed by a power law of the spectral distribution $S(f)$ like

$$S(f) = \frac{S_\alpha}{f^\alpha},$$

one finds for the correlation function at zero delay:

$$G(0) = 2 \cdot \int_0^\infty S(f) \cdot df \quad \rightarrow \quad 2 \cdot \int_{f_l}^{f_h} S(f) \cdot df \text{ finite.}$$

Obviously, one needs to introduce a finite upper cut-off frequency f_h for all $\alpha \leq 1$, otherwise the $G(0)$ would become infinite. (This is particularly important for white noise with $\alpha = 0$.) On the other hand, it is also obvious that a lower cut-off frequency is needed for all $\alpha \geq 1$. Only for $\alpha = 1$, both limits in frequency are needed. An infinite power is impossible for any real instrument, and it means, that in reality there exists no noise power spectrum with $\alpha \geq 1$ which behaves like $1/f^\alpha$ at very low frequencies. Instead, the noise power spectrum may be assumed as constant for $f < f_l$ for example. This is also important for the treatment of so called "flicker" noise with $\alpha = 1$.

From the assumptions above it can also be concluded that the noise power spectrum $S(f)$ has no singularities at any frequency, that it is a "smooth" function, positive valued, and going to zero for very high frequencies. In particular, since $G(0)$ is finite, $S(f)$ must also go to zero faster than $1/f$ for large f. This implies that flicker noise with pure $1/f$ characteristics at both ends of the spectrum cannot exist in reality.

In order to verify that the assumptions for the correlation function are also valid for spectral power laws as they are typically considered in the literature, the corresponding correlation functions of $1/f^\alpha$ spectra are calculated below. But it should be pointed out that the usage of such spectral power spectra is not really suitable in cases, where the actual spectral distribution might be a mixture of a large number of different power laws. The problems arise due to the fact that such power laws do not represent an orthogonal basis-set for a convenient and complete description of the spectral distribution.

For convenience we assume now a power law of the noise spectrum for $\alpha > 0$ like

$$S_\alpha(f) = \begin{cases} S_\alpha(0) & 0 \leq f < f_l \\ S_\alpha(0) \cdot \frac{1/f^\alpha - 1/f_h^\alpha}{1/f_l^\alpha - 1/f_h^\alpha} & f_l \leq f < f_h \\ 0 & f \geq f_h \end{cases}$$

f_h is the upper cut-off frequency above which the noise power spectrum is assumed to be zero. The constant term $1/f_h^\alpha$ is subtracted in order to avoid any sudden jump of the spectrum at frequency f_h. f_l is the lower cut-off frequency, which must be introduced to avoid a singularity at $f=0$. The sudden change of the slope of the spectral curve at the f_l seems to be rather rough, but one can show that the arguments below are still valid with a smoother behaviour when approaching zero frequency.

For $\alpha = 0$ we use:

$$S_\alpha(f) = \begin{cases} S_0(0) & 0 \leq f < f_h \\ 0 & f \geq f_h \end{cases}$$

$\alpha < 0$ we do not consider, because it would violate assumption 2.) above.

The coefficient $S_\alpha(0)$ may be replaced by the total power contained in the noise with the assumed frequency dependence. This is possible because the power always stays finite for spectra as described above. The power P_α which is contained in the spectrum $S_\alpha(f)$ is proportional to the value of the correlation function $G_\alpha(\tau)$ at time $\tau = 0$. The correlation function $G_\alpha(\tau)$ describes only that portion of the total correlation function, which is related to the assumed spectral distribution.

$$P_\alpha \propto G_\alpha(\tau = 0) = 2 \cdot \int_0^\infty S_\alpha(f) \cdot df$$

Therefore we can replace $S_\alpha(0)$ by the power P_α or by the value of the correlation function at $\tau=0$. If we assume for the two cut-off frequencies that $f_l \ll f_h$ then we have:

$$G_\alpha(0) = 2 \cdot S_\alpha(0) \cdot f_l \cdot \begin{cases} \frac{\alpha}{\alpha-1} & \alpha > 1 \\ \log(\frac{f_h}{f_l}) & \alpha = 1 \\ \frac{\alpha}{1-\alpha} \cdot [\frac{f_h}{f_l}]^{1-\alpha} & 0 < \alpha < 1 \\ \frac{f_h}{f_l} & \alpha = 0 \end{cases}$$

Due to the definition of $S(f)$ above it is guaranteed that $G_\alpha(0)$ is finite for all values of α considered here.

We are interested in the correlation function at all applicable times τ ($\tau \gg f_h^{-1}$)

with $G(\tau) = 2 \cdot \int_0^\infty S(f) \cdot \cos(2\pi f\tau) \cdot df$

We find then, using asymptotic and series expansions of Confluent Hypergeometric Functions with $\alpha > 0$ ($\alpha \neq 2k+1$, k = 0, 1, 2, ...) (see e.g. Footnote [2]) with $1/f_h \ll \tau \ll 1/f_l$:

$$G_\alpha(\tau) = 2 \cdot S_\alpha(0) \cdot f_l \cdot \left[\sum_{n\geq 0} \frac{(-1)^n}{(2n+1)!} \cdot \frac{\alpha}{\alpha - 2n - 1} \cdot (2\pi f_l \tau)^{2n} - \frac{\pi/2}{\sin(\nu)} \cdot \frac{(2\pi f_l \tau)^{\alpha-1}}{\Gamma(\alpha)} \right]$$

$\nu = (\alpha\text{-}1) \cdot \pi/2$ and $\Gamma(x)$ is the Gamma-function. The leading and non-constant terms are therefore proportional to τ^2 for any α larger than 3, and proportional to $\tau^{\alpha-1}$ for small α ($0 < \alpha < 3$). For α within a small interval around $\alpha = 3$ the first term of the sum and the last term with $\tau^{\alpha-1}$ compete, and both must be taken into account. Note that for α near 2k+1 the singularity of the last term caused by $1/\cos(\nu)$ is removed by the corresponding term in the sum for $n = k$.

For $\alpha = 2k+1$, k = 0, 1, 2, ..., and again for $1/f_h \ll \tau \ll 1/f_l$ we find, using asymptotic and series expansion of Exponential Integrals [2]:

$$G_\alpha(\tau) = 2 \cdot S_{2k+1}(0) \cdot f_l \cdot \left[\sum_{n\geq 0}^{n\neq k} \frac{(-1)^n}{(2n+1)!} \cdot \frac{2k+1}{2k-2n} \cdot (2\pi f_l \tau)^{2n} + \frac{(-1)^k}{(2k)!} \cdot \{h_{2k+1} + \log(1/2\pi f_l \tau)\} \cdot (2\pi f_l \tau)^{2k} \right]$$

with $h_{2k+1} = \sum_{s=1}^{2k+1} \frac{1}{s} - \gamma$, γ = 0.577 215 664 9.. Euler's constant.

For the case of $0 < \alpha < 1$ the leading term is shown below.

Only the leading terms of the expansion of the correlation function are of interest. Therefore, while replacing $S_\alpha(0)$ by $G_\alpha(0)$ as given above, we can write in first order approximation for $1/f_h \ll \tau \ll 1/f_l$:

$$G_\alpha(\tau) = G_\alpha(0) \cdot \begin{cases} 1 - \frac{\alpha-1}{\alpha-3} \cdot \frac{(2\pi f_l \tau)^2}{6} & \alpha > 3, \quad f_h \to \infty \\ 1 - \{1.2561 \ldots + log(1/2\pi f_l \tau)\} \cdot \frac{(2\pi f_l \tau)^2}{3} & \alpha = 3, \quad f_h \to \infty \\ 1 - \frac{\pi/2}{sin[(\alpha-1)\cdot\pi/2]} \cdot \frac{(2\pi f_l \tau)^{\alpha-1}}{\Gamma(\alpha+1)} & 1 < \alpha < 3, \quad f_h \to \infty \\ \frac{0.4228 \ldots + \log(1/2\pi f_l \tau)}{\log(f_h/f_l)} & \alpha = 1 \\ \frac{\pi/2}{\sin[(1-\alpha)\cdot\pi/2]} \cdot \frac{1}{\Gamma(\alpha+1)} \cdot \frac{1}{(2\pi f_h \tau)^{1-\alpha}} & 0 < \alpha < 1, \quad f_l \to 0 \\ \frac{sin[2\pi f_h \tau]}{2\pi f_h \tau} \quad \overrightarrow{f_h\ large} \quad \frac{\delta(\tau)}{2f_h} & \alpha = 0, \quad f_l \to 0 \end{cases} \qquad [7]$$

For values of α near 3 the terms given for $\alpha > 3$ and for $\alpha < 3$ must be considered at the same time. Note, that for $\alpha > 1$ the lower cut-off frequency f_l is decisive and for $\alpha < 1$ the higher cut-off frequency f_h. For $\alpha = 1$ we need both. Note also that the correlation function of flicker noise has logarithmic behaviour! The correlation still decreases with increasing delay, but very slowly. This behaviour is frequently not recognized in literature.

Thus, it is shown that for noise power spectra $S(f) \propto 1/f^\alpha$ the correlation function $G(\tau)$ behaves as assumed

[2] Handbook of Mathematical Functions, M.Abramowitz and I.A.Stegun, Dover Publications, Inc., New York

within first order approximation. In particular, it is a monotonically decreasing function, and it is also evident, that $G(\tau)$ remains finite for $\tau \to 0$. Therefore all conclusions derived are valid for the noise power spectra considered here.

Generating arbitrary noise power spectra

Sometimes it might be useful to ask, how a data stream with a particular noise power law should look like. For white noise this is simple, a good random number generator will do it. But it is a lot more tedious to construct a data set with other power law characteristics. Quite often one finds in literature analyses of data by means of a straight forward Fourier analysis claiming that one can clearly recognize a 1/f- or flicker-noise behaviour. This is rather misleading, since the Fourier-analysis of a data set and the characteristics of the noise power law are something completely different. It might be a good exercise, although not of much practical value, to construct data sets with statistical behaviour following a given power law. For this one should look for a suitable filter, which is applied to a set of random data representing white noise. If we consider a filter function $u(t)$, which is filtering the data of a white noise source, we have for the output of the filter:

$$y(t) = \int_{-\infty}^{\infty} x(t') \cdot u(t-t') \cdot dt'$$

Or, in some cases, we have an asymmetric filter function with

$$y(t) = \int_{-\infty}^{t} x(t') \cdot u(t-t') \cdot dt'$$

The filter function should be normalized so that

$$\int_{-\infty}^{\infty} u(t-t') \cdot dt' = 1$$

This guarantees that the mean input is identical with the mean output of the filter. The lower and upper limit of the integral may also be finite so that experimental data can be treated as well. In this case we have:

$$y(t) = \int_{t-a}^{t+a} x(t') \cdot u(t-t') \cdot dt'$$

This allows to process finite data sets on-line. Usually we have to deal with discrete data sets. Then we can write:

$$y_n = \frac{1}{R} \cdot \sum_{k=0}^{K-1} x_{n-k} \cdot u_k \quad \text{with} \quad R = \sum_{k=0}^{K-1} u_k$$

or, when preferring a more symmetric expression:

$$y_n = \frac{1}{R} \cdot \sum_{k=-K}^{K} x_{n-k} \cdot u_k \quad \text{with} \quad R = \sum_{k=-K}^{K} u_k$$

The correlation function

$$\Gamma(\tau) = < x(t) \cdot x(t+\tau) >$$

should now represent white noise as described by a delta function. The resulting new correlation function $G(\tau)$ of the data $y(t)$ is then:

$$G(\tau) = < \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \cdot u(t-t_1) \cdot u(t+\tau-t_2) \cdot x(t_1) \cdot x(t_2) > = \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \cdot u(t-t_1) \cdot u(t+\tau-t_2) \cdot < x(t_1) \cdot x(t_2) > =$$

$$= \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \cdot u(t-t_1) \cdot u(t+\tau-t_2) \cdot \Gamma(t_2-t_1) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} ds \cdot u(x) \cdot u(x+\tau+s) \cdot \Gamma(s)$$

For the last step we have set $t\text{-}t_1 = x$ and $t_2\text{-}t_1 = s$. For the case of white noise this converts into:

$$G(\tau) = \Gamma_0 \cdot \int_{-\infty}^{\infty} u(x) \cdot u(x+\tau) \cdot dx$$

While knowing the new correlation function $G(\tau)$ we can now determine the spectral power distribution of the filtered data with

$$S(f) = \int_{-\infty}^{\infty} G(\tau)\cdot\exp\{-2\pi i f\tau\}\cdot d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \cdot u(t-t_1)\cdot u(t+\tau-t_2)\cdot\Gamma(t_2-t_1)\cdot\exp\{-2\pi i f\tau\}\cdot d\tau$$

Again, we can substitute now $t - t_1 = x$, $t_2 - t_1 = s$, and therefore $t + \tau - t_2 = x + \tau + s$. This leads to

$$S(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} ds \cdot u(x)\cdot\exp\{+2\pi i f\cdot x\}\cdot u(x+\tau+s)\cdot\exp\{-2\pi i f\cdot(\tau+x+s)\}\cdot\Gamma(s)\cdot\exp\{+2\pi i f\cdot s\}\cdot d\tau =$$

$$= \int_{-\infty}^{\infty} u(x)\cdot\exp\{+2\pi i f\cdot x\}\cdot dx\cdot\int_{-\infty}^{\infty} u(y)\cdot\exp\{-2\pi i f\cdot y\}\cdot dy\cdot\int_{-\infty}^{\infty}\Gamma(s)\cdot\exp\{+2\pi i f\cdot s\}\cdot ds =$$

$$= U^*(f)\cdot U(f)\cdot G^*(f) = |U(f)|^2\cdot G^*(f)$$

where we have used that the power spectrum of the input noise is given by

$$G(f) = \int_{-\infty}^{\infty}\Gamma(\tau)\cdot\exp\{-2\pi i f\cdot\tau\}\cdot d\tau \quad \text{and} \quad U(f) = \int_{-\infty}^{\infty} u(t)\cdot\exp\{-2\pi i f\cdot t\}\cdot dt\cdot$$

G(f) is a real function, since the autocorrelation function $\Gamma(\tau)$ is always symmetric in time for real valued data. Therefore we can drop the conjugate complex indicator and have therefore:

$$S(f) = |U(f)|^2\cdot G(f)$$

If the input is white noise ($\Gamma(\tau) = \Gamma_0\cdot\delta(\tau)$), then the expression is simply:

$$S(f) = \Gamma_0\cdot|U(f)|^2$$

Thus we have, when comparing:

$$|U(f)|^2 = \int_{-\infty}^{\infty} G(\tau)\cdot\exp\{-2\pi i f\cdot\tau\}\cdot d\tau = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} u(x)\cdot u(x+\tau)\cdot dx\right)\cdot\exp\{-2\pi i f\cdot\tau\}\cdot d\tau$$

The value of *G(f)* has to be real, which is guaranteed because $\Gamma(\tau)$ is symmetric in τ. In case *u(t)* is also symmetric in time, *U(f)* itself becomes real. This means now that we can determine *u(t)* as the real valued Fourier transform of $\sqrt{S(f)}$, if we assume a particular power spectrum $S(f) \sim |U(f)|^2$. When using this we are able to simulate any kind of noise power spectrum, when applying such time filter to a white noise data set, as can be generated by random number generators. It should be noted that it is not sufficient to just apply a Fourier transformation to a data set and make conclusions about the noise power spectrum itself from this transformation. My impression is that the outcome of such transformation with $1/f$-characteristics is frequently misinterpreted as $1/f$-noise-power spectrum. In reality, it is a $1/f^2$-noise-power spectrum, when considering the above description. Such behaviour is usually found, if a nearly constant drift in time is present. But this violates the assumption for the statistical behaviour of such data, because the data set would never be long enough to represent the complete statistics of the data set.

The Fourier transform of the filtered data is on the other hand given by transforming the new signal

$$y(t) = \int_{-\infty}^{\infty} x(t')\cdot u(t-t')\cdot dt'$$

into the frequency domain. The output signal is the convolution of the input signal and the filter function. Thus, the spectral amplitude distribution of the output data is now

$$A_{out}(f) = \int_{-\infty}^{\infty} y(t)\cdot\exp\{-2\pi i f t\}\cdot dt = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(t-t')\cdot x(t')\cdot\exp\{-2\pi i f t\}\cdot dt'\cdot dt =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(t-t')\cdot\exp\{-2\pi i f(t-t')\}\cdot x(t')\cdot\exp\{-2\pi i f t'\}\cdot dt'\cdot dt =$$

$$= \int_{-\infty}^{\infty} u(\tau)\cdot\exp\{-2\pi i f\tau\}\cdot d\tau\cdot\int_{-\infty}^{\infty} x(t')\cdot\exp\{-2\pi i f t'\}\cdot dt'$$

Thus we have the well-known result that the spectral amplitude distribution of the output of the filter, i.e. the convolution of the input signal and the filter function, is given by the product of the amplitude spectrum of the input signal and the spectral amplitude distribution of the filter. But this is not the power spectrum!

$$A_{out}(f) = U(f)\cdot A_{In}(f)$$

If we take the absolute square of $A_{out}(f)$, we get the same result as derived before: The power spectrum of a

signal is the absolute square of the spectral filter function times the power spectrum of the input. As an example, let us first concentrate on an asymmetric exponential filter with

$$u(t-t') \;=\; \begin{cases} \dfrac{1}{\tau_s}\cdot\exp\{-(t-t')/\tau_s\}, & t \geq t' \\ 0 & t < t' \end{cases}$$

This corresponds to the behaviour of a simple RC-filter. When evaluating the Fourier-transform we get:

$$U(f) \;=\; \frac{1}{1+2\pi i\cdot f\cdot\tau_s}$$

The power spectrum on the other hand is given by the Fourier transform of the correlation function of the filtered white noise:

$$G(\tau) \;=\; \frac{\Gamma_0}{2\tau_s}\cdot\exp\{-|\tau|/\tau_s\}$$

And with white noise input we obtain:

$$|U(f)|^2 \;\propto\; S(f) \;=\; \frac{\Gamma_0}{1+(2\pi f\tau_s)^2} \;=\; \frac{\Gamma_0}{2\cdot\tau_s}\cdot\frac{\delta f/2\pi}{f^2+(\delta f/2)^2} \quad \text{with} \quad \delta f \;=\; 1/\pi\tau_s$$

δf is the full half width (FWHM) of the spectral distribution (including the negative frequency part!). The exponential filter leads to a $1/f^2$ spectral power law, but is finite at $f = 0$. This corresponds to the absolute square of the Fourier-transformed filter function which shows also the $1/f^2$ behaviour. By the way, the result justifies the above assumption of a finite value of *S(0)*, when considering drift noise as the result of a time constant in the system for example. It seems that the $1/f^2$ behaviour is rather common in experimental physics. If the input is just white noise, the output spectrum is the product of the input spectral distribution – i.e. a constant for white noise – with the filter function.

One simple conclusion can be found from the above mathematics: What happens, if one uses a series of such filters? This is now simple to answer, because we can use the convolution theorem iteratively. If we have a series of filters, each with its own time constant τ_n with $n = 1,2,\ldots,N$, then we can write:

$$S(f) \;=\; |U_1(f)|^2\cdot|U_2(f)|^2\cdot\ldots\cdot|U_N(f)|^2\cdot G(f)$$

If the input is again considered as white noise, we have therefore:

$$S(f) \;=\; \Gamma_0\cdot\frac{1}{1+(2\pi f\tau_1)^2}\cdot\frac{1}{1+(2\pi f\tau_2)^2}\cdot\ldots\cdot\frac{1}{1+(2\pi f\tau_N)^2} \xrightarrow{f>>\tau_n} \frac{1}{(\tau_1\cdot\tau_2\cdot\ldots\cdot\tau_N)^2}\cdot\frac{\Gamma_0}{(2\pi f)^{2N}}$$

Thus we can conclude, that various time constants in a system contribute to the noise power spectrum like a N-pole filter. This is also correct for LC-filters, as they are used in RF-technology. The roll-off of such filter combinations is equal to $(f-f_0)^{2N}$, if f_0 is the resonance frequency of the filters. This result seems rather trivial, but is the basis for all multi-pole filters in RF-technology..

It is interesting how the Fourier-transform of *u(t-t')* and the noise power spectrum look like in special cases. The case of an asymmetric RC-filter we have considered above. The result looks different for a symmetric exponential filter function. In this case we have:

$$u(t-t') \;=\; \frac{1}{\tau_s}\cdot\exp\{-|t-t'|/\tau_s\} \quad \text{for all } t \text{ and } t'.$$

The correlation function is then:

$$G(\tau) \;=\; \frac{\Gamma_0}{\tau_s}\cdot\exp\{-|\tau|/\tau_s\}\cdot(1+|\tau|/\tau_s)$$

This correlation function generates a power spectrum like

$$S(f) \;=\; \frac{4\cdot\Gamma_0}{\left[1+(2\pi f\tau_s)^2\right]^2},$$

while the amplitude Fourier transform is:

$$U(f) \;=\; \frac{2}{1+(2\pi f\tau_s)^2}.$$

When comparing with the asymmetric case the result looks a bit surprising, and it is a good demonstration, how important the small details may become, when discussing the effects of different filters. The asymmetric function generates a $1/f^2$ power spectrum while the symmetric function generates a $1/f^4$ power spectrum (at f

large).

Another example is a symmetric Gaussian filter with

$$u(t-t') \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\{-\frac{(t-t')^2}{2\sigma^2}\} \;=\; \frac{1}{\Delta}\cdot\exp\{-\pi\cdot\frac{(t-t')^2}{\Delta^2}\} \quad \text{with} \quad \Delta \;=\; \sqrt{2\pi\sigma^2}$$

Δ stands for the effective time width of the filter:

$$\Delta \;=\; \frac{\int_{-\infty}^{\infty} u(t)\cdot dt}{u_{Max}}$$

It is the equivalent width of a box-car filter with identical peak response of the filter function. We find here for the correlation function:

$$G(\tau) \;=\; \frac{\Gamma_0}{\sqrt{4\pi\sigma^2}}\cdot\exp\{-\frac{\tau^2}{4\sigma^2}\} \;=\; \frac{\Gamma_0}{\sqrt{2}\cdot\Delta}\cdot\exp\{-\pi\cdot\frac{\tau^2}{2\cdot\Delta^2}\}$$

The power spectrum is now:

$$S(f) \;=\; \Gamma_0\cdot\exp\{-(2\pi f\sigma)^2\} \;=\; \Gamma_0\cdot\exp\{-2\pi(f\cdot\Delta)^2\}$$

This spectral distribution rolls off faster than any power of $1/f^2$ (f large), but is again finite at f=0. The FWHM δf of the frequency distribution is given by:

$$\delta f \;=\; \sqrt{\frac{2\cdot\log(2)}{\pi}}\cdot\frac{1}{\Delta} \;=\; \frac{0.6643}{\Delta}$$

It should be pointed out again that it is not possible to describe random noise with an analytic amplitude function of time or else. Instead, it is only the spectral power distribution one can describe analytically, which is the Fourier-transform of the auto-correlation function of the amplitude. At the same time, the spectral power distribution is the absolute square of the spectral amplitude distribution, if such distribution could be established. This makes things a bit complicated. The easiest way to generate artificial data with a particular spectral power distribution is to generate first white noise data by using a good random number generator. Following is a Fourier-transform of these data, which leads to a unique constant, but noisy spectral distribution of the amplitude. When multiplying with a spectral function, which is equal to the square root of the desired power distribution, one obtains a new spectrum which represents the desired spectral power distribution. With a backward Fourier transformation we get finally a data set, which has the wanted characteristics. This can be proven by the calculation of the correlation function and its Fourier transform. The new amplitude function will look different for any new white noise input data set no matter what the length of the data set is. Therefore, any new set will be different from any other set. It is clear that a power law like $1/f^{\alpha}$ cannot be produced for arbitrary small values of f. Therefore, again, this makes it evident that something like a lower cut-off frequency or similar must exist.

The situation during "real" measurements

The arguments before are valid for the signals as they are measured by any instrument. In general, one does not investigate the behavior of the instantaneous signal function *x(t)*, but instead one pre-integrates the signal over some time interval "*T*". If we describe the statistical behavior of an instantaneous signal function *x(t)* by its original correlation function

$$G(\tau) \;=< x(t)\cdot x(t+\tau) >,$$

then we can determine the correlation function of the integrated signal. (For simplicity we skip the use of the sums in the formulas.) Usually, in a typical experiment, the measured signal is the average of the signal function *x(t')* taken over a time interval *T* beginning at time *t* and not the instantaneous signal voltage itself. Hence we have for the measured signal:

$$y(t,T) \;=\; \tfrac{1}{T}\cdot\int_{t-T}^{t} x(t')\cdot dt'$$

(Here we consider boxcar integration of the signal, an integrator with exponential characteristics does not change the following arguments.) For the expectation value of *y(t,T)* we have now for example:

$$< y(t,T) >_t \;=\; \lim_{t\to\infty}\tfrac{1}{t}\cdot\int_0^t y(t',T)\cdot dt' \;=\; < x(t) >_t$$

It means that the expectation of the integrated signal is identical with the expectation of the not integrated signal. For this "new" signal we can define a new correlation function $G_T(\tau)$ on the basis of the initial correlation function $\Gamma(\tau)$ with

$$G_T(\tau) = < y(t,T) \cdot y(t+\tau,T) >_t = < \tfrac{1}{T^2} \cdot \int_{t-T}^{t} dt_1 \int_{t+\tau-T}^{t+\tau} dt_2 \cdot x(t_1) \cdot x(t_2) >_t =$$

$$= \tfrac{1}{T^2} \cdot \int_{t-T}^{t} dt_1 \int_{t+\tau-T}^{t+\tau} dt_2 \cdot < x(t_1) \cdot x(t_2) >_t = \tfrac{1}{T} \cdot \int_{-T}^{T} (1 - \tfrac{|\tau'|}{T}) \cdot \Gamma(\tau+\tau') \cdot d\tau'$$

τ is the time separation between the begins of the two integrations with length T.

In order to describe the statistics of the integrated data the variance of the statistical distribution of the data can now be evaluated:

$$\sigma^2(T) = < (y(t,T) - < y(t',T) >_{t'})^2 >_t = < y^2(t,T) >_t - < y(t,T) >_t^2 =$$

$$= \tfrac{1}{T} \cdot \int_{-T}^{T} (1 - \tfrac{|\tau|}{T}) \cdot [\Gamma(\tau) - \lim_{s \to \infty} \Gamma(s \cdot \tau)] \cdot d\tau = G_T(0) - G_T(\infty)$$

$$= G_T(0) - < x(t) >^2$$

With a zero mean of $x(t)$ we have finally:

$$\sigma^2(T) = G_T(0)$$

If we are dealing with white noise, we have:

$$\Gamma(\tau) = \Gamma_0 \cdot \delta(\tau)$$

Thus, we get for the new correlation function:

$$\Gamma_T(\tau) = (1 - \tfrac{|\tau|}{T}) \cdot \frac{\Gamma_0}{T}$$

and for the variance:

$$\sigma^2(T) = \frac{\Gamma_0}{T} \qquad [8]$$

We have shown before that the correlation function of a simple exponential filter is given by:

$$\Gamma(\tau) = \frac{1}{2 \cdot \tau_s} \cdot \exp\{-|\tau| / \tau_s\} \cdot$$

In this case we find:

$$\sigma^2(T) = \frac{\Gamma_0}{T} \cdot \left[1 - \frac{1 - \exp\{-T/\tau_s\}}{T/\tau_s} \right]$$

In the limiting case of very small T as compared with τ_s we get:

$\sigma^2(T) = \frac{\Gamma_0}{2 \cdot \tau_s}$, and for T very large we have: $\sigma^2(T) = \frac{\Gamma_0}{T}$.

The second case is unaffected by the time constant in the system, whereas small integration times don't have an effect on the statistics when comparing with no integration.

Measuring Differences

During measurements with a scientific instrument it is quite normal that a "zero"-measurement must be made subtracting all contributions of the instrument or the system itself. The difference between the "signal"-measurement and the "zero"-measurement contains the experimental information needed. Thus the influence of the "zero"- or "reference"-measurement on the statistical error must be considered. In our case we are interested in the situation where the information derived from the difference of signal and the reference is only a very small fraction of the background as is measured during the reference phase.

Both, the signal- and the reference-measurement last for an integration time of length T. In addition, there is a delay time between the end of the first and the beginning of the second of length T_D, thus we have two measurements taken at different times t and $t+T+T_D$:

$$y_1 \;=\; y(t,T) \;=\; \frac{1}{T}\cdot\int_{t-T}^{t} x(t')\cdot dt',\quad y_2 \;=\; y(t+T+T_D,T) \;=\; \frac{1}{T}\cdot\int_{t+T_D}^{t+T+T_D} x(t')\cdot dt'.$$

We are interested in the result:

$$< D(t,T_D,T) > \;=\; < y_1 - y_2 > \;=\; < y(t,T) - y(t+T_D,T) >,$$

which we assume to be zero on average for the moment. For the statistics the variance of the statistical distribution of $D(t,T_D,T)$ must be considered:

$$\sigma_D^2(T_D,T) \;=\; < [D(t,T_D,T) - < D(t,T_D,T >)]^2 > \;=\; < D^2(t,T_D,T) > \;-\; < D(t,T_D,T) >^2$$

Since we have:

$< D(t,T_D,T) > \;=\; 0$, we can rewrite:

$$\sigma_D^2(T_D,T) \;=\; < [y_1 - y_2]^2 > \;=\; < [y(t,T) - y(t+T+T_D,T)]^2 > \;=\; 2\cdot[G_T(0) \;- G_T(T+T_D)] \quad [9]$$

When dealing with actual data, this establishes the connection between the expected statistical distribution and the correlation function of the data. It reflects the fact that the error of difference measurements is reduced, if there is correlation between the two data y_1 and y_2. Such correlation may be introduced by drift noise for example, and its influence stays small as long as the delay T_D is short compared ti the drift time constant.

The Allan-variance

In the past it has been demonstrated that the measurement of the so-called “Allan-variance” is an extremely useful tool to characterize the noise performance and thereby the quality of measuring instruments like radiometer systems[3]. The idea is to determine the errors one has to expect when taking data as differences of two independent measurements. The first may be a reference with the signal turned off and the second a measurement together with a signal. The difference takes away the undesired background which may be the major part of the output of the instrument. This scenario is typical when using Lock-In amplifiers, but is also standard during observations with a radio-telescope for example. The Allan variance is the ideal procedure to determine the error of differences of noisy data. The issue is here that data closely spaced in time (or space) are usually partly correlated due to common drifts in the system for example. It is therefore important to understand the influence of the correlation. When measuring a signal as the difference between two adjacent samples, it is clear that long term drift will not have much impact on the result, if the drift is slow in comparison to the time difference between the samples. The widely used method to understand the influence of the drifts is the measurement of the Allan variance as a function of time difference between two samples. The standard definition of the Allan variance is:

$$\sigma_A^2 \;=\; \tfrac{1}{2}\cdot < [y_1 - y_2]^2 >$$

This is identical with the usual definition of the variance using N=2 (see above).

$$\sigma_N^2 \;=\; \frac{1}{N-1}\cdot < \sum_{n=1}^{N}\left[y_n - \tfrac{1}{N}\cdot\sum_{m=1}^{N} y_m \right]^2 > \xrightarrow{N=2} \tfrac{1}{2-1}\cdot < \left\{ [y_1 - \tfrac{1}{2}\cdot(y_1+y_2)]^2 + [y_2 - \tfrac{1}{2}\cdot(y_1+y_2)]^2 \right\} > \;=\; \tfrac{1}{2}\cdot < [y_1 - y_2]^2 >$$

By the way, this is the route how Allan started his definition of the Allan variance for the characterization of time standards [4]. The two data samples y_1 and y_2, derived for example from one output channel of an instrument, are usually taken at different times.

In comparison, we have for the variance of difference measurements::

$$\sigma_D^2 \;=\; < [y_1 - y_2]^2 > \;=\; 2\cdot\sigma_A^2$$

Each data point is the result of an integration with time “T” beginning at time “t” and “$t+T_D$” as discussed above.

For the Allan variance we have therefore:

$$\sigma_A^2 \;=\; ½\cdot\sigma_D^2 \;=\; G_T(0) - G_T(T+T_D) \qquad [10]$$

Thus, the Allan variance describes the situation of difference measurements very accurately. Only the additional factor of ½ is to be considered. One could interpret this as the contribution of each data sample to the

---

[3] R.Schieder, G.Rau, B.Vowinkel, *Characterisation and Measurement of System Stability*, Proceedings of SPIE, Vol. 598, page 189 (1985)

[4] D.W. Allan, *Statistics of Atomic Frequency Standard*, Proceedings of the IEEE, 54, No. 2, 221-231 (1966)

error of the difference. Although we consider here the statistical error of difference measurements σ_D^2, we will use the Allan variance σ_A^2 for the further discussion, because it is already well established in literature. In addition, we assume for a meaningful definition of the Allan variance that T_D should be zero in order to make different measurements comparable. If we would introduce time delay between them, and If there would exist only white noise, the statistical error of the difference would be independent on the delay. In this case we find a constant value for the variance. But with the presence of drift noise the resulting variance starts to increase with increasing time delay between the samples, because the drift noise starts to contribute more and more. This explains why it is important to define the experimental method with a clear rule concerning the delay.

We can now continue with

$$\sigma_A^2 \;=\; \tfrac{1}{2}\cdot <[dy_1 - dy_2]^2> \;=\; \tfrac{1}{2}\cdot\left[< dy_1^2 > + < dy_2^2 >\right] - < dy_1 \cdot dy_2 >$$

while using: $dy_i \;=\; y_i - < y_i >,\; i = 1,2$

The expectation of the product $dy_1 \cdot dy_2$ can be converted by using the definition of the normalized first order correlation function g_{12}, as is defined above:

$$g_{12} \;=\; \frac{< dy_1 \cdot dy_2 >}{\sqrt{< dy_1^2 > \cdot < dy_2^2 >}}$$

With this we get now:

$$\sigma_A^2 \;=\; \tfrac{1}{2}\cdot\left[< dy_1^2 > + < dy_2^2 >\right] - \sqrt{< dy_1^2 > \cdot < dy_2^2 >}\cdot g_{12}$$

The variances σ_y^2 are defined as:

$\sigma_i^2 \;=\; < dy_i^2 >$, and we get therefore:

$$\sigma_A^2 \;=\; \tfrac{1}{2}\cdot\left(\sqrt{\sigma_1^2} - \sqrt{\sigma_2^2}\right)^2 + \sqrt{\sigma_1^2 \cdot \sigma_2^2}\cdot(1 - g_{12})$$

In most cases the variances of the two samples y_1 and y_2 are the same. Then we can drop the indices and get finally:

$$\sigma_A^2 \;=\; \tfrac{1}{2}\cdot\sigma_D^2 \;=\; \sigma_y^2\cdot(1 - g_{12}) \qquad [11]$$

with $\sigma_y^2 \;=\; \sigma_1^2 \;=\; \sigma_2^2$

From this expression we learn that the major distinction between the standard variance σ_y^2 and the Allan variance σ_A^2 is due to the influence of correlation between the data. The Allan variance is always smaller then or, at worst, equal to the ordinary variance. (We assume that anti-correlation does not exist, as is generally true for data which exhibit some kind of random statistics.)

The correlation between the data is decisive for the success of the application of a Lock-In amplifier for example. If g_{12} would be zero, the Lock-In would be useless! It is only the correlation, i.e. the drift characteristics of the system, which decides about the success of such instrumental effort. It also says that in case of pure white noise, where no correlation between adjacent data exists, there is no possibility to improve the signal to noise ratio! In other words: The success of the use of a Lock-In depends on the ratio of the amplitude of the signal modulation versus the amplitude of the noise fluctuations at the modulation frequency, i.e. at the inverse of the time separation between the two phases of the modulation. If the noise power spectrum rolls off with frequency, there is a chance to improve the signal to noise ratio (S/R). But in many cases the signal amplitude also becomes smaller at higher modulation frequencies so that the gain in *S/R* may be less than one is hoping for. It is frequently stated in literature that the application of a Lock-In has the additional advantage that one observes only those noise contributions, which are exactly in phase with the modulation. Therefore, one should gain some additional signal to noise. This is not correct, when assuming a purely statistical behavior of the noise. Just consider a rectangular modulation, where one sees noise in both phases of the modulation. Taking the difference of them results in the same noise contribution as one would see while no modulation is applied as long as white noise is concerned. Only the contributions of correlated noise, i.e. drift-noise will be suppressed, which is exactly the purpose of the Lock-In method. But noise phases cannot play any role here!

The question remains, what happens to the variance when introducing a delay T_D? For example, when doing difference measurements with a box-car integrator, it is typical that the instrument integrates for some short and fixed integration time T, and, after some delay time T_D, another sample is integrated with the same integration time T. The variance is then dependent on the time length T of the integration as well as eventually on the delay time T_D between the two samples. For very small T and T_D, only the white noise contribution is visible since the correlation of the drift contribution is at maximum. Consequently, the variance is only determined by the white noise level in the signal. With increasing T_D, that is with increasing influence of drifts the correla-

tion becomes smaller. Therefore, a plot of the variance as a function of T_D starts at some lower constant level and increases according to the decrease of the correlation.

Nevertheless, for a practical and unique definition of the experimental conditions for Allan variance measurements we assume that the delay T_D is zero, and we determine the Allan variance as a function of the integration time T only. This is related to the more practical situation, when modulating the signal with different modulation frequencies, which corresponds to different integration times T. When knowing the increase of σ_A^2 due to drifts with integration time T, it is then a straightforward matter to calculate the effect of additional delay T_D.

We have seen before that the correlation function of white noise is a Delta-function. According to Eq.[8], we have therefore the variance behaving like 1/T. How drift noise appears in the Allan plot can be calculated from the correlation function, as was found before for various noise spectra. Using the spectral power law $S(f) \sim 1/f^{\alpha}$ we have found the approximation for the correlation function (Eq.[7])

$$G_\alpha(\tau) \;=\; G_\alpha(0) - g_\alpha \cdot \tau^{\alpha-1} \quad \text{for } 1 < \alpha < 3 \qquad [12]$$

Drift noise shows as an increase of the variance, which can be calculated using Eq.[10]:

$$\sigma_A^2 \;=\; G_T(0) - G_T(T) \;=\; \frac{1}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)\cdot[G_\alpha(\tau)-G_\alpha(T+\tau)]\cdot d\tau \;=$$

$$=\; \frac{1}{T^2}\cdot\int_0^T (T-\tau)\cdot[2\cdot G_\alpha(\tau) - G_\alpha(T+\tau) - G_\alpha(T-\tau)]\cdot d\tau \;=\; g_\alpha\cdot\frac{4\cdot(2^{\alpha-1}-1)}{\alpha\cdot(\alpha-1)}\cdot T^{\alpha-1}$$

Together with white noise we have therefore:

$$\sigma_A^2 \;=\; a/T + b\cdot T^{\beta} \quad \text{with } \beta = \alpha\text{-}1$$

At short integration time the variance becomes less since the white noise contribution still dominates. At longer integration time the correlation due to low frequency noise (drifts) becomes less, the variance increases again. Consequently the variance passes through a minimum, and increases with some power of T behind the minimum. At very large T the drift contribution decreases again like 1/T after it has gone through a maximum at an integration time which corresponds to the mean correlation time of the signal output of the instrument, as long as it has statistical behavior. The time regime near or above the maximum is not of practical interest, since taking differences does not provide a significant advantage any more. The minimum in the Allan variance plot is now found at

$$T_A \;=\; \left(\tfrac{1}{\beta}\cdot\frac{a}{b}\right)^{\frac{1}{\beta+1}}$$

and we can rewrite:

$$\sigma_A^2 \;=\; a/T_A\cdot\left[\frac{1}{t}+\tfrac{1}{\beta}\cdot t^{\beta}\right] \quad \text{with } t = T/T_A \qquad [13]$$

Typically, the coefficient β has values between 1 and 2, most of the time it is near 2..

This expression shows that the statistical performance of a system can be described by only three parameters: the white noise level a, the Allan variance minimum time T_A, and the power coefficient β, which describes the increase due to the presence of drift noise. The value of a is determined by the white noise level and can be replaced by the corresponding factors in the radiometer equation, if radiometers are considered for example ($a = \langle y \rangle^2/B_{FI}$, B_{FI} is the "fluctuation bandwidth of the system). If one understands the characteristics of the drift noise performance, i.e. if one knows the coefficient β which describes the influence of the drift noise, one can now calculate the behavior of the system as a function of any additional delay T_D.

Long term drift automatically means correlation between the data as long as they are not very distant in time. Thus the Allan plot reveals, in what time regime the data are still sufficiently correlated so that the contribution due to the drift component is negligible when taking the difference. This information is precisely what is needed for the decision about the speed of an applied signal modulation when using Lock-In amplifiers or other observing mode strategies with radio-telescopes for example. The correlation function itself is always a function of the time separation between the two data. In general the correlation should disappear for very large distances, otherwise the statistics of the signal is not as one should expect for a reasonable experimental setup.

The example of the Allan variance indicates that correlation between data is a very common phenomenon in most experimental situations. Most of the time, signals are detected as differences between two "independent" measurements, one with the signal turned on and a second "zero-measurement". This procedure is only meaningful, if one expects some long-term change of the instruments output, which may be removed by the switching procedure. It is therefore important to understand the influence of the correlation on the outcome of

the statistical analysis of the experiment. The Allan variance is an excellent tool to learn about such problems. It should be understood that this procedure is applicable to any measuring system. It is not just a procedure suitable for radiometers!

Calculation of the Allan Variance for various noise power spectra

Despite the well-known behaviour of the variance with white noise only, one has to determine as well, how the different drift noise spectra influence the Allan variance. In order to evaluate this one has to find the values of (see Equ.[10]):

$$\sigma_A^2(T_D,T) \;=\; G_T(0) - G_T(T+T_D)\,.$$

The results obtained when using simple power series expansion for $G(\tau)$ are given earlier. Similar results one finds when using the power spectrum of the noise $S(f)$. Although most of these results are given already in Ref.[5]. we present the somewhat refined results here again. We have:

$$\sigma_A^2 \;=\; \frac{1}{T}\cdot\int_{-T}^{T}(1-\frac{|\tau|}{T})\cdot[\Gamma(\tau)-\Gamma(T+T_D+\tau)]\cdot d\tau \;=\; \frac{1}{T}\cdot\int_{0}^{T}(1-\frac{\tau}{T})\cdot[2\cdot\Gamma(\tau)-\Gamma(T+T_D+\tau)-\Gamma(T+T_D-\tau)]\cdot d\tau \;=$$

$$=\; \frac{2}{T}\cdot\int_{0}^{\infty}S(f)\cdot df\int_{0}^{T}(1-\frac{\tau}{T})\cdot[2\cdot\cos(2\pi f\tau)-\cos\{2\pi f\,(T+T_D+\tau)\}-\cos\{2\pi f\,(T+T_D-\tau)\}]\cdot d\tau \;=$$

$$=\; \frac{8}{T}\cdot\int_{0}^{\infty}S(f)\cdot\sin^2\{\pi f\,(T+T_D)\}\cdot\int_{0}^{T}(1-\frac{\tau}{T})\cdot\cos(2\pi f\tau)\cdot d\tau\cdot df \;=\; 4\cdot\int_{0}^{\infty}S(f)\cdot\sin^2\{\pi f\,(T+T_D)\}\frac{\sin^2(\pi fT)}{(\pi fT)^2}\cdot df$$

From this one obtains for the "ordinary" Allan variance (T_D = 0):

$$\sigma_A^2(T) \;=\; 4\cdot\int_{0}^{\infty}S(f)\cdot\frac{\sin^4(\pi fT)}{(\pi fT)^2}\cdot df$$

We set now $S(f) = S_\alpha(f)$, as defined above, and integrate first from 0 to f_l and then from f_l to f_h. When integrating we get for $\alpha > 0$, $\alpha \neq 2k+1$ (k = 0,1,2,...) and for $1/f_h \ll T \ll 1/f_l$

$$\sigma_A^2(T) \;=\; S_\alpha(0)\cdot 8\cdot f_l\cdot\left[\frac{2^{\alpha-1}-1}{\Gamma(\alpha+2)}\cdot\frac{\pi/2}{\sin(v)}\cdot(2\pi f_l T)^{\alpha-1}-\sum_{n\geq 1}\frac{(-1)^n}{(2n+2)!}\cdot\frac{2^{2n}-1}{2n+1}\cdot\frac{\alpha}{\alpha-2n-1}\cdot(2\pi f_l T)^{2n}\right]$$

again with $v = (\alpha-1)\cdot\pi/2$. $\Gamma(x)$ is here the well-known "Gamma-function". The variance develops with the square of the integration time at $\alpha > 3$ and with $T^{\alpha-1}$ for $1 < \alpha < 3$. For α close to the value of 3 both terms, $T^{\alpha-1}$ and the first term of the sum with n=1 must be taken into account. Again, note that for α near 2k+1 the singularity in the first term caused by $1/\sin(v)$ is removed by the corresponding term in the sum for n = k.

For $\alpha = 2k+1$, k = 0, 1, 2, ... we have:

$$\sigma_A^2(T) \;=\; S_{2k+1}(0)\cdot 8\cdot f_l\cdot\left[\sum_{n=1}^{n\neq k}\frac{(-1)^n}{(2n+2)!}\cdot\frac{2^{2n}-1}{2n-2k}\cdot\frac{2k+1}{2n+1}\cdot(2\pi f_l T)^{2n} \;+\right.$$

$$\left.+\;\frac{(-1)^{2k}}{(2k+2)!}\cdot[2^{2k}\cdot\log(2)\;-\;(2^{2k}-1)\cdot\{H_{2k+1}+\log(1/2\pi f_l T)\}]\cdot(2\pi f_l T)^{2k}\right]$$

with $\quad H_{2k+1} \;=\; \sum_{s=1}^{2k+2}\frac{1}{s}\;-\;\gamma\;+\;\frac{1}{2k+1}$ $\quad\gamma$ is again Eulers constant (see above).

In lowest approximation we have now:

---

[5] Barnes J A, Chi A R, Cutler L S, Healey D J, Leeson D B, McGunigal T E, Mullen J A, Smith W L, Sydnor R L, Vessot R F C and Winkler G M R 1971 Characterization of frequency stability IEEE Trans. IM-20 105-120 (1971)

$$\sigma_A^2(T) \;=\; G_\alpha(0)\cdot\begin{cases}\dfrac{\alpha-1}{\alpha-3}\cdot\dfrac{(2\pi f_l T)^2}{6} & \alpha>3,\quad f_h\to\infty\\ \{0.9153\ldots-\log(2\pi f_l T)\}\cdot\dfrac{(2\pi f_l T)^2}{3} & \alpha=3,\quad f_h\to\infty\\ \dfrac{2^{\alpha+1}-4}{\alpha\cdot\Gamma(\alpha+2)}\cdot\dfrac{\nu}{\sin(\nu)}\cdot(2\pi f_l T)^{\alpha-1} & 1<\alpha<3,\quad f_h\to\infty\\ \dfrac{2\cdot\log(2)}{\log(f_h/f_l)} & \alpha=1\\ \dfrac{4-2^{\alpha+1}}{\alpha\cdot\Gamma(\alpha+2)}\cdot\dfrac{\nu}{\sin(\nu)}\cdot\dfrac{1}{(2\pi f_h T)^{1-\alpha}} & 0<\alpha<1,\quad f_l\to 0\\ \dfrac{1}{2\,f_h T} & \alpha=0,\quad f_l\to 0\end{cases} \qquad [14]$$

The relation between $G_\alpha(0)$ and $S_\alpha(0)$ is given above. For values of α near 3 the terms given for $\alpha > 3$ and for $\alpha < 3$ must be considered at the same time. When comparing with Eq.[12] from above, we see that $\beta = \alpha - 1$. Thus, Equ.[12] describes the drift noise for all α with $0 < \alpha \le 3$. Above $\alpha = 3$ we have $\beta = 2$. In all ordinary experimental situations, we can expect that β is found somewhere between 1 and 2 with the exception of flicker noise, where β becomes zero.

From these formulas it is obvious, that for all $\alpha \ge 1$ and with $f_h^{-1} \ll T \ll f_l^{-1}$ the variance does not become smaller with increasing integration time. But, instead, it remains constant for flicker noise (α=1) or even increases for $\alpha > 1$. Therefore, as is frequently discussed in literature, it seems impossible for $\alpha \ge 1$ to integrate the noise down with longer integration times. Note that the problem is not only existent for flicker noise, but becomes even worse for higher order drift noise.

But, this is only true, if the total integration time T is short compared to $1/f_l$. For long integration times the result is different. We can rewrite for very large T ($T \gg f_l^{-1}$):

$$\frac{\sin^4(\pi f T)}{(\pi f T)^2} \;\to\; \tfrac{1}{2}\cdot\delta(f)$$

Thus we get for any spectrum $S_\alpha(f)$:

$$\sigma_A^2(T) \;=\; \frac{S_\alpha(0)}{2T}$$

Consequently, the variance decreases again like $1/T$ (as it does at any T for white noise) after having gone through a maximum. The result shows again, how important the requirement of a finite value of the spectral distribution at zero frequency is.

In consequence of the above assumptions, at time scales of $T \gg f_l^{-1}$ any noise can be integrated like white noise, otherwise the noise spectrum would be not in accordance to "realistic" experimental conditions. In case that $S_\alpha(0)$ is not assumed to be finite, that is, without the introduction of a lower cut-off frequency for a noise power spectrum like $1/f^\alpha$ the variance would behave completely different or undefined respectively.

Finally we also have to consider additional delays between the two data samples. When integrating, we find for $\alpha > 3$:

$$\sigma_A^2(T,T_D) \;=\; G_\alpha(0)\cdot\frac{\alpha-1}{\alpha-3}\cdot\frac{(2\pi f_l(T+T_D))^2}{6}$$

In the range $1 < \alpha < 3$ we have:

$$\sigma_A^2(T,T_D) \;=\; G_\alpha(0)\cdot\frac{1}{\alpha\cdot\Gamma(\alpha+2)}\cdot\frac{\nu}{\sin(\nu)}\cdot[(2+x)^{\alpha+1}+x^{\alpha+1}-2\cdot(1+x)^{\alpha+1}-2]\cdot(2\pi f_l T)^{\alpha-1}$$

$$=\; \sigma_A^2(T)\cdot\frac{(2+x)^{\alpha+1}+x^{\alpha+1}-2\cdot(1+x)^{\alpha+1}-2}{2^{\alpha+1}-4},\quad x=\tfrac{T_D}{T},\quad \sigma_A^2(T) \;=\; \sigma_A^2(T,T_D=0)$$

,

and for $\alpha = 1$ we have:

$$\sigma_A^2(T,T_D) \;=\; \tfrac{1}{2}\cdot G_1(0)\cdot\frac{(2+x)^2\cdot\log(2+x)+x^2\cdot\log(x)-2\cdot(1+x)^2\cdot\log(1+x)}{\log[f_h/f_l]} \;=$$

$$=\; \sigma_A^2(T)\cdot\frac{(2+x)^2\cdot\log(2+x)+x^2\cdot\log(x)-2\cdot(1+x)^2\cdot\log(1+x)}{4\cdot\log[2]},\quad x=\tfrac{T_D}{T},\; \sigma_A^2(T) \;=\; \sigma_A^2(T,T_D=0)$$

For $x = T_D/T \to 0$ the expressions approach the formulas given above. For white noise ($\alpha = 0$) there is no

change, i.e. the variance is independent on the delay. Thus with white noise, one observes a variance plot with a horizontal slope while keeping the integration time constant and varying the dead-time. Another case is the appearance of the plot at the presence of flicker noise. When integrating with very short time T while increasing the delay T_D, one sees a slow increase of the variance with a slope of $log(T_D/T)$; the limiting value of the variance at very large T_D is proportional to

$$\sigma_A^2(T,T_D) \underset{\alpha=1}{\overset{T_D>>T}{\rightarrow}} \frac{\frac{3}{2}+\log(\frac{T_D}{T})}{2\cdot\log(2)} .$$

These are examples how the interpretation of Allan variance measurements depends on the procedure of data taking. It is therefore important to use a generally accepted rule in order to guarantee comparable results. As was mentioned before, we therefore propose to use data without any dead-time while varying the integration time T as the standard.

One should mention here, that the results for $1 < \alpha < 3$ are identical with the outcome of the direct calculation with $G(\tau) = G_0 - \Gamma_\beta \cdot \tau^\beta$ (see Eq.[12]) with $\beta = \alpha$-1. We find here:

$$\sigma_A^2(T) = \Gamma_\beta \cdot \frac{2^{\beta+2}-4}{(\beta+1)\cdot(\beta+2)} \cdot T^\beta$$

If there is also delay T_D, then we find:

$$\sigma_A^2(T,T_D) = \Gamma_\beta \cdot \frac{(2+x)^{\beta+2}+x^{\beta+2}-2\cdot(1+x)^{\beta+2}-2}{(\beta+1)\cdot(\beta+2)} \cdot T^\beta$$

$$= \sigma_A^2(T)\cdot\frac{(2+x)^{\beta+2}+x^{\beta+2}-2\cdot(1+x)^{\beta+2}-2}{2^{\beta+2}-4}, \quad x = \frac{T_D}{T}$$

This result is valid for all exponents $\beta > -1$; also the range of $\beta > 2$ ($\alpha > 3$) is covered, but this has no equivalence with any 1/f$^\alpha$ power spectrum anyway. The exceptions are $\beta = 2k$ ($\alpha = 2k+1$), k = 1,2,…, while for $\beta = 0$ ($\alpha = 1$) the limiting expression for $\beta \rightarrow 0$ leads also to the correct result. This proves that the calculations with the Fourier transforms and the direct integrations are equivalent. In particular, the assumption of cut-off frequencies is again motivated. The variance is a monotonically increasing function with T_D at $\alpha \geq 1$, which reflects the obvious fact that longer dead time is always bad for the signal to noise ratio. It is a bit surprising that this is also valid for flicker noise. Therefore, when investigating the noise characteristics of a system, it is very important to avoid dead time between the data samples, or one has to consider the influence accordingly. The only exception is the white noise case.

Typical drift performance of most laboratory systems can be described by a slope β in the Allan variance plot between 1 and 2. While including the radiometric white noise to the variance as well, we get for the two extreme cases with $\beta = 1,2$ ($\alpha = 2,3$):

$$\sigma_A^2(T,T_D) = \sigma_{Rad}^2(T_A)\cdot\left[\frac{1}{t}+t+\tfrac{3}{2}\cdot d\right] \quad \text{for } \beta = 1 \quad \text{and}$$

$$\sigma_A^2(T,T_D) = \sigma_{Rad}^2(T_A)\cdot\left[\frac{1}{t}+\tfrac{1}{2}\cdot(t+d)^2\right] \quad \text{for } \beta = 2 \quad \text{with} \quad \sigma_{Rad}^2(T_A) = \frac{<y>^2}{B_{Fl}\cdot T_A} \qquad [15]$$

B_{Fl} is the “fluctuation bandwidth of the filter used in the system (see below). It is not immediately obvious, how a clean $\beta = 2$ slope can be constructed from 1/f$^\alpha$ drift spectra. On the other hand, it is rather frequent that such slopes appear in Allan variance plots. As was already mentioned before, pure 1/f$^\alpha$ spectra should be rare, and, when looking on the calculated drift variances above, it seems possible that a $\beta = 2$ slope can be assumed when using a mixture of various α including $\alpha = 3$ and $\alpha = 1$ spectra in order to remove the logarithmic contribution, as is inevitable with a pure $\alpha = 3$ noise spectrum.

Measurement of the Allan-variance

We should now introduce a procedure for the Allan Variance measurements: Sample data with constant (and short) integration time T_0 without any delay time between the samples. The data should be collected at constant rate and the total amount of data should cover a time interval which is at least similar to the time constants of the system under investigation. “Time constant” means here the average fluctuation time of the drift behavior of the system. If the total time is smaller, one must expect that the data set is not really representative for the statistical behavior of the system.

New samples for various integration times “T” are now constructed by adding the samples to multiples of the initial short integration time T_0 so that a full range of T can be established. For each T we have now $N(T)$ new

samples. The longest T_{Max} should still allow generating samples with sufficient credibility, e.g. $N(T_{Max}) \sim 20$. These data are now the basis for the calculation of the Allan variance as a function of the integration time T. The "new" data sets with the integration time $T_M = M \cdot T_0$ are then:

$$y_n(T_M) = \frac{1}{M} \cdot \sum_{m=1}^{M} y_{n \cdot M + m}(T_0) \quad \{n = 1,2,...;\ m = 1,2,...\}$$

We calculate now the Allan variance using:

$$\sigma_A^2 = \frac{1}{2} \cdot \frac{1}{N-1} \cdot \sum_{n=1}^{N} [y_n - y_{n-1}]^2 = \sigma_y^2 \cdot \left[1 - g_{1,2}(T_M)\right] \qquad [16]$$

The y_n are sums of the initial data at integration time T_M. The factor "½" must be applied since the data y_n are used twice in this formula.

At short intervals the drift noise is strongly correlated and only the white noise contributes to the statistics. At large time intervals, the drift correlation disappears and the variance increases therefore. From the plot it is easy to determine the time regime where the observing efficiency is optimum, i.e. where the inevitable white noise dominantly determines the errors. The standard procedure for Allan variance plots is now: collecting many data adjacent in time with short integration time, co-adding them to generate new data with increasing integration time, these are then used to calculate the variance as a function of integration time.

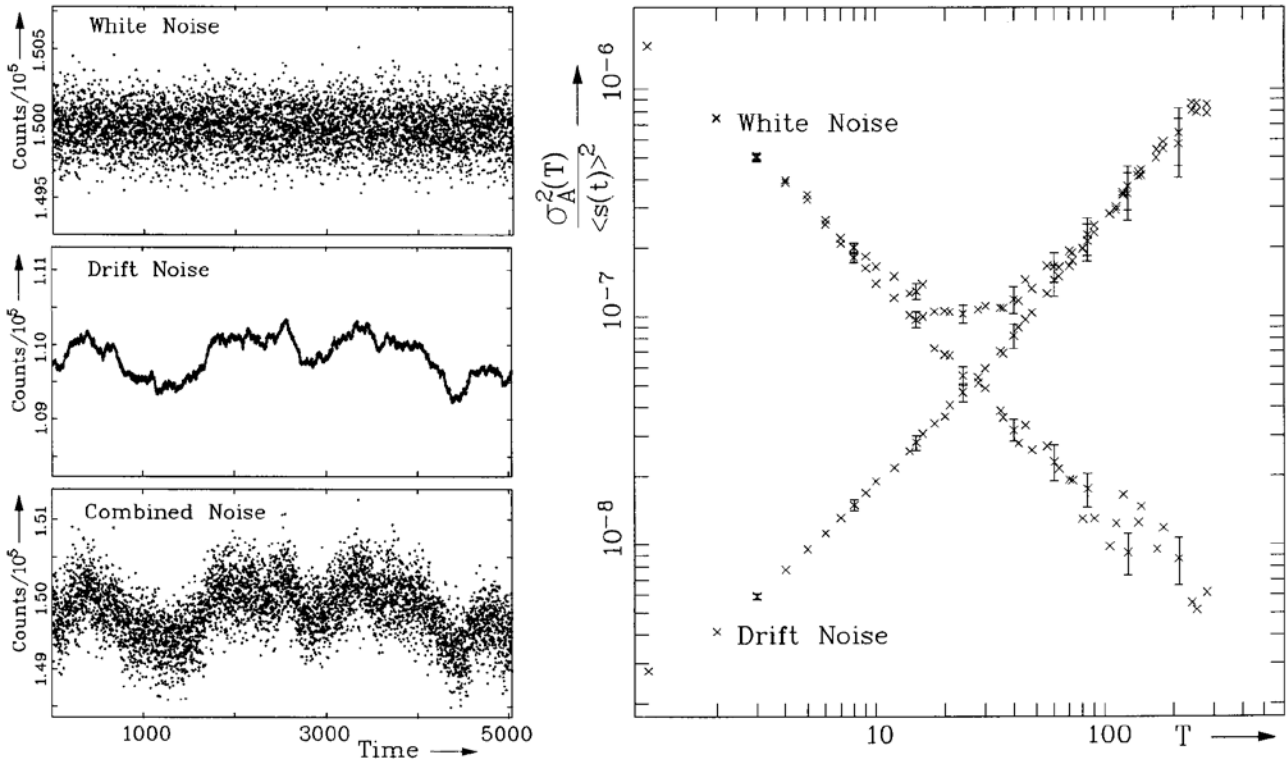


Fig.[1]:

Artificial data set generated by random numbers (left) with white noise of Gaussian distribution (top), drift noise (middle), and combined noise (bottom). Each data point corresponds to a sample integrated for 1 s while the fluctuation bandwidth was set to 600 kHz. The drift noise is calculated by filtering white noise with a sufficiently broad boxcar time-filter (width > T_{max} in the Allan variance plot). To the right the (relative) Allan variance plots of all three noise spectra are depicted. The white noise appears with a slope of −1, the drift noise with a slope of approximately +1. The combination of both results in a typical Allan plot with a minimum at some fairly well defined minimum time.

The total number of data is equal to P, while the number of data for each integration time T_M is equal to $N(T_M)$. Thus we have: $P = N(T_0) = M \cdot N(T_M)$. In order to avoid problems with different total number of data used for each N, it is advantageous to use only such N which are divisors of P. Thus, in order to have as many different values of N as possible, one should select a P, which has many divisors. Since it is unnecessary to make the plot with a very dense distribution of T_M, the selection of such values of P helps to avoid superfluous calculations without any additional valuable information. In addition, the density of points becomes almost constant in the logarithmic scale of the plot. This helps to reduce computing time. An example: Take $P = N(T_0) =$ 5040 data at, say, 1 sec integration time. This provides 35 divisors between 1 and 280. The timespan in inte-

gration time shown on the Allan-plot is therefore between 1 and 280 seconds (4.7 minutes). The total time length which is covered by the data set is 1.4 hours, which should be sufficient for most drift problems in the system. As it has turned out in the past, it is still useful to go up to an integration time of 280 seconds, although there are only 18 samples left. The disadvantage is that the error bars increase accordingly (see below), but one can still see the influence of the drifts at long integrations. If one wants to see the Allan variance at much longer integration time one should simply increase the initial integration time T_0 instead of increasing the number of collected data.

Spectroscopic Allan-variance

All multi-channel instruments have two types of drift noise. One component is common to all pixels and another is different for different pixels. Typically atmospheric fluctuations generate common changes in all frequency pixels in a radiometer; therefore they should not contribute to a ripple in the baseline of the spectrum. Instead they should generate a constant offset to all the pixels, which is of minor significance for spectroscopic work. The same is probably also true for the pixels of array bolometers. Baseline ripple seen with a spectrometer on the other hand is something which causes changing signal-differences between different pixels, and it may accidentally have similar appearance as the spectral line information to be observed with the spectrometer. A very special problem is the influence of unstable Laser "speckles" on the CCD detector in an AOS. They can be disastrous for the noise performance of the spectrometer. In order to distinguish between common and non-common offset drifts we introduced the "spectroscopic Allan variance" for the characterization of the noise performance of the spectrometer, which is the Allan variance of the normalized difference of two different pixels of the spectrometer. For this we calculate out of the individual data $y_n^{(1)}(T)$ and $y_n^{(2)}(T)$ of two pre-selected pixels 1 and 2 (see also in [6]):

$$y_n^{(0)}(T) = \left[\frac{1}{\sqrt{2}}\cdot\left\{\frac{y_n^{(1)}(T)}{< y^{(1)} >} - \frac{y_n^{(2)}(T)}{< y^{(2)} >}\right\} + 1\right] \cdot \frac{< y^{(1)} > + < y^{(2)} >}{2} \qquad [17]$$

$<y^{(1)}>$ and $<y^{(2)}>$ are the mean values of all collected data while these are numbered in time by the index n. The two channels should be selected under the assumption that their statistical behaviour is not correlated. The normalized difference removes all common gain drifts of the data of the two pixels, while the factor of $1/\sqrt{2}$ establishes the same relative white noise level as the initial data were subjected to. The additional "1" in the brackets guarantees that the relative level of white noise compared to the mean is preserved in the new data set. By multiplying the whole expression with the average of the means of the two initial data sets the statistical distribution and the mean of the new data becomes comparable in value with those of the initial data. It should be emphasized that this data manipulation does not remove the impact of common offset drifts completely, but it certainly reduces the common drift influence drastically. Therefore, when comparing the results of individual channels and the normalized differences one can nicely distinguish between the different drift contributions. This relates to problems connected to the generation of sometimes strong and unrealistic features in the spectrum which are introduced by the calibration procedure, particularly if the average signals at the ON- and the OFF-position are not the same.

This procedure was initially introduced in order to keep the amount of data small, since at that time, the computing power and storage capability of PCs was fairly low. The assumption was that the statistics of all pixels of the spectrometer should be more or less identical. In the meantime, the computer capabilities have drastically improved, and it is now possible to collect and store huge amounts of data and to process them within negligible amounts of time. Therefore it is now advisable to use all data of all channels of the spectrometer for the characterization of the stability of the instrument. This has the immediate advantage that the behaviour of all pixels is investigated simultaneously, and one does not depend on the accidental selection of only two spectrometer pixels. To do this, one should extend Eq.[17] to all K pixels of the spectrometer.

$$y_n^{(0)}(T) = \left\{\frac{1}{\sqrt{K}}\cdot\sum_{k=1}^{K}\frac{dy_n^{(k)}(T)}{< y^{(k)}(T) >} + 1\right\} \cdot \frac{1}{K}\cdot\sum_{k=1}^{K} < y^{(k)}(T) > = \qquad [18]$$

with $dy_n^{(k)}(T) = y_n^{(k)}(T) - < y_n^{(k)}(T) >$

k stands for the pixel number, k = 1,2,…K. n stands for the individual data samples of pixel k. The average $<y^{(k)}(T)> = <y^{(k)}>$ is identical for all T, since the total data amount is identical for all T (see above). At the same time we can replace $< dy_n^{(k)}(T)^2 >$ by $< (dy^{(k)})^2 > = \sigma_k^2$, since expectation of the variance should be

[6] The Cologne Acousto-Optical Spectrometers; V.Tolls, R.Schieder, G.Winnewisser; Experimental Astronomy 1, 101 (1989)

independent on *n*. When calculating the Allan variance with $\sigma_A^2(T) = ½ \cdot < \left[y_{n+1}^{(0)}(T) - y_n^{(0)}(T)\right]^2 >$, we can write as well:

$$y_n^{(0)}(T) = \left\{\frac{1}{\sqrt{K}} \cdot \sum_{k=1}^{K} \frac{y_n^{(k)}(T)}{< y^{(k)}(T) >} + 1\right\} \cdot \frac{1}{K} \cdot \sum_{k=1}^{K} < y^{(k)}(T) > \qquad [18a]$$

since $\frac{dy_n^{(k)}(T)}{<y^{(k)}(T)>} = \frac{y_n^{(0)}(T)}{<y^{(k)}(T)>} - 1$ and the "-1" disappears when taking the differences for the calculation of the Allan-variance. Using Eq.[16] one finds now:

$$\sigma_A^2(T) = \frac{1}{K} \cdot \left\{\sum_{k=1}^{K} \frac{\sigma_k^2(T)}{< y^{(k)} >^2} \cdot \left[1 - g_{k,k}(T)\right] + \sum_{i=1}^{K} \sum_{j \neq i}^{K} \frac{\sqrt{\sigma_i^2(T) \cdot \sigma_j^2(T)}}{< y^{(i)} > \cdot < y^{(j)} >} \cdot \left[g_{i,j}(0) - g_{i,j}(T)\right]\right\} \cdot \left[\frac{1}{K} \cdot \sum_{k=1}^{K} < y^{(k)} >\right]^2$$

If we assume that the statistics and the correlations between the frequency pixels of the spectrometer are all more or less identical, as should be always valid for a decent spectrometer, then we can simplify:

$$g_{i,i\pm k} = g_k, \; k = 0,1,2, \ldots \text{ and } \frac{\sigma_i^2(T)}{<y^{(i)}>^2} \approx \frac{\sigma_j^2(T)}{<y^{(j)}>^2} \approx \frac{\sigma_k^2(T)}{<y^{(k)}>^2}$$

$$\sigma_A^2(T) = \overline{\left\{\frac{\sigma_k^2(T)}{< y^{(k)} >^2}\right\}} \cdot \left\{[1 - g_0(T)] + 2 \cdot \sum_{k=1}^{K-1} \left(1 - \frac{k}{K}\right) \cdot [g_k(0) - g_k(T)]\right\} \cdot \overline{< y^{(k)} >}^2$$

with $\overline{\left\{\frac{\sigma_k^2(T)}{< y^{(k)} >^2}\right\}} = \frac{1}{K} \cdot \sum_{k=1}^{K} \frac{\sigma_k^2(T)}{< y^{(k)} >^2}$ and $\overline{< y^{(k)} >} = \frac{1}{K} \cdot \sum_{k=1}^{K} < y^{(k)} >$

With white noise only one has $g_k(T) = 0$ and $\sigma_k^2(T) / < y^{(k)} >^2 = 1/(B_{Fl} \cdot T)$. Thus, one obtains finally:

$$\sigma_A^2 = \frac{\overline{< y^{(k)} >}^2}{B_{Fl} \cdot T} \cdot \frac{K}{K_{eff}} \quad \text{with}$$

$$K_{eff} = \frac{K}{1 + 2 \cdot \sum_{k=1}^{K-k} \left(1 - \frac{k}{K}\right) \cdot g_k(0)} = \frac{K}{1 + 2 \cdot \left(1 - \frac{1}{K}\right) \cdot g_1(0) + 2 \cdot \left(1 - \frac{2}{K}\right) \cdot g_2(0) + \cdots}$$

The expression of K_{eff} is already well known from the error of the mean of co-added data (see Eq.[5]). It is not surprising that the correlation between adjacent pixels plays some role here. The result shows that the variance, as defined above, overestimates the Allan-variance. Therefore, Eq.[18] should be modified to:

$$y_n^{(0)}(T) = \left\{\frac{\sqrt{K_{eff}}}{K} \cdot \sum_{k=1}^{K} \frac{dy_n^{(k)}(T)}{< y^{(k)}(T) >} + 1\right\} \cdot \frac{1}{K} \cdot \sum_{k=1}^{K} < y^{(k)}(T) > \qquad [18b]$$

The needed information about the values of the $g_k(0)$ can be derived from a measurement of the filter-curve of the frequency pixels provided that they are all identical. Instead one could also use Eq.[2] to determine the values of correlation. A lot simpler path would be to preselect frequency pixels, e.g. by choosing only pixels which are separated by, say, 4 pixels in order to avoid mutual correlation. In this case the value of K_{eff} is exactly identical with the number of preselected pixels. At the same time we have found here the justification, why pixels should be chosen far enough apart for the use of Eq.[17].

In total the expression provides a good impression of the overall performance of the instrument. It shows the same behaviour with drift noise, as was discussed before. Nevertheless, experience has proven that the outcome of a two-channel and a full-channel Allan variance plot is very similar, which confirms the expectation that all spectrometer channels behave identical in a statistical sense. It is only essential to use two pixels which are separated enough so that any correlation between the channels becomes irrelevant. But one should know that using many spectrometer pixels instead of two does not mean that the statistical relevance of the analysis is improved. It is still necessary to collect data for a much longer time interval then the Allan-plot is showing. It is essential that the data set covers a reasonable time span in comparison with the typical time constants of the drifts. Otherwise the plot would not represent a good picture of the behaviour of the instrument.

Ratio Variance

It is not unusual that the gain of an instrument is time-dependent. To consider the difference of adjacent data is therefore sometimes a bit problematic. In order to be independent on such gain variations it is advantageous to calculate the statistics of the ratio of two data-packages instead of the difference, as it is done for the

calculation of the normal Allan variance. Therefore, the statistics of the ratio

$$Q_n = \frac{X_n(T)}{X_{n+1}(T)}$$

This removes automatically the gain of the system in the statistical analysis. We assume that the standard deviation of the data X_n is very small in comparison to their mean. Then we can write:

$$Q_n(T) = \frac{X_n(T)}{X_{n+1}(T)} = \frac{<X>+\delta X_n}{<X>+\delta X_{n+1}} = \left[1+\frac{\delta X_n}{<X>}\right]\cdot\left[1-\frac{\delta X_{n+1}}{<X>}\pm\ldots\right] \approx$$

$$\approx 1+\frac{\delta X_n}{<X>}-\frac{\delta X_{n+1}}{<X>} = 1+\frac{\delta X_n-\delta X_{n+1}}{<X>} = 1+\frac{X_n-X_{n+1}}{<X>} = 1+\frac{D_n}{<X>}$$

For the statistical investigation we define now:

$$\sigma_Q^2 = \tfrac{1}{2}\cdot\sigma^2(Q_n(T)) = \tfrac{1}{2}\cdot\sigma^2\left(\frac{D_n}{<X>}\right) = \tfrac{1}{2}\cdot\frac{\sigma_D^2}{<X>^2} = \frac{\sigma_A^2}{<X>^2}$$

Therefore we have:

$$\frac{\sigma_A^2}{<X>^2} = \sigma_Q^2$$

As is stated above, this is only correct for small fluctuations δX. When comparing with the radiometer formula, it turns out that this is a very convenient expression, since we have:

$$\sigma_D^2 = 2\cdot\frac{<X>^2}{B_{Fl}\cdot T} \quad \text{and therefore} \quad B_{Fl} = \frac{2\cdot<X>^2}{\sigma_D^2\cdot T} = \frac{<X>^2}{\sigma_A^2\cdot T} = \frac{1}{\sigma_Q^2\cdot T}$$

This is only meaningful, if only white (radiometric) noise is present, but with drift noise the appearance is not different from the standard Allan variance plot. In conclusion one can state that the ratio variance is very similar to the Allan variance, and the arguments about the Allan variance remain the same for the new definition. On the other hand, this new definition is better applicable to all situations where the gain of the system is not stable.

All formulas above are valid in principle only if true averages can be evaluated. Since this is not possible within finite time, the error analysis of finite data sets must be applied. Therefore, in all cases we use a modified Allan variance definition, which is derived from the ordinary error analysis for an experimental data set of N data D_n with $1 \le n \le N$ (see Eq.[4]). The D_n on the other hand are differences of two independent signal measurements X_n. We have therefore:

$$\sigma'^2_A(T) = <\tfrac{1}{2}\cdot\frac{1}{N-1}\cdot\sum_{n=1}^{N}[D_n-D_N]^2> = \tfrac{1}{2}\cdot\frac{N}{N-1}\cdot[<D_N^2>-<D_N>^2] \qquad [19]$$

with $D_n = D_n(T) = X(t_n,T) - X(t_n+T,T)$ and

$$D_N = \frac{1}{N}\cdot\sum_{n=1}^{N}D_n \quad \text{and} \quad D_N^2 = \frac{1}{N}\cdot\sum_{n=1}^{N}D_n^2$$

(The index n indicates that the samples are measured with the integration beginning at the times t_n and t_n+T). This definition is not exactly identical with the initial definition given by Allan because of the denominator "N-1" instead of "N" and the subtraction of the mean of all D_n, which should approach zero anyway.

In reality the effect on the values of the variance is mostly marginal. The meaning of the new definition is identical to that of statistical errors of any finite data set, if the difference measurements are considered as the individual elements of the data. Therefore the slight modification of the Allan variance should represent the real experimental conditions more closely. On the other hand, the additional, and in principal superfluous factor of 1/2 is still used in order to preserve the context to original definition by Allan and Barnes.

Error of the Allan-variance

For the presentation of the Allan variance plot it is also important to get an additional impression about the reliability and credibility of the information shown. For this some estimate of the errors of the calculated Allan variance values should be presented as is seen in Fig.1 for example. Each of the measured pairs of signal and reference measurements may be assumed to represent a single value of a newly defined function f_n:

$$f_n = \tfrac{1}{2}\cdot\frac{N}{N-1}\cdot[D_n-D_N]^2$$

The mean value of all *N* individual measurements is the final result of the complete experiment and one gets the modified Allan variance:

$$f \;=\; \tfrac{1}{N}\cdot\sum_{n=1}^{N} f_n \;=\; \tfrac{1}{2}\cdot\frac{1}{N-1}\cdot\sum_{n=1}^{N}[\,D_n - D_N\,]^2$$

The error of this mean value is easily found when using the common definition in error analysis (neglecting correlation):

$$\delta f^{\,2} \;=\; \frac{1}{N\cdot(N-1)}\cdot\sum_{n=1}^{N}[\,f_n - f\,]^2 \quad \text{with} \quad f^{\,k} \;=\; \tfrac{1}{N}\cdot\sum_{n=1}^{N} f_n^{\,k}$$

Inserting everything one gets:

$$<\delta f^{\,2}> \;=\; \frac{1}{2^2}\cdot\left[\frac{N}{N-1}\right]^2\cdot\frac{1}{N-1}\cdot[<(D_n - D_N)^4> - <(D_n - D_N)^2>^2\,]$$

For large *N* one can neglect the ratio *N/(N-1)* and one has:

$$\delta\sigma'^2_A(T_D,T) \;=\; [\delta f^{\,2}]^{1/2} \;\approx\; \tfrac{1}{2}\cdot\left[\frac{1}{N-1}\cdot[<(D_n - D_N)^4> - <(D_n - D_N)^2>^2\,]\right]^{1/2} \qquad [20]$$

This is the formula one can use for the calculation of the error bars in the Allan-plot. It provides a very good impression about the credibility of the calculated Allan variance values. If the errors become too large, the meaning of the apparent position of the Allan plot minimum for example cannot be taken very seriously. This happens typically, if the total sampling time of the complete data set is too small as compared to the typical drift time constant of the system so that it does not provide a true estimate of the drift behaviour of the system.[7]

Noise reduction while co-adding

Co-adding of a couple of pixels of a real time spectrometer in order to improve the signal to noise ratio is general practice when dealing with noisy spectra, and, it is also not quite as trivial as most people seem to believe. This rather simple case can now easily be evaluated similar to the treatment before. We have to determine the expected statistics of the mean of data y_n with:

$$z_k \;=\; \tfrac{1}{N}\cdot\sum_{n=1}^{N} y_{n+k}$$

with *N* the number of co-added pixels. We get now identical to the derivation before:

$$\sigma_N^2 \;=\; \frac{\Sigma_{n=1}^{N}\left[y_n - \frac{1}{N}\cdot\Sigma_{m=1}^{N} y_m\right]^2}{N\cdot(N_{eff}-1)} \quad \text{again with} \quad N_{eff} \;=\; \frac{N}{1+2\cdot\Sigma_{k=1}^{N-k}(1-\frac{k}{N})\cdot g_k} \qquad [21]$$

This is the same expression for the variance of the average of *N* data, as was discussed earlier. The effective number of pixels N_{eff} is obviously smaller than the number of co-added pixels. In the limiting case of very large *N* we get:

$$N_{eff} \;=\; \frac{N}{1 + 2\cdot g_1 + 2\cdot g_2 + \cdots}$$

The values of the correlation g_k can now be calculated if the filter-curves of the spectrometer are known. Usually one can easily calculate them when measuring the single frequency response of one spectrometer pixel as a function of frequency.

If we consider the situation when using a spectrometer like a filter-bank, an auto-correlator, or an AOS; then the filter-curve of co-added pixel-data with large *N* should be approximately equivalent to a square shaped filter with about the same frequency width as the total of co-added pixels represents. Therefore, the "fluctuation bandwidth" B_N, which now determines the residual noise by means of the radiometer equation, should become very closely identical to N times the pixel spacing "*d*" in the spectrometer. We have therefore:

$$\sigma_N^2 \;=\; \frac{C}{B_N\cdot t} \quad \text{and} \quad \sigma_y^2 \;=\; \frac{C}{B_{Fl}\cdot t}$$

[7] A detailed discussion of applications of Allan-variance tests can be found in: “Optimization of Heterodyne Observations Using Allan Variance Measurements; R.Schieder, C.Kramer; A&A 373, 746-756 (2001)"

"C" is an irrelevant constant which depends on the observing mode, and "*t*" is the integration time used for the observation. "B_{Fl}" is the fluctuation bandwidth of one single pixel of the spectrometer. In case of large *N*, this leads to:

$$B_N = \frac{N}{1 + 2 \cdot g_1 + 2 \cdot g_2 + ...} \cdot B_{Fl} \Rightarrow N \cdot d$$

$$=> \quad B_{Fl} = d \cdot \{1 + 2 \cdot g_1 + 2 \cdot g_2 + ... \}$$

This relation is only valid, if the pixel separation is small as compared to the resolution bandwidth of each pixel.

As an example, for the SWAS AOS we have the following experimental values:

separation of pixels: $d = 1.03$ MHz

Cross-correlation between pixels:
$g_1 = 0.44$
$g_2 = 0.10$
$g_3 = 0.02$
$g_k = 0$ for $k > 3$

This leads to:

$$B_{Fl} = 2.12 \cdot d = 2.2 \text{ MHz}$$

The value of B_{Fl} for the single pixel fluctuation bandwidth is exactly what was measured for the SWAS AOS during the ground tests and also later in orbit by means of the Allan variance test. At the same time, the measurement of the single pixel filter-curve delivers the same value.

Similarly, we can now calculate the effective number of pixels in the limiting case of large *N*:

$$N_{eff}(N\ large) = N / 2.12.$$

This means, the *rms* noise reduction is only about 69% of the naively expected value when just considering the number of co-added pixels. These results show very nicely, how significant the cross-correlation between pixels is when dealing with the baseline noise of the spectra taken with such an instrument. It is interesting to note that, according to Eq.[21], the discrepancy between the effective number of pixels and *N* becomes smaller at small *N*. We have for example for the SWAS-AOS:

$$N_{eff}(N=2) = 2 / 1.44 = 1.39$$
$$N_{eff}(N=3) = 3 / 1.65 = 1.81$$
$$N_{eff}(N=4) = 4 / 1.77 = 2.26$$

Similar are the effects on the effective fluctuation bandwidth. These results demonstrate how significant the influence of the correlation for the application of co-adding is.

Noise reduction when repeating measurements

Before, we have investigated the variances of different noise spectra, but we have not yet verified how the errors develop when repeating measurements several times in order to improve the error budget. The question is how the resulting errors will look like when starting from a particular value of the Allan variance even in the drift domain at arbitrary integration time *T* after many measurements. The outcome is obvious with white noise, but still not yet clear for the various drift noise spectra considered above.

When doing difference measurements with a box-car integrator, it is typical that the instrument integrates for some short and fixed integration time *T*, and, after some delay time T_D, another sample is integrated with the same integration time *T*. The variance is then dependent on the time length *T* of the integration as well as on the delay time T_D between the two samples. With pure white noise, one should expect that the noise is independent on the delay time T_D, and the standard deviation should decrease like $1/\sqrt{T}$ when varying the integration time *T*. This calls for long *T* in order to have low noise. On the other hand, one usually has limited time for doing the experiment; let us call the total time for the observation T_{Tot}, which is given by:

$$T_{Tot} = 2 \cdot N \cdot T + (2 \cdot N - 1) \cdot T_D$$

N is the number of pairs of samples taken within the total time T_{Tot}. Clearly, the optimum strategy for the measurement would be to reduce the influence of T_D by making it zero or choose $N = 1$. But this is only true of we do not have to deal with drift noise. In this case we are not allowed to integrate for arbitrarily long *T*, but instead we have a limit before the drift contributes significantly to the overall statistics. This limit needs to be found. The Allan variance tells us, how the error of one individual pair of data looks like when neglecting any delay time T_D. In this case the total variance of one single pair measurement is given by:

$$\sigma_1^2 \;=\; \sigma_D^2(T) \;=\; 2\cdot\sigma_A^2(T)$$

The factor of 2 arises because we take the difference of two data. When repeating the measurement the final variance should develop like 1/*N*. Thus we have:

$$\sigma_N^2 \;=\; \frac{\sigma_1^2(T)}{N} \;=\; \frac{4\cdot\sigma_A^2(T)\cdot T}{T_{Tot}}$$

In case of pure white noise, σ_A^2 develops like 1/*T* so that $\sigma_A^2\cdot T$ becomes independent on *T*. This confirms the usual result that the rms goes like $1/\sqrt{T_{Tot}}$. It shows that the value of the Allan variance determines what error budget one has to deal with after a sometimes lengthy data collection. The question is now, whether the actual value of the Allan variance for a given integration time *T* still determines the final outcome of an experiment. When assuming that the drift noise does also behave in a pure statistical manner, then we should expect that the arguments above apply under all circumstances.

The question is now how the drift noise behaves when averaging. This question is not trivial, since it is rather likely that the drift movements are still strongly correlated within the time period of a couple of difference measurements. Therefore, when running an Allan variance plot, can we assume that at any integration time *T* the average reduces like 1/*N*, if *N* is the number of samples taken? This means in consequence that any rms-error found for individual samples with Allan variance measurements is the starting point of the averaging procedure even if severe drift noise is present. But, frequently one gets the impression in the literature that this should not apply for flicker noise for example. In fact, this conclusion would be valid for any drift noise with $\alpha \geq 1$, and it holds certainly for the rather theoretical assumption that a lower cut-off frequency f_l is missing. But from an experimental point of view, as is considered here, one should reinvestigate the problem.

The final error of the complete measurement is the error of the mean D_N of all *N* difference measurements $D_n = S_n - R_n$ with

$$D_N \;=\; \tfrac{1}{N}\cdot\sum_{n=1}^{N} D_n$$

For simplicity we assume, that there is no dead-time involved ($T_D = 0$), and that *N* is very large. We write:

$$S_n(t) \;=\; \tfrac{1}{T}\cdot\int_{t+2n\cdot T}^{t+(2n+1)\cdot T} x(t+t')\cdot dt', \quad R_n(t) \;=\; \tfrac{1}{T}\cdot\int_{t+(2n+1)\cdot T}^{t+(2n+2)\cdot T} x(t+t')\cdot dt'$$

For the variance we have to evaluate the expectation values:

$$\sigma_N^2(T) \;=\; <D_N^2>_t \;-\; <D_N>_t^2$$

The suffix $_t$ stands for the average over very long time *t*. Since we are interested in the noise only, we may assume that the average of *S* and *R* are the same so that $<D_N>$ is exactly zero. We have now finally:

$$\sigma_N^2(T) \;=\; <D_N^2>_t \;=\; <\tfrac{1}{N^2}\cdot\left[\sum_{n=1}^{N}[S_n(t)-R_n(t)]^2\right]>_t$$

and we get, using the definition of the correlation function $g(\tau)$:

$$\sigma_N^2(T) = \tfrac{1}{(N\cdot T)^2}\cdot\sum_{n=1}^{N}\sum_{m=1}^{N}\int_0^T\int_0^T d\tau\cdot d\tau'\cdot\left[2\cdot g[2(n-m)\cdot T+\tau-\tau'] - g[(2(n-m)+1)\cdot T+\tau-\tau'] - g[(2(n-m)-1)\cdot T+\tau-\tau']\right]$$

While introducing the Fourier transform *S(f)* of the correlation function, and summing all terms in the double sum we get after some tedious, but straightforward manipulations:

$$\sigma_N^2(T) \;=\; \tfrac{1}{N}\cdot\int_0^\infty S(f)\cdot df \;\cdot\; \frac{\sin^2(\pi fT)}{(\pi fT)^2}\cdot\frac{\sin^2(2\pi NfT)}{N\cdot\cos^2(\pi fT)}$$

The integrand remains finite at $f \to 0$, because *S(f)* is assumed to be finite at zero. But instead, the integrand develops poles for $f\cdot T = k+½$ ($k = 0, \pm1, \pm2, \ldots$), if *N* becomes very large.

For the following we assume again that $1/f_h \ll T \ll 1/f_l$. It is now useful to rewrite (see e.g. [Footnote [2]]):

$$\frac{\sin^2(2\pi NfT)}{N\cdot\cos^2(\pi fT)} \;=\; \begin{cases} \displaystyle\sum_{k=-\infty}^{\infty}\frac{\sin^2(2\pi NfT)}{N\cdot[\pi fT-(k+\frac{1}{2})\cdot\pi]}, & f \neq (k+\frac{1}{2})/T \\ 4N & f \;=\; (k+\frac{1}{2})/T \end{cases}$$

From this we find therefore for large *N*:

$$\frac{\sin^2(2\pi NfT)}{N\cdot\cos^2(\pi fT)} \approx \tfrac{2}{T}\cdot\sum_{k=-\infty}^{\infty}\delta[f-(k+\tfrac{1}{2})/T] \quad \text{using} \quad \lim_{m\to\infty}\frac{\sin^2(m\cdot x)}{m\cdot x^2} = \pi\cdot\delta(x)$$

with $\delta(x)$ the Dirac Delta-function. This approximation is valid as long as $S(f)$ can be considered as practically constant within a frequency interval of $\Delta f \approx 1/(NT)$ around the poles at $(k+½)/T$. This determines a lower limit for the number of samples N. If N is large enough this condition is always true, since $S(f)$ does not diverge at any frequency and is a "smooth" function of f. This is assured by the assumptions for the correlation function $g(\tau)$. Thus we have, when integrating from f_l to ∞ while neglecting the interval between 0 and f_l for the moment:

$$\tfrac{1}{N}\cdot\int_{f_l}^{\infty}S(f)\cdot df\cdot\frac{\sin^2(\pi fT)}{(\pi fT)^2}\cdot\frac{\sin^2(2\pi NfT)}{N\cdot\cos(\pi fT)} \approx \frac{1}{N\cdot T}\cdot\frac{8}{\pi^2}\cdot\sum_{k=0}^{\infty}\frac{S((k+\tfrac{1}{2})/T)}{(2k+1)^2}$$

On the other hand, for the integration within the frequency interval $0 \le f < f_l$ and with $T \ll 1/f_l$, $N\cdot T \gg 1/f_l$ ($\sin^2(2\pi NfT) < 1$!) we find:

$$\tfrac{1}{N}\cdot\int_{0}^{f_l}S(f)\cdot df\cdot\frac{\sin^2(\pi fT)}{(\pi fT)^2}\cdot\frac{\sin^2(2\pi NfT)}{N\cdot\cos(\pi fT)} \le \frac{S(0)\cdot f_l}{N^2} = O(1/N^2)$$

since $S(0)$ is finite and nearly constant between 0 and f_l, while $f\cdot T$ is very small compared to unity within this interval so that $\sin^2(\pi fT) \approx (\pi fT)^2$, $\cos(\pi fT) \approx 1$, and $\langle\sin^2(2\pi NfT)\rangle = ½ < 1$. Thus, we get finally:

$$\sigma_N^2 \approx \tfrac{1}{N}\cdot\frac{1}{T}\cdot\frac{8}{\pi^2}\cdot\sum_{k=0}^{\infty}\frac{S((k+\tfrac{1}{2})/T)}{(2k+1)^2}$$

Until here we did not assume any particular spectral distribution of the noise power, therefore the result applies for any $S(f)$ as long as the conditions for the correlation function are fulfilled. Since $S(f)$ is assumed to remain finite for all f and to approach zero at $f \to f_h$, the sum is definitely convergent, independent on the actual shape of $S(f)$. Thus, the result shows that the variance develops like $1/N$ for any drift noise as required for the success of averaging.

Considering a specific spectral power law we use for $f_l \le f \le f_h$:

$$S(f) = S_\alpha(0)\cdot\frac{f^{-\alpha}-f_h^{-\alpha}}{f_l^{-\alpha}-f_h^{-\alpha}}$$

From this we get:

$$\sigma_N^2 \approx \frac{S_\alpha(0)}{NT}\cdot\frac{(2f_lT)^\alpha}{1-(f_l/f_h)^\alpha}\cdot\frac{8}{\pi^2}\cdot\left[\sum_{k=0}^{K}\frac{1}{(2k+1)^{\alpha+2}}-\frac{1}{(2f_hT)^\alpha}\cdot\sum_{k=0}^{K}\frac{1}{(2k+1)^2}\right]$$

K is the next integer below $f_h\cdot T$-½, which is large compared to unity but still finite since f_h is finite. Nevertheless, the sum can be extended to $K \to \infty$ without too much error because of the rapid decrease of the terms in the sum. The second sum may be neglected because $f_h\cdot T$ is very large, and we get finally for $\alpha > 0$:

$$\sigma_N^2 \approx \frac{2}{N}\cdot S_\alpha(0)\cdot\frac{f_l}{\pi^{\alpha-1}}\cdot\frac{\lambda(\alpha+2)}{\lambda(2)}\cdot(2\pi f_lT)^{\alpha-1} = \frac{2}{N}\cdot G_\alpha(0)\cdot\frac{\alpha-1}{\pi^{\alpha-1}\cdot\alpha}\cdot\frac{\lambda(\alpha+2)}{\lambda(2)}\cdot(2\pi f_lT)^{\alpha-1}$$

$\lambda(x) = (1 - 2^{-x})\cdot\zeta(x)$, and $\zeta(x)$ is the Riemann Zeta function (see e.g. [Footnote 2]). $\lambda(2)$ is equal to $\pi^2/8$ and decreases monotonically to the limiting value of 1 for $x \to \infty$ so that we have: $8/\pi^2 \approx 0.81 \le \lambda(\alpha+2)/\lambda(2) \le 1$. It is interesting to note that the variance of the average increases like $T^{\alpha-1}$ ($\alpha > 1$) without any upper limit for α, which is different from the result found for the Allan variance.

For $1 < \alpha \le 3$ the Allan variance is also proportional to $T^{\alpha-1}$, and we can write:

$$\sigma_N^2 = \tfrac{2}{N}\cdot C_\alpha\cdot\sigma_A^2(T)$$

The factor of 2 appears because of the difference of two independent measurements S and R. Using the expression for $\sigma_A^2(T)$ one finds, that the value of C_α is 1 for $\alpha = 0$ and decreases monotonically with increasing α. Therefore we can assume, that with any "normal" drift noise with $1 < \alpha \le 3$ the statistical error of the averaged measurements behaves like

$$\sigma_N^2 \le \tfrac{2}{N}\cdot\sigma_A^2(T) \qquad [22]$$

as long as we have for the complete observing time T_{Obs}:

$$T_{Obs} = 2\cdot N\cdot T \gg f_l^{-1} \quad \text{and} \quad T \ll f_l^{-1}$$

It means that the total observing time needs to be long enough in comparison with the longest drift time con-

stants, but the individual integration times T need to be very small. For large $\alpha > 3$ the derivation above becomes questionable since the leading term of $\sigma_A^2(T)$ is proportional to T^2, and the term $T^{\alpha-1}$ grows faster than T^2. But such exponents are probably not of interest during standard experimental situations.

This result is important, because it proves that, when using an Allan variance measurement a prediction of the final error after averaging is possible. This is true for all drift noise contributions as well as for flicker noise! Usually this is only assumed for the case of white noise! For sufficiently long observing time the noise reduction is at least the same as for white noise. It is a straightforward matter to derive the identical result also when introducing an additional delay time between the signal- and reference-measurement. It can therefore be concluded that the result above holds under any relevant experimental circumstances. Therefore, taking a value at a particular integration time T in the Allan plot, it is the starting value for any following averaging.

Error estimate of arbitrary linear functions of correlated data

Frequently it is very important to investigate, how the usual error estimate is affected by correlation. Let us assume the standard procedure to evaluate errors from a function $F(y_1,y_2,....y_N)$ of the data set $\{y_n: n = 1,2,\ldots,N\}$ is:

$$\sigma_F^2 = \sum_{n=1}^{N}\left[\frac{\partial F}{\partial y_n}\right]^2 \cdot < dy_n^2 > \qquad [23]$$

with $dy_n = y_n - <y_n>$

$$F = \sum_{n=1}^{N} f_n \cdot y_n \qquad [24]$$

We assume now again, that the statistics of the data points is the same for all n. When inserting now the derivatives of F, we get:

$$\sigma_F^2 = \sum f_n^2 \cdot < dy_n^2 > = \sigma_y^2 \cdot \sum f_n^2$$

On the other hand, if we use the general definition of the variance, we have:

$$\sigma_F^2 = < F^2 > - < F >^2 = <[\sum_{n=1}^{N} f_n \cdot y_n]^2 > - <\sum_{n=1}^{N} f_n \cdot y_n >^2 =$$

$$= <[\sum_{n=1}^{N} f_n \cdot dy_n]^2 > = \sum_{n=1}^{N}\sum_{m=1}^{N} < dy_n \cdot dy_m > =$$

$$= \sum_{m=1}^{N} f_n^2 \cdot < dy_n^2 > + \sum_{m=1}^{N-1} f_n \cdot f_{n+1} \cdot < dy_n \cdot dy_{n+1} > + ...$$

When comparing this with the standard expression for errors, it becomes clear that the error estimate, as we all have learned it during the beginners laboratory education, may be incorrect. The correlation between the data changes the picture completely. In all ordinary cases the values of the correlation function g_k are non-negative, so that the correct value is always larger than the simpler estimate from above. What we find here is that the correct formula for the error analysis is:

$$\sigma_F^2 = <[\sum_{n=1}^{N} \frac{\partial F}{\partial y_n} \cdot dy_n]^2 >, \qquad [25]$$

which is **not** the standard expression! There is only agreement, if there is no correlation in the data.

When assuming the errors of all y_n to be identical, we get now:

$$\sigma_F^2 = <\sum_{n=1}^{N}\sum_{m=1}^{N} \frac{\partial F}{\partial y_n} \cdot \frac{\partial F}{\partial y_m} \cdot dy_n \cdot dy_m > = \sigma_y^2 \cdot \left\{ \sum_{m=1}^{N}\left[\frac{\partial F}{\partial y_n}\right]^2 + 2 \cdot g_1 \cdot \sum_{m=1}^{N-1} \frac{\partial F}{\partial y_n} \cdot \frac{\partial F}{\partial y_{n+1}} + ... \right\}$$

The first term corresponds to the usual "Gaussian" error estimate of Eq.[23], while the following terms in the brackets take into account the mutual correlation between the data points. Assuming a linear relationship between the y_n and the fit function F while inserting we get:

$$\sigma_F^2 = \sigma_y^2 \cdot \left\{ \sum_{n=1}^{N} f_n^2 + 2 \cdot g_1 \cdot \sum_{n=1}^{N-1} f_n \cdot f_{n+1} + 2 \cdot g_2 \cdot \sum_{n=1}^{N-2} f_n \cdot f_{n+2} + ... \right\} \qquad [26]$$

Error Estimate of Fit-Parameters of an Arbitrary Fit-Function

The standard procedure for fitting a set of K data $y_k = y(x_k)$ with $1 \le k \le K$ to a predefined function $F_n = F(a_1,\ldots,a_P;x_n)$ uses a linearization algorithm. The function F is a function of P parameters a_p, $p = 1,\ldots,P$, and of a single variable x. For the fit we want to minimize the differences between the fit function and the data. In case the fit function describes fully the reality, we expect a normal statistical distribution of the differences between data and fit. Thus we use the standard definition of the variance of this distribution and try to minimize it in order to find the optimum parameters of the fit function. Thus, the following expression is minimized:

$$\sigma^2 \;=\; \frac{1}{N-1}\cdot\sum_{n=1}^{N}[y_n - F(a_1,a_2,...,a_P;x_n)]^2 \;=\; \frac{1}{N-1}\cdot\sum_{n=1}^{N}[y_n - F_n]^2 \;=\; Min$$

The fit function might be the mean value of all data, as was discussed before for example. One should be aware of the fact that frequently the fit function does not fully describe reality. This destroys the assumption of the statistical distribution, or, in other words, the differences between data and fit function become strongly correlated.

The minimum requirement leads to P equations with:

$$\sum_{n=1}^{N}\frac{\partial F_n}{\partial a_p}\cdot[y_n - F_n] \;=\; 0, \quad p = 1,2,\ldots,P$$

We use an iterative process. At the k^{th} step we have a set of parameters $a_p^{(k)}$ (which is not yet the correct one), and we can determine corrections of the $a_p^{(k)}$ by expanding $F(a_1,\ldots,a_P,x_k)$ with

$$F(a_1^{(k+1)},a_2^{(k+1)},...,a_P^{(k+1)};x_n) \;\approx\; F(a_1^{(k)},a_2^{(k)},...,a_P^{(k)};x_n) \;+\; \sum_{p=1}^{P}\frac{\partial F(a_1^{(k)},a_2^{(k)},...,a_P^{(k)};x_n)}{\partial a_p}\cdot da_p^{(k+1)}$$

where $da_p^{(n+1)}$ stands for: $da_p^{(n+1)} = a_p^{(n+1)} - a_p^{(n)}$

Accordingly, this defines an iteration for the evaluation of the parameters a_p. Inserting above we get:

$$\sum_{n=1}^{N}\frac{\partial F_n^{(k)}}{\partial a_p}\cdot\left\{y_n \;-\; F_n^k \;-\; \sum_{q=1}^{P}\frac{\partial F_n^{(k)}}{\partial a_q}\cdot da_q^{(k+1)}\right\} \;=\; 0\,, \quad p,q = 1,\ldots,P$$

We rearrange this:

$$\sum_{q=1}^{P}\left[\sum_{n=1}^{N}\frac{\partial F_n^{(k)}}{\partial a_p}\cdot\frac{\partial F_n^{(k)}}{\partial a_q}\right]\cdot da_q^{(k+1)} \;=\; \sum_{n=1}^{N}\frac{\partial F_n^{(k)}}{\partial a_p}\cdot[y_n - F_n^{(k)}]\,, \quad p = 1,\ldots,P$$

and define a matrix $\boldsymbol{A}^{(k)}$ with the elements $\quad A_{p,q}^{(k)} \;=\; \sum_{n=1}^{N}\frac{\partial F_n^{(k)}}{\partial a_p}\cdot\frac{\partial F_n^{(k)}}{\partial a_q}$

and a vector $b^{(k)}$ with the components $\quad b_p^{(k)} \;=\; \sum_{n=1}^{N}\frac{\partial F_n^{(k)}}{\partial a_p}\cdot[y_n - F_n^{(k)}]\,, \quad p = 1,\ldots,P$

The P equations above can now be combined to:

$$\mathbf{A}^{(k)}\cdot da^{(k+1)} \;=\; \mathbf{b}^{(k)}$$

The vector $\mathbf{da}^{(k+1)}$ consists of the components $da_p^{(k+1)}$ with $p = 1,\ldots,P$. We are interested in the vector $\underline{\mathbf{da}}^{(k+1)}$, which we can determine by introducing the inverse matrix $\mathbf{A}^{-1}$ with the components A_{pq}^{-1} ($\mathbf{A}\cdot\mathbf{A}^{-1} \;=\; \mathbf{A}^{-1}\cdot\mathbf{A} \;=\; \mathbf{1}$, the $\mathbf{1}$ stands for the unity matrix.). We get now:

$da^{(k+1)} \;=\; A^{-1(k)}\cdot b^{(k)}$ or as components:

$$da_p^{(k+1)} \;=\; \sum_{q=1}^{P}A_{p,q}^{-1(k)}\cdot b_q^{(k)}$$

Since $\quad b_p^{(k)} \;=\; \sum_{n=1}^{N}\frac{\partial F_n^{(k)}}{\partial a_p}\cdot[y_n - F_n^{(k)}]\,,$

it represents an inhomogeneous set of linear equations of the experimental data y_k, which justifies the linearity assumption before for any fit-function F in this approximation. Repeating the calculation while using the above equation many times leads finally to the desired values of the parameters a_p, when the $da_p^{(k+1)}$ approach zero.

The total errors of the fit-parameters a_p are determined by the errors of the b_q, so that we have:

$$\sigma_p^2 \;=\; <\left[\sum_{q=1}^{P} A_{p,q}^{-1}\cdot \delta b_q\right]^2> \;=\; \sum_{q=1}^{P}\sum_{r=1}^{P} A_{p,q}^{-1}\cdot A_{p,r}^{-1}\cdot < db_q \cdot db_r >$$

(Note that we neglect here the iteration indicator "*(k)*") The errors δb_q of the b_q are completely determined by the errors of the experimental data y_k:

$$\delta b_q \;=\; \sum_{n=1}^{N}\frac{\partial F_n}{\partial a_q}\cdot dy_n\;, \quad \text{which leads to:}$$

$$\sigma_p^2 \;=\; \sum_{q=1}^{P}\sum_{r=1}^{P} A_{p,q}^{-1}\cdot A_{p,r}^{-1}\cdot \sum_{n=1}^{N}\sum_{m=1}^{N}\frac{\partial F_n}{\partial a_q}\cdot\frac{\partial F_m}{\partial a_r}\cdot < dy_n dy_m >$$

Assuming that all errors of the y_k are identical we can write:

$$<dy_n\cdot dy_{n+s}> \;=\; \sigma_y^{\,2}\cdot g_s$$

with g_s the normalized correlation function of the data dy_n. In reality, a reasonable error estimate is mostly rather difficult. It is therefore practical to use for the error estimate:

$$dy_n \;=\; y_n - F_n \;=\; y_n - F(a_1,...,a_P;x_n)$$

This includes now the eventually present errors due to an incorrect fit function *F* as well. The variance is then calculated by:

$$\sigma_y^2 \;=\; \tfrac{1}{N-1}\cdot\sum_{n=1}^{N}[y_n - F(a_1,...,a_P;x_n)]^2$$

The calculation g_s of the correlation function follows the same rules as described earlier. Yet, if the fit-function is not appropriate, it is rather likely that the differences $y_n - F_n$ have a non-statistical behaviour. It is therefore important to investigate the difference carefully and to introduce an improved fit-function if necessary. If things are "perfect", then we can write:

$$\sigma_p^2 \;=\; \sum_{q=1}^{P}\sum_{r=1}^{P} A_{p,q}^{-1}\cdot A_{p,r}^{-1}\cdot\left[\sum_{n=1}^{N}\frac{\partial F_n}{\partial a_q}\cdot\frac{\partial F_n}{\partial a_r} + \sum_{s=1}^{N-1}\left\{\sum_{n=1}^{N-s}\frac{\partial F_n}{\partial a_q}\cdot\frac{\partial F_{n+s}}{\partial a_r} + \sum_{n=s+1}^{N}\frac{\partial F_n}{\partial a_q}\cdot\frac{\partial F_{n-s}}{\partial a_r}\right\}\cdot g_s\right]\cdot\sigma_y^2 \qquad [27]$$

This is the complete and final equation for the error estimate of the fit parameters a_p. It is obvious that correlation of data makes the error estimate a lot more complicated. But with modern computers it should not cause too much trouble. What is left is the evaluation of the correlation g_s. It can be calculated by using:

$$g_s \;=\; \frac{\frac{1}{N-s-1}\cdot\sum_{n=!}^{N-s}\delta y_n\cdot\delta y_{n+s}}{\sqrt{\frac{1}{N-s-1}\cdot\sum_{n=!}^{N-s}\delta y_n^2\cdot\frac{1}{N-s-1}\cdot\sum_{n=s+!}^{N}\delta y_{n+s}^2}} \quad \text{with} \quad \delta y_n \;=\; y_n - F_n \;=\; y_n - F(a_1,...,a_P;x_n)$$

The fit parameters a_p are those found with the fit after a sufficient number of iterations. The correlation becomes influenced by the validity of the fit-function *F*! If it is not well chosen, the correlation values g_k become larger. In particular, e.g. large baseline ripples in a spectrum result in strong correlation even at large *k*. This has consequently enormous influence on the error estimation.

Sometimes it may be quite tedious to calculate these quantities, but, if all data are uncorrelated (the fit is perfect!), the equation simplifies to:

$$\sigma_p^2 \;=\; \sum_{q=1}^{P}\sum_{r=1}^{P} A_{p,q}^{-1}\cdot A_{p,r}^{-1}\cdot\sum_{n=1}^{N}\frac{\partial F_n}{\partial a_q}\cdot\frac{\partial F_n}{\partial a_r}\cdot\;\sigma_y^2$$

Here we can use that

$$A_{q,r} \;=\; \sum_{n=1}^{N}\frac{\partial F_n}{\partial a_q}\cdot\frac{\partial F_n}{\partial a_r}\,,$$

and we get therefore:

$$\sigma_p^2/\sigma_y^2 \;=\; \sum_{q=1}^{P}\sum_{r=1}^{P} A_{p,q}^{-1}\cdot A_{p,r}^{-1}\cdot A_{q,r} \;=\; \sum_{q=1}^{P}\sum_{r=1}^{P}(A_{p,q}^{-1}\cdot A_{q,r})\cdot A_{p,r}^{-1} \;=\; \sum_{r=1}^{P}\delta_{p,r}\cdot A_{p,r}^{-1} \;=\; A_{p,p}^{-1}$$

Without correlation between the individual experimental data y_k the well-known and rather simple error estimate is valid:

$$\sigma_p^2 = A_{pp}^{-1} \cdot \sigma_y^2 \qquad [28]$$

This is the mostly used formula for the error of the fit parameters. Nevertheless, it is only a crude estimate of the error, and frequently this estimate is too small. If we remember the influence of correlation on the error estimate, we should use the correlated fit error from above together with the general error estimate:

$$\sigma_y^2 = \frac{\sum_{n=1}^{N}[y_n - F_n]^2}{N-1-2\cdot\sum_{s=1}^{N-1}(1-\frac{s}{N})\cdot g_s}$$ with the g_s as found above.

This considers now all the deviations between data and fit as well the correlation between the differences of data and fit, which appears with an insufficient fit function. In general, the error estimate becomes significantly larger, since the g_s will probably contribute strongly even at large values of *s*. (Remember, with full correlation between all data, i.e. at $g_s = 1$ for all *s*, the error estimate becomes undefined, since the denominator and eventually also the numerator in the formula become zero!) Again, it might be difficult to derive a reliable estimate of the g_s. In some cases it should be possible to repeat the measurement several times so that the error bars become smaller. Otherwise, one has to live with the fact that error estimates are estimates and nothing else.

An Example: Fit of a "Gaussian"

A good example for the application of the results of the last chapter is the fit of a Gaussian with predetermined width to a noisy spectrum. If we call "P_n" the values of a Gaussian at spectrometer pixel # *n* with

$$P_n = e^{-\left[\frac{\nu_n - \nu_0}{D}\right]^2 \cdot 4 \cdot \log(2)}$$

D = *FWHM* of the Gaussian,
v_n = frequency of pixel *n*, and
v_0 = frequency of the center of the Gaussian,

then we have for the fit:

$$\sum_{n=1}^{N}[y_n - a \cdot P_n]^2 = Min$$

with y_n the value of pixel *n* of the spectrometer data. The interval used for the fit is given by the finite pixel number *N*. From this it follows for the peak amplitude "*a*" of the Gaussian:

$$a = \frac{\sum_{n=1}^{N} P_n \cdot y_n}{\sum_{n=1}^{N} P_n^2}$$

The area of the Gaussian is then:

$$F = a \cdot \sum_{n=1}^{N} P_n \cdot d = \frac{\sum_{m=1}^{N} P_m}{\sum_{m=1}^{N} P_m^2} \cdot \sum_{n=1}^{N} P_n \cdot y_n \cdot d$$

with *d* the (constant) spacing between the pixels. The f_n, introduced in the last chapter (Eq.[24]), are now:

$$f_n = \frac{\sum_{m=1}^{N} P_m}{\sum_{m=1}^{N} P_m^2} \cdot d \cdot P_n$$

In case the *FWHM* is large compared to the pixel spacing d, we can replace the sums by the equivalent integrals and get:

$$\frac{\sum_{m=1}^{N} P_m}{\sum_{m=1}^{N} P_m^2} \Rightarrow 2^{1/2}$$

The coefficients become now:

$$f_n = 2^{1/2} \cdot d \cdot P_n$$

For the variance of the error distribution of F - which is the square of the *rms* - we can use now Eq.[26]. If there is only noise and no line, <F> is zero, so we have now:

$$\sigma_F^2 = 2 \cdot \sigma_y^2 \cdot d^2 \cdot \left\{ \sum_{n=1}^{N} P_n^2 + 2 \cdot g_1 \cdot \sum_{n=1}^{N-1} P_n \cdot P_{n+1} + 2 \cdot g_2 \cdot \sum_{n=1}^{N-2} P_n \cdot P_{n+2} + ... \right\}$$

σ_y^2 is the variance of the statistical distribution of the initial data set y_n, which is assumed to be identical for all pixels n.

The sums over the values of P_m may be replaced again by the integrals:

$$\sum_{n=1}^{N-k} P_n \cdot P_{n+k} \cdot d = \sqrt{\frac{\pi}{8 \cdot \log(2)}} \cdot D \cdot \exp\{-(\tfrac{k \cdot d}{D})^2 \cdot 2 \cdot \log(2)\} =$$

$$= \sqrt{\frac{\pi}{8 \cdot \log(2)}} \cdot D \cdot P_k^{1/2}$$

with π = 3.14... . From this we get now:

$$\sigma_F^2 = \sigma_Y^2 \cdot \sqrt{\frac{\pi}{2 \cdot \log(2)}} \cdot d \cdot D \cdot \{1 + 2 \cdot g_1 \cdot P_1^{1/2} + 2 \cdot g_2 \cdot P_2^{1/2} + ...\} \qquad [29]$$

This is the square of the *rms* error of the fit to the data. The more common error estimate according to Eq.[23] leads to a different variance $\sigma_F'^2$:

$$\sigma_F'^2 = \sigma_Y^2 \cdot \sqrt{\frac{\pi}{2 \cdot \log(2)}} \cdot d \cdot D$$

As one can see again in this particular case, with correlation the error becomes much larger then usually expected. The correct

When analyzing the noise performance by means of the statistical distribution of the ratio of the fitted Gaussian area and the conventional *rms*, one obtains very good information about the validity of the statistical assumptions used for the mathematical treatment here. Or, the other way around, the careful analysis of a data set from a spectrometer might reveal some non-statistical performance problems, when comparing with the mathematical analysis given here.

Signal to Noise Ratio of the Fit

It is rather common, unfortunately, that radio-astronomical spectra consist of noise predominantly, and it is often not quite obvious, whether there is a line signal or if the noise just pretends to show a line. For a clear identification of a signal it is therefore very important to understand the appearance of the noise in the spectrum very accurately. In order to deal with the problem, we will determine here, what Gaussian line signal is being "seen" by a fit procedure in case there is plain noise only. This in turn determines how large the error of the fit will be. It is clear that the expectation of the Gaussian area, as described in the previous chapter, will be zero. Thus, an estimate of the amplitude of the fit result can only be given by using

$$S = \langle F^2 \rangle^{1/2},$$

since <F> is zero in our case. The expectation value of F^2 we have already determined above based on an the definition given by Eq.[25] (see Eq.[26]). It is evident that the Gaussian area determined by the fit becomes larger, if there is correlation between the initial data. This result is not surprising, since, due to the correlation, the values of adjacent pixels become dragged by pixels which accidentally deviate from the mean significantly. This somehow might appear as a "line signal", although there is no line in reality.

Typically, the noise of a spectrum is determined by the *rms* of the baseline in regions where there is no signal present. Such regions need to be determined by the observer. There is certainly quite some subjectivity in-

volved, and one has to be rather careful when doing this. All routines determining *rms* values in spectra are based on the assumption of non-correlation, therefore Eq.[23] is usually applied. This means the rms of the spectrum would be calculated to:

$$rms^2 \;=\; \sigma_y^2 \cdot \sqrt{\frac{\pi}{2\cdot\log(2)}} \cdot d \cdot D$$

For the signal to noise ratio we find therefore:

$$S/N \;=\; \frac{S}{rms} \;=\; \{1 + 2\cdot g_1\cdot P_1^{½} + g_2\cdot P_2^{½} + \ldots\}$$

The correct value of the *S/N* should be equal to "1" on the other hand, since we are dealing with a "zero-signal" situation. This can easily be verified when using Eq.[26] for the *rms* error estimate of the baseline noise. Therefore one has to increase the noise *rms* accordingly. This has the consequence that the level of confidence for e.g. a "3-sigma detection" of a line, as usually given, is reduced significantly. The conclusion must be that one has to be very careful when dealing with noisy data, in particular, if correlation is involved.

The discussion here evolved during the SWAS mission when searching for a line signal of molecular Oxygen. Since the molecule seemed to be undetectable, it was evident to define some method for a clear distinction of just noise in the spectra and/or an eventual line signal. The above discussion helped to find a clear distinction between noise and line.

Smoothing of data

The general procedure, when smoothing data, can be described by Eq.[23]. The task is to generate smoothed data by co-adding data while applying a weighting function

$$F_n \;=\; \sum_{m=1}^{M} f_m \cdot y_{n+m}$$

In order to leave any dc-offset or "slow" structures in the data unchanged, the coefficients f_m have to fulfill the additional normalization condition:

$$\sum_{m=1}^{M} f_m \;=\; 1$$

The effect of the procedure is a reduction of the *rms* noise as given by Eq.[25],.

$$\sigma_F^2 \;=\; \sigma_y^2 \cdot \left\{ \sum_{m=1}^{M} f_m^2 + 2\cdot g_1 \sum_{m=1}^{M-1} f_m \cdot f_{m+1} + 2\cdot g_2 \sum_{m=1}^{M-2} f_m \cdot f_{m+2} \;+ \ldots \right\}$$

It is now simple to define an effective number of pixels according to

$$\sigma_F^2 \;=\; \frac{\sigma_y^2}{M_{eff}} \quad \text{with}$$

$$M_{eff} \;=\; \frac{1}{\sum_{m=1}^{M} f_m^2 + 2\cdot\sum_{m=1}^{M-1} f_m \cdot f_{m+1} + 2\cdot\sum_{m=1}^{M-2} f_m \cdot f_{m+2} + \cdots}$$

One example with a simple "box-car" function $f_n = 1/M$ leads to

$$M_{eff} \;=\; \frac{M}{1 + \sum_{s=1}^{M-1}(1 - \frac{s}{M})\cdot g_s}$$

Another, frequently used function is a Gaussian. In this case we have:

$$f_n \;=\; \frac{P_n}{\sum_{m=1}^{M} P_m} \;\rightarrow\; \sqrt{\frac{4\cdot\log(2)}{\pi}} \cdot \frac{d}{D} \cdot P_n$$

We get for the noise of the smoothed data (for details see above):

$$M_{eff} \;=\; \frac{\sqrt{\frac{\pi}{2\cdot\log(2)}}}{1 + 2\cdot g_1\cdot P_1^{½} + 2\cdot g_2\cdot P_2^{½} + \cdots} \cdot \frac{D}{d}$$

Using the SWAS-AOS data with D = 3 km/sec we find:

$$M_{eff} \;=\; 0.75 \cdot D/d \;=\; 3.58$$

In case of very large width D of the Gaussian we obtain the limiting value of M_{eff} :

$M_{eff} \Rightarrow 0.71 \cdot D/d$

This is to be compared with the effective number of pixels without any correlation. Then we have:

$M_{eff} = 1.51 \cdot D/d = 7.19$

It is rather unpleasant to discover that the effectiveness of the smoothing procedure is only half of that what one would probably hope for when forgetting about the correlation.

Direct Data Analysis (Smoothing of derivatives)

In laboratory spectroscopy it is rather common to observe the positions of atomic or molecular line signals by looking for zeros of first, sometimes third derivative signals as seen for example when using a Lock-In amplifier together with a frequency modulated source. This method is rather effective but needs the possibility to apply frequency modulation on the monochromatic source. The digital storage of data offers a rather simple method to extract information like position of lines in measured spectra etc. It is based on a purely mathematical treatment of data streams, which leads to the same information as the usage of a Lock-In amplifier. Assumption is that a (DC-coupled) detector signal is directly digitized and processed afterwards in the computer. Signals like first or second derivative are easily derived. For this we use a polynomial fit of order P to the data set $\{y_n: 1 \le m \le M\}$ so that:

$$\sum_{m=1}^{M} [y_m - (a_0 + a_1 \cdot x_m + a_2 \cdot x_m^2 + ... + a_P \cdot x_m^P)]^2 = Min$$

The optimisation leads to a set of $P+1$ equations like

$$a_0 \cdot \sum_{m=1}^{M} x_m^p + a_1 \cdot \sum_{m=1}^{M} x_m^{p+1} + a_2 \cdot \sum_{M=1}^{M} x_m^{p+2} + ... + a_P \cdot \sum_{m=1}^{M} x_m^{p+P} = \sum_{m=1}^{M} x_m^p \cdot y_m$$

with $0 \le p \le P$.

The coefficients a_p are related to the derivatives of the data set y_m, since we can write:

$$y_m = y(x_m) = y(0) + \frac{dy}{dx}(0) \cdot \frac{x_m}{1!} + \frac{d^2y}{dx^2}(0) \cdot \frac{x_m^2}{2!} + ...$$

The a_p can be therefore identified by

$$\frac{d^p y}{dx^p} = p! \cdot a_p$$

Certainly, we do not use a polynomial of higher degree than is necessary for the calculation of the corresponding derivative. Thus, for a first derivative we use P=1, or for a second derivative we use P=2. The expansion coefficients are now determined for windows of width "M", which one can chose according to the needs of smoothing. Certainly, the success of the algorithm depends on the scanning speed across the spectrum. It is essential that within the time needed to scan through the window of M data points the noise is not dominated by drift noise. Therefore, scanning needs to be fairly rapid. This is different from most modulation methods like with a Lock-In amplifier.

When applying this to a data stream, one gets immediately the spectrum of the derivatives, while the center of the window is shifted point by point. When writing this properly, we have as before:

$$< y_n >_{[M]} = \frac{1}{M} \cdot \sum_{m=1}^{M} y_{n+m}$$

For the first derivative assuming equidistant x_m ($x_{m+1} - x_m = \delta$) we have:

$$< \frac{dy_n}{dx} >_{[M]} = \frac{12/\delta}{(M-1) \cdot M \cdot (M+1)} \cdot \sum_{m=1}^{M} [m - \tfrac{1}{2} \cdot (M+1)] \cdot y_{n+m}$$

With a Window width of M = 2 we get the usual formula for the derivative:

$$\frac{dy}{dx_{[2]}} = \frac{y_2 - y_1}{\delta}.$$

Larger M lead to a saw-tooth window. For instance, for M = 6 we get:

$$\frac{dy}{dx_{[6]}} = \frac{-5 \cdot y_1 - 3 \cdot y_2 - 1 \cdot y_3 + 1 \cdot y_4 + 3 \cdot y_5 + 5 \cdot y_6}{35 \cdot \delta}$$

For the second derivative we have:

$$<\frac{d^2y_n}{dx^2}>_{[M]} = \frac{60/\delta^2}{(M-2)\cdot(M-1)\cdot M\cdot(M+1)\cdot(m+2)}\cdot\sum_{m=1}^{M}[(M+1)\cdot(M+2)-6\cdot(M+1)\cdot m+6\cdot m^2]\cdot y_{n+m}$$

For the smallest possible M = 3 we have here:

$$\frac{d^2y}{dx^2}_{[3]} = \frac{y_1-2\cdot y_2+y_3}{\delta^2}$$

This again is the standard expression, when calculating a second derivative from a data set. With M = 6 we have for example:

$$\frac{d^2y}{dx^2}_{[6]} = \frac{5\cdot y_1-1\cdot y_2-4\cdot y_3-4\cdot y_4-1\cdot y_5+5\cdot y_6}{28\cdot\delta^2}$$

The method may be continued to higher orders. The advantage of the procedure is that it leads directly to the smoothed values of the derivatives, which can be very useful. In addition, one does not need sophisticated hardware, modulation schemes, or integrators to extract the desired information like line profiles of molecular absorptions for example. Just a computer and an ADC is required.

In many cases it is practical to use a symmetric arrangement of the data points with $-K \le m \le +K$ instead of $1 \le m \le N$. The above formula becomes in this case:

$$\sum_{m=-K}^{+K}[y_m-(a_0+a_1\cdot x_m+a_2\cdot x_m^2+...+a_P\cdot x_m^P)]^2 = Min$$

In consequence, when calculating the coefficients a_p, all sums over the x_m^p become zero for p odd, provided the x_m are equidistant. The total number of data points is odd ($M = 2\cdot K+1$), when using this formula. The first derivative is in this case:

$$<\frac{dy_n}{dx}>_{[2K+1]} = \frac{3/\delta}{K\cdot(K+1)\cdot(2K+1)}\cdot\sum_{m=-K}^{+K}m\cdot y_{n+m}$$

Similar, we have for the second derivative:

$$<\frac{d^2y_n}{dx^2}>_{[2K+1]} = \frac{30/\delta^2}{K\cdot(K+1)\cdot(2K-1)\cdot(2K+1)\cdot(2k+3)}\cdot\sum_{m=-K}^{+K}[3\cdot m^2-K\cdot(K+1)]\cdot y_{n+m}$$

With modern computers it is easy to generate the desired information online while taking data, since the required mathematical procedures are rather simple. The width of the window, i.e. the values of M (or K), determine, how much smoothing of the data is applied. At the same time, the width should match the structure of the expected signals, i.e. it should be near or slightly smaller than the half-width of a molecular line for example. At the same time, the time between taking individual data should fit to the time constant of the analogue system. (For this, see below.) By the way, the algorithm can be improved in speed, when using an iterative procedure for the data handling. It consists more or less in co-adding the next new value and subtracting older ones. The details we will not discuss here. Another question is, how much shift one should use between individual smoothed data values. It is obvious that with M large at a shift of Δn = 1, the adjacent data, are strongly correlated. This is generally not very desirable. A good proposal might be to shift by half a window width, which corresponds more or less to the so called “Nyquist Sampling”. This is a compromise between correlation and amount of data.

It is important to know, how much improvement of the signal to noise one can achieve as a function of the window width N. According to the previous consideration we have for the resulting noise level:

$$\sigma^2 = \sigma_y^2\cdot\left\{\sum_{m=1}^{M}f_m^2+2\cdot g_1\cdot\sum_{m=1}^{M-1}f_m\cdot f_{m+1}+2\cdot g_2\cdot\sum_{m=1}^{M-2}f_m\cdot f_{m+2}+...\right\}$$

when assuming identical noise for all data points y_n. The coefficients f_n are

$$f_m = \frac{1}{M} \quad \text{for} \quad <y>_M$$

$$f_m = \frac{12/\delta}{(M-1)\cdot M\cdot(M+1)}\cdot[m-\tfrac{1}{2}\cdot(M+1)] \quad \text{for} \quad <\frac{dy}{dx}>_M, \text{ and}$$

$$f_m = \frac{60/\delta^2}{(M-2)\cdot(M-1)\cdot M\cdot(M+1)\cdot(M+2)}\cdot[(M+1)\cdot(M+2)-6\cdot(M+1)\cdot m+6\cdot m^2] \quad \text{for} \quad <\frac{d^2y}{dx^2}>_M$$

Neglecting any correlation ($g_k = 0$, k > 0) we can now calculate the resulting variance using

$$\sum_{m=1}^{M} f_m^2 = \frac{1}{M} \quad \text{for} \quad < y >_{[M]}$$

$$\sum_{m=1}^{M} f_m^2 = \frac{12/\delta^2}{(M-1)\cdot M \cdot (M+1)} \quad \text{for} \quad < \frac{dy}{dx} >_{[M]}, \text{ and}$$

$$\sum_{m=1}^{M} f_m^2 = \frac{720/\delta^4}{(M-2)\cdot(M-1)\cdot M \cdot (M+1)\cdot(M+2)} \quad \text{for} \quad < \frac{d^2y}{dx^2} >_{[M]}$$

This assumes that the timing of the data stream is appropriate and constant. The other sums are more complicated, but can also be calculated, if correlation needs to be included in the calculation.

As is visible in Table [1], the *rms* noise of the first and second derivatives reduces drastically with the width of the window. With large M it develops like $1/M^{3/2}$ for the first derivative and like $1/M^{5/2}$ for the second. The performance compares nicely with the function of the RC-filter in the output of a Lock-In amplifier. But the algorithm has the advantage, that a memory effect 0f more than the width of the window is absent, what is certainly very different from that what a standard RC filter or most of the published digital filters are doing.

M \ $(\Sigma f_n^2)^{1/2}$	$<y>_{[M]}$	$<dy/dx>_{[M]}$	$<d^2y/dx^2>_{[M]}$
1	1.00	-	-
2	0.71	1.41	-
3	0.58	0.71	2.45
5	0.45	0.32	0.53
10	0.32	0.11	0.087
20	0.22	0.039	0.015

Table [1]: *rms* noise amplitude after application of the derivative algorithm.

Conclusion:

The amazing fact is that one does not need sophisticated hardware to extract signal information from a data stream as is usually implemented by means of Lock-In amplifiers or box-car integrators. It is fully sufficient to sample with proper speed, which is adjusted to the expected signal variation in the data stream. (How one should use an AD-converter efficiently is discussed below.) Any information can then be extracted by digital means in the computer. In addition, when adjusting the window width, final filtering is also available with the advantage that one must not deal with nasty time constants, which may asymmetrically distort the signals after all. The procedure is simple enough to do everything on-line, so that one does not loose immediate access to the information. In principle, the algorithm may be implemented into a hard-wired FPGA with the advantage that time delay does not occur.

Other modulation schemes

The method described above could be used for signal detection in laboratory experiments. On the other hand, the rapidly developing digital technology allows considering other schemes for signal detection than a simple Lock-In amplifier for example uses. The disadvantage of a Lock-In is that any strong modulated interference close to any harmonics or sub-harmonics of the modulation frequency in use causes severe perturbations in the output. A typical example is the “hum” of the power line (50 or 60 cycles per second), which is often present in systems with very sensitive signal detectors. But it is by no means necessary to use a well-defined modulation frequency for the experimental procedure. Using computers one can easily detect signals when implementing any arbitrary sequence of signal switching. Let us assume to use a modulator which turns a signal On and Off in any desired fashion. Instead of periodically switching it might be useful to apply some random switching scheme, as long the detection method knows about the switching sequence. An example is the nowadays intensively used “Direct Sequence Spread Spectrum” known as ‘DSSS’ (see. e.g. in Wikipedia http://en.wikipedia.org/wiki/Direct-sequence_spread_spectrum). The basic idea is to distribute the signal over a large bandwidth around a center-frequency instead of using only one particular frequency – i.e. the modulation frequency. In consequence the signal becomes more resistant to interference problems. The method is widely used in the communication field, particularly for the radio connection of cell-phones. But in earlier days it was well known as the “Woodpecker” in the short wave ham radio bands. It was an over-the-horizon-Radar in Russia, which operated with enormous power by using quasi-randomly switched phases of the pulsed

emission. In consequence it became more or less insensitive to the radio-emissions of all stations operating in the same radio band.

The idea is to use a random modulation; in case of the Russian Radar it was a multiple phase switch within the duration of each emitted pulse of 3,1 msec based on a pseudo-random switch generator. In particular, its signal contained a clearly recognizable structure in each pulse, which was identified as a 31-bit pseudo-random binary sequence, with a bit-width of 100 µs resulting in the final pulse length. The pseudo-random number generator can be simply provided by linear shift registers who's output and some intermediate positions are coupled back using XOR-gates or similar to the input while the shift is clocked by an external clock. This scheme could easily be adapted for detecting noisy signals in laboratory experiments instead of ordinary Lock-In amplifiers while using computers controlling the signal modulation and the signal collection by means of an ADC. The advantage is that certain interference problems originating from power lines or motors etc. could be suppressed very effectively. (For further details of the pseudo-random-number generators look in http://en.wikipedia.org/wiki/Linear_feedback_shift_register. But it would be advisable, to look for generators with an even number of shifts, because otherwise an offset of the result would become very likely.) When considering schemes as are described above one has to think about a proper implementation of the analogue-to-digital converters. This is discussed below.

II. About Analogue to Digital Conversion

Today it is standard that one has to deal with digital instruments, it may be a simple voltmeter or more sophisticated instruments like digital storage oscilloscopes or else. The question is how the analogue-to-digital conversion should be applied in order to avoid significant losses of important information. It is not very common that there are significant differences in the behavior of analogue and digital devices. One of the typical effects one can observe with digital oscilloscopes when looking at noise of high frequency signals. More common is this effect visible on TV, when dense patterns cause strange "moiré fringes" on the screen. The "digital" patterns of the screen pixels are responsible for this effect. The "aliasing" effect is sometimes a quite annoying phenomenon. It is the purpose of the following articles to help understand and overcome typical problems of digitized data.

The "ideal" ADC

All Analogue-to-Digital Converters (ADC) add noise to the signal, although it is a completely different mechanism than just simple statistics. Let us assume that we have an ADC, which is perfect in response to the input. This means that we have regular and completely identical steps in input voltage generating the next bit combination. If we apply a noiseless input signal to the ADC in the voltage interval $n\cdot\Delta u$ and $(n+1)\cdot\Delta u$, we get an ADC answer "n" at the output. Δu is the voltage step which increases the ADC-reading by "1". Therefore, we will not be able to determine the voltage with accuracy smaller than Δu. In fact, it is a systematic error we have to deal with.

To illustrate this, let us assume that the probability for the occurrence of a voltage u between $n\cdot\Delta u$ and $(n+1)\cdot\Delta u$ is uniform. Thus we assume a probability distribution for the occurrence of the voltage u like

$$p(u) = \begin{cases} 1/\Delta u & n\cdot\Delta u < u \le (n+1)\cdot\Delta u \\ 0 & elsewhere \end{cases}$$ so that the integral over all values of u is unity.

This means that the average of the voltage at the input for an output reading "n" is given by:

$$< u > \; = \int_{n\cdot\Delta u}^{(n+1)\cdot\Delta u} u\cdot p(u)\cdot du \; = \; \frac{1}{\Delta u}\cdot \int_{n\cdot\Delta u}^{(n+1)\cdot\Delta u} u\cdot du \; = \; \tfrac{1}{2}\cdot[(n+1)^2 - n^2]\cdot\Delta u = \; (n+\tfrac{1}{2})\cdot\Delta u$$

The expectation for u^2 is:

$$< u^2 > \; = \int_{n\cdot\Delta u}^{(n+1)\cdot\Delta u} u^2\cdot p(u)\cdot du \; = \; \frac{1}{\Delta u}\cdot \int_{n\cdot\Delta u}^{(n+1)\cdot\Delta u} u^2\cdot du \; = \; \tfrac{1}{3}\cdot[(n+1)^3 - n^3]\cdot\Delta u^2 \; = \; (n^2+n+\tfrac{1}{3})\cdot\Delta u^2$$

The variance is then

$$\sigma_u^2 \; = \; < u^2 > - < u >^2 \; = \; [n^2+n+\tfrac{1}{3}-(n+\tfrac{1}{2})^2]\cdot\Delta u^2 \; = \; \Delta u^2/12$$

The standard deviation is therefore

$$rms \; = \; \sqrt{\sigma^2} \; = \; \Delta u/\sqrt{12} \; \approx \; 0.2887\cdot\Delta u$$

The uncertainty of the measured voltage is consequently:

$$\delta u \; = \; \pm\, 0.2887\cdot\Delta u$$

It sounds a bit strange that the uncertainty range is less than half a LSB, but this is the consequence of the definition of the standard deviation, which is based on the statistics of a Gaussian distribution. It is important to note that the error here is systematic! Therefore, repeating the measurement and averaging does not improve the accuracy!

It is rather amazing that the situation improves, if we add noise to the input. We may describe the input signal by its probability distribution:

$$p(u)\cdot du \; = \; \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\exp\left(\frac{(u-u_0)^2}{2\cdot\sigma^2}\right)\cdot du$$

The probability to read a value of "n" with the ADC leads then to the Error Function $erf(x)$:

$$P(n) \; = \; \int_{n\cdot\Delta u}^{(n+1)\cdot\Delta u} p(u)\cdot du = \frac{1}{2}\cdot\left\{ erf\left(\frac{(n+1)\cdot\Delta u - u_0}{\sqrt{2\cdot\sigma^2}}\right) - erf\left(\frac{n\cdot\Delta u - u_0}{\sqrt{2\cdot\sigma^2}}\right)\right\} \qquad [30]$$

If the noise amplitude is large enough, i.e. if $\Delta u/\sqrt{(2\cdot\sigma^2)} \ll 1$, then we can approximate

$$erf\left(\frac{(n+1)\cdot\Delta u - u_0}{\sqrt{2\cdot\sigma^2}}\right) = erf\left(\frac{n\cdot\Delta u - u_0}{\sqrt{2\cdot\sigma^2}}\right) + \frac{2}{\sqrt{\pi}}\cdot\exp\left(-\frac{(n\cdot\Delta u - u_0)^2}{2\cdot\sigma^2}\right)\cdot\frac{\Delta u}{\sqrt{2\cdot\sigma^2}} \pm \ldots$$

and we get:

$$P(n) \approx \frac{\Delta u}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\exp\left(-\frac{(n\cdot\Delta u - u_0)^2}{2\cdot\sigma^2}\right)$$ while neglecting all higher order terms.

The expectation value of the ADC output is now given by:

$$< n > = \sum_{n=0}^{N-1} n\cdot P(n) \approx \frac{\Delta u}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\sum_{n=0}^{N-1} n\cdot\exp\left(-\frac{(n\cdot\Delta u - u_0)^2}{2\cdot\sigma^2}\right)$$

and the expected voltage reading is

$$< u > = \sum_{n=0}^{N-1} n\cdot\Delta u\cdot P(n) \approx \frac{\Delta u}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\sum_{n=0}^{N-1}(n\cdot\Delta u)\cdot\exp\left(-\frac{(n\cdot\Delta u - u_0)^2}{2\cdot\sigma^2}\right)$$

The sum we can consider as the Riemann-sum of an integral

$$< u > \approx \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\int_0^{N\cdot\Delta u} u\cdot\exp\left(\frac{(u-u_0)^2}{2\cdot\sigma^2}\right)\cdot du \approx \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\int_{-\infty}^{\infty} u\cdot\exp\left(\frac{(u-u_0)^2}{2\cdot\sigma^2}\right)\cdot du = u_0$$

The result is not valid close to the edges of the ADC range, where parts of the voltage amplitudes are cut off.

Similar, we have for the expectation of u^2:

$$< u^2 > = < (n\cdot\Delta u)^2 > \approx \frac{\Delta u}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\sum_{n=0}^{N-1}(n\cdot\Delta u)^2\cdot\exp\left(-\frac{(n\cdot\Delta u - u_0)^2}{2\cdot\sigma^2}\right) \approx$$

$$\approx \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\int_0^{N\cdot\Delta u} u^2\cdot\exp\left(-\frac{(u-u_0)^2}{2\cdot\sigma^2}\right)\cdot du \approx \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot\int_{-\infty}^{\infty} u^2\cdot\exp\left(-\frac{(u-u_0)^2}{2\cdot\sigma^2}\right)\cdot du = u_0^2 + \sigma^2$$

Thus, the resulting variance is:

$$\sigma_{tot}^2 = < u^2 > - < u >^2 = \sigma^2$$

This looks like a trivial result but it is only correct, as long as the approximations above are valid. This is only possible for large σ^2 and far away from the ends of the input range of the ADC!

We have now learned about the two extreme cases $\sigma^2 \to 0$ and σ^2 large. What happens in between? This problem cannot be solved analytically, and we have to ask the computer for a correct answer. For an estimate of the systematic error we have to calculate the expectation <*n*> and <n^2> using the correct formula Eq.[30] while using as input $u_0 = (m+s)\cdot\Delta u$ with $0 \le s < 1$. When varying *s* we expect that the expectation <*u*> = <*n*>·Δu does not exactly agree with u_0. The result, i.e. the difference δu_0 = <*u*> - u_0, is plotted in Fig[2]. The structure repeats over the range of the ADC with the exception that near zero and *N* we have strong errors because of the missing voltage coverage of the ADC at the edges. When neglecting the edges we can now calculate the average variation of $\sigma_u^2 = <\delta u_0^2> - <\delta u_0>^2$, which should be equal to $\Delta u/12$ for $\sigma^2 = 0$. The result of the computation is plotted in Fig.[3] as a function of σ^2. For values of $\sqrt{\sigma^2}$ larger than $0.4\cdot\Delta u$ the remaining error is less than 1% of Δu, and the ADC operates with an accuracy of nearly 5 additional Bit when comparing with zero input noise (*rms* ~ $0.01\cdot\Delta u$ instead of $\Delta u/\sqrt{12}$).

It should be understood that one has to repeat the measurement many times in order to reduce the statistical error accordingly, since the discussion here concerns only the systematic error of the ADC. One can estimate the statistical error by using

$$\sigma_K^2 \approx 1/K\cdot\sigma^2.$$

Once σ_K^2 has reached the value of the systematic error, it is not worthwhile to continue sampling. Nevertheless, we have found the important information that fairly moderate noise amplitudes at the input improve the ADC resolution significantly, but many data samples are needed in order to reach this accuracy. For example, to get to the 1% level of one LSB one needs at least 1600 samples (=$[0.4/0.01]^2$), which is an issue for the available observing time. By the way, the KOSMA AOS use a sampling interval of 10 msec. The one percent level would therefore be reached after 16 seconds. But, don't forget, this is correct for a noise rms of 0.4 LSB. Typically one has much higher noise in the range of more than 10 LSB, when using a 12 Bit ADC. Therefore, one needs at least 3 hours integration time for 1% accuracy. This is not always an unusual observing time; therefore, the arguments here apply in reality indeed.

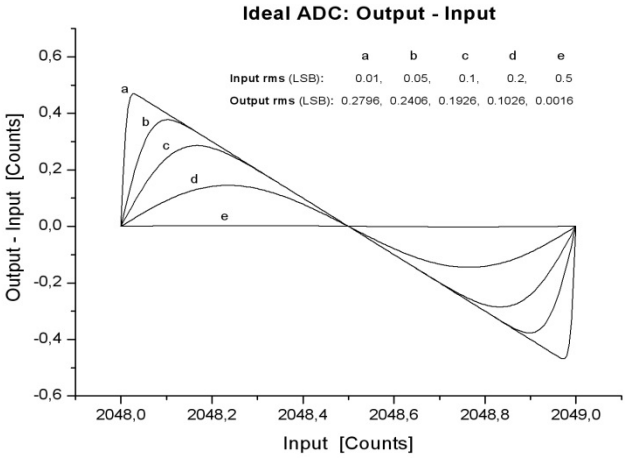


Fig.[2]:

Deviation of expectation of the ADC-output from input with various noise amplitudes. The mean deviation between average input and output is below 1% of one LSB at an input *rms* noise amplitude larger than 0.4 LSB. Thus, rather low noise levels suffice already for small systematic errors.

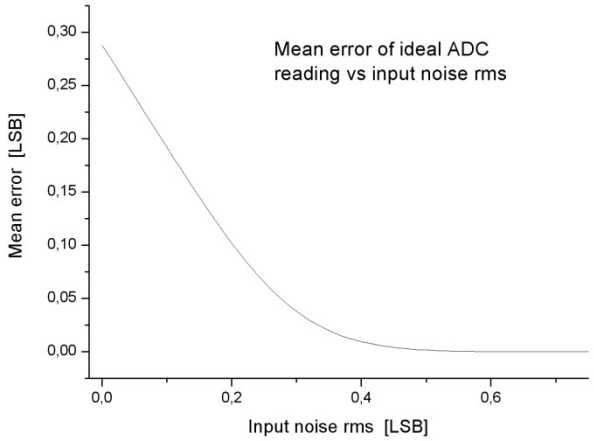


Fig.[3]:

Mean deviation of ADC output from input vs input noise rms.

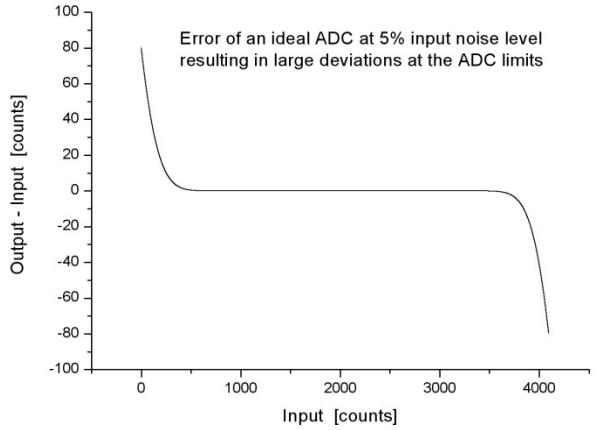


Fig.[4]:

Problem areas of an ideal ADC with 12 Bit at the edges of the input range. The input noise *rms* is 5% of the ADC range (205 counts). For an error of less than 1 LSB one has to stay away from the edges of the input range by more than 450 counts. This corresponds to a reduction of the useful range of the ADC by more than 20%!

We have mentioned above that near the edges of the ADC range an additional error appears, since the voltage fluctuations have a chance to lie outside the input range of the ADC. It depends now on the width of the fluctuations, how close one can operate at the edges. The Plot in Fig.[4] shows how the error develops near the edges for a fairly large standard deviation of the input noise. In general, one has to stay away from the edges by significantly more than 2 times the *rms* of the input noise to keep the error below 1 LSB. This tells us, that it is not of much use to increase the analogue gain in front of the ADC in order to improve the resolution. The loss in useful operating range is quite drastic. A good compromise should be to keep the noise *rms* at the input of the ADC in the range of a few percent of the ADC range.

The “real” ADC, Integral and Differential Non-Linearity,

Above, we have treated the performance of an ideal ADC, which unfortunately does not exist in reality. Any ADC has two problem areas: one is the integral non-linearity (INL) and the other the differential non-linearity (DNL). There are different interpretations of these two effects. We call integral non-linearity the deviation of the conversion curve from strict linearity. Such behavior is nothing new, if we remember that any amplifier has usually also some compression problems with similar consequences. The differential non-linearity is of different quality. It reflects the inevitable effect that the various ADC steps are not of equal size. The situation is sketched in Fig.[5]. On average the ADC curve never leaves the linearity curve completely, but instead it fluctuates around it. To illustrate the effect, we consider a partially ideal ADC with one step at some digital value “*m*”, where the step size is not Δu_0, but $\Delta u = \Delta u_0 \cdot (1+\varepsilon)$. We use again some co-added noise at the input, so that the step function of the ADC should not be significant. If the mean voltage at the input is now slowly var-

ied across the m^{th} position, we find for the output a deviation from the expected value like is plotted in Fig.[6]. How should we understand this?

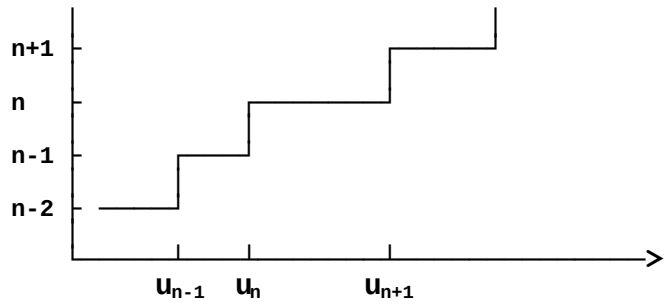


Fig.[5] Response of a "real" ADC to the input voltage.

We describe again the response of the ADC by using the probability function *p(u)*. The output range of the ADC is between 0 and *N*-1. We have now:

$$< n > \;=\; \sum_{n=0}^{N-1} n \cdot P(n) \;=$$

$$= \; 0 \cdot \int_{-\infty}^{\Delta u} p(u) \cdot du \; + \; \sum_{n=1}^{m-1} n \cdot \int_{n \cdot \Delta u}^{(n+1) \cdot \Delta u} p(u) \cdot du \; + \; m \cdot \int_{m \cdot \Delta u}^{(m+1+\varepsilon) \cdot \Delta u} p(u) \cdot du \; +$$

$$+ \; \sum_{n=m+1}^{N-2} n \cdot \int_{(n+\varepsilon) \cdot \Delta u}^{(n+1+\varepsilon) \cdot \Delta u} p(u) \cdot du \; + \; (N-1) \cdot \int_{(N-1+\varepsilon) \cdot \Delta u}^{\infty} p(u) \cdot du \;\; =$$

$$= \; \sum_{n=1}^{N-2} n \cdot \int_{n \cdot \Delta u}^{(n+1) \cdot \Delta u} p(u) \cdot du \; + \; (N-1) \cdot \int_{(N-1) \cdot \Delta u}^{\infty} p(u) \cdot du \; - \; \sum_{n=m+1}^{N-1} \int_{n \cdot \Delta u}^{(n+\varepsilon) \cdot \Delta u} p(u) \cdot du \;\; =$$

$$= \; < n >_{ideal} \; - \; \sum_{n=m+1}^{N-1} \int_{n \cdot \Delta u}^{(n+\varepsilon) \cdot \Delta u} p(u) \cdot du$$

For $\sqrt{\sigma^2} \gg \varepsilon$ we can approximate: $\int_{n \cdot \Delta u}^{(n+\varepsilon) \cdot \Delta u} p(u) \cdot du \;\approx\; \varepsilon \cdot P(n)$

with $P(n) \;=\; \int_{n \cdot \Delta u}^{(n+1) \cdot \Delta u} p(u) \cdot du$

Then we can combine:

$$\sum_{n=m+1}^{N-1} \int_{n \cdot \Delta u}^{(n+\varepsilon) \cdot \Delta u} p(u) \cdot du \;\approx\; \varepsilon \cdot \sum_{n=m+1}^{N-1} \int_{n \cdot \Delta u}^{(n+1) \cdot \Delta u} p(u) \cdot du \;=\; \varepsilon \cdot \int_{(m+1) \cdot \Delta u}^{N \cdot \Delta u} p(u) \cdot du \;=$$

$$= \; \frac{\varepsilon}{2} \cdot \left[erf\left(\frac{N \cdot \Delta u - u_0}{\sqrt{2 \cdot \sigma^2}} \right) - erf\left(\frac{(m+1) \cdot \Delta u - u_0}{\sqrt{2 \cdot \sigma^2}} \right) \right] \;\approx\; \frac{\varepsilon}{2} \cdot \left[1 - erf\left(\frac{(m+1) \cdot \Delta u - u_0}{\sqrt{2 \cdot \sigma^2}} \right) \right]$$

The last step is allowed because we assume that u_0 is far away from the edges of the ADC. Thus we have now:

$$< n > \;\approx\; < n >_{ideal} \; - \; \frac{\varepsilon}{2} \cdot \left[1 - erf\left(\frac{(m+1) \cdot \Delta u - u_0}{\sqrt{2 \cdot \sigma^2}} \right) \right]$$

If $[(m+1) \cdot \Delta u - u_0] \,/\, \sqrt{2 \cdot \sigma^2}$ is negative and large in absolute value, the Error Function approaches a value of -1 so that the error becomes equal to $\varepsilon \cdot \Delta u$. This dependence is plotted in Fig.[6] (with $\varepsilon < 0$), and one can see that at lower input voltage far below $m \cdot \Delta u$ the reading of the ADC is correct. Above we have an error which is determined by ε, which was set to -0.4 in our case. A too small step leads to a too large reading. In fact, we have constructed quasi an integral non-linearity, since the output deviates from strict linearity on average.

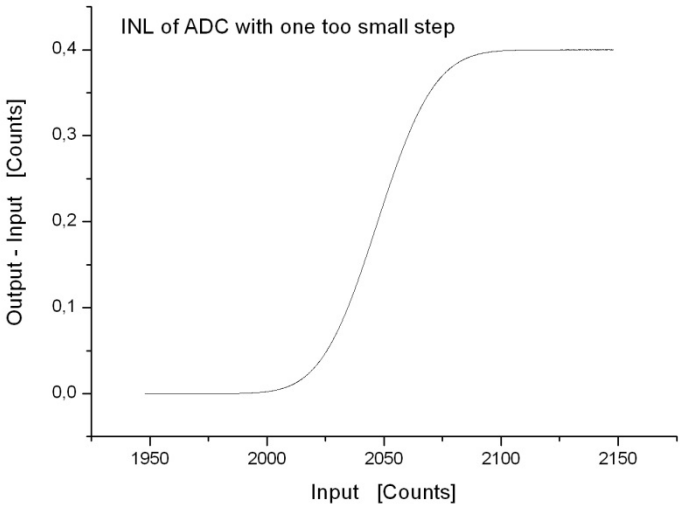


Fig.[6]:

INL of a nearly ideal 12 Bit ADC with an -0.4 LSB error of the highest Bit. The input noise *rms* is set to 10 LSB. Plotted is the difference between output and input.

In general, standard ADCs do not have just one irregular step, but instead, they have many of them. It is now dependent on the ADC type, how these errors look like. A standard ADC works with the principle of successive approximation, which means that each of the *K* bit has some minor error. Thus we have 2^K output values generated with *K* ADC parameter data. The output of the ADC is then

$$n = \sum_{k=0}^{K-1} p_k(n) \cdot 2^k$$

The p_k can be either 0 or 1. It represents a voltage:

$$u_n = \sum_{k=0}^{K-1} p_k(n) \cdot (2^k + \varepsilon_k) \cdot \Delta u\,, \quad \text{but is considered as} \quad u_n' = \sum_{k=0}^{K-1} p_k(n) \cdot 2^k \cdot \Delta u$$

Thus, the ADC does not provide an exact representation of the input voltage. Depending on the values of ε_k the deviation from ideal the ADC is characterized by a mean deviation from ideal described typically by a parameter *q* in the ADC specification sheets, which is the average (or sometimes maximum) deviation of the ADC output from ideal in units of 1 LSB (Least Significant Bit). Here we talk now about the "differential nonlinearity" (DNL). In case the ε_k are too large, the ADC may have "missing codes", i.e. there are output values which never can occur, because some step sizes become negative. Such ADCs are un-usable.
One can calculate now the performance of an ADC with DNL using some properly chosen values of ε_k. An example is given in Table.[2] for a 12-bit ADC. Certainly, also a Flash-ADC may be considered, which means that each output value has its own deviation from ideal, independent on the binary values.

K	0	1	2	3	4	5	6	7	8	9	10	11
ε_k	-0.011461	-0.037152	0.074245	-0.169669	-0.20269	0.23362	0.16646	0.2040	-0.0480	0.1252	-0.0560	-0.2210

Table.[2]: Sample for the deviations of the 12 conversion values of a sequential approximation ADC.

The ADC has now many such uneven step sizes so that the voltages, which are represented by the output values, fluctuate everywhere. It should be understood that these deviations from ideal are systematic, so that it does not help to average. Therefore one reaches fairly soon the ultimate limit in sensitivity which such ADC. This is particularly disastrous when trying to measure small signal differences with a spectrometer. Since each frequency pixel of the spectrometer usually sees different power, the resulting error looks different for each pixel. In consequence, the spectrum, particularly the baseline of the spectrum becomes much more "noisy" than anticipated. But, as was mentioned already, it is caused by a systematic error of the ADC, which one cannot correct afterwards as long as noise is involved.

A simple way to increase the accuracy of the system is to use ADCs with higher number of bit, since generally the absolute values of ε_k must stay significantly below 1 for all binaries, particularly for the lowest ones. This is the most important argument for ADCs with high number of Bit! Fig.[7] shows an example of the output signal plotted as the difference to the best linear fit (which represents the original input voltage) in units of 1 LSB. There is serious fluctuation, although it seems to have some regularity, which is correlated with the changes of individual Bit in the ADC. The calculation for the four traces in Fig[8] uses the data in Table [2] at noise levels of $\sqrt{\sigma^2}$ = 3, 10, and 30 LSB. The corresponding output *rms* converges to smaller values than without noise at a *rms* = 0.3806 LSB, which is significantly more than $1/\sqrt{12}$ = 0.2887 LSB of an ideal ADC, but it needs very long sampling times (~ 200, 3500 and 65,000 samples!). It demonstrates that the DNL effect is dependent on the width of the voltage fluctuations at the input. The larger the variations, the less visible the DNL becomes.

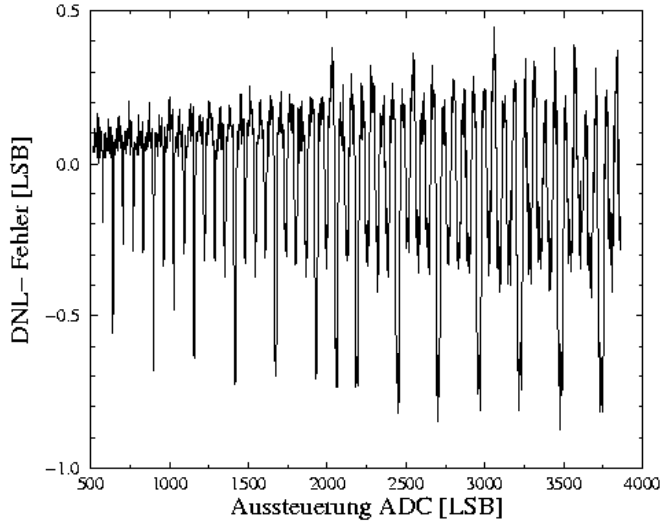


Fig.[7]:

Typical DNL of a commercial 12-Bit ADC (Experiment by Oliver Siebertz, PhD Thesis, Cologne 1998). Plotted is the difference between output and input of the ADC.

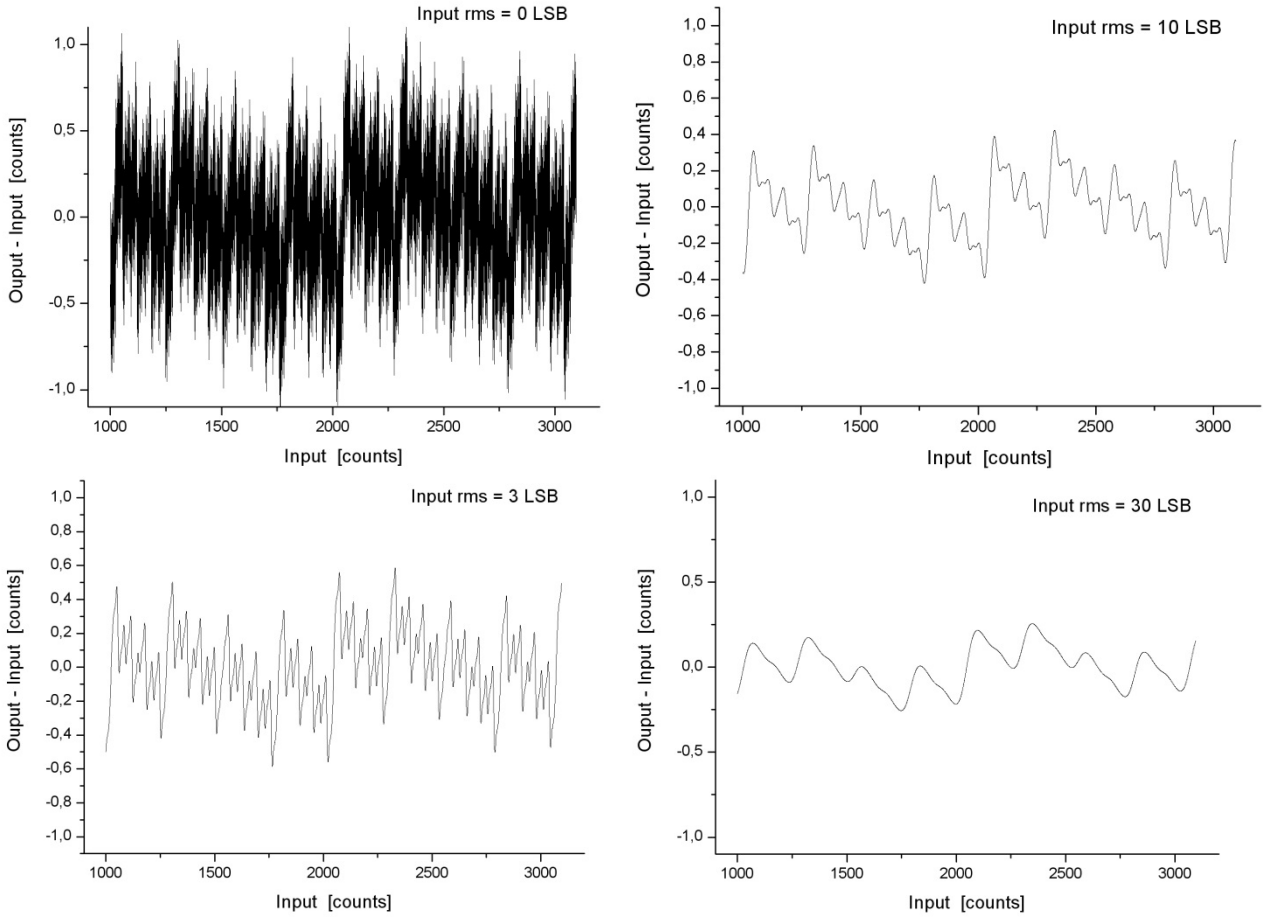


Fig.[8]: Calculated DNL of a 12-Bit sequential approximation model-ADC with different input noise levels. Plotted is the difference between output and input. The rms of the difference output-input is:
Top left (zero noise): 0.3806 LSB Bottom left (3 LSB noise): 0.2159 LSB
Top right (10 LSB noise): 0.1682 LSB Bottom right (30 LSB noise): 0.1168 LSB

This is a hint, how one can improve the situation. One can add a dither voltage to the input, it may be a regular saw-tooth signal or else. If one can measure the dither voltage independently while turning off the input signal, one artificially averages over a certain ADC range. With each reading the input voltage is therefore varied, so that in effect, one increases artificially the width of the statistical distribution of the input. But one does not have to deal with real statistics, since one knows about the offset for each single measurement. After subtraction of the measured dither value, one recovers the original signal voltage again. In effect the DNL errors are flattened so that one increases the accuracy

after all, but it is on the costs of some dynamic range of the ADC. A linear ramp of the dither voltage is already adequate for reducing DNL-effects of the ADC. Typically, a few percent of full ADC range dither amplitude is sufficient, so that the loss in dynamic range of the ADC is less than with equivalent noise amplitude, because the dither has no statistical behavior. Experience is that one is able to increase the accuracy of an ADC by a factor of three to four, which is often a sufficient improvement. This method was implemented in the WBS-system (Wide-Band-Spectrometer) of the HIFI-instrument on the Herschel-observatory successfully. The performance of a 14-Bit ADC was improved close to the accuracy of a 16-Bit ADC. Such accuracy is particularly important, when detecting weak signals with a heterodyne system at larger offsets between On- and Off-positions, as are typical at the presence of continuum offsets. A good example is the situation while observing atmospheric signals.

Data acquisition with an AD-converter

In general, AD-converters probe the applied signal voltage at a certain instant in time, and they do not integrate like a box-car integrator for example. The typical acquisition time of an ADC is in the range of µsec or sometimes much less, which means that the noise amplitude seen is much higher than assumed when averaging over certain time intervals. This is especially true for an input which contains white noise. Also, the well-known aliasing problem becomes visible when taking data at a speed comparable with the period of an oscillating signal at the input. It would be much more informative, if one could integrate with the device. This can be achieved by either using an analogue box-car integrator before sampling or by sampling at high speed and co-adding afterwards. The co-adding frequency should be faster than the time constant, or better the correlation time of the input signal. Let us assume that we have a low pass filter to smooth the data according to a time constant τ. The signal *y(t)* at the output of the filter is then:

$$y(t) = \frac{1}{\tau} \cdot \int_{-\infty}^{t} x(t') \cdot \exp\{-(t-t')/\tau\} \cdot dt'$$

x(t) is the signal coupled to the filter. We assume now that we sample the signal *y(t)* with the ADC at a sampling rate of 1/δ samples per second. We take *N* samples and co-add them, so that we get finally as the final data value:

$$z(t) = \frac{1}{N} \cdot \sum_{n=0}^{N-1} y(t + n\delta)$$

It is obvious that the result approaches that of a box-car integrator as long as
$\delta \ll \tau$ and $N \cdot \delta \gg \tau$.
In reality, we are never close to this. The question is therefore, how the three parameters *N, δ,* and τ can be optimized. When co-adding the ADC output over *N* samples, the co-added signal is then:

$$z(t) = \frac{1}{N \cdot \tau} \cdot \sum_{n=0}^{N-1} \int_{-\infty}^{t_n} x(t') \cdot exp\left\{-\frac{t_n - t'}{\tau}\right\} \cdot dt', \quad t_n = t + \vartheta + n \cdot \delta$$

Due to the time constant, the data are now partly correlated. ϑ stands for a "wait time" before data accumulation begins.

We consider an application like with a Lock-In amplifier. We assume that the output of our signal detector provides a modulated signal, which is of rectangular shape. The modulation amplitude we call ΔS. We assume that we have a change in signal by ΔS at *t*=0. Then we have for the response Δz:

$$\Delta z(t) = \Delta S \cdot \frac{1}{N} \cdot \sum_{n=0}^{N-1} \left[1 - e^{-\frac{t_n}{\tau}}\right] = \Delta S \cdot \left\{1 - \frac{1}{N} \cdot e^{-\eta} \cdot \frac{1 - e^{-N \cdot \zeta}}{1 - e^{-\zeta}}\right\}, \quad \zeta = \frac{\delta}{\tau}, \quad \eta = \frac{\vartheta}{\tau}$$

The time ϑ stands here for a waiting time after the change ΔS has occurred in order to allow some response of the filter to the signal change before sampling. With very long ϑ the response Δz becomes identical with ΔS, but one would waste a lot of observing time. On the other hand, at short ϑ, it is clear that the time constant τ leads to some memory of the previous signal when starting with the ADC sampling too early.

We assume that the noise of the signal input is purely white. In this case we can describe the noise of the unfiltered signal by a correlation function g_x like:

$$g_x(T) = < x(t+T) \cdot x(t) > = < x(t) >^2 + c^2 \cdot \delta(T)$$

<*x(t)*> = <*x*> is the mean signal value. The constant "c" describes the amount of noise and is not important here. The correlation function of the filtered signal g_y is now:

$$g_y(T) \;=\; < y(t+T)\cdot y(t) > \;=\; < \frac{1}{\tau^2}\cdot \int_{-\infty}^{t+T} dt' \int_{-\infty}^{t} dt''\cdot x(t')\cdot x(t'')\cdot e^{-(2\cdot t+T-t'-t'')/\tau} > \;=$$

$$= \;< x >^2 + \frac{c^2}{2\cdot\tau}\cdot e^{-T/\tau} \quad (T \geq 0), \quad g_y(-T) \;=\; g_y(T)$$

This means that the output data of the filter are now correlated with a time constant τ. The variance of the noise fluctuations behind the filter is determined by the value g_y at T=0 and at T= ∞, and we have for the noise:

$$\sigma_y^2 \;=\; g_y(0) - g_y(\infty) \;=\; \frac{c^2}{2\cdot\tau}$$

For the noise of the co-added signal we have Eq.[5] from above:

$$\sigma_z^2 \;=\; \frac{1}{N}\cdot\sigma_y^2\cdot\left\{1 \;+\; 2\cdot\sum_{k=1}^{N-1}\left(1-k/N\right)\cdot g_k\right\}, \quad g_k \;=\; e^{-\delta_k/\tau}, \quad \delta_k \;=\; k\cdot\delta$$

Therefore we get now:

$$\sigma_z^2 \;=\; \frac{c^2}{2\cdot N\cdot\tau}\cdot\left\{1+2\cdot\sum_{k=1}^{N-1}\left(1-k/N\right)\cdot e^{-k\cdot\xi}\right\}, \quad \xi = \frac{\delta}{\tau}$$

The sum can be evaluated and we get finally:

$$\sigma_z^2 \;=\; \frac{c^2}{2\cdot N\cdot\tau}\cdot\left\{\frac{1+e^{-\zeta}}{1-e^{-\zeta}} - {}^{2}\!/_{N}\cdot e^{-\zeta}\cdot\frac{1-e^{-N\cdot\zeta}}{[1-e^{-\zeta}]^2}\right\}$$

At $N\cdot\zeta \ll 1$ ($N\cdot\delta \ll \tau$) the sum in the brackets approaches N, and we have then:

$$\sigma_z^2 \;=\; \frac{c^2}{2\cdot\tau}$$

This means, that we have zero improvement of the noise when repeating the signal acquisition N times.

On the other hand, if τ is very small ($\zeta \gg 1$, $\delta \gg \tau$), the variance becomes

$$\sigma_z^2 \;=\; \frac{c^2}{2\cdot N\cdot\tau}$$

But, since τ is supposed to be small as compared to δ, the value of σ_z^2 is still larger than expected.

If $\zeta \ll 1$, i.e. $\delta \ll \tau$ and $N\cdot\zeta \gg 1$, i.e. $N\cdot\delta \gg \tau$, then we get:

$$\sigma_z^2 \;=\; \frac{c^2}{N\cdot\delta},$$

since N is very large in this case. This is practically identical with the ideal case of a true box-car integrator, which would be:

$$\sigma_z^2 \;=\; \frac{c^2}{\vartheta + (N-1)\cdot\delta}$$

If we put the waiting time ϑ equal to δ, the results are identical. It makes sense that we have to wait after the occurrence of the signal change, since the response is delayed due to the time constant τ. A sample close to ϑ = 0 would contribute practically nothing to the signal reading anyway. The last condition shows, how one has to operate: Sample in short intervals, short compared to the time constant of the analogue system, and do it much longer than the time constant suggests.

It is not the noise itself alone, which is important, but instead, it is the signal to noise ratio (S/R) one is usually interested in. We consider the signal to noise as the ratio of signal response Δz and the standard deviation $\sqrt{\sigma_z^2}$. The optimum is achieved with a box-car integrator. S/R_{Box} is the theoretical signal to noise ratio obtainable with a box-car integrator, while integrating for the full time interval $t_{Int} = \vartheta+(N-1)\cdot\delta$.

$$S/R_{Box} = \;\frac{\Delta z}{c}\cdot\sqrt{\vartheta + (N-1)\cdot\delta}$$

We discuss the S/R in comparison with that of a boxcar integrator. The ratio R of both is given by:

$$R = \frac{S/R}{S/R_{Box}} = \sqrt{\frac{2\cdot N}{\eta + (N-1)\cdot\zeta}} \cdot \frac{1 - \frac{1}{N}e^{-\eta}\cdot\frac{1-e^{-N\cdot\zeta}}{1-e^{-\zeta}}}{\sqrt{\frac{1+e^{-\zeta}}{1-e^{-\zeta}} - \frac{2}{N}e^{-\zeta}\cdot\frac{1-e^{-N\cdot\zeta}}{[1-e^{-\zeta}]^2}}}, \quad \zeta = \frac{\delta}{\tau}, \quad \eta = \frac{\vartheta}{\tau}$$

The expression can be optimised, if we determine the maximum for the Signal-to-Noise as a function of the number of co-added samples *N*. This is plotted in Fig.[9]. As it turns out, the optimum is always found at a delay near $\eta \approx 1$, i.e. $\vartheta \approx \tau$, which means that longer waiting for further settling of the exponential response does not help. For various *N* the best values are given in Table [3].

As one can see in the table, the signal to noise improves with large *N*. Also, the sampling interval δ becomes smaller relative to the time constant τ with increasing number of co-added samples. There exists also an optimum wait time before one starts to sample the signal with the ADC. This is understandable, since at the very beginning the signal has not yet come close to the final value, so that early samples cannot contribute significantly to the total. Also, In order to understand the table or the plot respectively, a comment should be made: If one wants to sample a signal change ΔS only with one reading of an ADC, the optimum signal to noise is provided if one starts to sample after the signal has developed to 1.26 times the applied time constant τ. But the reading Δz will only be 71% of ΔS and the signal to noise will be about 90% of that of an ideal box-car integrator. In general, the observed signal amplitude Δz is strongly dependent on *N*. With $N \leq 5$ the error of

N	η_{Opt}	ζ_{Opt}	R_{Opt}	$\Delta z/\Delta S$ at Opt
1	1.2564	-	90.3%	71.5%
2	1.3344	0.7200	92.0%	80.4%
5	1.2611	0.7696	94.0%	89.7%
10	1.1454	0.6787	95.6%	93.6%
20	1.0450	0.5627	96.9%	95.9%
50	0.9452	0.4254	98.1%	97.8%
100	0.8891	0.3410	98.8%	98.5%
200	0.8457	0.2723	99.2%	99.1%
500	0.8031	0.2015	99.5%	99.5%
1000	0.7792	0.1603	99.7%	99.7%

Table [3]: Optimum conditions for integration as a function of the number of samples. η is the ratio of the waiting time ϑ and time constant τ, ζ is the ratio of the repetition time δ of ADC readings and time constant.

Δz at optimum signal to noise is more than 10%, which might be already acceptable. An error of less than 1% requires an *N* in the range of 200 while maintaining best signal to noise. Therefore, one should sample with high speed with the ADC and select the number of samples according to the modulation frequency. The highest modulation frequency is determined by the minimum number of samples needed for an acceptable signal to noise. A proposal might be that *N* should be always larger than or equal to 10. This guarantees a signal to noise level above 95% in comparison to an ideal analogue box-car integrator and signal efficiency above 90%. By the way, for the range $N \geq 5$ a reasonable estimate is near $\eta = 1.0$ (wait time $\vartheta \approx \tau$) and near $\zeta = 0.5$ (sampling interval $\delta \approx \frac{1}{2}\cdot\tau$). The loss in comparison with the ideal signal to noise is not more than about 1% for all *N* in this range. For practical reasons it is therefore recommended to use these values for a computer controlled system.

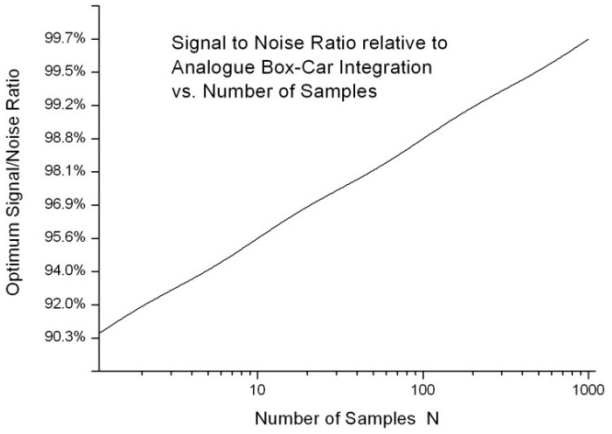


Fig.[9]:

Signal to Noise ratio of the ADC box-car simulator
In percent of an ideal box-car integrator as a function of the number of samples.

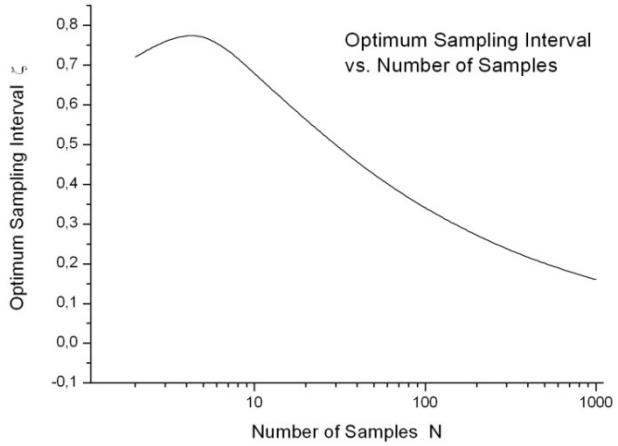


Fig.[10]:

Optimum sampling interval δ in units of the analogue time constant τ as a function of the number of samples *N*.

When operating an AD-converter with the performance near an analogue box-car integrator, one has to consider the signal to noise as well as the error of the measurement. Finally, one usually averages over many modulation periods of the signal and get a box-car integrated answer of the system to the modulation. The discussed method is much faster in response than a normal RC-filter could ever provide; therefore, the described method should be advantageous in comparison. When using an analogue Lock-In amplifier with subsequent analogue to digital conversion, the situation is not different than described above. When reading the output signal, one needs to adjust the sampling rate of the ADC to the time constant of the Lock-In accordingly.

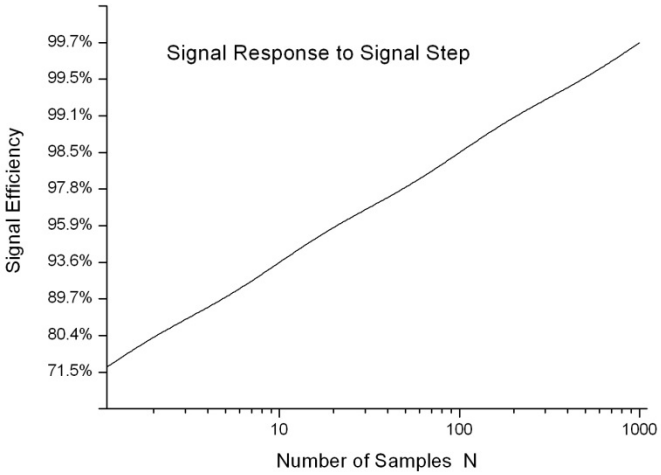


Fig.[11]:

Response efficiency of the ADC box-car simulator to a signal step at optimum signal to noise as a function of the number of samples *N*.

Usually, there is always some residual ripple left from the signal modulation at the output of the Lock-In, which can only be reduced by means of a long time constant when using the RC-circuitry of the Lock-In. This is a particular disadvantage of the analogue Lock-In. When demodulating with the proposed scheme, there is no such left-over ripple, which makes the method very different in behaviour. The final integration, which corresponds to the operation of the RC-filter in the analogue Lock-In, can then easily be done in the computer just by co-adding the results of several modulation periods. The next sampling can then start without any superfluous memory effect of the applied time constant. This means that one can save experiment time, because no precious time is wasted while waiting for the settling of the RC-filter. This should be seen in the context of much better signal accuracy. What we have learned from the above exercise is that one should strongly oversample a signal, if one wants to achieve high efficiency and high signal to noise. And, in addition, the use of a usually rather expensive high quality analogue Lock-In-amplifier becomes obsolete. Actually, there are commercial digital Lock-In's available, but they all do not use the scheme described above, consequently their performance should not be optimal. By the way, the method was implemented in the Gornergrat Continuum-AOS, which was used for measuring broad-band continuum signals as well as a narrow-band signal of an external transmitter for the telescope surface alignment.

Handling of noisy digital data

Usually, when collecting data with some digitizing equipment, one rarely considers the economy of data storage. The recent developments in digital technology make it look superfluous to think about eventual problems of storage capacity or redundancy of the data. But there are still some areas, where such considerations remain quite important like in space programs for example. The following paragraphs should help to make everybody become more aware of such problems and suggest eventually some ideas how to deal with them. By the way, resampling of data, as is discussed below, is also a tool to reduce any data amount to a bearable level. Quite often, the output of the instrumentation carries a lot of redundant information, which is not worth to be kept.

How to store noisy data

For the calculation of the number of Bit required for the data storage and transmission of spectrometer data – in our case AOS data – which are subject to radiometric noise, one has to compare the maximum number of counts at the longest desired integration time versus the minimum absolute noise at minimum integration time and minimum signal level. For this the desired limit for an eventual noise increase due to digitization has to be defined. Her and in the following we assume that the relative noise level on all data is constant, so that the amplitude of the fluctuations is directly proportional to the mean signal level.

It is well known that the digitization error of an ideal ADC is given by (see also above):

$$\delta = 1/\sqrt{12}\ LSB$$

(We neglect here the DNL-problems of ADCs.) For a given digital resolution the total noise relative to the radiometric noise level is then given by:

$$\delta s = \sqrt{1 + x^2/12}$$

with *x* the ratio of digital resolution (1 LSB) and the noise *rms*. δs may be for example 1.02 (2% noise increase) and *x* becomes in this case 1/√2 = 0.707 (one count of the ADC corresponds to 70.7% of the *rms*).

According to the radiometer equation the noise amplitude of the signal is:

$$\delta S = S/\sqrt{B_{Fl} \cdot t_{Int}}$$

B_{Fl} is the fluctuation bandwidth of the spectrometer pixel. The total signal *S* is comprised of *n* read-outs of a CCD in an AOS for example, each integrated with a "frame time" t_0 (t_0 = 0.01 sec in our case). The integration time t_{Int} is therefore given by:

$$t_{Int} = n \cdot t_0$$

The signal *S* is co-added from *n* read-outs, so that we have:

$$S = n \cdot s = n \cdot \alpha \cdot s_{Max}$$

with $s_{Max} = 2^p - 1$ the maximum number of counts per frame time provided by the ADC (*p* Bit), and α the input level to the ADC relative to s_{Max}. The spectrometer shall be used within a dynamic range *D*, so that the minimum drive level α is then:

$$\alpha_{Min} = 1/D.$$

The minimum signal allowed is therefore:

$S_{Min} = n_{Min} \cdot (2^p - 1)/D$ with n_{Min} the assumed minimum number of read-outs

With this we have for the minimum noise amplitude:

$$\delta S_{Min} = n_{Min} \cdot \frac{(2^p - 1)}{D} / \sqrt{B_{Fl} \cdot t_0 \cdot n_{Min}}$$

The required digital resolution is equal to $x \cdot \delta S_{Min}$ in order to achieve the desired total noise level.

The maximum number of counts S_{Max} is given by the maximum number of read-outs n_{Max} times the maximum of ADC counts:

$S_{Max} = n_{Max} \cdot (2^p - 1)$ with $n_{Max} = t_{Max}/t_0$, t_{Max} is the assumed maximum integration time.

This value as well as the noise of the minimum signal must both be described by a predefined maximum number of Bit. We have therefore for the required dynamic range of the digital data representation:

$$r = S_{Max}/\delta S_{Min} = t_{Max}/t_{Min} \cdot \frac{D}{x} \cdot \sqrt{B_{Fl} \cdot t_{Min}} = t_{Max}/\sqrt{t_{Min}} \cdot \frac{D}{x} \cdot \sqrt{B_{Fl}} = n_{Max}/\sqrt{n_{Min}} \cdot \frac{D}{x} \cdot \sqrt{B_{Fl} \cdot t_0}.$$

The number of required Bit to describe all possible data values within the range S_{Min} and S_{max} with sufficient accuracy is:

$$m = \log[r]/\log[2]$$

It is interesting to note that the number of ADC Bit does not contribute here directly as long as the total number of Bit after co-adding is covering the whole range needed for the data representation. As an example we consider the Wide-Band-Spectrometer (WBS) of the Heterodyne-Instrument-for-the Infrared (HIFI) of the Herschel Satellite. The maximum integration time should be 160 sec, which makes n_{Max} = 16000. B_{Fl} is 1.6 MHz, and t_0 is 10 msec. We want to keep the increase in noise at 2% maximum, thus we have x = 0.707. The dynamic range of the WBS Acousto-Optical-Spectrometer is expected to be 13 dB so that D is 20. The minimum integration time is assumed to be 80 msec, and we find for the required number of Bit:

$$m = log\left[160/\sqrt{0.08} \cdot 20/0.707 \cdot \sqrt{1.6 \cdot 10^6}\right]/log[2] \approx 24$$

This means that 24 Bit are sufficient to describe all possible data values of WBS within the given dynamic range and for integration times between 0.08 and 160 seconds. Given a 16 Bit ADC the lowest 6 Bit of the accumulated data (or of the ADC output) can be truncated without risking a significant degradation of the data accuracy. (This does not mean that a 10-Bit ADC would be sufficient, because the DNL problems would become very significant.) For integration times between 0.02 and 80 seconds the same number of Bit is appropriate, but in this case the lowest 5 Bit must be truncated. It is interesting to note that the ratio of maximum and minimum integration time becomes larger for smaller n_{Max}!

The example of the HIFI spectrometer demonstrates how an efficient data handling can be established on the basis of the noise amplitudes of the data. The procedure is not limited to spectrometer data, but any other sequence of noisy data from an infrared detector for example can be treated similar.

Reduction schemes for noisy data

Introduction

It is one of the very important and difficult tasks in space or on ground to deal with the typically overwhelming amount of data generated on the course of the lifetime of an experiment. For this it is essential to develop compression schemes which are designed for the particular data structure in the experiment. The situation is rather special in case of spectral investigations at radio-frequencies, because the major content of the data is white and uncorrelated noise. This prohibits in principal any effective compression scheme, but, on the other hand, the noisy character of the data may suggest considering data transmission and data storage with slightly reduced signal to noise characteristics so that more efficient compression algorithms might become applicable.

The typical situation in radio-astronomy is that data are generated at rather high speed in a multi-channel real time spectrometer with sometimes several thousand channels in parallel. On a satellite, the data provided by the back-ends should be stored in rather small time segments in order to achieve highest data reliability in case some data become affected by radiation effects for example. If single small scans are degraded, it then would not hurt too much to throw one scan away. On the other hand, smaller time segments require more storage volume, therefore some reasonable compromise or data compression algorithms must be found. Most compression algorithms are based on some kind of subtraction scheme, so that only the changes of the data from one scan to the next are stored. This increases the risk of data loss, because any signal distortion in one data set destroys the following and therefore dependent data as well. Therefore there are limits for the maximum compression efficiency.

Direct Compression

As a first estimate for the amount of Bit needed to transmit one complete set of spectrometer data we need to define the minimum amplitude resolution to establish a reliable representation of the radiometric noise. From the radiometer equation we know that the *rms* of the fluctuations is given by:

$$rms_{Rad} = \frac{<x>}{\sqrt{B_{Fl} \cdot t_{Int}}}$$

with <*x*> the mean of the data *x*, B_{Fl} the "fluctuation bandwidth", and t_{Int} the integration time. The minimum *rms* we get at the minimum power level applied to the spectrometer (the minimum of <*x*>), which is defined by the required dynamic range of the spectrometer. If D is the dynamic range (e.g. 13 dB, D = 20), we have:

$$rms_{Min} = \frac{<x_{Min}>}{\sqrt{B_{Fl} \cdot t_{Int}}} = \frac{1}{D} \cdot \frac{<x_{Max}>}{\sqrt{B_{Fl} \cdot t_{Int}}}$$

In order to resolve these fluctuations sufficiently, the LSB (Least-Significant-Bit) of the digital representation of the data should correspond to some fraction of the *rms* of the original data. The resolution of the analogue to digital conversion is therefore determined for example by:

$$LSB = rms_{Min}/\sqrt{2}$$

This estimate corresponds to an additional sampling noise of

$$\sigma_{Samp}^2 = \left[\frac{1}{\sqrt{2}} rms_{Min}/\sqrt{12}\right]^2 = [0{,}204 \cdot rms_{Min}]^2,$$

since it is well known that the digitization noise is 1 LSB/√12 (see above). Thus, with this digital resolution we would have an increase of the total noise at minimum signal level like:

$$rms_{tot} = \sqrt{\sigma_{Rad}^2 + \sigma_{Samp}^2} = \sqrt{1 + \cdot 1/24} \cdot rms_{Rad} = 1.0206 \cdot rms_{Rad}$$

A 2% increase of the noise is probably acceptable in most cases, and it should be kept in mind that for higher signal levels it becomes less. Due to the unavoidable quantization noise it is inevitable that with digitization of analogue data one has to accept some increase in noise anyway, how much can be determined and is a key question for the compression algorithms. For a correct understanding of the treatment of compression and related issues, this effect should always be included in order to obtain a reasonable idea about the consequences of "added" noise when digitizing analogue signals.

The maximum digital value M_{Max} of the data becomes now:

$$M_{Max} = \frac{<x_{Max}>}{LSB} = Int\{2 \cdot D \cdot \sqrt{B_{Fl} \cdot t_{Int}}\}$$

For WBS of the HIFI-instrument on Herschel we have a fluctuation bandwidth per pixel $B_{Fl} \approx 1.6$ MHz, which corresponds to about 1 MHz resolution bandwidth. Desirable is a data dump every, say, 5 seconds. This implies that the data must be stored with 17 Bit accuracy according to above equation. The resulting number of Bit to describe the 8 GHz total bandwidth at Nyquist sampling with 1 MHz resolution is therefore

$$N_{Bit} = \frac{log[M_{Max}]}{log[2]} \cdot 8 \cdot 2000 = 272\ kBit$$ in 5 seconds or

$$n_{Bit} = \frac{N_{Bit}}{t_{Int}} = 54.4\ kBit/sec$$

This is certainly far too much for the storage capabilities on a satellite like HERSCHEL because it would result in about 4,7 GBit data per day which would have to be stored and, even more problematic, to be down-linked to ground in preferably within very short time slots. Therefore all possibilities for data compression need to be implemented.

Some improvement can be achieved by applying a fit to the shape of the spectrum using some set of simple orthogonal polynomials or other orthogonal functions (low order coefficients of a Fourier transform for example, as was used during the SWAS mission). It is then sufficient to transmit the fit parameters and the differences of the data to the fit need to be handled only. Depending on the accuracy of the fit the compression can become very significant, but the details of the spectra will decide how much improvement is really possible. The polynomials chosen may be some specially defined "discrete polynomials" [8]. In principle, this method can be practically "loss-less", if no reduction of Bit is applied, but compression with some reduced signal to noise may also be considered.

Ratio Data

It is one of the frequently used procedures when applying compression to consider the differences of contiguous sets of data in case the expected changes of the data are marginal. In case of radiometric noise, the noise fluctuations of such differences will vary with varying gain of the different frequency pixels of a spectrometer. This is not a very good starting point for a simple and efficient compression algorithm, since the number of required Bit to represent these fluctuations is not uniform across the bandwidth of the spectrometer. The situation may be different, when considering data from an infrared detector which are typically detec-

[8] see e.g. Handbook of Mathematical Functions, M.Abramowitz and I.A.Stegun, Ninth Printing, Orthogonal Polynomials of a Discrete Variable, p.788

tor-noise limited and not influenced by the signal level itself. In such cases differences of adjacent scans would always have the same *rms* so that an efficient compression using differences may be rather successful.

The way how radio-astronomical data usually appear offers some alternative approach. In general, one wants to split all data into small chunks in order to reduce the danger loosing too much information if some glitches occur from time to time. For example, during total power measurements with a telescope (differences of On- and Off- integrations on a signal- and a reference-position) we can expect that we have to integrate for, say, 60 seconds on each position in order to deal with the rather slow slew capabilities of the telescope or, even worse, of a spacecraft. 60 seconds can be composed of twelve 5-second scans for example. During this integration we do not expect to see a significant change of the signal amplitude, if only radiometric noise is involved. This can (and must) be verified by the spectroscopic Allan variance plot or similar. If the plot is still in the regime with a slope of -1 during 60 seconds integration time, we know that there are no other significant signal fluctuations than those described by the radiometer equation. Under this assumption it is sufficient to transmit the full data amplitude of the first scan, and afterwards, only the differences to the first scan need to be stored.

The transmission of differences of data has the disadvantage that the noise to be resolved is dependent on the signal level on each of the pixels which might vary by much more than 3 dB. It is therefore much better to provide the data of the ratios of scans, because the *rms* of the fluctuations will then be more or less constant and independent on the actual signal level. For this it is assumed that the "zero-level" of the spectrometer output is already subtracted, which may be an important requirement. Thus let us assume that we calculate the ratio of two spectra pixel by pixel:

$$r_q(n) \;=\; \frac{x_q(n)}{x_q(1)},\;\; n = 2,3,\ldots N$$

q indicates the pixel number of the spectrometer ($1 \le q \le Q$), and *n* the sequence of the 5-second scans. $x_q(1)$ is the first scan in a sequence. All the values of $r_q(n)$ will be non-negative since the fluctuations of the initial data cannot lead to negative power whatsoever.

x_1 and x_2 may have the same mean values, but this is not essential for the discussion. We assume that both data x_i have the same relative statistical distribution, therefore we describe the relative variance of both data with

$$< (x_i - \langle x_i\rangle)^2 > / \langle x_i\rangle^2 \;=\; rms_{Rad}^2 / \langle x_i\rangle^2 \;=\; \delta_{Rad}^2 \quad i = 1,2.$$

δ_{Rad} is assumed to be small as is normal for data with radiometric noise. We want to represent the ratio *r* by numbers *m*, which guarantee that the resolution is sufficient for the noise fluctuations of *r* in units of fractions of δ_{Rad}. Usually *r* is on average equal to “1”, but its digital representation by *m* should be larger in order to provide sufficient resolution. We write:

$$r \;=\; m \cdot \delta_{Rad}/q$$

We therefore describe *r* by multiples of fractions of δ_{Rad} as defined by *q*, and the result *m* should represent now the ratio *r*.

The variance of the new variable *m* is now:

$$\sigma_m^2 \;=< (m - \langle m\rangle)^2 > \;=\; g^2 \cdot q^2$$

with $\;g^2 \;=\; \frac{<(r-\langle r\rangle)^2>}{\delta_{Rad}^2} \;=\; \frac{\sigma_r^2}{\delta_{Rad}^2}$

For small noise amplitudes of the data x_i, we find for σ_r^2:

$$\sigma_r^2 \;=\; \langle\left(\tfrac{\langle x_1\rangle+\delta x_1}{\langle x_2\rangle+\delta x_2}\right)^2\rangle - \langle\tfrac{\langle x_1\rangle+\delta x_1}{\langle x_2\rangle+\delta x_2}\rangle^2 \;\approx\; \left(\tfrac{\langle x_1\rangle}{\langle x_2\rangle}\right)^2 \cdot \langle\left[\left(1+\tfrac{\delta x_1}{\langle x_1\rangle}\right)\cdot\left(1-\tfrac{\delta x_2}{\langle x_2\rangle}\right)\right]^2\rangle - \langle\left(1+\tfrac{\delta x_1}{\langle x_1\rangle}\right)\cdot\left(1-\tfrac{\delta x_2}{\langle x_2\rangle}\right)\rangle^2 \;\approx$$

$$\approx\; \langle r\rangle^2 \cdot \left[\langle\left(\tfrac{\delta x_1}{\langle x_1\rangle}\right)^2\rangle + \langle\left(\tfrac{\delta x_2}{\langle x_2\rangle}\right)^2\rangle\right] \;=\; 2\cdot\langle r\rangle^2\cdot\delta_{Rad}^2 \;=\; 2\cdot\delta_{Rad}^2$$

for uncorrelated x_1 and x_2, $< x_1> = <x_2 >$ and $\delta_{Rad} \ll 1$

The δx_i are the fluctuations around $<x_i>$. (By the way, the expectation of $<r> \;=\; <(x_1/x_2)>$ is not exactly identical with the value of $<x_1>/<x_2>$, but in our case the difference is negligible.) Therefore, we find that *g* is equal to $\sqrt{2}$.

The probability distribution of *m* is now given by a Gaussian distribution:

$$w(m)\cdot dm = \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_m^2}}\cdot exp\left\{-\frac{(m-<m>)^2}{2\cdot\sigma_m^2}\right\}\cdot dm$$

For a digital representation of the signal we consider only the (rounded) integer values of *m*. The probability to find *m* within an interval of the width 2^p ($-2^{p-1} \le m-<m> \le 2^{p-1}$) is now given by:

$$w_p = \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_m^2}}\cdot \int_{<m>-2^{p-1}}^{<m>+2^{p-1}} exp\left\{\frac{(m-<m>)^2}{2\cdot\sigma_m^2}\right\}.dm = erf\left\{\frac{2^{p-1}}{q\cdot g\cdot\sqrt{2}}\right\}$$

(*erf{x}* is the "Error-function" as is defined in many textbooks.) In our case we have $g = \sqrt{2}$ (see above) and therefore:

$$w_p(r) = erf\{2^{p-2}/q\}$$

With $q = \sqrt{2}$ we get now:

$$w_p(r) = erf\{2^{p-5/2}\}$$

From this we have now the probability to find the actual value of $r_q(n)$ within the assumed interval:

$$w_p\{-2^{p-1} \le m - \langle m\rangle \le 2^{p-1}\} = erf\{2^{p-5/2}\}.$$

The values of $w_p = erf\{2^{p-5/2}\}$ are given in the following table.

It is obvious that the data are practically described perfectly well using only four Bit. But, in case of the Herschel-HIFI AOS each of the spectrometer scans is already made of Q = 16,000 pixel data, thus we have to define the probability that a complete data set can be stored with a limited but fixed number of Bit.

Number of Bit *p*	w_p
2	0.6827
3	0.9545
4	$1-6.333\cdot10^{-5}$
5	$1-1.244\cdot10^{-15}$

Table [4]:

Probability of full representation of spectrometer data with *p* Bit.

In order to find such probabilities we have to calculate the Q^{th} power of the single data probability, in order to obtain the probability that **all** Q data are sufficiently resolved with *p* Bit at the same time.

Number of Bit *p*	w_p^Q (Q = 16,000)
2	0
3	0
4	0.363
5	$1-2\cdot10^{-11}$

Table [5]

Probabilities for 16,000 spectrometer pixels using *p* Bit.

Thus, it is evident that the probability to transmit 16,000 pixel data as ratios correctly with 5 Bit is extremely close to 100%. There is a 36.3% chance that already four Bit are sufficient, but for safety reasons one should stay with 5 Bit. From this we can conclude that we need for one total power scan with 60 seconds integration time on the average a total of maximum

$$N_{Total} = 16{,}000\cdot(1\cdot17 + 11\cdot5) = 1.15\cdot10^6 \text{ Bit per position.}$$

The first 17 Bit in the bracket are needed for the first scan (see above). This results to

$$n_{Total} = N_{Total} / t_{Int} = 1.15\cdot10^6 / 60 = 19.2 \text{ kBit/sec}$$

The exact value of <*r*> is also needed for the recovery of the initial data, which adds a few more Bit. To improve the data quality one might use a smaller fraction *q* of the *rms* than assumed above, but the gain in accuracy is quite marginal. The 2% additional noise is probably already negligible for nearly all applications.

The scenario described before assumes that all 16,000 AOS pixel data are transported as one string with identical number of Bit per Pixel. As is easily found, it is not really advantageous to adjust the number of Bit for each of the transmitted data values, since the transmission of the Bit information will consume more Bit than the eventual saving could possibly provide. Normally, one has to consider the possibility of glitches once in a while. If one expects them more frequently, the scheme can easily be converted for smaller sub-portions of data. This has the advantage that a lower number of Bit may be sufficient for several of the smaller sub-portions, but it complicates the handling. Nevertheless, one should consider these problems early enough in order to be prepared for such unfavorable situations. Similar, "Zero-" and "Comb-" measurements, as they are

needed for the correct interpretation of AOS-spectra for example, must also be included in this calculation, which needs another amount of additional Bit, but it is not very likely that these numbers become very significant.

Since we have set the *rms* of the ratios to $\sqrt{2}\cdot$LSB, the number of Bit needed for each of the sub-scans becomes independent on the value of the fluctuation bandwidth or even the integration time used, which is rather surprising. Therefore, the compression efficiency is the same for any resolution of the spectrometer. In order to have 100% data reliability, it is important that each of the transmitted data sets has the mean of the ratio $\langle r(n) \rangle$ and the actual number of Bit for the following values as leading data included, which adds another 20 Bit per scan or so.

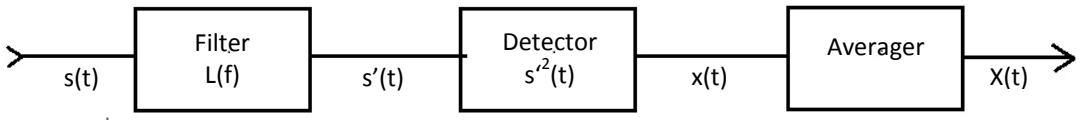
The ratio-method described before was proposed for the data handling of the HIFI data on Herschel, since the normal amount of generated data on board was already more than the down-link capacity could provide. Finally the decision was made not to apply it, since the method was not yet well approved. Instead it was decided to reduce the amount of data by neglecting obviously unnecessary information. Nevertheless, intensive and successful tests were performed as proof for the reliability of the algorithm.

III. About "Radiometric Noise"

In radio-astronomy the noise at the output of a heterodyne receiver system is determined by the so called "fluctuation bandwidth" B_{Fl} as is used in the well-known "radiometer equation". It is not directly understandable, why the fluctuation bandwidth is different from the so called "resolution bandwidth" δ_{Res}, which stands for the resolution capability of the system. In order to clarify this, the following paragraph should provide a consistent derivation of the radiometer equation together with some further analysis.

For the detection of RF in a receiver one usually uses a detector with quadratic response, the output is therefore proportional to the average of the square of the input voltage. This is true for RF-detectors as well as optical detectors like a CCD in an acousto-optical spectrometer (AOS) for example. The number of generated photo electrons is proportional to the average power of the incident light, i.e. to the square of the amplitude of the incident radiation field. (The situation is different with a standard rectifier, since it provides a voltage proportional to the absolute value of the AC-signal. We will not discuss this here because it does not provide any new insight into the matter.)

Consider a „simple" apparatus for the detection of radio signals:



The RF-signal *s(t)* represents „white noise", which is characterized by the correlation function

$$G(\tau) \;=\; < s(t)\cdot s(t+\tau) >_t \;=\; \Gamma_0 \cdot \delta(\tau)$$

Usually, the amplitude *s(t)* itself cannot be described by a simple analytical function. We therefore have to deal with the correlation function.

In a spectrometer, the RF is filtered by a narrow-band filter; there may exist several hundred to a couple of thousands of such filters each operating at different frequencies. The function of the filter is characterized by the (real valued) filter function *L(f)*, which is defined for non-negative frequencies only. The center frequency of the filter we call f_0.

The so called „resolution bandwidth" of the filter δ_{Res} is the width of an equivalent box-car filter, which transmits the same power as the real filter. Its peak transmission is identical with the peak transmission of the real filter. We therefore define:

$$\delta_{\mathrm{Re}\,s} \;=\; \frac{\int_0^\infty L(f)\cdot df}{L_{Max}} \qquad [31]$$

This width has nothing to do with the "FWHM" (Full Width at Half Mean) of the filter, as is frequently used in RF-technology. This 3-dB width is a bit problematic when comparing different filters, because it does not characterize eventual broad and flat wings of a filter curve. Therefore, any information, which does not include some information about such wings, is not very useful. This is the reason why we generally use the resolution bandwidth as one, but not the only filter parameter.

Frequently, the so called "Rayleigh criterion" is also used for the description of the resolution of an optical spectrograph for example. It describes two monochromatic Frequency components as „resolved", if the diffraction maximum of one component is coincident with the first zero of the diffraction pattern of the second component. The diffraction pattern is simply the square of a Sinc-function. The sum of the two patterns of equal strength results in a curve with two maxima with a valley in between at a power of 81% (=8/π^2) of the maxima. In a more generalized picture this value may be used for a characterization of the resolution even if there are no diffraction minima visible. This is a rather artificial construction, and we therefore are not going to use it either.

The quadratic detector delivers a DC voltage as well as high frequency components, as we can see when squaring a single carrier signal.

$$s(t) \;=\; s_0\cdot\cos[2\pi f t] \qquad s^2(t) \;=\; \frac{s_0^2}{2}\cdot\{1+\cos[4\pi f t]\}$$

Only the DC- component is of interest here. (We are using the frequency *f* instead of the circular frequency ω because it removes some of the nasty constants in the following formulas.) The subsequent averaging makes

the high frequency component disappear anyway. Usually we have to consider a full conglomerate of various frequency components. In this case we also get slowly varying components as mixing products due to the quadratic detection scheme. This we can see easily when considering the detection of two monochromatic signals.

$$s(t) = s_1 \cdot \cos[2\pi f_1 t] + s_2 \cdot \cos[2\pi f_2 t)]$$

$$s^2(t) = \frac{s_1^2}{2} \cdot \{1 + \cos[4\pi f_1 t]\} + \frac{s_2^2}{2} \cdot \{1 + \cos[4\pi f_2 t]\} + s_1 \cdot s_2 \cdot \{\cos[2\pi(f_1 - f_2) \cdot t] + \cos[2\pi(f_1 + f_2) \cdot t]\}$$

The high frequency Cos-terms in the brackets are equivalent to the frequency doubling term above and are therefore not relevant. The first Cos-term in the last bracket is at low frequency if both carrier frequencies are close enough to be passed by the filter. There are many of such difference frequencies, if the signal consists of many individual components, which give rise to relatively slow fluctuations at the detector output. If the input is white noise, these fluctuations will also be erratic. The maximum difference frequency is determined by the width of the filter. By the way, one should be aware of the fact that a real quadratic detector does not exist. Usually one is dealing with a diode which has on one arm of the current to voltage plot (I-V-curve) the quadratic response, but on the (normally negative) arm the response is zero. It is therefore only “half of a quadratic detector”. Certainly, when applying a bias voltage, one can generate the situation of a simple quadratic response, which makes use of the local non-linear curvature of the I-V-curve of the device, but this is not exactly a quadratic behaviour in the classical sense. Nevertheless, the real detector characteristics do not change the following arguments.

Generally, one is not interested in these fluctuations, but instead in the average of the detector output. The fluctuations determine now the signal to noise ratio of the detected signal, which is mainly determined by the width of the filter. The averaging at the end reduces the fluctuations again, and we describe the averaging by the function of a box-car integrator:

$$s'(t) = \frac{1}{T} \cdot \int_{t-T}^{t} s(t') \cdot dt'$$

T is the integration time if the integrator, and, by dividing by *T*, we get the average of the signal within the time interval *T*. A different averaging, for example by means of an RC-filter, could also be discussed, but is not of much interest right now.

The Radiometer Equation

We now want to deduce the noise performance of a receiver system as is plotted above. The incident signal voltage *s(t)* is modified when passing through the filter. In order to describe the function of the filter we introduce the Fourier transform of the signal of *s(t)*:

$$s(t) = \int_{-\infty}^{\infty} u(f) \cdot \exp[2\pi i f t] \cdot df \quad \text{and} \quad u(f) = \int_{-\infty}^{\infty} s(t) \cdot \exp[-2\pi i f t] \cdot dt \qquad [32]$$

(The existence of the integral requires that the integral over the absolute value of *s(t)* exists, which requires that the integration interval should be limited. If *s(t)* represents white noise, we need to introduce “band-limited” white noise which has an upper cut-off frequency. But this has practically no influence on the following discussion. It should also be clear that the Fourier transform of *s(t)* is not a regular and analytic function, since for example the phases of *u(f)* are completely random. When using the real valued Sin- and Cos-Fourier transforms, it means that they are completely undefined.) We are using here the complex Fourier transform, and, since the signal voltage is supposed to be real, we have[9]

$$u(-f) = u^*(f).$$

Behind the filter we get now the new signal voltage

$$s'(t) = \int_{-\infty}^{\infty} A(f) \cdot u(f) \cdot \exp[2\pi i f t] \cdot df$$

[9] If one prefers to deal with real numbers only, then we have: *u(-f) = u(f)* and *s(-t) = s(t)* and one can also write:

$$s(t) = 2 \cdot \int_0^{\infty} g(t) \cdot \cos[2\pi \cdot f \cdot t] \cdot df \quad \text{and} \quad u(f) = 2 \cdot \int_0^{\infty} s(t) \cdot \cos[2\pi \cdot f \cdot t] \cdot dt$$

A(f) is the amplitude filter function of the filter, which may be a complex function. (The connection between *A(f)* and the power transmission function *L(f)* of the filter will be introduced further down.) The quadratic detector now delivers the low frequency part as

$$x(t) \;=\; s'(t)\cdot s'^{*}(t) \;=\; \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df'\cdot A(f)\cdot A^{*}(f')\cdot u(f)\cdot u^{*}(f')\cdot \exp[2\pi i(f-f')\cdot t]$$

When averaging we get then:

$$X(t) \;=\; \tfrac{1}{T}\cdot \int_{t-T}^{T} dt' \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df'\cdot A(f)\cdot A^{*}(f')\cdot u(f)\cdot u^{*}(f')\cdot \exp[2\pi i(f-f')\cdot t]$$

Of interest is now only the expectation value of the output voltage, which we are indicating by brackets "<>". We are not going to average over time, but instead, we are averaging over the "ensemble". This means that we understand the signal voltage s(t) as the result of the contributions of many individual oscillators, all oscillating with different phases and/or different frequencies. We can write therefore:

$$X(t) \;=\; \tfrac{1}{T}\cdot < \int_{t-T}^{T} dt' \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df'\cdot A(f)\cdot A^{*}(f')\cdot u(f)\cdot u^{*}(f')\cdot \exp[2\pi i(f-f')\cdot t] > \;=$$

$$=\; \tfrac{1}{T}\cdot \int_{t-T}^{T} dt' \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df'\cdot A(f)\cdot A^{*}(f')\cdot < u(f)\cdot u^{*}(f') > \cdot \exp[2\pi i(f-f')\cdot t]$$

The average is to be taken only over the noise amplitudes *u(f)* and not over the filter amplitudes *A(f)*, since they do not contribute anything to the statistics. We can separate *u(f)* into the absolute value $|u(f)|$ and the phase factor $exp(i\varphi)$. Since the phases φ are completely random, the average over the ensemble can only deliver a non-zero result, if the phases of *u(f)* and $u^{*}(f')$ are conjugate identical, i.e. if $f = f'$: Thus we have:

$$< X(t) > \;=\; < s'^{2}(t) > \;=\; \int_{-\infty}^{\infty} |A(f)|^{2}\cdot < |u(f)|^{2} > \cdot df$$

The result of the ensemble average is independent on the time *t*, as is to be expected. If we deal with white noise, the spectral power $<|u(f)|^{2}>$ should be independent on frequency. We therefore set $<|u(f)|^{2}> = c$ and write

$$< X(t) > \;=\; c\cdot \int_{-\infty}^{\infty} |A(f)|^{2}\cdot df$$

In reality we are using non-negative frequencies only, thus we should find the connection between *A(f)* and *L(f)*, and we have:

$$L(f) \;=\; |A(f)|^{2} + |A(-f)|^{2} \;=\; 2\cdot |A(f)|^{2} \qquad [33]$$

If we use the definition of the resolution bandwidth, and if we call ρ_{in} the spectral power density (= power per frequency interval), then we can write:

$$< X(t) > \;=\; < s'^{2}(t) > \;\sim\; L_{Max}\cdot \rho_{In}\cdot \delta_{Res} \;\text{ and }\; < s^{2}(t) > \;\sim\; \rho_{In}\cdot \delta_{Res}$$

Thus we have: $c \sim \rho_{In}$. $\rho_{In}\cdot\delta_{Res}$ is the mean input power within the bandwidth of the filter, and $L_{Max}\cdot\rho_{In}\cdot\delta_{Res}$ stands for the input power as is transmitted through the filter. Here one can see the purpose of the definition of the resolution bandwidth as the width of an equivalent box-car filter, which transmits the same power as the real filter. In case the spectral power density at the input side can be assumed as constant within the filter width the formula is precisely valid even if we do not deal with white noise.

For the derivation of the expected noise we have to evaluate the variance of the signal

$$\sigma^{2}(T) \;=\; <[X(t) - < X(t) >]^{2} > \;=\; < X(t)^{2} > - < X(t) >^{2}$$

The expectation value of $<X(t)>$ we know already; left is the calculation of $<X^{2}(t)>$. For this we have:

$$< X^{2}(t) > \;=\; \frac{1}{T^{2}}\cdot \int_{t-T}^{t} dt' \int_{t-T}^{t} dt'' \int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \int_{-\infty}^{\infty} df_3 \int_{-\infty}^{\infty} df_4 \cdot A(f_1)\cdot A^{*}(f_2)\cdot A^{*}(f_3)\cdot A(f_4)\cdot < u(f_1)\cdot u^{*}(f_2)\cdot u^{*}(f_3)\cdot u(f_4) > \cdot$$

$$\cdot \exp\{2\pi i\cdot[(f_1 - f_2)\cdot t' - (f_3 - f_4)\cdot t'']\}$$

The average over the phases of *u(f)* can only be non-zero if either $f_2 = f_1$ and $f_4 = f_3$, or $f_3 = f_1$ and $f_4 = f_2$, or $f_4 = -f_1$ and $f_3 = -f_2$. (The last follows from $u^{*}(f) = u(-f)$.) We therefore get three terms

$$< X^2(t) > \;=\; \frac{c^2}{T^2} \cdot \int_{t-T}^{t} dt' \int_{t-T}^{t} dt'' \int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \cdot \left|A(f_1)\right|^2 \cdot \left|A(f2)\right|^2 \cdot \{1 + 2 \cdot \exp\{2\pi i \cdot (f_1 - f_2) \cdot (t'-t'')\}$$

The integration over the first term in the brackets over time is trivial, the second delivers with $x = f_1\text{-}f_2$:

$$\int_{t-T}^{t} \exp\{2\pi i \cdot x \cdot t'\} \cdot dt' \;=\; \exp\{2\pi i \cdot x \cdot (t - T/2)\} \cdot \frac{\sin\{\pi \cdot x \cdot T\}}{\pi \cdot x}$$

The second integration over time results in the conjugate complex. Thus we get now:

$$< X^2(t) > \;=\; c^2 \cdot \int_{-\infty}^{\infty} df_1 \int_{\infty}^{\infty} df_2 \cdot \left|A(f_1)\right|^2 \cdot \left|A(f_2)\right|^2 \cdot \left[1 + 2 \cdot \frac{\sin^2\{\pi \cdot (f_1 - f_2) \cdot T\}}{\{\pi \cdot (f_1 - f_2) \cdot T\}^2}\right]$$

Is the averaging time T large enough, i.e. if $T \gg 1/\delta_{Res}$, then the second term in the brackets approaches a definition of the Dirac Delta-function:

$$\frac{\sin^2\{\pi \cdot (f_1 - f_2) \cdot T\}}{T \cdot \{\pi \cdot (f_1 - f_2)\}^2} \xrightarrow{T \to \infty} \delta(f_1 - f_2)$$

This means that $\left|A(f_2)\right|^2$ should not significantly differ from $\left|A(f_1)\right|^2$ at a frequency separation of $\left|f_1\text{-}f_2\right| \le 1/T$. We get then

$$< X^2(t) > \;=\; \left[c \cdot \int_{-\infty}^{\infty} \left|A(f)\right|^2 \cdot df\right]^2 \;+\; \frac{2 \cdot c^2}{T} \cdot \int_{-\infty}^{\infty} \left|A(f)\right|^4 \cdot df$$

If we go now to non-negative frequencies, we can replace

$$L^2(f) \;=\; \{\left|A(f)\right|^2 + \left|A(-f)\right|^2\}^2 \;=\; 4 \cdot \left|A(f)\right|^4$$

and we get for non-negative frequencies:

$$< X^2(t) > \;=\; \left[c \cdot \int_{0}^{\infty} L(f) \cdot df\right]^2 \;+\; \frac{c^2}{T} \cdot \int_{0}^{\infty} L(f)^2 \cdot df$$

The first term is identical with the square of the average detected power, and we get therefore for the variance:

$$\sigma^2(T) \;=\; < X^2(t) > - < X(t) >^2 \;=\; \frac{c^2}{T} \cdot \int_{0}^{\infty} L^2(f) \cdot df$$

The constant c we can replace

$$c \;=\; \frac{< X(t) >}{\int_0^\infty L(f) \cdot df}$$

and we finally get the famous “radiometer equation”:

$$\sigma^2(T) \;=\; \frac{< X(t) >^2}{B_{Fl} \cdot T} \quad \text{with} \quad B_{Fl} \;=\; \frac{\left[\int_{0}^{\infty} L(f) \cdot df\right]^2}{\int_{0}^{\infty} L^2(f) \cdot df} \quad \text{or} \quad \sigma_P^2(T) \;=\; \frac{< P(f) >^2}{B_{Fl} \cdot T} \qquad [34]$$

$<P(f)>$ is the power as is transmitted through the filter with width δ_{Res}. B_{Fl} is called the “fluctuation bandwidth” and it describes the width of an equivalent box-car filter, which transmits the same fluctuations as the actual filter. The larger B_{Fl}, the smaller are the resulting fluctuations. B_{Fl} is not identical with the resolution bandwidth δ_{Res}, although it is frequently used for an estimate of the expected noise of a radio receiver. In principle this is incorrect. Both definitions, the resolution bandwidth and the fluctuation bandwidth together provide much better insight into the properties of the filter, because, for example, extended wings will make the fluctuation bandwidth very different from the resolution bandwidth, whereas a filter with rectangular shape makes them identical.

Relation between fluctuation and resolution bandwidth

Considering the fact that the power transmission curve *L(f)* never can be negative we can assume that

$$\int_0^\infty L^2(f)\cdot df \;\le\; L_{Max}\cdot\int_0^\infty L(f)\cdot df$$

Consequently we can write:

$$\int_0^\infty L^2(f)\cdot df \;=\; \frac{L_{Max}}{\kappa}\cdot\int_0^\infty L(f)\cdot df \quad \text{with } \kappa \ge 1$$

This leads to: $B_{Fl} = \kappa\cdot\delta_{Res}$

κ can only be equal to unity, if the filter function has one or several box-car shaped structures and is zero otherwise. In all other cases κ is always larger than "1"! It is therefore some kind of a quality indicator for the filter. The smaller κ the "better" is the filter.

For a single pole filter (simple LC circuit) we have for example: $\kappa = 2$
For a Gaussian filter we have: $\kappa = \sqrt{2} = 1.414$
For digital and other correlators we have: $\kappa = 1.5$

The Gaussian filter is therefore a "better" filter. In AOSs we deal approximately with a Gaussian filter, since the laser illumination of the Bragg-cell automatically is Gaussian, the Gaussian image on the CCD is superimposed with some wings caused by diffraction at the edges of the Bragg-cell. Large values of κ indicate that the filter function has large and extended wings, which certainly are not very advantageous.

One should be aware of the fact that during broad-band reception one usually does not have well-defined filter curves. This is particularly true for direct detection receivers as well as for heterodyne receivers, which are used for continuum measurements. Typically one has to consider large gain variations in the band so that one finds relatively large values of κ. In narrow band applications one has typically values of κ between 1 and 2.

Very short averages

Of interest is also, what happens at very short integration time *T*. The assumption of the *Sinc²* function approaching the Delta-function is not valid anymore. We are therefore now investigating the case of

$$T \;>>\; 1/\delta_{\mathrm{Res}}$$

In this case we have for practically all frequencies f_1 and f_2 within the filter curve

$$\frac{\sin^2\{\pi\cdot(f_1-f_2)\cdot T\}}{\{\pi\cdot(f_1-f_2)\cdot T\}^2} \;\approx\; 1$$

The expectation of $X^2(t)$ is then:

$$< X^2(t) > \;=\; c^2\cdot\int_{-\infty}^{\infty} df_1\int_{\infty}^{\infty} df_2\cdot\left|A(f_1)\right|^2\cdot\left|A(f_2)\right|^2\cdot\left[1+2\cdot\frac{\sin^2\{\pi\cdot(f_1-f_2)\cdot T\}}{\{\pi\cdot(f_1-f_2)\cdot T\}^2}\right] \;\approx$$

$$\approx\; c^2\cdot\int_{-\infty}^{\infty} df_1\int_{\infty}^{\infty} df_2\cdot\left|A(f_1)\right|^2\cdot\left|A(f_2)\right|^2\cdot 3 \;=\; 3\cdot < X(t) >^2$$

The variance is now:

$$\sigma^2(T) \;=\; < X(t)^2 > - < X(t) >^2 \;=\; 2\cdot < X(t) >^2 \qquad [35]$$

This result is very different from the usual radiometer equation! There is obviously a regime, where the above "radiometer equation" does not hold. We therefore need a more general expression.

The "True" Radiometer Equation

Instead of applying a longtime integration with the final box-car integrator we can also determine how the instrument output will look like when using an arbitrary low pass filter instead of the integrator. The route to follow is to define the spectral distribution of the detector output by means of the Fourier transform of the output correlation function. The product of this spectrum with the spectral characteristics of the final averager will then provide the final spectrum. When integrating over all frequencies we finally get the variance of the output fluctuations.

First we are therefore interested in the characteristics of the fluctuations of the signal *x(t)* at the output side of the detector. For this we evaluate the (not normalized) first order correlation function $\Gamma(\tau)$ of the output as a function of relative delay time τ, which is given by

$$\Gamma(\tau) = \langle x(t)\cdot x(t+\tau)\rangle = \int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \int_{-\infty}^{\infty} df_3 \int_{-\infty}^{\infty} df_4 \cdot A(f_1)\cdot A^*(f_2)\cdot A^*(f_3)\cdot A(f_4) \cdot$$
$$\cdot \langle u(f_1)\cdot u^*(f_2)\cdot u^*(f_3)\cdot u(f_4)\rangle \cdot \exp\{2\pi i\cdot(f_1 - f_2 - f_3 + f_4)\cdot t\}\cdot\exp\{-2\pi i\cdot(f_3 - f_4)\cdot\tau\}$$

Again, as before, we get a non-zero result only if either $f_2 = f_1$ and $f_4 = f_3$, or if $f_3 = f_1$ and $f_4 = f_2$, or if $f_4 = -f_1$ and $f_3 = -f_2$. Thus we have now:

$$\Gamma(\tau) = c^2 \cdot \int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \cdot |A(f_1)|^2 \cdot |A(f_2)|^2 \cdot \left[1 + 2\cdot\exp\{-2\pi i\cdot(f_1 - f_2)\cdot\tau\}\right]$$

Setting $\mu = f_2 - f_1$ we can modify this to

$$\Gamma(\tau) = c^2 \cdot \left\{ \left[\int_{-\infty}^{\infty} |A(f')|^2 \cdot df' \right]^2 + 2 \cdot \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |A(f')|^2 \cdot |A(f'+\mu)|^2 \cdot df' \right) \cdot \exp\{2\pi i\cdot\mu\cdot\tau\}\cdot d\mu \right\}$$

While replacing the first term in the brackets we get finally:

$$\Gamma(\tau) = \langle x(t)\rangle^2 \cdot \left\{ 1 + \frac{1}{B_{Fl}} \cdot \int_{-\infty}^{\infty} \Phi(\mu)\cdot\exp\{2\pi i\cdot\mu\cdot\tau\}\cdot d\mu \right\} \quad \text{with} \quad \langle x(t)\rangle^2 = c^2 \cdot \left[\int_{-\infty}^{\infty} |A(f')|^2 \cdot df' \right]^2$$

and
$$\Phi(\mu) = \frac{\int_{-\infty}^{\infty} |A(f')|^2 \cdot |A(f'+\mu)|^2 \cdot df'}{\int_{-\infty}^{\infty} |A(f')|^4 \cdot df'} = \frac{\frac{1}{2}\cdot\int_{0}^{\infty} L(f')\cdot[L(f'+\mu) + L(f'-\mu)]\cdot df'}{\int_{0}^{\infty} L^2(f')\cdot df'} \qquad [36]$$

B_{Fl} is defined as above. The filter overlap integral $\Phi(\mu)$ is unity at $\mu = 0$ and smaller than 1 otherwise. It approaches zero when μ becomes much larger than the filter width δ_{Res}. The relation leads also to a different definition of the fluctuation bandwidth with:

$$B_{Fl} = \int_{0}^{\infty} \Phi(\mu)\cdot d\mu = \frac{1}{2}\cdot\int_{-\infty}^{\infty} \Phi(\mu)\cdot d\mu$$

The fluctuation bandwidth is therefore the integral over the overlap of the filter curve with itself. Since $\Phi(\mu)$ may be called the “filter correlation function”, we can conclude that it represents the normalized spectral power density of the noise fluctuations of filtered white noise.

The variance of the output fluctuations directly behind the detector is given by the value of $\Gamma(0)$-$\Gamma(\infty)$. Thus we have here

$$\sigma_x^2 = 2\cdot\langle x(t)\rangle^2 ,$$

since the integral of $\Phi(\mu)$ over positive and negative frequencies equals $2\cdot B_{Fl}$. This is exactly the same result as has been found above (Eq.[35]) for the case of a large post-detection bandwidth.

The Fourier transform of the correlation function $\Gamma(\tau)$ provides us now with the spectral distribution of the output power of the detector as a function of the post-detection frequency *F*, which is proportional to the square of the detector current:

$$\langle x^2(F)\rangle = \int_{-\infty}^{\infty} \Gamma(\tau)\cdot\exp\{-2\pi i\cdot F\cdot\tau\} = \langle x(t)\rangle^2 \cdot\{\delta(F) + \Phi(F)/B_{Fl}\}$$

$\delta(F)$ is the Dirac delta function. The first term in the bracket describes the square of the DC-value of the current, while the second is responsible for the current fluctuations.

Usually there is also a post-detection filter behind the detector, which is characterized by an audio frequency amplitude transmission function *b(f)*. The fluctuations behind this filter are then given by:

$$\langle \Delta X^2(F)\rangle = \frac{\langle x(t)\rangle^2}{B_{Fl}} \cdot \Phi(F)\cdot|b(F)|^2$$

(Note that this relation is defined for both, negative and positive frequency components *F*. Note also that we talk about the square of the signal, i.e. the power, so that the absolute square of the post-detection filter function applies here.) When integrating this now over all frequencies, we get the variance of the current fluctuations with

$$\sigma^2 \;=\; <\Delta X^2> \;=\; \int_{-\infty}^{\infty} <\Delta X^2(F)> \cdot dF \;=\; \frac{<x(t)>^2}{B_{Fl}} \cdot \int_{-\infty}^{\infty} \Phi(F) \cdot |b(F)|^2 \cdot dF \qquad [37]$$

This is the exact formulation of the radiometer equation!

In case the post-detection filter width is very small as compared to the pre-detection filter width, we can set $\Phi(F)$ equal to $\Phi(0)$ = 1 within the complete frequency range, where $b(F)$ is non-zero, and we get finally:

$$\sigma^2 \;=\; \frac{<x(t)>^2}{B_{Fl}} \cdot \int_{-\infty}^{\infty} |b(F)|^2 \cdot dF$$

For a box-car integrator $b(F)$ is given by:

$$|b(F)|^2 \;=\; |b_0|^2 \cdot \frac{\sin^2\{\pi FT\}}{\{\pi FT\}^2}$$

with T the integration time, so that the integral leads to a factor $1/T$. In this case we end finally with the well-known result:

$$\sigma^2(T) \;=\; \frac{<x(t)>^2 \cdot |b_0|^2}{B_{Fl} \cdot T} \;=\; \frac{<X(t)>^2}{B_{Fl} \cdot T}$$

This describes the standard case during all radio-astronomical observations and is usually considered as the radiometer equation, although it is only a special case of the general expression.

With the above formulas it is now simple to calculate also the response of the system with other types of post detection filters. For example, for a simple RC-filter with time constant τ the integral provides a factor of $1/(2\cdot\tau)$, as long as this number is small as compared to the width δ_{Res} of the pre-detection filter. In other applications, such as the reception of audio signals with a telecommunication radio receiver for example, the post-detection bandwidth is practically the same as the pre-detection bandwidth, and one has to evaluate the integral in Eq.[37] accordingly. (It typically approaches there a value of about $1\cdot B_{Fl}$.) The limiting case is reached at very large post-detection bandwidth, where one finds for the current fluctuations:

$$\sigma^2 \;=\; 2 \cdot <X(t)>^2$$

(This results from the fact that the integral over $\Phi(F)$ has a value of twice the fluctuation bandwidth in case $b(f)$ is constant over the full range where $\Phi(F)$ is non-zero (see Eq.[36]). This result is independent on the actual shape of the pre-detection filter as well as on the full width of the post-detection filter! Since the mean post-detection current is proportional to the pre-detection filter width, we have finally that the *rms* background noise is proportional to the square root of the resolution bandwidth in the radiometric case (since $B_{Fl} \approx \delta_{Res}$), and proportional to the resolution bandwidth itself in the case of commercial radio receivers. This is the decisive difference between the two cases as far as the noise is concerned. It follows that it should be advantageous to use small pre-detection filters for commercial receivers instead of narrow post-detection filters. The reason is that noise fluctuations distant from the reception frequency can contribute to the noise output, if the input filter width is large. They contribute also at small output frequencies, so that a small output filter does not reduce their contributions. It should be noted that the filters at the intermediate frequency in a heterodyne receiver are working like an input filter. Important is, that it also reduces the contributions of the noise provided by the input amplifiers. Therefore, in a good receiver one should first amplify the signal and mix then down to the IF frequency, unless the noise produced at the input frequency is much higher than the noise generated at the IF frequency (including the mixer losses).

Determining the fluctuation bandwidth

There exist two possibilities to determine the fluctuation bandwidth: The first is clear from the mathematics above. One measures the filter curve $L(f)$ and evaluates the integrals as shown above. Both, the resolution bandwidth δ_{Res} and the fluctuation bandwidth B_{Fl} are found easily. The second possibility is to use the radiometer equation. When measuring an Allan variance plot, the first part, which exhibits a slope of -1, should exactly follow the radiometer equation. Thus, when using the value of the variance in this regime together with the mean value of the data, one can directly determine the fluctuation bandwidth.

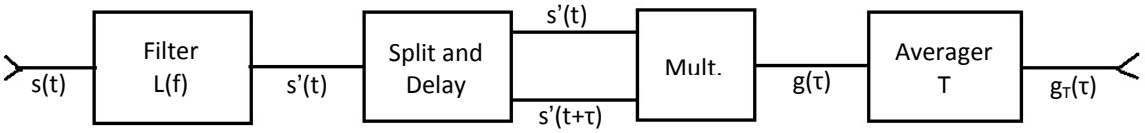
$$B_{Fl} \;=\; \frac{< X(T) >^2}{\sigma_A^2(T)\cdot T}$$

When comparing the outcome of both methods, the results of B_{Fl} should be identical. But one should not forget that the spectrometer itself may generate additional noise, like the final detector, the CCD, does in an AOS. This will affect the statistical result but not the filter curve data. It might be interesting to determine such additional contribution from a comparison of both. If there is additional noise, the resulting fluctuation bandwidth from the Allan plot will become too small! Experience shows that the two results should agree within a few percent. Otherwise there are problems in the spectrometer itself.

Actually, the arguments here are not correct for Auto-Correlators, Digital Fourier-Transform Spectrometers, or Chirp-Transform Spectrometers. The assumption that the power response should always be non-negative is not fulfilled with these instruments. Therefore, the above formulas for the resolution as well as for the fluctuation bandwidth do not apply! (See below.)

Noise of an analogue correlator

Somewhat different is the situation when dealing with a correlator to measure spectra of radio sources. The scheme of an auto-correlator is shown below.



To detect the auto-correlation of a signal at one particular delay-time τ the signal is split into two channels, one with delay τ, and the other without. Typically, the delay is done in a digital shift register, but an analogue technique, like e.g. in systems operating in the infrared or visible region, can also be applied. The calculation for the evaluation of the variance in such system one has to introduce the delay in the above formulas. One finds for the expectation of the correlation function:

$$g(\tau) \;=\; \overset{s(t)}{\int_{-\infty}^{\infty}} S(f)\cdot e^{2\pi i\cdot f\cdot\tau}\cdot df \;=\; 2\cdot\int_0^{\infty} S(f)\cdot\cos[2\pi\cdot f\cdot\tau]\cdot df \;=$$

$$=\; c\cdot\int_{-\infty}^{\infty}|A(f)|^2\cdot e^{2\pi i\cdot f\cdot\tau}\cdot df \;=\; c\cdot\int_0^{\infty} L(f)\cdot\cos[2\pi\cdot f\cdot\tau]\cdot df$$

(Remember that $L(f) = 2\cdot|A(f)|^2$.) *S(f)* is the filtered power spectrum as is fed to the multiplier, which is proportional to the square of the Fourier transform of the signal function *s'(t)*. At delay $\tau = 0$ we have:

$$g(0) \;=\; < s'^2(t) > \;=\; c\cdot\int_0^{\infty} L(f)\cdot df \;=\; \int_{-\infty}^{\infty} S(f)\cdot df$$

$$c \;=\; \frac{g(0)}{\int_0^{\infty} L(f)\cdot df} \;\sim\; \rho_{In}$$

g(0) is practically the power of the RF-signal, which is transmitted by the filter and is determined at delay zero. ρ_{In} is the input spectral power distribution (see above). We can write now:

$$g(\tau) \;=\; g(0)\cdot\frac{\int_0^{\infty} L(f)\cdot\cos[2\pi\cdot f\cdot\tau]\cdot df}{\int_0^{\infty} L(f)\cdot df} \qquad [38]$$

The delay-time dependent correlation function is therefore the normalized Fourier-transform of the filter-function at the input side. Eventual signals, as could be present in the signal, can be interpreted as part of the filter. *g(τ)* vanishes with increasing τ, the broader the filter function *L(f)*, the faster it disappears.

The expectation value of $g^2(\tau)$, averaged over time *T*, is now found similar as above:

$$< g_T^2(\tau) > \;=\; c^2\cdot\int_{-\infty}^{\infty} df_1\int_{-\infty}^{\infty} df_2|A(f_1)|^2\cdot|A(f_2)|^2\cdot\left\{e^{2\pi i\cdot(f_1-f_2)\cdot\tau} \;+\; \left[\frac{\sin[\pi\cdot(f_1-f_2)\cdot T]}{\pi\cdot(f_1-f_2)\cdot T}\right]^2\cdot[1+\;e^{2\pi i\cdot(f_1+f_2)\cdot\tau}]\right\}$$

At long integration time *T* we have again:

$$\frac{1}{T}\cdot\frac{\sin^2\{\pi\cdot(f_1-f_2)\cdot T\}}{\{\pi\cdot(f_1-f_2)\}^2} \;\rightarrow\; \delta(f_1-f_2)$$

With this we get finally after some manipulation:

$$< g_T^2(\tau) > = \left| c \cdot \int_{-\infty}^{\infty} |A(f)|^2 \cdot e^{2\pi i \cdot f \cdot \tau} \cdot df \right|^2 + \frac{c^2}{T} \cdot \int_{-\infty}^{\infty} |A(f)|^4 \cdot \{1 + e^{4\pi i \cdot f \cdot \tau}\} \cdot df =$$

$$= \left[c \cdot \int_0^{\infty} L(f) \cdot cos[2\pi \cdot f \cdot \tau] \cdot df \right]^2 + \frac{c^2}{T} \cdot \int_0^{\infty} L^2(f) \cdot cos^2[2\pi \cdot f \cdot \tau] \cdot df$$

The factor *c* we can replace again. If we use now the fluctuation bandwidth as known from the radiometer equation, then we find for the variance of the correlation function:

$$\sigma^2(T,\tau) = < g_T^2(\tau) > - < g_T(\tau) >^2 = \frac{< g_T(0) >^2}{B_{Fl} \cdot T} \cdot \frac{\int_0^{\infty} L^2(f) \cdot \cos^2\{2\pi \cdot f \cdot \tau\} \cdot df}{\int_0^{\infty} L^2(f) \cdot df}$$

When considering the fluctuations of the power at the output of the correlator, we can also write:

$$\sigma_P^2(T,\tau) = \frac{P_{In}^2}{B_{Fl} \cdot T} \cdot \frac{\int_0^{\infty} L^2(f) \cdot \cos^2\{2\pi \cdot f \cdot \tau\} \cdot df}{\int_0^{\infty} L^2(f) \cdot df} \qquad [39]$$

with P_{In} the input power within the filter bandwidth. The standard deviation of the correlation function is therefore proportional to the power fed to the detector. The fluctuation bandwidth B_{Fl} is large in this case contrary to the situation with a filter-bank, since the bandwidth of the filter includes the whole band the correlator is working in. It is usually not much different from the full bandwidth of the filter. It is interesting that the expression leads for large delays to

$$\sigma_P^2(T, \tau \text{ large}) = \frac{P_{In}^2}{2 \cdot B_{Fl} \cdot T},$$

since the $cos^2(x)$ has an average of ½. This is found as well when considering white noise only, so that $L^2(f) = L_0$ = const. for all $f < f_{Max}$. (Here and for the following we assume that the filter transmits up to a maximum frequency f_{Max}.) For τ=0 on the other hand one finds:

$$\sigma_P^2(T, \tau = 0) = \frac{P_{In}^2}{B_{Fl} \cdot T}.$$

Here the variance is twice as large as for long delays. But still, the fluctuations are rather small what requires that the electronics has to be very low noise! The resolution of a digital sampling electronics must resolve the low level noise at large delays with at least one Bit, but at high dynamic range for the detection of the zero-lag signal, which can be a difficult undertaking when considering very large bandwidth of the correlator. This is a particular problem for the frequently used one- or two-Bit digital correlators, because their resolution is not adequate for the needed accuracy of the zero-lag measurement. Therefore, these systems need an independent method for the power measurement in order to provide the zero-lag information. This is one of the serious difficulties with so-called “Hybrid-Correlators”, where a broad band is split in several smaller bands covered by individual correlators which are operating at lower speed. To measure the power in all sub-bands with sufficient high accuracy is mostly nearly impossible. Therefore, the zero-offset in the spectra as derived by the Fourier-transformation is rather uncertain, so-called “platforming” becomes almost inevitable.

There is one peculiar property of the correlator lags, as is the correlation between the lags. This is found by a rather similar method as in the above consideration. One has:

$$\begin{aligned} < \delta g_T(\tau_1) \cdot \delta g_T(\tau_2) > &= < [g_T(\tau_1) - < g_T(\tau_1) >] \cdot [g_T(\tau_2) - < g_T(\tau_2) >] > = \\ &= < g_T(\tau_1) \cdot g_T(\tau_2) > - < g_T(\tau_1) > \cdot < g_T(\tau_2) > = \\ &= \frac{c^2}{T} \cdot \int_{-\infty}^{\infty} |A(f)|^4 \cdot \{e^{2\pi i \cdot f \cdot (\tau_1 - \tau_2)} + e^{2\pi i \cdot f \cdot (\tau_1 + \tau_2)}\} \cdot df = \qquad [40] \\ &= \frac{c^2}{2 \cdot T} \cdot \int_0^{\infty} L^2(f) \cdot \{\cos[2\pi \cdot f \cdot (\tau_1 - \tau_2)] + \cos[2\pi \cdot f \cdot (\tau_1 + \tau_2)]\} \cdot df = \\ &= \frac{c^2}{T} \cdot \int_0^{\infty} L^2(f) \cdot \cos[2\pi \cdot f \cdot \tau_1] \cdot \cos[2\pi \cdot f \cdot \tau_2] \cdot df \end{aligned}$$

It looks like the Fourier-transformation of the square of the filter-function (with arguments τ_1-τ_2 and τ_1+τ_2). Since we have to assume that the filter-function $L(f)$ is slowly changing within the input band, we expect only contributions at small τ, if any. This is obvious, if we assume a box-car shaped $L(f)$ with $L(f) = L_0$ in the band $0 < f \leq 1/2\tau_0$. (τ_0 is the increment between the time lags $f_{Max} = 1/2\tau_0$ is the maximum frequency, the correlator can observe accordingly.) In this case the correlation is minimal and we get:

$$< \delta g(\tau_1) \cdot \delta g(\tau_2) > = \frac{c^2 L_0^2}{2 \cdot T} \cdot \left\{ \frac{\sin[2\pi \cdot (\tau_1 - \tau_2) \cdot f_{Max}]}{2\pi \cdot (\tau_1 - \tau_2)} + \frac{\sin[2\pi \cdot (\tau_1 + \tau_2) \cdot f_{Max}]}{2\pi \cdot (\tau_1 + \tau_2)} \right\}$$

Together with f_{Max}=1/2τ_0 and $\tau_n = n\cdot\tau_0$ we find that the cross-correlation between all lags is zero! This is to be expected in case of pure white noise at the input side. But, if the input band is not uniform, this changes; the more there is structure, the more correlation appears.

The Spectrum

With the correlation function *g(τ)* it is now straight forward to calculate the spectrum by means of a Fourier-transformation. We have:

$$S(f) = \int_{-\infty}^{\infty} g(\tau)\cdot e^{2\pi i\cdot f\cdot\tau}\cdot d\tau = 2\cdot\int_{0}^{\infty} g(\tau)\cdot\cos[\,2\pi\cdot f\cdot\tau]\cdot d\tau$$

$$g(\tau) = \int_{-\infty}^{\infty} S(f)\cdot e^{-2\pi i\cdot f\cdot\tau}\cdot df = 2\cdot\int_{0}^{\infty} S(f)\cdot\cos[2\pi\cdot f\cdot\tau]\cdot df$$

With a correlator with discrete samples one can replace the integral by the equivalent sum:

$$S_c(f) = \tau_0\cdot\sum_{n=-N}^{N} g(\tau_n)\cdot \mathrm{e}^{2\pi i\cdot f\cdot\tau_n} = 2\cdot\tau_0\cdot\sum_{n=0}^{N}\left(1-\tfrac{1}{2}\delta_{n,0}\right)\cdot g(\tau_n)\cdot\cos[2\pi\cdot f\cdot\tau_n] \qquad [41]$$

τ_0 is the (constant) delay difference of adjacent correlator lags, and $g(\tau_n)$ is identical for positive and negative arguments. We use "$S_c(f)$" for the reconstructed spectrum in order to distinguish it from the input spectrum *S(f)*. (It should be repeated again that the filter should not transmit above the maximum frequency f_{Max} = 1/2τ_0.) The frequencies of the recalculated spectrum must also be within this range $0 < f \le f_{Max}$. Inserting $g(\tau_n)$ in the above sum one gets:

$$S_c(f) = \tau_0\cdot c\cdot\int_{0}^{\infty} L(f')\cdot\left\{\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f)\cdot\tau_0]}{\sin[\pi\cdot(f'-f)\cdot\tau_0]}+\frac{\sin[\pi\cdot(2N+1)\cdot(f'+f)\cdot\tau_0]}{\sin[\pi\cdot(f'+f)\cdot\tau_0]}\right\}\cdot df'$$

To simplify things we introduce again negative frequencies and we can write:

$$S_c(f) = \tau_0\cdot c\cdot\int_{-\infty}^{\infty} L(f')\cdot\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f)\cdot\tau_0]}{\sin[\pi\cdot(f'-f)\cdot\tau_0]}\cdot df'$$

A problem is here again the denominator in the formula, since it generates an infinite number of poles. But we can replace the Sin-function in the denominator by:

$$\tau_0\cdot\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f)\cdot\tau_0]}{\sin[\pi\cdot(f'-f)\cdot\tau_0]} = \frac{1}{\pi}\cdot\sum_{n=-\infty}^{\infty}(-1)^n\cdot\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f)\cdot\tau_0]}{f'-f+n/\tau_0]} = \frac{1}{\pi}\cdot\sum_{n=-\infty}^{\infty}\cdot\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f+n/\tau_0)\cdot\tau_0]}{f'-f+n/\tau_0]}$$

Thus we have:

$$S_c(f) = \frac{c}{\pi}\cdot\int_{-\infty}^{\infty} L(f')\cdot\sum_{n=-\infty}^{\infty}\frac{\sin\left[\pi\cdot(2N+1)\cdot\left(f'-f+n/\tau_0\right)\cdot\tau_0\right]}{f'-f+n/\tau_0]}\cdot df' =$$

$$= \frac{c}{\pi}\cdot\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty} L(f'-n/\tau_0)\cdot\frac{\sin\left[\pi\cdot(2N+1)\cdot\left(f'-f\right)\cdot\tau_0\right]}{f'-f}\cdot df'$$

(Here we have used that *L(-f) = L(f)*.) If it is secured that *L(f)* outside the interval $-1/2\tau_0 < f < 1/2\tau_0$ is zero everywhere, then we can define again a "pseudo-function" $\Lambda(f)$ with:

$$\Lambda(f') = L(f'-\frac{n}{\tau_0}) \quad \text{with} \quad (n\text{-}½)/\tau_0 < f' < (n\text{+}½)/\tau_0,\ n = 0, \pm1, \pm2, \ldots$$

By this we construct a periodic continuation of the initial input spectrum. Thus we write now:

$$S_c(f) = \frac{c}{\pi}\cdot\int_{-\infty}^{\infty}\Lambda(f')\cdot\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f)\cdot\tau_0]}{f'-f}\cdot df'$$

With *N* large we can now replace

$$\frac{1}{\pi}\cdot\frac{\sin[\pi\cdot(2N+1)\cdot(f'-f)\cdot\tau_0]}{f'-f} \rightarrow \delta(f'-f)$$

and we get:

$$S_c(f) = c\cdot<L(f)> = <S(f)>.$$

<*L(f)*> is the average value of *L(f)* in the interval of width $1/[(N+½)\cdot 2\tau_0]$, and this is identical with the value of the input spectrum at the frequency *f*. (One should recognize that the observed frequency f should not lie outside of the filter width of *L(f)*.) This proofs that the auto-correlator indeed reproduces the input spectrum. By the way, the effective frequency width of the reconstructed frequency pixel defines also that the number of available frequency pixels is identical with the number of lags of the correlator.

In case the input spectrum consists of a single monochromatic line: $S(f) = S_0 \cdot \delta(f\text{-}f_0)$, then we find the answer of the correlator to such line with

$$S_c(f) \;=\; S_0 \cdot l(f), \quad l(f) \;=\; \tau_0 \cdot \frac{\sin[\pi\cdot(2N+1)\cdot(f-f_0)\cdot\tau_0]}{\sin[\pi\cdot(f-f_0)\cdot\tau_0]} \qquad [42]$$

l(f) we call the "filter-function" of the spectrometer. The response is therefore a Sinc-function when neglecting the higher order poles of the Sin-function in the denominator, and it converts to a Delta-function at *N* very large.

This expression is slightly unrealistic at finite *N*, since it becomes negative at $(f\text{-}f_0)\cdot\tau_0 = -1/2, +3/2, \ldots$, but a power spectrum cannot be negative! Also some other aspects of the auto-correlator are not quite realistic. For example, the value of the resolution bandwidth becomes:

$$\delta_{Res} \;=\; \frac{1}{l_{Max}} \cdot \int_0^\infty l(f)\cdot df \;=\; \frac{1}{(N+1/2)\cdot 2\cdot\tau_0} \;=\; \frac{f_{Max}}{N+1/2}$$

This is exactly the width of a spectral pixel, which would suggest, that the response on each pixel is equivalent to a box-car shaped filter, but the Sinc-function represents a completely different filter-function. Consequently we have to state that the general definition of the resolution bandwidth does not provide a meaningful value for the frequency resolution of the correlator.

Noise of the spectrum

If we are interested in the noise performance of the correlator we have to determine the value of the square of *P(f)*.

$$<\delta S_c(f)^2> \;=\; <S_c(f)^2> \;-<S_c(f)>^2 \;=$$

$$=\; \frac{c^2}{T}\cdot\sum_{n=-N}^{N}\sum_{m=-N}^{N}[<g(\tau_n)\cdot g(\tau_m)> -<g(\tau_n)>\cdot<g(\tau_m)>]\cdot e^{2\pi i\cdot f\cdot\tau_n}\cdot e^{2\pi i\cdot f\cdot\tau_m}$$

Inserting now the expressions for $g(\tau)$ we obtain:

$$<\delta S_c(f)^2> \;=\; \frac{c^2}{T}\cdot\int_0^\infty L^2(f')\cdot df'\cdot\left[\frac{\sin[(2N+1)\cdot\pi\cdot(f'-f)\cdot\tau_0]}{\sin[\pi\cdot(f'-f)\cdot\tau_0]}+\frac{\sin[(2N+1)\cdot\pi\cdot(f'+f)\cdot\tau_0]}{\sin[\pi\cdot(f'+f)\cdot\tau_0]}\right]^2 \;=$$

$$=\; \frac{c^2}{T}\cdot\int_{-\infty}^{\infty}L^2(f')\cdot df'\cdot\frac{\sin^2[(2N+1)\cdot\pi\cdot(f'-f)\cdot\tau_0]}{\sin^2[\pi\cdot(f'-f)\cdot\tau_0]} \;\rightarrow\; \frac{c^2}{T}\cdot<L^2(f)>\cdot(2N+1)\cdot\tau_0 \;=$$

$$\rightarrow\; \frac{<P(f)>^2}{B_{Fl}\cdot T} \quad \text{with} \quad B_{Fl} \;=\; \frac{1}{(N+1/2)\cdot 2\tau_0} \;=\; \frac{f_{Max}}{N+1/2} \qquad [43]$$

<*P(f)*> is the power within the effective bandwidth of one frequency pixel assuming that the spectral power density is constant within the filter width so that $<P(f)^2> = <P(f)>^2$. (The cross-term between the two sin-terms in the brackets vanishes.) Again, the radiometer-equation is valid here, but the fluctuation bandwidth B_{Fl} is now identical with the resolution-bandwidth δ_{Res} which is only possible for a box-car filter as is obviously not present here. The conclusion must be that a correlator system, as is discussed here, generates more radiometric noise as it is supposed to do. The performance of the correlator is not as ideal as a normal spectrometer would perform as far the fluctuation bandwidth is concerned.

The reason for this must be related to the filter-function – the Sinc-function – which has unrealistic negative values in the side lobes although the power detected cannot be negative! In order to avoid this one frequently uses a method called "apodization". Instead of using the direct Fourier-transform formula one "apodizes" the correlator data by:

$$P(f) \;=\; \sum_{n=-N}^{N}\left[1-\frac{|\tau_n|}{\tau_N}\right]\cdot g(\tau_n)\cdot e^{2\pi i\cdot f\cdot\tau_n}$$

The corresponding filter-function is now:

$$l(f) \;=\; \tau_0^2 \cdot \frac{sin^2[N\cdot\pi\cdot(f-f_0)\cdot\tau_0]}{sin^2[\pi\cdot(f-f_0)\cdot\tau_0]} \qquad [44]$$

as one can find after some tedious calculation. This filter-function is non-negative everywhere and it seems to be more plausible. For the resolution and the fluctuation bandwidth we get now:

$$\delta_{Res} \;=\; 2\cdot\frac{f_{Max}}{N}, \quad B_{Fl} \;=\; 3\cdot\frac{f_{Max}}{N}.$$

These expressions are certainly more meaningful as the ones above, but the price is that one has only half the resolution, and one gets only half the number of useful pixels in comparison. The filter-function has smaller side-lobes as before, which is preferable. The apodization method is generally used in most of the correlator-spectrometers, and with lower and lower prices for the digital electronics it is nowadays not a major problem to implement.

Frequently it is stated in literature, that a Fourier-transform spectrometer is more sensitive than an ordinary spectrometer with the argument that one observes the full power of the whole band with the detector instead of small portions seen in a conventional spectrometer. This must be seen in the context of the noise power contributed by the detector itself. When considering additional detector noise, only that part of the detector noise contributes to the noise which is transmitted by the effective filter width of each frequency channel (see above). This is different in conventional spectrometers, since the detector noise is fully present on each frequency channel of the spectrometer. But, only the dark-noise makes mainly the difference, since the shot-noise originating from the signal current itself is in case of the classical spectrometer nearly zero, but a lot higher in case of the FTS. The difference between the two spectrometer methods depends obviously on the detector dark-noise alone. In the visible frequency range detector dark-noise may be more or less neglected, there should be no difference. When going to the far-infrared, Fourier-transform spectroscopy is obviously advantageous because of the generally rather high detector dark-noise in this frequency range.

The Digital Fourier-Transform-Spectrometer (DFT)

(The effect of finite sampling)

A nice example how correlation can lead to surprising outcomes of particular techniques is the Digital-Fourier-Transform-Spectrometer. This fairly new technology is obviously a very attractive method for future coverage on growing spectrometer needs in various research areas. With rapid sampling of the input voltage received from a radiometer output by means of a fast A/D converter and subsequent Fast Fourier-Transformation (FFT) a spectrum is directly calculated and averaged afterwards. Nowadays bandwidths of a couple of GHz are achieved, which is the consequence of the newest developments of very fast digital electronics.

Signal output

The DFT determines the frequency amplitude of the input amplitudes of a signal, which are rapidly sampled with an ADC in constant time intervals τ_0. These data are then Fourier transformed via a FFT (Fast Fourier Transformation) like

$$U_k(f,t_0) \;=\; \tau_0 \cdot \sum_{n=0}^{N-1} s(t_0+(kN+n)\tau_0)\cdot\exp\{-2\pi i\cdot f\cdot(kN+n)\tau_0\}$$

k indicates here the numbering of the individual data sets, which are separately Fourier transformed. N is the length of each data set, which should be equal to some power of 2 when applying the Fast-Fourier-Transform method. n is the numbering of the data within each data set. The data $s(t)$ are sampled at times $t_0+(kN+n)\cdot\tau_0$. Each set consists of N such samples, which are then repeated K times (index k). In general, we can write:

$$s(t) \;=\; \int_{-\infty}^{\infty} u(f')\cdot\exp\{2\pi i\cdot f'\cdot t\}\cdot df'$$

$u(f')$ is the "true" frequency amplitude of the signal $s(t)$. Since $s(t)$ is supposed to be real, we have $u^*(f) = u(-f)$. The calculated power spectrum $|U_k(f)|^2$ is then given by:

$$|U_k(f)|^2 \;=\; \tau_0^2 \cdot \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} s(t_0+(kN+n)\cdot\tau_0)\cdot s(t_0+(kN+m)\cdot\tau_0)\cdot\exp\{2\pi i\cdot f\cdot[(kN+n)-(kN+m)]\cdot\tau_0\}$$

We distinguish between the calculated amplitude spectrum $U(f)$ and the true amplitude spectrum $u(f)$, since they can differ under certain circumstances.

The measurement is repeated K times, and the average represents the final spectrum. The expectation value of the average is given by:

$$<\overline{|U_K(f)|^2}> = \frac{1}{KN\tau_0}\cdot\sum_{k=0}^{K-1}<|U_k(f)|^2> =$$

$$= \frac{1}{KN\tau_0}\cdot\sum_k\sum_n\sum_m\int_{-\infty}^{\infty}df_1\int_{-\infty}^{\infty}df_2\cdot<u(f_1)\cdot u^*(f_2)>\cdot\exp\{2\pi i\cdot[f(n-m)-(f_1\cdot(kN+n)-f_2\cdot(kN+m))]\cdot\tau_0-(f_1-f_2)\cdot t_0\}$$

$KN\tau_0$ is the total integration time of the spectrum. Since the phases of *u(f')* are supposed to be completely random, the expectation value of this product will be zero with the exception of $f_1 = f_2$. Thus we can write:

$$<u(f_1)\cdot u^*(f_2)> = <|u(f_1)|^2>\cdot\delta(f_2-f_1)$$

with $\delta(x)$ the Dirac Delta-function. When summing up we get now:

$$<\overline{|U_K(f)|^2}> = \frac{\tau_0}{N}\cdot\int_{-\infty}^{\infty}<|u(f')|^2>\cdot\frac{\sin^2[\pi(f'-f)N\tau_0]}{\sin^2[\pi(f'-f)\tau_0]}\cdot df'$$

We can expand now again the denominator with

$$\frac{1}{\sin^2[\pi(f'-f)\tau_0]} = \frac{1}{\pi^2\tau_0^2}\cdot\sum_{s=-\infty}^{\infty}\frac{1}{[f'-f-s/\tau_0]^2}$$

Therefore we have:

$$<\overline{|U_K(f)|^2}> = \frac{1}{N\pi^2\tau_0}\cdot\int_{-\infty}^{\infty}<|u(f')|^2>\cdot\sum_{s=-\infty}^{\infty}\frac{\sin^2[\pi(f'-f)N\tau_0]}{[f'-f-s/\tau_0]^2}\cdot df' =$$

$$= \frac{1}{N\pi^2\tau_0}\cdot\int_{-\infty}^{\infty}\sum_{s=-\infty}^{\infty}<|u(f'+s/\tau_0)|^2>\cdot\frac{\sin^2[\pi(f'-f)N\tau_0]}{[f'-f]^2}\cdot df' = \frac{1}{N\pi^2\tau_0}\cdot\int_{-\infty}^{\infty}\Lambda(f')\cdot\frac{\sin^2[\pi(f'-f)N\tau_0]}{[f'-f]^2}\cdot df'$$

$\Lambda(f')$ is the periodically continued power spectrum with

$$\Lambda(f') = \sum_{s=-\infty}^{\infty}<|u(f'+s/\tau_0)|^2>$$

Note that this transformation is only possible, if the spacing τ_0 between the data is constant. In order to avoid aliasing problems we have to assume that there is no overlap from different orders s. It is therefore important that the input power spectrum does not contain components at frequencies above $f_{Max} = 1/(2\tau_0)$ and below $f_{Min} = -1/(2\tau_0)$. At the same time, the observing frequency f should not lie outside this frequency interval as well. Then we can write:

$$\Lambda(f') = <|u(f'+s/\tau_0)|^2> \quad \text{for} \quad -(s+½)/\tau_0 < f' < -(s-½)/\tau_0$$

This removes now the sin^2-function in the denominator under the integral, but it requires that the input to the spectrometer is properly filtered. By the way, if one uses a filter which transmits at frequencies in the interval $(s-½)/\tau_0 < |f'| < (s+½)/\tau_0$ with $s \neq 0$, one can observe the spectrum at frequencies f' within the same interval without pre-processing the input by means of an additional frequency converter. Certainly, a suitable filter at the input side must be implemented. (Note that this is only going to work properly, if the AD-converter does not average over an appreciable fraction of the sampling interval τ_0. Otherwise, the response to the input frequencies becomes reduced and the overall efficiency is degraded.)

For white noise we can write:

$<|u(f')|^2> = u^2 = const$ within the allowed frequency range, and therefore

$$\Lambda(f') = u^2 = const$$

Thus, we get finally when integrating:

$$<\overline{|U_K(f)|^2}> = <|u(f)|^2>$$

The calculated power spectrum is identical with the true power spectrum. This is also valid for realistic spectra as long as $<|u(f')|^2>$ does not vary within the relevant part of the $Sinc^2$ function in the numerator.
If we have a coherent carrier as input $<|u(f')|^2> = u_0^2\cdot\delta(f'-f_0)$, then we have:

$$S(f) = <[U(f)]^2> = \frac{u_0^2\cdot\tau_0}{N}\cdot\frac{sin^2[\pi(f-f_0)\cdot N\cdot\tau_0]}{sin^2[\pi(f-f_0)\cdot\tau_0]} \approx \frac{u_0^2}{\delta_{Res}}\cdot\frac{sin^2[\pi(f-f_0)/\delta_{Res}]}{[\pi(f-f_0)/\delta_{Res}]^2} \qquad [45]$$

with $\delta_{Res} = 1/(N\tau_0)$. This $Sinc^2$-function we call the filter function $S(f)$ of the DFT-spectrometer which we can determine by experiment while scanning a synthesizer line in small steps through the position of one frequency pixel of the spectrometer.

The filter function should be used to calculate the fluctuation bandwidth, which is the bandwidth needed for the radiometer equation. We find $B_{Fl} = 3/2 \cdot \delta_{Res}$, which does actually not agree with the experimental results. So far, so good! No doubt, the DFT provides a correct representation of the spectrum, but the calculation of the noise seems to reveal some problem.

Noise output

For the noise seen with the spectrometer we have to find the value of

$$\sigma_K^2 = <\overline{|U_K(f)|^2}^2> - <\overline{|U_K(f)|^2}>^2$$

The second term we know already, for the first term we have:

$$<\overline{|U_K(f)|^2}^2> = \frac{\tau_0^2}{(KN)^2} \cdot \sum_k \sum_l \sum_n \sum_m \sum_p \sum_q \int df_1 \int df_2 \int df_3 \int df_4 <u(f_1)u^*(f_2)u^*(f_3)u(f_4)> \cdot$$
$$\cdot \exp\{2\pi i \cdot [f \cdot (t_{k,n} - t_{k,m} - t_{l,p} + t_{l,q}) - (f_1 \cdot t_{k,n} - f_2 \cdot t_{k,m} - f_3 \cdot t_{l,p} + f_4 \cdot t_{l,q})]\}$$

with $t_{k,n} = (k \cdot N + n) \cdot \tau_0$. Since all phases of $u(f_i)$ are random, we can use:

$$<u(f_1) \cdot u^*(f_2) \cdot u^*(f_3) \cdot u(f_4)> = <u(f_1) \cdot u^*(f_2)> \cdot <u^*(f_3) \cdot u(f_4)> \cdot \delta(f_2 - f_1) \cdot \delta(f_4 - f_3) +$$
$$+ <u(f_1) \cdot u^*(f_3)> \cdot <u^*(f_2) \cdot u(f_4)> \cdot \delta(f_3 - f_1) \cdot \delta(f_4 - f_2) + <u(f_1) \cdot u(f_4)> \cdot <u^*(f_2) \cdot u^*(f_3)> \cdot \delta(f_4 + f_1) \cdot \delta(f_3 + f_2)$$

The first term delivers the square of the expectation value of the average. For the last term it should be remembered that $u^*(f) = u(-f)$. Therefore, we have now:

$$\sigma_K^2(f) = \frac{\tau_0^2}{(KN)^2} \cdot \int_{-\infty}^{\infty} <|u(f_1)|^2> \cdot df_1 \cdot \int_{-\infty}^{\infty} <|u(f_2)|^2> \cdot df_2 \cdot \frac{\sin^2[\pi(f_1 - f_2)KN\tau_0]}{\sin^2[\pi(f_1 - f_2)N\tau_0]} \cdot \frac{\sin[\pi(f_1 - f)N\tau_0]}{\sin[\pi(f_1 - f)\tau_0]} \cdot \frac{\sin[\pi(f_2 - f)N\tau_0]}{\sin[\pi(f_2 - f)\tau_0]} \cdot$$
$$\cdot \left[\frac{\sin[\pi(f_1 - f)N\tau_0]}{\sin[\pi(f_1 - f)\tau_0]} \cdot \frac{\sin[\pi(f_2 - f)N\tau_0]}{\sin[\pi(f_2 - f)\tau_0]} + \frac{\sin[\pi(f_1 + f)N\tau_0]}{\sin[\pi(f_1 + f)\tau_0]} \cdot \frac{\sin[\pi(f_2 + f)N\tau_0]}{\sin[\pi(f_2 + f)\tau_0]} \right]$$

For the moment, let us consider the case K=1, where just one set of N sample data is being used. In this case, the first term under the integrals is equal to unity, and we have:

$$\sigma_{K=1}^2(f) = \left\{ \frac{\tau_0}{N} \cdot \int_{-\infty}^{\infty} <|u(f')|^2> \cdot \frac{\sin^2[\pi(f' - f)N\tau_0]}{\sin^2[\pi(f' - f)\tau_0]} \cdot df' \right\}^2 +$$
$$+ \left\{ \frac{\tau_0}{N} \cdot \int_{-\infty}^{\infty} <|u(f')|^2> \cdot \frac{\sin[\pi(f' - f)N\tau_0]}{\sin[\pi(f' - f)\tau_0]} \cdot \frac{\sin[\pi(f' + f)N\tau_0]}{\sin[\pi(f' + f)\tau_0]} \cdot df' \right\}^2$$

Again, we can replace the denominators $sin[x]$ by x when limiting f and f' to the allowed frequency range while using the periodic continuation of the input spectrum. For white noise and $f = \lambda/(N\tau_0)$ ($\lambda = \pm1, \pm2, \pm3, \ldots, |\lambda| < N/2$) we have therefore for the first term:

$$\sigma_{K=1}^2(\lambda / N\tau_0) = 1 \cdot u^4 = \frac{u^4}{N\tau_0 \cdot 1/N\tau_0} = \frac{<|U(\lambda / N\tau_0)|^2>^2}{t_{Int} \cdot B_{Fl}} \quad \text{with } t_{Int} = N\tau_0 \text{ and } B_{Fl} = 1/N\tau_0$$

The second term above does not contribute here. When comparing with the radiometer equation it means that the noise is determined by a fluctuation bandwidth $B_{Fl} = 1/N\tau_0$ and not by the result found with the $Sinc^2$-filter function. But, this is not really conclusive because the usual radiometer equation is not supposed to be fully valid here (t_{Int} has to be very large compared to $1/B_{Fl}$! See above.). On the other hand, since each of K data sets is statistically independent on the others, one should expect that $\sigma_K^2(\lambda/N\tau_0) = 1/K \cdot \sigma_1^2(\lambda/N\tau_0)$.

Nevertheless, we have to determine with some more mathematics, how the noise is developing at long integration time, i.e. when repeating the measurement many times ($t_{Int} = K \cdot N\tau_0$, $K \gg 1$). Then we can rewrite the first factor in the expression for $\sigma^2_{K\gg1}(f)$ with:

$$\frac{\sin^2[\pi(f_1-f_2)KN\tau_0]}{\sin^2[\pi(f_1-f_2)N\tau_0]} = \sum_{s=-\infty}^{\infty}\frac{\sin^2[\pi(f_1-f_2)KN\tau_0]}{(\pi N\tau_0)^2\cdot(f_1-f_2-\frac{s}{N\tau_0})^2} \Rightarrow \frac{K}{N\tau_0}\cdot\sum_{s=-\infty}^{\infty}\delta(f_1-f_2-\tfrac{s}{N\tau_0})$$

The expression is valid for K very large. When using this we get

$$\sigma^2_{K>>1}(f) = \left(\frac{\tau_0}{KN}\right)^2\cdot\sum_{s=-\infty}^{\infty}\frac{K}{N\tau_0}\cdot\int_{-\infty}^{\infty}\Lambda(f'+\tfrac{s}{N\tau_0})\cdot\Lambda(f')\cdot\frac{\sin[\pi(f'-f)N\tau_0]}{\pi(f'-f)\tau_0}\cdot\frac{\sin[\pi(f'-f+\frac{s}{N\tau_0})N\tau_0]}{\pi(f'-f+\frac{s}{N\tau_0})\tau_0}\cdot$$
$$\cdot\left[\frac{\sin[\pi(f'-f)N\tau_0]}{\pi(f'-f)\tau_0}\cdot\frac{\sin[\pi(f'-f+\frac{s}{N\tau_0})N\tau_0]}{\pi(f'-f+\frac{s}{N\tau_0})\tau_0}+\frac{\sin[\pi(f'+f)N\tau_0]}{\pi(f'+f)\tau_0}\cdot\frac{\sin[\pi(f'+f+\frac{s}{N\tau_0})N\tau_0]}{\pi(f'+f+\frac{s}{N\tau_0})\tau_0}\right]\cdot df'$$

Again, we consider white noise only so that $\Lambda(f')$ becomes constant = u^2. For frequencies $f=\lambda/N\tau_0$ the integral can be evaluated. For $s=0$ we get:

$$\sigma^2_{K>>1}(\lambda/N\tau_0, s=0) = \frac{(u/\pi)^4}{K(N\tau_0)^3}\cdot\int_{-\infty}^{\infty}\frac{\sin^4[\pi f'N\tau_0]}{(f'-\frac{\lambda}{N\tau_0})^2}\cdot\left[\frac{1}{(f'-\frac{\lambda}{N\tau_0})^2}+\frac{1}{(f'+\frac{\lambda}{N\tau_0})^2}\right]\cdot df' = \sigma^{2(I)}_{K>>1}(\lambda/N\tau_0) + \sigma^{2(II)}_{K>>1}(\lambda/N\tau_0)$$

The first term leads to:

$$\sigma^{2(I)}_{K>>1}(\lambda/N\tau_0) = \frac{(u/\pi)^4}{K(N\tau_0)^3}\cdot\int_{-\infty}^{\infty}\frac{\sin^4[\pi(f'-\frac{\lambda}{N\tau_0})N\tau_0]}{(f'-\frac{\lambda}{N\tau_0})^4}\cdot df' = \frac{2}{3}\cdot\frac{u^4}{K}$$

This is equivalent to the expectation with the Sinc^2-function. But there are other terms to consider as well. The second term gives:

$$\sigma^{2(II)}_{K>>1}(\lambda/N\tau_0) = \frac{(u/\pi)^4}{K(N\tau_0)^3}\cdot\int_{-\infty}^{\infty}\frac{\sin^4[\pi f'N\tau_0]}{(f'-\frac{\lambda}{N\tau_0})^2\cdot(f'+\frac{\lambda}{N\tau_0})^2}\cdot df' = \frac{(u/\pi)^4}{K(N\tau_0)^3}\cdot\int_{-\infty}^{\infty}\frac{\sin^4[\pi f'N\tau_0]}{(\frac{2\lambda}{N\tau_0})^2}\cdot\left(\frac{1}{f'-\frac{\lambda}{N\tau_0}}-\frac{1}{f'+\frac{\lambda}{N\tau_0}}\right)^2\cdot df' =$$
$$= \frac{(u/\pi)^4}{KN\tau_0\cdot(2\lambda)^2}\cdot\int_{-\infty}^{\infty}\left(\frac{\sin^4[\pi(f'-\frac{\lambda}{N\tau_0})N\tau_0]}{(f'-\frac{\lambda}{N\tau_0})^2}+\frac{\sin^4[\pi(f'+\frac{\lambda}{N\tau_0})N\tau_0]}{(f'+\frac{\lambda}{N\tau_0})^2}\right)\cdot df' = \frac{u^4}{K}\cdot\frac{1}{(2\pi\lambda)^2}$$

(The cross term from the square of the bracket is zero when integrating.) In total we have consequently for $s=0$:

$$\sigma^2_{K>>1}(\lambda/N\tau_0, s=0) = \frac{u^4}{K}\cdot\left(2/3+\frac{1}{(2\pi\lambda)^2}\right)$$

Similar, we find for s ≠ 0, s = ±1, ±2. ±3, …:

$$\sigma^2_{K>>1}(\lambda/N\tau_0, s\neq 0) = \frac{u^4}{K}\cdot\left(\frac{1}{(\pi\cdot s)^2}-\frac{1}{2\cdot(2\pi\lambda)^2}\cdot(\delta_{s,2\lambda}+\delta_{s,-2\lambda})\right)$$

The second term in the bracket is introduced again by the f'+f terms in the equation above, and it cancels with the result for s = 0 when summing. Thus, we have finally:

$$\sigma^2_{K\gg1}(\lambda/N\tau_0) = \sigma^2_{K\gg1}(\lambda/N\tau_0, s=0) + \sum_{s\neq0}\sigma^2_{K\gg1}(\lambda/N\tau_0, s\neq0) = \frac{u^4}{K}\cdot\left(\frac{2}{3}+\frac{2}{\pi^2}\cdot\sum_{s\geq1}s^{-2}\right)$$

The sum is infinite because of the periodic extension of the input spectrum, as explained before. It leads now to the Riemann Zeta-function with

$$\zeta(2) = \sum_{s=1}^{\infty}s^{-2} = \frac{\pi^2}{6}$$

and we get finally:

$$\sigma^2_{K>>1}(\lambda/N\tau_0) = \frac{u^4}{K}\cdot(2/3+1/3) = \frac{u^4}{K} = \frac{1}{K}\cdot\frac{u^4}{KN\tau_0\cdot1/N\tau_0} = \frac{1}{K}\cdot\frac{<|v_K(\lambda/N\tau_0)|^2>^2}{t_{Int}\cdot B_{Fl}} = \frac{1}{K}\cdot\sigma^2_{K=1}(\lambda/N\tau_0)$$

with $B_{Fl}=1/(N\tau_0)$, since the total integration time is given by $t_{Int}=K\cdot N\tau_0$.

This rather complicated derivation is essential for a full understanding of the noise performance of the DFT. The result of the average is equivalent to that of one single sample, which reflects the fact that the Fourier transforms of each data package are completely un-correlated. It is remarkable that the definition of the fluctuation bandwidth does not represent the pure radiometric noise, and the conclusion is that there must be a reason for the additional noise, which seems to be hidden behind the mathematics.

Correlation function

A similar calculation can be done for the correlation function of two frequency pixels of the spectrometer at frequencies $f_\lambda = \lambda/N\tau_0$ and $f_\mu = \mu/N\tau_0$, $\lambda, \mu = \pm1, \pm2, \pm3, \ldots$. For this, we need to evaluate the following expression:

$$G(\lambda-\mu) \;=\; \langle\, [\,\overline{|U(f_\lambda)|^2} - \langle\overline{|U(f_\lambda)|^2}\rangle\,]\cdot[\,\overline{|U(f_\mu)|^2} - \langle\overline{|U(f_\mu)|^2}\rangle\,]\,\rangle$$

The treatment of this expression is very similar to that above. As it turns out, there are only a few significant terms of the sum of the Delta-functions remaining. We have for the numerator while neglecting a couple of unimportant factors:

$$I \;=\; \sum_s \int u^4 \cdot \frac{\sin^4[\pi\, f' N\tau_0]}{(f'-\frac{\lambda-s}{N\tau_0})\cdot(f'-\frac{\lambda}{N\tau_0})\cdot(f'-\frac{\mu-s}{N\tau_0})\cdot(f'-\frac{\mu}{N\tau_0})}\cdot df'$$

For $s \neq 0$ we have:

$$I_{s\neq0} \;=\; (N\tau_0)^2\cdot u^4 \cdot \int \frac{\sin^4[\pi\, f' N\tau_0]}{s^2}\cdot[\frac{1}{f'-\frac{\lambda}{N\tau_0}} - \frac{1}{f'-\frac{\lambda-s}{N\tau_0}}]\cdot[\frac{1}{f'-\frac{\mu}{N\tau_0}} - \frac{1}{f'-\frac{\mu-s}{N\tau_0}}]\cdot df' \;=$$

$$= \;-(N\tau_0)^2\cdot\frac{u^4}{(\lambda-\mu)^2}\cdot\int\left[\frac{\sin^4[\pi(f'-\frac{\lambda}{N\tau_0})N\tau_0]}{(f'-\frac{\lambda}{N\tau_0})^2}\cdot\delta_{s,\lambda-\mu} + \frac{\sin^4[\pi(f'-\frac{\mu}{N\tau_0})N\tau_0]}{(f'-\frac{\mu}{N\tau_0})^2}\cdot\delta_{s,\mu-\lambda}\right]\cdot df' \;=$$

$$= \;-\frac{u^4}{2}\cdot\frac{\pi^2\cdot(N\tau_0)^3}{(\lambda-\mu)^2}\cdot(\delta_{s,\lambda-\mu}+\delta_{s,\mu-\lambda})$$

All other cross-terms disappear. For $s = 0$ we get:

$$I_{s=0} \;=\; u^4\cdot\frac{\pi^2\cdot(N\tau_0)^3}{(\lambda-\mu)^2}$$

In total, when summing, the contribution is zero. Similar results are found for the $(f'+f)$ terms, so that we have zero correlation for all pixels with the exception of $\lambda = \mu$. By definition, the correlation function is then equal to the variance of the statistical distribution of the pixel. It is therefore clear that the correlation function is not that of the *Sinc*^{2}-filter function, as one could expect. It becomes understandable when considering the input amplitude at a frequency of $\lambda/(N\tau_0)$, which has a simple *Sinc(*f-f_0*)*-response function, is affected by other frequency components in the neighborhood. For example, a contribution at a distance of $½/(N\tau_0)$ contributes with negative 62% ($-2/\pi$) to the amplitude at the pixel frequency. Thus, we have anti-correlation to such input components. It means, that, if the amplitude of a frequency component at this distance moves accidentally upwards, the amplitude of the center component will move into the opposite direction. This behavior does not disappear when taking the square of the amplitude. Thus, the anti-correlation persists, but the correlation between adjacent pixels seems to vanish when integrating over all frequency contributions. Therefore, it follows that there is no correlation between adjacent pixels, which is only correct for white noise input. This result for the correlation function corresponds to the findings for the fluctuation bandwidth. One should be aware that the baseline of the noise spectrum of the spectrometer looks different than that of a filter-bank or an AOS, because these exhibit correlation between neighbored pixels so that their fluctuations appear less erratic.

Conclusions

The mathematical treatment above supports the experimental finding that the noise output of the DFT is higher than one would expect from the filter function of the spectrometer. The missing correlation between neighbored pixels is difficult to interpret, when considering the fact that the filter curves of adjacent pixels overlap. But consequently, the noise decreases faster than usually when binning several frequency pixels. Therefore, for large bins, the deviation of the noise output from expectation vanishes. This compensates for the increased noise level at the beginning, as long as very high frequency resolution is not required. It seems, all this is a fundamental property of the DFT, which is introduced by the principle how the DFT functions. By the way, a CTS (Chirp Transform Spectrometer) should have the identical problem. It also Fourier transforms chunks of data by using the chirp technology. Each chunk has exactly the time length as given by the resolution similar to the DFT. Therefore, the fluctuation bandwidth is also reduced by a similar factor.

The outcome of the mathematics above is surprising when comparing with the situation in an AOS or any other spectrometer. In principle, the AOS does exactly the same as the DFT. In the Bragg-cell a time segment of the signal amplitude is present as travelling acoustic wave within a time window as is determined by the

acoustic velocity and the length of the aperture of the Bragg-cell. The imaging behind the Bragg-cell establishes the Fourier transform of the signal, which the CCD then detects. The photodiodes in the CCD are square law detectors, and the generated charges represent the square of the Fourier-transformed amplitude of the signal. So far, the characteristics of an AOS should be similar to that of the DFT. Experience tells that this is not correct. The filter function of the AOS leads directly to the observed fluctuation bandwidth, and the observed pixel to pixel correlation function is following the expectation as well. Similar arguments one may introduce for other spectrometer types, since any filtering is usually based on interference, i.e. correlation between many waves. Why is the performance different?

There is one fundamental difference between the two instruments, which is the way, how the signals are averaged. In the AOS this is not done in blocks of N signal data, but instead it is a continuous averaging of the square of the Fourier-transform, which takes place during the total integration time. This means that, instead of summing over K Fourier transforms, we have to sum over $K{\cdot}N$ Fourier transforms. Therefore, the term

$$\frac{SIN^2[\pi(f_1-f_2)KN\tau_0]}{K^2\cdot SIN^2[\pi(f_1-f_2)N\tau_0]} \;=\; \frac{1}{(\pi KN\tau_0)^2}\cdot\sum_{s=-\infty}^{\infty}\frac{SIN^2[\pi(f_1-f_2)KN\tau_0]}{[f_1-f_2-s/N\tau_0]^2} \;\Rightarrow\; \frac{1}{KN\tau_0}\cdot\sum_{s=-\infty}^{\infty}\delta(f_1-f_2-s/N\tau_0)$$

must be replaced by (K very large!)

$$\frac{SIN^2[\pi(f_1-f_2)KN\tau_0]}{(KN)^{2\cdot}\cdot SIN^2[\pi(f_1-f_2)\tau_0]} \;=\; \frac{1}{(\pi KN\tau_0)^2}\cdot\sum_{s=-\infty}^{\infty}\frac{SIN^2[\pi(f_1-f_2)KN\tau_0]}{[f_1-f_2-s/\tau_0]^2} \;\Rightarrow\; \frac{1}{KN\tau_0}\cdot\sum_{s=-\infty}^{\infty}\delta(f_1-f_2-s/\tau_0)$$

With proper frequency filtering of the input only one term with $s = 0$ survives. Therefore, when following the derivation above, the terms with $s \neq 0$ do not contribute anymore, so that the fluctuation bandwidth becomes

$$B_{FL} \;=\; \tfrac{3}{2}\cdot\frac{1}{N\tau_0} \;=\; \tfrac{3}{2}\cdot\delta_{\mathrm{Res}}$$

The conclusion is that it is not the discrete Fourier transformation, which is responsible for the characteristics; It is the averaging scheme, which causes the differences. This points to a somewhat theoretical solution of the problem: The FFT should be implemented after each single reading of the input signal $s(t_n)$. This is probably not a viable option, since such calculation would be needed every 0.5 nsec for a 1 GHz DFT for example. On the other hand, exactly this is done in the AOS automatically so that the difference between the two spectrometer types becomes now understandable. By the way, the AOS would behave comparably to the DFT, if the laser source would be pulsed with a repetition rate corresponding to the traveling time of the acoustic wave through the Bragg-cell aperture. The noise problem of the DFT reduces significantly when calculating the Fourier transform of a data set after sampling N/q data so that all data are used q times. The resulting additional contribution reduces then to $1/3{\cdot}1/q^2$. Thus, already for $q = 2$ one has a reduction of the excess noise by a factor of ¼! This level corresponds to an increase of the rms noise by 6.1% in comparison to the pure radiometric behavior instead of 22.5% with $q = 1$. Correspondingly, the observing time increases only by 12.5% instead of 50%.

Correlation between frequency pixels in a spectrometer

One of the important issues during the development of a spectrometer is how the contents of adjacent pixels are eventually correlated. For an estimate we calculate the difference of the contents of both pixels. The amplitude filter functions of both pixels are $A_1(f)$ and $A_2(f)$. For the expectation of the difference we have:

$$D_T(t) \;=\; <X_1(t)> - <X_2(t)> \;=\; c\cdot\int_{-\infty}^{\infty}|A_1(f)|^2\cdot df \;-\; c\cdot\int_{-\infty}^{\infty}|A_2(f)|^2\cdot df$$

We have assumed that the noise amplitudes within the range of both filter functions are constant, i.e. that we are dealing with pure white noise at the input. For the variance of the difference we have then:

$$\sigma_D^2(T) \;=\; <D_T^2(t)> \;-\; <D_T(t)>^2$$

Let us consider first the expectation of D_T^2:

$$<D_T^2(t)> \;=\; <\left[\tfrac{1}{T}\cdot\int_{t-T}^{t}dt'\left|\int_{-\infty}^{\infty}df\cdot A_1(f)\cdot u(f)\cdot\exp\{2\pi i\cdot f\cdot t'\}\right|^2 \;-\; \tfrac{1}{T}\cdot\int_{t-T}^{t}dt'\left|\int_{-\infty}^{\infty}df\cdot A_2(f)\cdot u(f)\cdot\exp\{2\pi i\cdot f\cdot t'\}\right|^2\right]^2$$

$$= \;<X_1^2(t)> \;+\; <X_2^2(t)> \;-\frac{2}{T^2}\cdot\int_{t-T}^{t}dt'\int_{t-T}^{t}dt''\int_{-\infty}^{\infty}df_1\int_{-\infty}^{\infty}df_2\int_{-\infty}^{\infty}df_3\int_{-\infty}^{\infty}df_4\cdot A_1(f_1)\cdot A_1^*(f_2)\cdot A_2^*(f_3)\cdot A_2(f_4)\cdot$$

$$\cdot\,<u(f_1)\cdot u^*(f_2)\cdot u^*(f_3)\cdot u(f_4)>\cdot\exp\{2\pi i\cdot[(f_1-f_2)\cdot t'-(f_3-f_4)\cdot t'']\}$$

For the lengthy integral expression we can use the same arguments as before about the non-zero contribution of *u(f)*, and we get now:

$$< D_T^2(t) > \;=\; < X_1^2(t) > \;+\; < X_2^2(t) > \;-\; 2\cdot c\int_{-\infty}^{\infty}|A_1(f)|^2\cdot df\;\cdot c\int_{-\infty}^{\infty}|A_2(f)|^2\cdot df$$

$$-\;\; 4\cdot c^2\cdot\int_{-\infty}^{\infty}df_1\int_{-\infty}^{\infty}df_2\cdot A_1(f_1)\cdot A_2^*(f_1)\cdot A_1^*(f_2)\cdot A_2(f_2)\cdot\frac{\sin^2\{\pi\cdot[f_1-f_2]\cdot T\}}{\{\pi\cdot[f_1-f_2]\cdot T\}^2}\;\;=$$

$$=\;< X_1^2(t) > \;+\; < X_2^2(t) > \;-2\cdot< X_1(t) >\cdot< X_2(t) > \;-$$

$$-\;\; 4\cdot c^2\cdot\int_{-\infty}^{\infty}df_1\int_{-\infty}^{\infty}df_2\cdot A_1(f_1)\cdot A_2^*(f_1)\cdot A_1^*(f_2)\cdot A_2(f_2)\cdot\frac{\sin^2\{\pi\cdot[f_1-f_2]\cdot T\}}{\{\pi\cdot[f_1-f_2]\cdot T\}^2}$$

The *Sinc*^{2}-function becomes again a Delta-function for large *T* so that we get now

$$< D_T^2(t) > \;=\; < X_1^2(t) > \;+\; < X_2^2(t) > -\,2\cdot< X_1(t) >\cdot< X_2(t) > \;-\;\frac{4}{T}\cdot c^2\cdot\int_{-\infty}^{\infty}|A_1(f)|^2\cdot|A_2(f)|^2\cdot df$$

On the other hand we have:

$$< D_T(t) >^2 \;=\; [< X_1(t) > \;-\; < X_2(t) >]^2 \;=\; < X_1(t) >^2 + < X_2(t) >^2 \;-2\cdot< X_1(t) >\cdot< X_2(t) >$$

This leads to

$$\sigma_D^2(T) \;=\; < D_T^2(t) > - < D_T(t) >^2 \;=\; < S_1^2 > - < S_1 >^2 + < S_2^2 > - < S_2 >^2 \;-\; 2\cdot[< S_1\cdot S_2 > - < S_1 >\cdot< S_2 >]$$

$$=\;< \delta S_1^2 > + < \delta S_2^2 > - 2\cdot< \delta S_1\cdot\delta S_2 >$$

and finally

$$\sigma_D^2(T) \;=\; \sigma_1^2(T) + \sigma_2^2(T) - 4\cdot\frac{c^2}{T}\cdot\int_{-\infty}^{\infty}|A_1(f)|^2\cdot|A_2(f)|^2\cdot df$$

We replace again $|A_1(f)|^2$ and $|A_2(f)|^2$ by the power filter function $L_1(f)$ and $L_1(f)$ and use the relation between the constant *c* and the average power and we get

$$\sigma_D^2(T) \;=\; \sigma_1^2(T) + \sigma_2^2(T) - 2\cdot\sqrt{\sigma_1^2(T)\cdot\sigma_2^2(T)}\cdot\Phi_{12}$$

with $\Phi_{12} \;=\; \dfrac{< \delta S_1\cdot\delta S_2 >}{\sqrt{< \delta S_1^2 >\cdot< \delta S_2^2 >}} \;=\; \dfrac{< \delta S_1\cdot\delta S_2 >}{\sqrt{\sigma_1^2\cdot\sigma_2^2}} \;=\; \dfrac{\int_0^{\infty}L_1(f)\cdot L_2(f)\cdot df}{\sqrt{\int_0^{\infty}L_1^2(f)\cdot df\cdot\int_0^{\infty}L_2^2(f)\cdot df}}$ and

$$\sigma_i^2(T) \;=\; 2\cdot\frac{< X_i(t) >^2}{B_{Fl}\cdot T},\;\; i=1,2$$

The X_i are the signals of the ith pixel of the spectrometer. The additional factor “2” results from the fact that the signals are the difference of two independent measurements “On” and “Off”.

It is obvious that the correlation between the two pixels is independent on the integration time *T*. Its value is readily found when knowing the two filter curves of the pixels. Differences in gain cancel through the expressions in the denominator. The correlation is small in case the two filter curves have little overlap. It is at maximum (=1), if $L_1(f)$ and $L_2(f)$ are proportional. In this case the variance of the differences vanishes.

To simplify things further we assume also that we have two similar pixels at identical power level. Then we have

$$< S_1^2(t) > \;=\; < S_2^2(t) > \;=\; < S^2(t) >,\;\; B_{Fl,1} = B_{Fl,2} \;=\; B_{Fl}$$

$$\sigma_1^2(T) \;=\; \sigma_2^2(T) \;=\; \sigma^2(T) \;=\; 2\cdot\frac{< X(t) >^2}{B_{Fl}\cdot T}$$

In this case we have now

$$\sigma_D^2(T) \;=\; 2\cdot\sigma^2(T)\cdot[1-\Phi_{12}] \qquad [46]$$

This derivation is of particular interest, if one wants to compare the noise of two different spectrometers. In case the filter functions of the pixels are not exactly identical, one observes some leftover noise in the difference. The same is true if the pixels are shifted in frequency with respect to each other. One should keep in mind that the above derivation does not apply for correlator- as well as digital-correlator-spectrometers, since their filter-functions do not reflect the mutual correlation as is shown above. Therefore, in such cases Eq.[46] is not applicable.

Correlation at additional spectrometer noise

The previous derivation is only exact, if the observed noise is generated exclusively by the input signal to the receiver. If the spectrometer is adding some noise, then one has to consider this separately. For this we are using a simple statistical consideration. The variance of the difference measurements is defined as:

$$\sigma_D^2(T) = <[D_T(t) - <D_T(t)>]^2> =$$

$$= <[S_{1T}(t) - <S_{1T}(t)>]^2> + <[S_{2T}(t) - <S_{2T}(t)>]^2> - 2\cdot<[S_{1T}(t) - <S_{1T}(t)>]\cdot[S_{2T}(t) - <S_{2T}(t)>]> =$$

$$= \sigma_1^2(T) + \sigma_2^2(T) - 2\cdot\sqrt{\sigma_1^2(T)\cdot\sigma_2^2(T)}\cdot\frac{<[S_{1T}(t) - <S_{1T}(t)>]\cdot[S_{2T}(t) - <S_{2T}(t)>]>}{\sqrt{<[S_{1T}(t) - <S_{1T}(t)>]^2>\cdot<[S_{2T}(t) - <S_{2T}(t)>]^2>}} =$$

$$= \sigma_1^2(T) + \sigma_2^2(T) - 2\cdot\sqrt{\sigma_1^2(T)\cdot\sigma_2^2(T)}\cdot g_{12}$$

g_{12} is the normalized first order correlation function of the two signals S_1 and S_2. When comparing with the above treatment, we find for the pure radiometric noise case

$$g_{12} = \Phi_{12}$$

In a spectrometer like an AOS the detected noise consists of two contributions: the radiometric noise and the shot- and dark-noise of the detecting CCD. In the AOS we have therefore:

$$\sigma^2(T) = \sigma_{Rad}^2(T)\cdot q^2$$

q is always larger or equal then unity since the noise is at least that of the incoming RF. For example, we can write for the CCD noise:

$$q^2 = q^2(\alpha) = 1 + \frac{B_{Fl}\cdot\delta}{n_0\cdot\beta^2}\cdot\frac{1}{\alpha} + \frac{B_{Fl}\cdot\delta}{r^2\cdot\beta^2}\cdot\frac{1}{\alpha^2}$$

B_{Fl} is the fluctuation bandwidth of the spectrometer, δ is the frame time for the read-out of the CCD, n_0 is the full well capacity of each of the CCD pixels, r is the ratio of full well and dark noise rms, α is the signal level at the CCD relative to the full well capacity of each pixel, and β is the ratio of ADC maximum to full well of the CCD. A typical value for β is near 0.9, while α can be anywhere between 0.05 and 1.0. The second term is caused by the photo-electron shot noise and the third by the dark current. (Similar expressions should apply for the detector noise of other spectrometer types as well, since shot noise and dark- or read-out-noise are common phenomena of practically all detectors.)

The expectation of the product S_1 and S_2 is now influenced by the two different contributions. For this we write:

$$\delta S_i(t) = S_i(t) - <S_i(t)>, \quad \delta S_i(t) = \delta S_{i,Rad}(t) + \delta S_{i,spectr}(t), \quad i = 1,2$$

Then we get

$$<\delta S_1(t)\cdot\delta S_2(t)> = <[\delta S_{1,Rad}(t) + \delta S_{1,Spectr}(t)]\cdot[\delta S_{2,Rad}(t) + \delta S_{2,Spectr}(t)]> = <\delta S_{1,Rad}(t)\cdot\delta S_{2,Rad}(t)>$$

All non-radiometric noise contributions from the spectrometer pixels are uncorrelated. The same is valid for the products of radiometric and spectrometer noise. Therefore, only the radiometric term survives.

The cross term above considered only the radiometric part, so that we have to modify it. We had:

$$\Phi_{12} = \frac{<\delta S_{1,Rad}(t)\cdot\delta S_{2,Rad}(t)>}{\sqrt{\sigma_{1,Rad}^2\cdot\sigma_{2,Rad}^2}} = \frac{\int_0^\infty L_1(f)\cdot L_2(f)\cdot df}{\sqrt{\int_0^\infty L_1^2(f)\cdot df\cdot\int_0^\infty L_2^2(f)\cdot df}}$$

Consequently we have now:

$$g_{12} = \frac{<\delta S_{1,Rad}(t)\cdot\delta S_{2,Rad}(t)>}{\sqrt{\sigma_1^2\cdot\sigma_2^2}} = \sqrt{\frac{\sigma_{1,Rad}^2}{\sigma_1^2}}\cdot\sqrt{\frac{\sigma_{2,Rad}^2}{\sigma_2^2}}\cdot\frac{<\delta S_{1,Rad}(t)\cdot\delta S_{2,Rad}(t)>}{\sqrt{\sigma_{1,Rad}^2\cdot\sigma_{2,Rad}^2}} = \frac{\Phi_{12}}{q_1\cdot q_2}$$

With identical radiometric noise from both pixels we get

$$\sigma_D^2(T) \;=\; \sigma_{Rad}^2(T)\cdot[q_1^2+q_2^2-2\cdot\Phi_{12}]$$

In case the two spectrometer pixels are seeing the same signal level we have $q_1 = q_2 = q$, and we have then:

$$\sigma_D^2 \;=\; 4\cdot\frac{<S(t)>^2}{B_{Fl}\cdot T}\cdot[q^2-\Phi_{12}] \qquad [47]$$

The impact of correlation can directly be calculated by means of the correlation function as is already defined above. One can evaluate it by building products of the content of the frequency noise output of the spectrometer. The spectrum should be shifted pixel by pixel and the product of all available pixels should be calculated. One obtains the correlation function of the frequency output of the spectrometer:

$$g_{12}(k) \;=\; \frac{\frac{1}{N-k-1}\cdot\sum_{n=1}^{N-k}S_{1,n}\cdot S_{2,n}}{\sqrt{\frac{1}{N-k-1}\cdot\sum_{n=1}^{N-k}S_{1,n}^2\cdot\frac{1}{N-k-1}\cdot\sum_{n=k+1}^{N}S_{2,n}^2}}$$

k is the number of the shift in terms of frequency pixels. 1 and 2 stands for the two spectra – here the identical spectra - which should be correlated and *N* is total number of available pixels. In case there is unwanted structure in the baseline of the spectra it is recommended to apply a low order baseline fit to the data in order to avoid the influence of "ripple" to the calculation of the correlation. The $S_{i,n}$ (i=1,2) are the pixel data, as they are for example derived from (ON-OFF)/(OFF-Z) or other calibrated observations (Z is the Zero-measurement of the spectrometer.) minus the baseline fit. $g_{12}(0)$ is identical with the above defined correlation g_{12}. The appearance of the correlation function as a function of *k* contains now all information, if different spectra are suspected to be correlated. In this case the expectation value of $g_{12}(k)$ will become unequal to zero for some values of *k*. Certainly, also the autocorrelation can be found when setting $S_{1,n} = S_{2,n}$. The appearance of the auto-correlation function is a very useful tool to find out about particular problems in the spectrometer such as electronics, speckles and their stability in the AOS, read-out of CCDs etc. As is mentioned above, also here the situation with correlator- or digital correlator-spectrometers is different.

Noise when adding two spectrometer pixels

A frequent question is how the statistics of co-added pixels looks like. For this we can follow the identical mathematics as before. For the variance of two pixels, in particular of two adjacent pixels one has:

$$S(T) \;=\; [S_1(t)+S_2(T)]/2$$

$$\sigma_S^2(T) \;=\; \left\{\sigma_1^2(T)+\sigma_2^2(T)+2\cdot\sqrt{\sigma_1^2(T)\cdot\sigma_2^2(T)}\cdot g_{12}\right\}/4$$

Assuming that

$$\sigma_1^2(T) \;=\; \sigma_2^2(T) \;=\; \sigma^2(T),$$

then we get:

$$\sigma_S^2(T) \;=\; \sigma^2(T)\cdot[1+g_{12}] \;=$$
$$=\; \sigma^2(T)\cdot[1+\phi_{12}/q^2]$$

Φ_{12} is the mutual overlap of the two frequency pixels (see above). With zero correlation one obtains the expected result: the variance becomes half. But with correlation the result is larger! In such cases it is impossible to reduce the noise of experimental data to the expected level by means of co-adding pixels.

Usually it is advisable to keep the correlation, i.e. the value of g_{12} of adjacent pixels as small as possible. Only then the value of the content of the other pixel provides something new in terms of statistics. At large overlap of the filter functions of pixels one generates largely the identical information in both pixels, as is certainly undesirable. At the same time, the statistical fluctuations are partly identical, which means that there is only little improvement of the signal to noise ratio. Therefore it is important to determine the frequency spacing of neighboured pixels on the basis of the expected correlation of the pixels.

Noise of many co-added frequency pixels

Frequently, it is useful to co-add several neighboured frequency pixels, as is needed for example when detecting the continuum level of a source or when observing broad signals from external Galaxies. According to the radiometer equation one should expect an appreciable reduction of the noise. As we have seen above for

the case of co-adding two pixels, it may be not as effective as one might hope for. Again, the outcome depends strongly on the correlation between the pixels. (The consideration is also valid for the co-addition of data taken in a time series, if there is correlation due to a time constant for example.) We assume to co-add *N* data S_k and we write:

$$Y_N(k) \;=\; \frac{1}{N}\cdot\sum_{n=1}^{N} S_{k+n}$$

The variance of this co-added signal is then:

$$\begin{aligned}\sigma_N^2 \;&= < Y_N(k)^2 > - < Y_N(k) >^2 \;= \\ &= \;\frac{\sigma_1^2}{N^2}\cdot\left[N\cdot g(0) + 2\cdot N\cdot\textstyle\sum_{k=1}^{N-1}(1-\frac{k}{N})\cdot g\,(k)\right] \;= \\ &= \;\frac{\sigma_1^2}{N}\cdot\left[1 + 2\cdot\left(1-\frac{1}{N}\right)\cdot g(1) + \;2\cdot\left(1-\frac{2}{N}\right)\cdot g(2) + \cdots\right]\end{aligned} \qquad [48]$$

$g_{12}(k)$ is the correlation function of the initial data S_k. The result does not depend simply on 1/*N* as one might expect. Only at *N* very large it becomes simpler, if the values of the g_{12} approach zero fast enough. Without correlation we find the ordinary result:

$$\sigma_{\mathrm{N}}^2 \;=\; \frac{\sigma_1^2}{N} \quad \text{with} \quad \sigma_1{}^2 \text{ the variance of the initial data.}$$

How much the outcome differs depends on the actual values of g_{12}. For example, if all data are fully correlated, i.e. $g_{12}(k) = 1$ for all *k*, then one gets:

$$\sigma_{\mathrm{N}}^2 \;=\; \sigma_1^2$$

and the improvement of the signal to noise becomes zero. Normally the filter curves of neighboured pixels in AOS overlap, and we found with our spectrometers by experiment values like

$g_{12}(1) \approx 0.45,\; g_{12}(2) \approx 0.2,\; g_{12}(k>2) \approx 0$

Such values depend in an AOS on the quality of the optics, the appropriate adjustment, the Bragg-cell itself, and the characteristics of the CCD. One should always try to make the values of the correlation as small as possible. If one considers Nyquist-sampling for example one should have a value for $g_{12}(1)$ below 0.5.

As an example we consider a co-add of 10 pixels. Using the above values for the correlation we get then:

$$\sigma_{\mathrm{N}}^2 \;=\; \frac{\sigma_1^2}{10}\cdot\left[1 + 2\cdot\left(1-\frac{1}{10}\right)\cdot 0.45 + 2\cdot\left(1-\frac{2}{10}\right)\cdot 0.2\right] \;=\; \frac{\sigma_1^2}{4.7}$$

The rms-error reduces here to

$$\frac{1}{\sqrt{4.7}} \;\approx\; 0.46 \quad \text{instead of} \quad \frac{1}{\sqrt{10}} \;\approx\; 0.32$$

It means that the noise level is now about 46% higher than expected. It is essential that one considers eventual correlation between the pixel data before expecting too much.

One should also look into the outcome when co-adding very large numbers of pixels. In the ideal case of identical filter-functions and identical gain one should expect that the resulting equivalent filter-curve has practically the shape of a box-car filter and the resolution bandwidth becomes accordingly:

$$\delta_{\mathrm{Res}} \;\approx\; \mathrm{N}\cdot\delta$$

with δ the pixel frequency spacing. This is only correct if the pixel spacing is small as compared to the resolution bandwidth of the individual pixels. (Otherwise the result of the co-addition would not result in a box-car like filter curve.) Because of the near box-car shape the fluctuation bandwidth B_{Fl} should be practically identical with the resolution bandwidth δ_{Res}. Thus we have also:

$$B_{Fl}(N) \;\approx\; N\cdot\delta$$

On the other hand, we have for *N* large:

$$\sigma_{\mathrm{N}}^2 \;=\; \frac{\sigma_1^2}{\mathrm{N}}\cdot\left[1 + 2\cdot\mathrm{g}_{12}(1) + 2\cdot\mathrm{g}_{12}(2) + \cdots\right] \;=\; \frac{\sigma_1^2}{\mathrm{N}_{\mathrm{eff}}}$$

In addition we have the radiometer equation:

$$\sigma_{\mathrm{N}}^2 \;=\; \frac{<y_{\mathrm{N}}>^2}{\mathrm{B_{Fl}(N)}\cdot\mathrm{T}}$$

$<y_N>$ is identical with $<S>$ and we have therefore:

$$\sigma_N^2 = \frac{<S>^2}{B_{Fl}(N)\cdot T} = \frac{<S>^2}{B_{Fl}(1)\cdot T}\cdot\frac{1}{N^*} \quad \text{with}$$

$$N^* = N/[1 + 2\cdot g_{12}(1) + 2\cdot g_{12}(2) + \cdots]$$

The resulting fluctuation bandwidth is therefore under the above assumptions:

$$B_{Fl} = N^*\cdot B_{Fl}(1) = N\cdot\delta$$

Thus we have:

$$B_{Fl}(1) = \delta\cdot[1 + 2\cdot g_{12}(1) + 2\cdot g_{12}(2) + \cdots]$$

We have found a simple method to compare the fluctuation bandwidth as found by means of the Allan variance plot with above formula using the values of the correlation function as found from the noise data of spectrometer spectra[10]. A different outcome is a hint for additional noise in the spectra, which might reduce the correlation. (Again, the situation with correlator- and digital correlator spectrometers is not the same.)

[10] This formula is only approximately correct. Exactly we have:

$$B_{Fl} = \left[\int_0^\infty L(f)\cdot df\right]^2 / \int_0^\infty L^2(f)\cdot df = \frac{1}{2}\cdot\int_{-\infty}^{\infty}\Phi(\nu)\cdot d\nu$$

with $\Phi(\nu) = \frac{1}{2}\cdot\int_0^\infty L(f)\cdot[L(f+\nu) + L(f-\nu)]\cdot df / \int_0^\infty L^2(f)\cdot df$

This becomes identical with above formula, if $\Phi(\nu)$ is slowly varying so that the integral can be replaced by its Riemann sum on the basis of the interval δ replacing the integral. It is only correct as long as δ is small enough. Experience confirms that above formula is still valid when applying Nyquist-sampling. Consequently, the procedure remains rather useful.

V. Miscellaneous

Modulation-Transfer-Function (MTF)

In spectroscopy it should be a standard question, what spectrometer, i.e. what resolution of the spectrometer should be applied for the particular needs of the research. In many cases this is probably not much considered because there is little choice between different spectrometers with appropriate resolution. But still, if there is for example a spectrometer available with too high resolution one can still apply the co-adding of frequency pixels in order to adjust the effective resolution to the needs of the scientific task. The same is actually true for observing maps of signals from sky with a telescope in order to learn about the distribution of particular signals in space. A rarely used method to investigate such problems is the use of the so-called “Modulation-Transfer-Function” (MTF). One finds it in descriptions of CCD line- or matrix-arrays, but otherwise it is mostly ignored.

When considering the properties of a spectrometer the MTF is easily calculated by the normalized Fourier-transform of the pixel-filter-function *I(f)*. It provides information how the spectrometer shows eventual periodic structures in the spectrum. If we Fourier-analyze the spectral distribution, the various contents at particular Fourier-components *F* become visible as are given by this expression:

$$M(F) = \frac{\int I(f)\cdot\cos[2\pi F\cdot f]\cdot df}{\int I(f)\cdot df} \qquad [50]$$

M(F) is always equal to unity for the frequency *F*=0. It means that a constant offset in the spectrum is fully visible. (In some cases instead of a Fourier-component the response to a box-car-function is considered, as is meaningful for example in case of CCDs, since they represent a rectangular filter-function for each pixel of the CCD.)

In general it is more important to determine the Signal-to-Noise (S/N) of the Fourier-component at frequency F seen with the particular spectrometer. In case of radiometry we can describe this by the expression

$$S/N = \frac{M(F)}{rms} \propto \sqrt{B_{FI}}\cdot M(F) = \frac{\int I(f)\cdot\cos[2\pi F\cdot f]\cdot df}{\sqrt{\int I^2(f)\cdot df}}$$

It provides information, how well spectral features of spatial frequency *F* can be detected with a particular spectrometer with filter-function *I(f)*.

For a constant offset signal (F=0) the S/N is obviously determined by $\sqrt{B_{FI}}$. The function allows now to find out what frequency resolution is best for the detection of particular features in the spectra. Certainly, the higher the resolution the better the amplitude of the features at given frequency *F* can be determined, but the noise on the other hand is increasing with higher resolution. Therefore, a compromise can be found, where the signal can be seen at best signal-to-noise. In cases, where one wants to detect special features like Gaussian line-profiles for example, a different approach like a wavelet calculation instead of the MTF can be applied (see below). Similar, if the noise in the system is not purely radiometric, the fluctuation-bandwidth should be replaced by an appropriate figure. But in general, the MTF is very useful in most cases. For instance, if one has a box-car like filter, the MTF and the S/N are calculated as

$$M(F) = \frac{SIN[\pi F\delta_{Res}]}{[\pi F\delta_{Res}]} \qquad S/N = \frac{SIN[\pi F\delta_{Res}]}{\pi F\cdot\sqrt{\delta_{Res}}}$$

For $F = 1/\delta_{Res}$ the response is here for example zero, since a full period of the spatial frequency *F* falls into the width of one pixel.

For a Sinc^2-filter characteristics of an apodized correlator we have:

$$M(F) = 1-\frac{k}{N} = 1-F\cdot\delta_{Res} \quad F = k\tau_0, \quad S/N = \sqrt{\tfrac{3}{2}\cdot\delta_{Res}}\cdot[1-F\cdot\delta_{Res}]$$

for a component of the spectral distribution with period $F = k\cdot\tau_0$, k = 1,2,3,… It means that a spatial component, which amounts to exactly one resolution bandwidth ($F = 1/\delta_{Res}$), cannot be observed. Therefore, Nyquist-sampling, i.e. one resolution element per half the resolution bandwidth, is not useful. One should remember that the idea of Nyquist-sampling is based on the structure of Delta-function-like signals, and not of distributed features (see below).

For the Sinc-function one finds:

$$S/N = \sqrt{\delta_{Res}}$$

Thus we have: $M(F) = 1$ für $0 \le F < 1/(2\delta_{Res})$, which is remarkable, but not surprising, since the Fourier-transform of the Fourier-transform of a rectangular filter is the same filter again.

With a Gaussian filter-function, as we have e.g. with an AOS or rather closely with a telescope beam on sky, we get:

$$l(f) = \frac{1}{\delta_{Res}} \cdot exp\{-\pi \cdot f^2/\delta_{Res}^2\}, \quad M(F) = exp\{-\pi \cdot F^2 \cdot \delta_{Res}^2\}$$

$$S/N = \sqrt{\sqrt{2} \cdot \delta_{Res}} \cdot exp\{-\pi \cdot F^2 \cdot \delta_{Res}^2\}$$

For $F = 1/\delta_{Res}$ we see still 4.3% of the amplitude of the signal.

A Lorentzian filter-function leads to:

$$l(f) = \frac{\frac{\delta_{Res}}{\pi^2}}{f^2 + \left(\frac{\delta_{Res}}{\pi}\right)^2}, \quad M(F) = exp\{-2 \cdot F \cdot \delta_{Res}\}$$

$$S/N = \sqrt{2 \cdot \delta_{Res}} \cdot exp\{-2 \cdot F \cdot \delta_{Res}\}$$

At $F = 1/\delta_{Res}$ one can still see 13,5% of the modulation amplitude.

Sometimes it might be more useful to apply some kind of a wavelet- instead of a Fourier-analysis. In many cases a molecular line has the appearance of a Gaussian, for instance, if Doppler-broadening is involved. It is therefore useful to calculate

$$M_G(\Delta_{½}) = \int_{-\infty}^{\infty} l(f) \cdot \exp\{-[\tfrac{f}{\Delta_{½}}]^2 \cdot 4 \cdot \ln(2)\} \cdot df \cdot$$

$\Delta_{½}$ is the full half-width (FWHM) of the Gaussian distribution, which can now be varied to describe the responsivity to Gaussian-like features in the spectra. In addition, the filter-curve of acousto-optical spectrometers (AOS) is more or less a Gaussian. Due to the characteristics of the Gaussian shape, it becomes evident that the modified MTF rolls off rather slowly at smaller line-widths. How this roll-off looks like depends on the filter-function of the spectrometer. It is not easy to present correct formulas for the $M_G(\Delta_{1/2})$ unless the spectrometer filter-function is also a Gaussian. In this case we have:

$$l(f) = \sqrt{\frac{4 \cdot log[2]}{\pi}} \cdot \frac{1}{\delta_{½}} \cdot exp\left\{-4 \cdot log[2] \cdot \frac{f^2}{\delta_{½}^2}\right\} = \frac{1}{\delta_{Res}} \cdot exp\left\{-\pi \cdot \frac{f^2}{\delta_{½}^2}\right\}$$

The modified MTF becomes now:

$$M_G(\Delta_{½}) = \frac{\Delta_{½}}{\sqrt{\Delta_{½}^2 + \delta_{½}^2}} = \frac{\Delta_{½}}{\sqrt{\Delta_{½}^2 + \frac{4 \cdot \ln(2)}{\pi} \cdot \delta_{\mathrm{Res}}^2}}$$

The roll-off at small $\Delta_{1/2}$ definitely slower than otherwise. The Signal-to-noise is now:

$$S/N = \sqrt{\sqrt{2} \cdot \delta_{\mathrm{Res}}} \cdot \frac{\Delta_{½}}{\sqrt{\Delta_{½}^2 + \delta_{½}^2}}$$

Although this discussion looks a bit theoretical, it might be worthwhile to consider it carefully. The new Digital-Fourier-Transform-Spectrometers (DFT) usually provide a lot more frequency resolution than is needed in practically all cases. Therefore, nearly any desired resolution can be generated by co-adding without serious loss in information. It is then a matter of data processing afterwards in order to find the best possible S/N for the discussion of the results. Here one can eventually improve the data quality appreciably by means of the tools presented above.

Nyquist Sampling

In many cases Nyquist sampling is recommended when measuring complete maps of a source on sky or fully sampled spectra with a spectrograph. The initial idea about Nyquist sampling was that one has data which are derived by means of a "Delta-comb" so that there is zero overlap of the spatial or frequency response function of the instrument. The assumption is therefore that with each position one obtains 100% information about the signal strength at that position. This is only a (nearly) correct assumption, if the spacing between the data points is large in comparison with the filter width of the instrument. The situation changes when dealing with a

dense sampling grid. The main question is how much additional information one can derive when increasing the density of the grid. For an estimate we can use the MTF, which we define, as before, as the convolution of the instrument filter function *l(f)* with a sinusoidal power distribution of the signal at a (spectral or spatial) frequency *v*. This could be a Fourier-component of the spectral distribution for example.

For the MTF we need to calculate the integral

$$M(\nu) \;=\; \int_{-\infty}^{\infty} cos(2\pi \cdot f \cdot \nu) \cdot l(f) \cdot df$$

For an AOS or a telescope with Gaussian illumination we can approximate the filter function by a Gaussian:

$$l(f) \;=\; \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot exp[-f^2/2 \cdot \sigma^2] \;=\; \frac{1}{\delta_{Res}} \cdot exp[-\pi \cdot f^2/\delta_{Res}^2]$$

Here again, the so called "resolution bandwidth" δ_{Res} of a filter (or "beam solid angle" of a telescope) is given by :

$$\delta_{Res} \;=\; \frac{1}{l_{Max}} \cdot \int_{-\infty}^{\infty} l(f) \cdot df$$

as we have already used earlier. With this we get now:

$$M(\nu) \;=\; \frac{1}{\delta_{Res}} \cdot \int_{-\infty}^{\infty} \cos(2\pi \cdot f \cdot \nu) \cdot exp[-\pi \cdot f^2/\delta_{Res}] \;=\; exp[-\pi \cdot \delta_{Res}^2 \cdot \nu^2]$$

The MTF is 1 at very small spatial frequency *v* and approaches zero with large *v*.

Nyquist sampling defines the distance between adjacent data points, and one usually assumes that one should have two data samples per maximum observable frequency period ν_{Max}. If the instrument has a resolution δ_{Res}, it sounds plausible that the spacing of the data points should be $1/2 \cdot \delta_{Res}$. In this case we have $\nu = 1/\delta_{Res}$, and we get for the MTF at this spatial frequency:

$$M(\nu = 1/\delta_{Res}) \;=\; \exp(-\pi) \;=\; 4.3\%\ !!$$

Thus, the measured amplitude of a signal contribution at spatial frequency $1/\delta_{Res}$ is very small. Consequently, it is not worthwhile to consider fully sampled maps, because the derived additional information is practically close to zero, but is very costly on the other hand in terms of observing time. If we increase the grid spacing to δ_{Res} ($\nu = (2\delta_{Res})^{-1}$), the response becomes 45.6%, which is still not very satisfying but already useful. Certainly, real filter functions are not Gaussian, but in all standard cases the approximation is close enough to reality. The discussion here concerns only one dimension, but things don't change significantly, when talking about two-dimensional maps.

The result above can be compared with the value of the correlation function of two adjacent pixels. If we assume two identical pixels spaced by a distance s, we have for the correlation between the two data:

$$g_{12}(s) \;=\; \frac{\int_{-\infty}^{\infty} l(x + s/2) \cdot l(x - s/2) \cdot dx}{\sqrt{\int_{-\infty}^{\infty} l^2(x + s/2) \cdot dx \;\cdot \int_{-\infty}^{\infty} l^2(x - s/2) \cdot dx}} \;=\; exp\left[-\pi \cdot \frac{s^2}{2 \cdot \delta_{Res}^2}\right]$$

Thus, the correlation of two data points at Nyquist sampling is

$g_{12}(s = \delta_{Res}/2) \;=\; 67.5\%$

Correlation means, that the statistical fluctuations seen with the second data point are by more than two thirds identical with that of the first point. With $s = \delta_{Res}$, the correlation is only 20.8%.[11]

As a rule of the thumb one might consider an optimized Nyquist pixel spacing at a MTF near 0.5. This suggests a spacing of $s \approx \delta_{Res}$. The correlation has there a value of 20.1%. It means that a Fourier-component of a period of $\nu \approx (2 \cdot \delta_{Res})^{-1}$ will be observed with nearly 50% of the real amplitude. As one can see, the information provided by the MTF is not exactly identical with that of the correlation, which is not a surprise, since the MTF provides information about signal sensitivity while the correlation describes the noise behavior. It is a

[11] One should keep in mind that the noise argument, as is considered by correlation, applies only if the signals seen with the pixels are observed simultaneously, as is valid for example in a standard spectrometer. Even with array-receivers, the situation is not equivalent since the noise is independently generated in the mixers of the individual receivers, so that noise-correlation does not occur.

matter of the user to decide what one considers as more important. With the KOSMA AOS we have typical values of the correlation between adjacent pixels of about g_{12} = 0.45 (see above), which fits quite well to the pixel spacing of 1.03 MHz. The MTF is therefore about 50% when setting $\nu \approx (2 \cdot \delta_{Res})^{-1}$. When measuring maps with a telescope it might be worthwhile to consider the observing economy as most important, i.e. the MTF is decisive. It is also clear that correlated noise between adjacent pixels is not of importance, for instance, if the data are taken subsequently. In that case the noise is not correlated, since the data are not taken at the same time.

Resampling

Resampling procedures are useful when reducing the amount of data which are delivered by high resolution spectrometers for example in order to simplify the handling and to reduce the amount of the partly superfluous data. The same might apply for large maps measured with telescopes on sky. Resampling is also needed, if higher resolution spectra are co-added or resampled in order to reduce resolution or to make spectra of different spectrometers comparable. In general, just do-adding an integer number of pixels might be sufficient, but things look different, when trying to make spectra of different sources, i.e. spectrometers or telescopes, comparable. In that case one has to consider resampling of non-integer numbers of pixels. Just when co-adding fractions of pixels one ends with different effective new pixel resolutions and/or fluctuation bandwidths, depending on the accidental positions of the center-frequencies of the new pixels on the initial grid. In the particular case of Acousto-Optical Spectrometers a proper resampling procedure is essential when linearizing the frequency scale of the spectrometer, because it is usually non-linear within 1 or 2 %. In the latter case the resampling algorithm should reproduce the characteristics of the channel profiles of the spectrometers, i.e. the filter curves should look similar after resampling. In case of filter-banks, the filter curve is a Lorentzian or a multiple of it in case of multi-pole filters. With AOS one has a curve which looks very similar to a Gaussian. DFT or Chirp-Transform-Spectrometers (CTS) have usually $Sinc^2$-characteristics or similar, but their resolution is typically very high, so that it is rather unlikely that other spectra need to be resampled to the resolution of the spectra of DFTs or CTSs.

When considering the situation of AOS, one could propose, to use a linearized new grid with the pixel width of the average of the pixel separation of the initial spectrometer grid. In some cases the new pixel position just coincides with the position of the initial grid with the consequence that the resolution and fluctuation bandwidth is identical with the old one. But, if the position of one of the new pixels is just halfway between two old ones, one needs to co-add the content of two adjacent old pixels with 50% of their content. In consequence, one ends up with about twice the resolution and fluctuation bandwidth of the old pixels. This now practically undefined situation is rather unfavorable and the resampling procedure should therefore be modified. Actually, for maps taken on sky at a non-regular grid one would have the same kind of difficulties. In order to reduce the problems in the particular in case of AO'S-spectra we therefore propose a resampling procedure using a Gaussian weight function. The first step to resample is to identify the frequency calibration of the spectrometer. This is usually no issue with filter-banks of DFT, but AOS or CTS may exhibit some deviations from linearity so that the exact pixel to frequency calibration has to be found by comb-spectra for example. In case the frequency calibration is well established, the resampling routine can be applied. The next step is to determine the pixel spacing of the resampled spectrum. It might be advisable to use optimized Nyquist sampling, i.e. the new pixel spacing D could be about equal to the resolution bandwidth of the new resampled frequency pixels (see above, but other spacing is also possible). We call δ the resolution bandwidth of the initial spectrum and Δ the width of the resampling function (in our case a Gaussian), then we can write:

$$D = \sqrt{\Delta^2+\delta^2} \quad \text{or} \quad \Delta = \sqrt{D^2 - \delta^2}$$

This defines now the width of the resampling function. Here we have assumed that the width of the initial and the resampling function co-add like Gaussians, i.e. the resulting width equals the square-root of the sum of the squares. In case of AOS-spectra this is a fairly good approximation. Gaussians have the advantage that their role-off in the wings of the filter function is fairly steep so that correlation with other pixels becomes rather small. By the way, the Gaussian resampling method might also be rather useful when applied to maps of the intensity distribution of extended sources taken with a telescope, since the beam profile of radio-telescopes is usually also not very different form Gaussian profiles.

The resampling procedure uses now the following algorithm:

The new frequencies are given by:

$$f_j = f_0 + j \cdot D, \; j = 0,1,2,\ldots,j_{Max,} \quad f_{Min} \leq f_j \leq f_{Max}$$

f_{Min} and f_{Max} are the minimum and maximum frequencies of the input spectrum. The resampling of the initial data x_n at frequencies ν_n to new data y_j at frequencies f_j is now given by:

$$y_j = \frac{1}{N} \cdot \sum_{|\nu_n - f_j| < k \cdot \Delta} exp\left\{-\pi \cdot \frac{(\nu_n - f_j)^2}{\Delta^2}\right\} \cdot x_n \cdot d\nu_n$$

$d\nu_n$ is the frequency separation between adjacent pixels of the input data, which can slightly vary in an AOS, but are constant with Filter-banks for example. The constant k should be set to $k \approx 1.7$; it confines the sampling interval to a width of $3.4 \cdot \Delta$. (The exponential function has there a value of about 0.0001.) The normalization-factor N makes the result to "1" if all the x_n are constant and unity. Thus we have:

$$N = \sum_{|\nu_n - f_j| < k \cdot \Delta} exp\left\{-\pi \cdot \frac{(\nu_n - f_j)^2}{\Delta^2}\right\} \cdot d\nu_n$$

In case of very narrow spacing of the input data one has approximately: $N \approx \Delta$ for $d\nu_n \ll \Delta$. It should be mentioned, that near the two ends of the frequency band the sampling does not provide reasonable results, since the number of available data within the sampling interval shrinks. But this would happen with any resampling procedure.

In order to obtain some feeling of the characteristics of the resampling procedure, it may be worthwhile to investigate the resulting filter-curve, the effective resolution bandwidth and the fluctuation bandwidth as well. The new filter-curve $L(f)$ can be constructed by means of a measured filter-curve $l(f)$ of the initial data. The center-frequency of the original filter-curve we call f_0 (it should be identical with the frequency of a pixel of the initial data), f_c is the center-frequency of the resulting filter-curve, which does not necessarily coincide with the frequency of the new pixel, but should not deviate by more than half a pixel width from the center-frequency f_0 of the initial filter-curve. f is the frequency of the new resampled pixel. Then we have:

$$L(f) = \frac{1}{N} \cdot \sum_{|f_n - f_c| < p \cdot \Delta} exp\left\{-\pi \cdot \frac{(f_n - f_c)^2}{\Delta^2}\right\} \cdot l(f + f_n - f_0)$$

f_0 is now the frequency of the pixel next to the center of the Gauss-distribution. Depending on the distance, the resulting resolution bandwidth

$$\delta_{New} = \int L(f) \cdot df / L_{Max}$$

will vary slightly across the band for the new pixels. In order to keep this effect small the resampling width Δ should be at least twice the pixel frequency distance of the input data. The same is true for the new fluctuation bandwidth. A too small width is not advisable.

Noise at Resampling

Resampling is also a linear function of data, and the resulting data are most likely also correlated. We assume that the variance of all input data are the same and equal to σ_y^2. We had earlier for the error of arbitrary linear functions of correlated data (see Eq.[26])

$$\sigma_y^2 = \sigma_x^2 \cdot \left\{\sum_{n=1}^{N} p_n^2 + 2 \cdot g_1 \cdot \sum_{n=1}^{N-1} p_n \cdot p_{n+1} + 2 \cdot g_2 \cdot \sum_{n=1}^{N-2} p_n \cdot p_{n+2} + \ldots\right\}.$$

The function p_n is in this case:

$$p_n = \frac{1}{N} \cdot exp\left\{-\pi \cdot \frac{(\nu_n - f_j)^2}{\Delta^2}\right\} \cdot d\nu_n \quad \text{with}$$

$$N = \sum_{|\nu_n - f_j| < k \cdot \Delta} exp\left\{-\pi \cdot \frac{(\nu_n - f_j)^2}{\Delta^2}\right\} \cdot d\nu_n$$

The spacing $d\nu_n$ we can replace by the locally more or less constant frequency spacing d within the resampling interval $2 \cdot k \cdot \Delta$. With the approximation that the sum above can be replaced by the integral, as is allowed if Δ sufficiently large, we find: $N \cong \Delta$.

For the first term in the brackets we have:

$$\sum p_n^2 \sim d \cdot \int p^2(\nu - f_j) \cdot d\nu = \frac{d}{\sqrt{2} \cdot \Delta}$$

For the correlation terms we find:

$$\sum p_n \cdot p_{n+q} \;\sim\; \frac{d}{\sqrt{2}\cdot\Delta}\cdot exp\left\{-\pi\cdot\frac{(q\cdot d)^2}{2\cdot\Delta^2}\right\} \;=\; \frac{d}{\sqrt{2}\cdot\Delta}\cdot S_q^{½} \quad \text{with} \quad S_q \;=\; exp\left\{-\pi\cdot\frac{(q\cdot d)^2}{\Delta^2}\right\}$$

Thus we have:

$$\sigma_y^2 \;=\; \sigma_x^2 \cdot \left\{1 + g_1\cdot S_1^{½} + \; g_2\cdot S_2^{½} + \cdots\right\}\cdot\frac{d}{\sqrt{2}\cdot\Delta}.$$

The g_m are the values of the correlation between adjacent or more pixels. Again, the reduction of noise is strongly dependent on the correlation between the pixels of the input data. (See also above: “Fit of a Gaussian”, Eq.[29].)