FFT Use in NI DIAdem™

Contents

What You Always Wanted to Know About FFT...

The FFT (Fast Fourier Transformation) is a frequently used mathematical evaluation function, which transforms time-related oscillation data into the frequency range. With this function it is easy to determine to what extent the various frequencies are contained in a signal.

Unfortunately the FFT is not only used very often but, out of ignorance of the practical usage, also used very often incorrectly. However, the reason for this is not usually a lack of basic theoretical knowledge. The user is usually familiar with terms such as frequency range, amplitude, phase, sampling rate, and window function. The problem is usually how to apply this basic knowledge to achieve accurate practical results.

This document intends to answer the questions that are often posed in the practical application of the FFT and that usually remain unanswered because there is never enough time to deal with questions in detail. The mathematical background will only play a minor part. The primary focus will be on practical examples that demonstrate what the FFT does and how it behaves under different conditions. The document will show how to achieve results that are more accurate and/or more significant. In addition, the document will present methods with which to evaluate results and to detect errors early on.

FFT Basics

The classic FFT algorithm has the name Fast Fourier Transformation, because it calculates considerably faster than other Fourier Transformations. However the disadvantage is that the number of values must be a power of 2 (which means 8, 16, 32,..., 1024,..., 65536,...). DIAdem uses a new FT algorithm that is not only faster than the classic FFT but can also operate with any number of values.

Because the term FFT has established itself, this Fourier Transformation is also called FFT.

The FFT transforms time-related measurement data (oscillations) into the frequency domain. First one needs measurement data of a wave signal that is acquired over a period of time with a constant sampling rate. From this data the FFT calculates a frequency response which shows how strongly the various frequencies are contained in the signal.

For the following calculations it is important to know how the sampling rate, the measurement duration, the frequency domain, and the resolution are connected.

⇨ The frequency domain always ranges from zero to half the sampling rate. The half-value of the sampling rate itself is not contained in the frequency domain.

⇨ In the frequency domain one receives always half as many values as in the original time channel. However, one receives two results per frequency: Amplitude and phase.

Example:

If the sampling rate is 1000 Hz with 500 measurement values, one receives a frequency channel of 0 to nearly 500 Hz (498 Hz to be exact) with 250 values.

Increasing the sampling rate, results in a frequency channel with higher frequencies. A measurement with 2000 Hz and 500 measurement values results in 250 frequencies of 0 to 1000 Hz-- twice the frequency range as before.

If the measurement duration is increased while the sampling rate is maintained, one receives a frequency channel with more values. A measurement with 1000 Hz and 1000 values results in 500 frequencies of 0 to 1000 Hz-- twice the frequency resolution as before.

Consequently several principles determine the frequency analysis.

To inspect high frequencies, the sampling rate must be appropriately fast. For technical reasons the sampling rate must be at least 2.5 times higher than the highest frequency in the signal that you want to analyze.

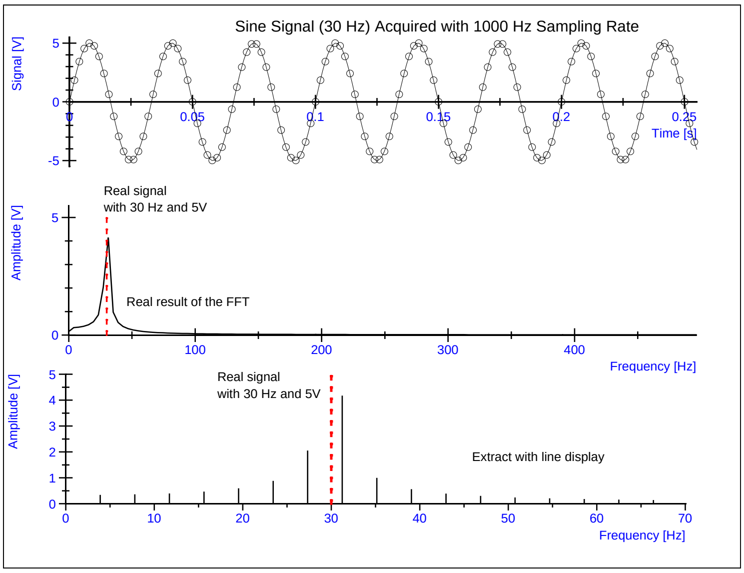
To resolve frequencies accurately the measurement must be accordingly long in duration. The necessary measurement duration (in seconds) must be equal to or exceed the inverse of the desired frequency resolution (in Herz).

A Simple Example

The following example analyzes a "clean" sine wave with an amplitude of 5 Volts and a frequency of 30 Hz. The signal is acquired with a sampling rate of 1000 Hz over a period of 0.255 seconds, acquiring in total 256 values.

The result should be an amplitude frequency response with a peak of 30 Hz and an amplitude of 5. Other frequencies should have amplitudes of exact zeros.

The following figure shows the actual result of this simple FFT.



At first sight the calculation result does not meet the expectations. Instead of one amplitude of 5 Volts at 30 Hz the result is a number of amplitudes, of which the greatest is 31.25 Hz and has a value of approximately 4.15 Volts.

A more accurate look at the figure reveals two details that are responsible for the deviation from the expected value.

➩ The time signal (top) contains slightly more than seven oscillations (approximately 7.6). The FFT algorithm on the other hand assumes periodic signals that can be repeated any number of times. This means the FFT rises at the end because it "thinks" that after the last value the signal starts over again at 0. The FFT does not "know" that in actual fact the sine curve was truncated in the middle of a wave.

➩ The extract with a line display (bottom) shows that the actual frequency of 30 Hz does not appear as a line in the result. Therefore the actual frequency can only be inferred from its nearest neighboring frequency lines. The frequency axis is divided into equidistantly spaced samples. If one counts the lines, the frequency 30 Hz is at about sample 7.6.

These two effects seem to be related. We will see later that the FFT returns exactly the theoretically expected line when the measured sine curve contains an exact integral number of periods. So, with some skill it is possible to find a sampling rate that calculates the expected result accurately. In practice however, when waves are measured, it does not make sense to adjust the sampling rate to the frequency of the already known sine.

Incidentally: If the same calculation is executed with sine curves that have slightly different frequencies or that do not start exactly at zero, the result is each time a differently formed peak and different amplitudes.

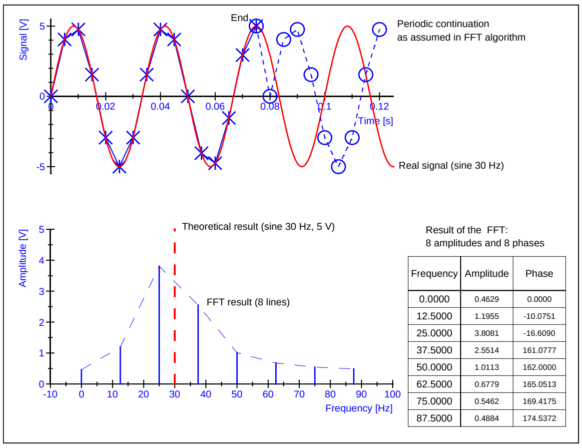
In principle, the superposition of all calculated FFT amplitudes is always less than or equal to the amplitude of the original wave signal (5 Volts in the case of the above example).

The sum of the amplitudes always differs slightly, because the truncated rests of the sine curve always have different sizes.

FFT under Scrutiny

To understand in more detail how FFT works, the above example is repeated with very few acquired data points. The sampling rate is reduced to 200 Hz and only 16 values are acquired.

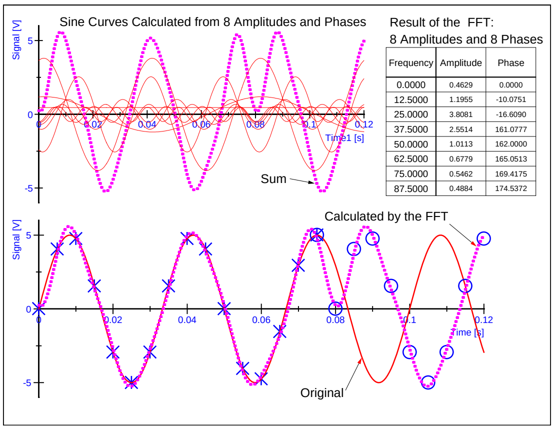
The time signal displayed below shows that the wave cannot be completely displayed with so few values. At the end of the acquisition the signal is truncated in the middle of the wave, just like in the original example. The FFT algorithm assumes that the signal is periodic and that after the last data point it starts again at the first data point. The dashed line with the circles below displays the waveform as assumed by the FFT.



Frequency	Amplitude	Phase
0.0000	0.4629	0.0000
12.5000	1.1955	-10.0751
25.0000	3.8081	-16.6090
37.5000	2.5514	161.0777
50.0000	1.0113	162.0000
62.5000	0.6779	165.0513
75.0000	0.5462	169.4175
87.5000	0.4884	174.5372

With 16 measurement values, the FFT returns the result 8 amplitudes and 8 phases. The following diagram shows the 8 amplitudes as vertical lines (spikes) at the associated frequencies. The dashed curve illustrates that the FFT results must not be connected with a continuous line. The 8 amplitudes indicate that the signal contains exactly these 8 individual components. The FFT does not reveal anything about the frequencies between the 8 lines.

One can calculate the resulting 8 sine curves from these 8 FFT components. If the sine curves are superposed, they create a waveform that nearly approximates the original input channel. To be exact, the wave below is a cosine not a sine wave.



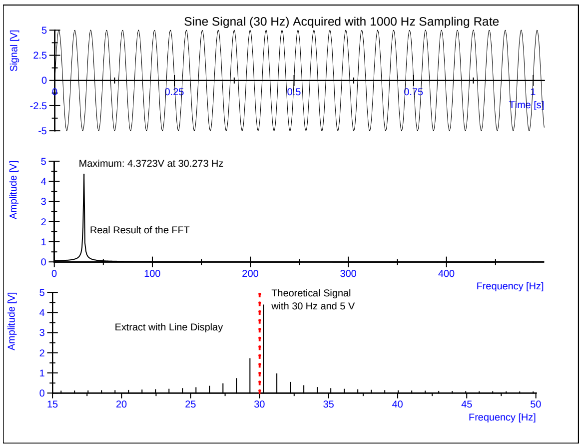
Frequency	Amplitude	Phase
0.0000	0.4629	0.0000
12.5000	1.1955	-10.0751
25.0000	3.8081	-16.6090
37.5000	2.5514	161.0777
50.0000	1.0113	162.0000
62.5000	0.6779	165.0513
75.0000	0.5462	169.4175
87.5000	0.4884	174.5372

This figure shows the 8 calculated solid curves at the top (one constant and 7 cosine curves with different frequencies), while the dotted line shows the summation of the 8 curves. In the bottom diagram one can see that the summation wave calculated by the FFT converges closely with the individual measurement values (identified by asterisks) of the original signal. After the last acquired data point, though, the calculated FFT signal is required to repeat periodically and deviates from the original signal. It is mathematically impossible to find 8 amplitudes and 8 phases that describe the curve better than the result of the FFT for the 8 frequencies that the sampling rate determines. Thus the FFT must be seen as the only correct solution for the 16 equations with 16 unknowns.

FFT with Many Interpolation Points

The main problem of the FFT is, that the FFT does not calculate a frequency that a measured signal contains, rather the FFT calculates for specified frequencies the part that the frequencies have in the total signal. The comparison of the two tests shown above shows that a higher number of measurement values normally leads to better results.

To test this assumption, the first measurement is repeated over a duration 4 times longer than the original measurement. The frequency response of 0 to 500 Hz is now divided much finer and the original frequency of 30 Hz can be inferred more accurately because the neighboring FFT lines are much closer. It is now easier to determine accurately which frequency range of the measured signal has the greatest amplitude.



The extract shows that one FFT frequency lies relatively close to the actual signal frequency. Even so, the maximum amplitude read out is approximately 12.5 percent smaller than the theoretically correct amplitude in the original signal. If the signal is exactly sinusoidal, this difference is not due to the number of measurement values but depends on whether the correct frequency coincides exactly with one of the frequency lines of the FFT.

This example shows that by increasing the number of values the results can be improved. However a remainder error always remains. The impact of the error depends on whether the "correct frequency" lies close to the frequency line by coincidence or whether it lies exactly between two calculated frequencies.

In practice signals often are not exactly sinusoidal and the wave frequencies change during the measurement.

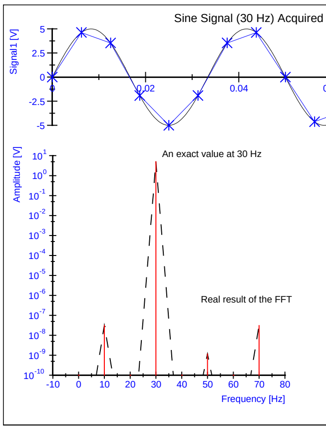
An Exact Result

One can only obtain an exact result if one adjusts the sampling rate exactly to the actual frequency. To do so, the sampling rate must be an integer multiple of the oscillation time. The following example will sample exactly 3 waves which will take exactly 0.1 seconds at 30 Hz. For 16 measurement values this results in a sampling rate of 160 Hz.

One should not forget though that in practice it is not possible to know the frequency beforehand. Normally the wave contains a mix of different frequencies and the FFT is implemented to detect which frequencies these are and how strongly each contributes.

The figure below shows the FFT of an exactly matched sine curve. The 16 sampled data points describe exactly 3 wave periods. Value 17 would lie at exactly 0.1 seconds and zero volt. From there the signal can be repeated periodically.

At 30 Hz an amplitude of 5 Volts was calculated. As the logarithmic display shows, all other amplitudes lie under 10^{-7}, which is practically zero considering the accuracy of the calculation.



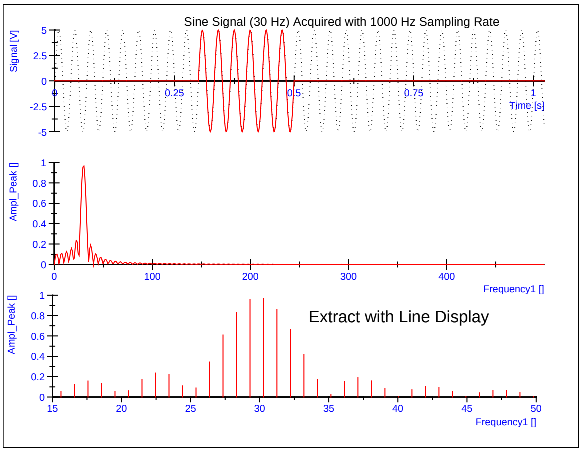
Frequency	Amplitude	Phase
0.0000	0.0000000000	0.0000
10.0000	0.0000000315	90.0000
20.0000	0.0000000000	0.0000
30.0000	5.0000000005	-90.0000
40.0000	0.0000000000	0.0000
50.0000	0.0000000012	-90.0000
60.0000	0.0000000000	0.0000
70.0000	0.0000000309	90.0000

The figure illustrates clearly that FFT results should be displayed as spikes. The dashed line is wrong, because only one amplitude at exactly 30 Hz is of any significance.

After this short outline of theoretically exact FFT, the next chapter will describe how to obtain more exact amplitudes in practice.

Transient Signals

Frequently a good resolution is required but the signal is only temporarily available. However, a high resolution requires a long measurement. One might want to execute a long measurement in any case.



The sine wave with the frequency 30 Hz is only available for approximately a fifth of the measurement time. The FFT now determines a high resolution but the exact frequency and amplitude are still not revealed.

Results cannot be improved by adding zeros. Adding short signals repeatedly to one another does not help either. This does not improve the result.

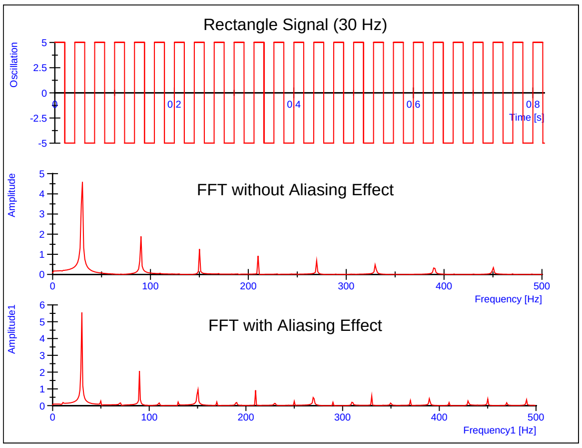
Typical FFT Results

Some FFT results for certain standard signals characterize the behavior of FFT functions. The results for sine curves have been described in detail.

The following examples will show some other typical results.

Square Wave

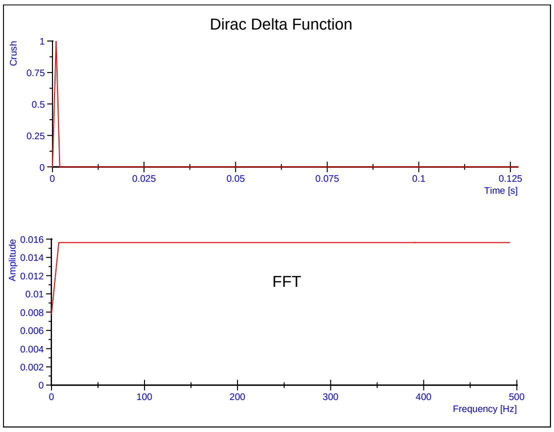
Square waves do not contain all frequencies. They contain the main frequency and multiples (3, 5, 7, ...) of this frequency with a decreasing amplitude.



Theoretically this leads to an infinite series. Therefore aliasing effects always occur with square waves. A later chapter will deal with this in more detail. In the middle diagram the effects are prevented through various measures (higher sampling rate, filter, reduction).

Dirac Delta Function

An important signal for the FFT is a theoretically infinitesimally short pulse.



This ideal pulse has an evenly distributed spectrum. In later examples we will see how a short impact (from a hammer) triggers an oscillation that contains all frequencies.

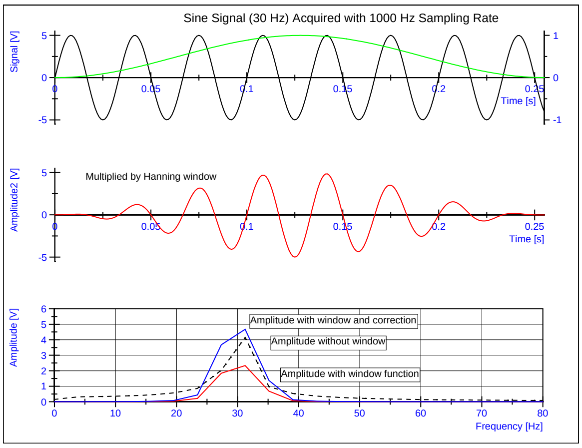
A typical example for such a pulse is the ringing of a tuning fork. The initial pulse contains all frequencies. After the initial pulse the tuning fork continues to oscillate with its own particular resonant frequency.

FFT Application

In practice the actual signal frequency is unknown. Normally the signal is not a pure sine signal but rather a mix of different signals. These signals are not totally sinusoidal and not distributed evenly over the measurement period. The fluctuations in the result of several measurements are usually greater than the system-related fluctuations that the FFT algorithm causes when it is used correctly.

Using the Window Function

The main problem of the FFT is that some frequencies do not fit exactly into the sampling time and are truncated in the middle of the wave (as discussed earlier in this document). This has the effect that the “correct” frequency cannot be calculated and that an abrupt jump occurs in the signal. However, jumps in the signal lead to a corruption of the amplitudes. A short pulse has no specific frequency but increases all frequencies evenly. A simple method against such jumps is the use of a window function.



This figure shows the simple sine curve from the first example. In the top axis system the Hanning window curve is plotted by which the raw data is multiplied. In the middle one can see the result. The measured curve is intentionally attenuated at the edges. Therefore there is no jump at the beginning of the next period.

The window function amplitudes are smaller than the raw data amplitudes because the calculation attenuated the signal. It is important that the amplitudes close to the original frequency of 30 Hz are only attenuated whereas all frequencies that are farther away are practically zero.

The average attenuation of the amplitudes can be calculated for each window. This enables an approximated correction of the amplitudes.

Because the window function does not have an impact on the calculated frequencies this method does not promise a “correct” calculation of the frequency either. The greatest corrected amplitude is a lot closer to the theoretical amplitude than is the case with the FFT without a window function (rectangular window function).

Two variations are available with the correction. To improve the accuracy of the greatest amplitude, periodic correction is the best method. In a Hanning window the factor is 2, because the amplitude of an exact sine is exactly half. If the frequency does not fit exactly, the error is about half as big as the error in a square window. However, the sum of the amplitudes that are necessary, for example, for sound measurements, is no longer correct.

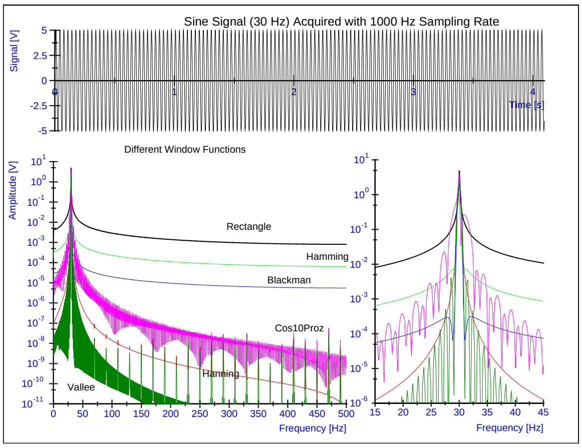
The random correction corrects all amplitudes “in the sum correctly”. The factor is 1.63299 (sqrt (8/3)). The sum of all amplitudes is considerably more accurate, because the error at the boundaries is completely eliminated.

Different Window Functions

DIAdem contains a number of different window functions that lead to different results. The question is which window function is the right one to use in each practical application. Every window function has its special features. In some cases the noise alongside the sine frequencies in the signal are attenuated better, in other cases the original signal is only slightly changed. The DIAdem Help and specialized literature on frequency analysis offer more information on this topic.

The different window functions provide different results without one result automatically being more correct than the others.

However, for many applications the behavior of the various window functions is only of academic interest, because the first two windows (square or Hanning) are usually completely sufficient. Nevertheless, the large number of window functions in DIAdem is justified. For example, one can try out different window functions and compare the results. This then quickly reveals which parts of the result are impacted by the algorithm and which are not. Additionally standardized applications require certain window functions because only they guarantee reproducible results. Thus an "optimal" window was constructed and only this window can provide the standardized results.



This figure shows a selection of the available window functions. The Hanning window achieves a very large attenuation that increases evenly over the entire frequency range. The Vallée window shows the largest attenuation but is very uneven. All other windows have their specific characteristics. The characteristics of the window functions can change considerably with different frequencies.

The bottom part of the graphic (Vallee window) illustrates disturbances that are due to the algorithm but occur through the inaccurate use of the floating point number calculation.

All windows with the exception of the exponential window focus on the center of the signal and attenuate the edges. Events that occur at the start or the end of signals which do not run evenly during the measurement are damped.

A typical example is a pulse that occurs at the beginning of the measurement and which leads to a decaying wave. In this case a Hanning window causes an error that cannot be corrected. Such a signal can only be processed by the exponential window, if at all.

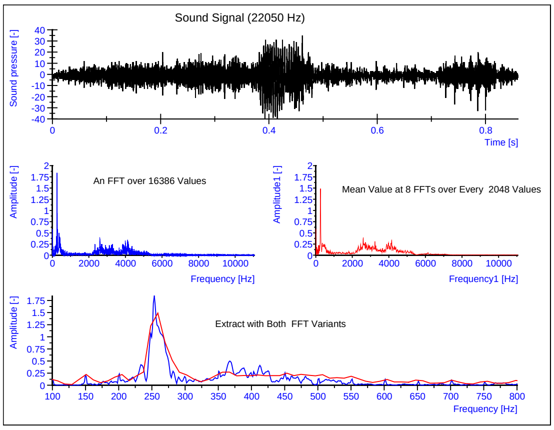
Time Intervals

In reality, measurements can never be repeated exactly. The measurement values of each run are always slightly different. As a result, the FFT results are also different for each run. For example, when measuring engine vibration slight engine speed changes cause the RPM frequency to be a line in the FFT spectrum one second and a second later to lie between two lines, which leads to differences in FFT amplitude of 10 or even 20 percent.

This is why it is best to execute several measurements or one especially long measurement, so that you can compare several partial FFT results with each other. Based on the differences of the various partial measurements, one can decide whether individual results are more important or whether averaging all results would lead to a better representation of the total vibration.

Averaging with Intervals

Averaging the FFTs from several measurements can improve the stability of the results. These multiple measurements can either be separate data set or multiple run events within one longer measurement.



This figure shows a sound signal of 19,000 data points acquired with a sampling rate of 22,100 Hz. Two FFT calculations with a Hanning window were executed with the data. The left hand side shows the result of a large FFT with 16386 values. The frequency resolution at 1.35 Hz is very high. The right hand side shows the averaged result of 8 FFT calculations with 2048 values each. The frequency resolution is 10.7 Hz. The characteristics of these two curves are roughly the same, however the amplitudes and the form of the peaks are very different as the depicted detail shows.

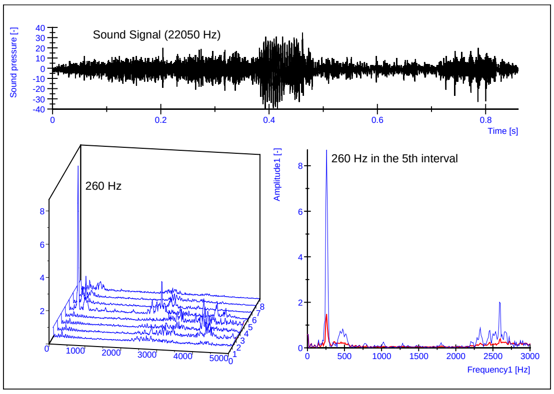
Because the number of values in the FFT must always be a power of 2, a part of the data is not used. A partitioning into several intervals leads to more values being used because one can get closer to the number of 19,000 values. It is also possible to calculate overlapping intervals and to integrate all values in the calculation; however some values are then used twice. The advantage of using overlapping intervals in window functions such as Hanning is that vibrations, which occur only briefly and are attenuated randomly, are included in the neighboring interval even stronger.

Appropriate application areas exist for all algorithms. For example, to inspect a sound wave to find out which of two sound sources that are close together holds the larger share of the total sound, an evaluation with as many values as possible is the right choice. In contrast, the above example shows one second of a human voice signal, which naturally does not contain any exact constant frequencies. The peak of about 260 Hz is random and can be at a different position in the next second.

If one looks at the time signal closely, one can see that it changes radically in the time period of just one second. A sound analysis can only either be an average of the total time whereby the frequencies cannot be exactly determined, or the exact frequencies are determined at a certain time point of the measurement. In both cases an FFT with as many values as possible does not make sense. The right approach for this signal is therefore the averaged FFT or a Third/Octave analysis.

Joint-Time-Frequency Analysis

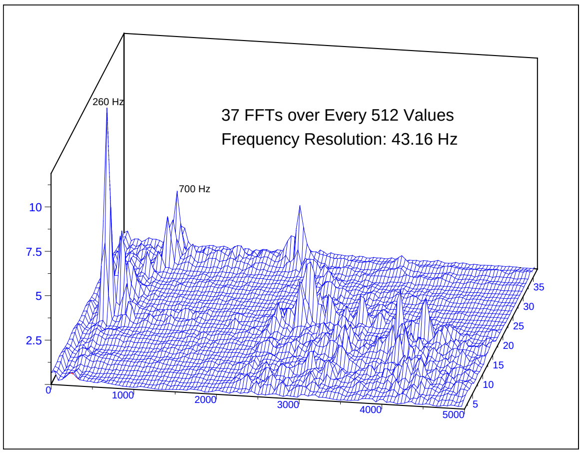
A sound signal seems to contain different amplitudes and frequencies at different times. To inspect this in detail the FFT is executed with intervals without averaging, sometimes called a Joint-Time-Frequency Analysis. IF you divide the above voice signal into 8 time segments and take the FFR of each, the results is 8 FFT curves—one for each time interval.



The calculation result shows that there are substantial differences between the individual FFT curves. The largest peak at 260 Hz, for example, only occurs briefly in interval 5 (in the time signal after approximately 0.4 seconds). Because the frequency occurs accurately only for a very short period, the amplitude peak is high and very narrow.

The signal can change considerably from one tenth of a second to the next. So the question is whether it might not be better to inspect shorter periods of time. The following figure shows an evaluation with 37 intervals with 512 values each.

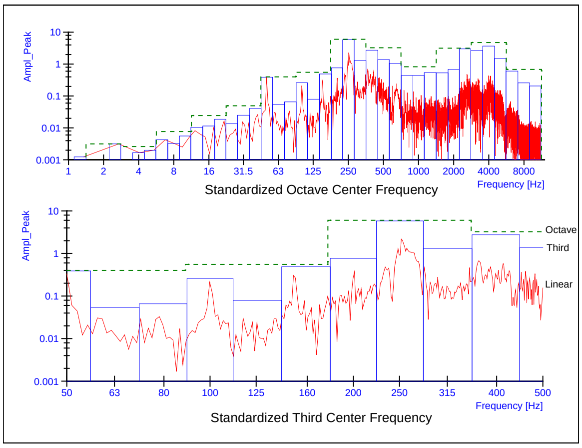
This illustrates a fundamental problem of the FFT. The more accurate one wants to determine the time point at which a certain frequency occurs, the more inaccurate the frequency resolution becomes. In this example the resolution is only 43 Hz. As already described, more accurate frequency resolutions require longer measurement intervals.



For each application a compromise between exact time resolution and exact frequency resolution must be found.

Third/Octave Analysis

The evaluation of sounds that can be heard by humans covers a very large frequency range-- from about 20 Hz to 20 kHz. The FFT provides equidistant (evenly distributed) frequency steps, which is impractical for such a large frequency range. For sound data it is not so much the exact signal frequency that is important but the volume in different frequency ranges. The total frequency domain is divided into standardized ranges. The octave analysis divides the frequencies into octaves (frequency is duplicated), the third analysis divides the frequencies into thirds (three thirds are octave). The following example shows the third/octave analysis together with the original amplitude after the FFT calculation.



The top axis system shows the total frequency domain of the signal. The logarithmic division of the frequency axis highlights the importance of low frequencies. The dashed line shows the octave analysis that summates all amplitudes that occur in the octave. 13 amplitudes, whose mean frequencies are at the frequency axis, represent the total frequency domain. The bottom axis system shows a detail. Here the mean frequencies of the third are at the frequency axis. The mean frequencies are standardized and enable the comparison of different measurements.

Amplitudes (Peak) are summed by adding the squares and taking the square root of the result. Each of the third/octave amplitudes shown above contains the sum of all the single amplitudes in its band. If the frequencies are very low, there are not enough FFT values and some thirds are not used.

If the FFT is executed in several intervals, the results differ. However, the amplitudes of the individual thirds and octaves remain approximately the same, because the sum of amplitudes in a frequency band remains stable regardless of the number. In the lower frequency domain it can happen quicker that the FFT provides no or only few amplitudes for the individual frequency bands. Therefore the result of the third/octave analysis improves when the calculation uses large intervals.

In each octave band or third band the sum of the individual amplitudes is established. Therefore the amplitude of the band is always greater than the greatest individual amplitude in that band.

FFT Functions

You can use a number of different FFT functions to evaluate the resulting amplitudes of the FFT calculation. The functions are explained in the following text:

The FFT algorithm first determines two channels which are named real part (R) and imaginary part (I) and which are not relevant here. You can determine the amplitudes and the phase shifting from the real part and the imaginary part.

The amplitude, which is the most important result of the simple FFT, is calculated by the square root (R^2+I^2). This result is called the peak amplitude. All previous examples showed peak amplitudes.

Apart from the peak amplitude, the RMS amplitude is also relevant. The RMS value is the “quadratic integral mean value“ of an oscillation. The RMS value can be calculated directly from the time signal. A typical application is the calculation of the RMS sound pressure from the measured sound vibration. In absolute sine waves the RMS value corresponds to the amplitude divided by the square root of two. This relation also applies to the FFT results because the FFT divides time signals into sine amplitudes.

The square of the peak amplitude is called the Autospectrum and is also relevant in some calculations. The same applies for the square of the RMS amplitude (RMS^2), which is also called the Power Spectrum.

A further calculation method is the PSD amplitude (Power-Spectrum-Density), in which the power spectrum is divided by the interval of the frequency lines.

Summary:

Peak amplitude:	$\sqrt{R^2 + I^2}$
RMS amplitude:	$\frac{\sqrt{R^2 + I^2}}{\sqrt{2}}$
Autospectrum:	$R^2 + I^2$
Power spectrum:	$RMS^2 = \frac{R^2 + I^2}{2}$
Power spectrum density:	$PSD = \frac{R^2 + I^2}{2 * \Delta F}$

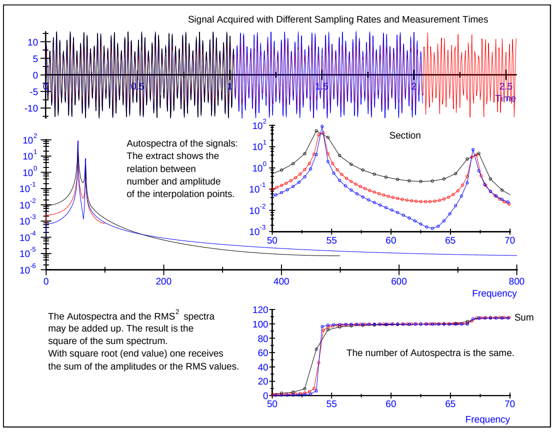
Why Autospectra?

A typical question is: Why do we need these different FFT functions?

In practice the different functions all have their relevance. It is important to remember which function was used, because otherwise the data might be incorrectly interpreted.

The following example shows a typical application with Autospectra.

The example analyzes a signal that consists of two sine frequencies. The results of three measurements with different sampling rates and measurement periods are compared.



At the top one can see the signal with the different measurement periods. The red measurement was executed for an especially long period, and the blue measurement was executed with an especially high sampling rate.

In the middle one can see very clearly that there are great differences especially next to and between the two signal frequencies. The quality of the results also depends on coincidence. The blue measurement is “lucky” and meets the actual frequency quite accurately. This makes the read out maximum amplitude more accurate.

In the lower part of the figure one can see that always the same sum results, regardless of the different amplitude development. However, the summation may only be executed with the two squared FFT functions, the power spectra, or the Autospectra, as shown here.

With sound data the spectrum shows a large number of peaks, but it is difficult to establish which frequency domains occupy the largest part in the total level.

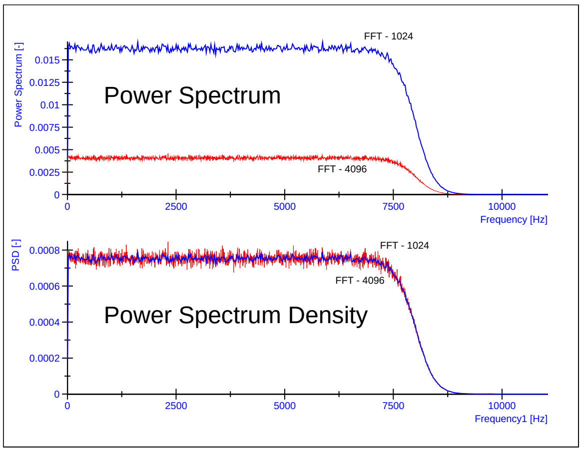
In this case the cumulative curve offers a very good overview of the various parts of the different ranges of a spectrum of the total level.

When does a PSD Make Sense?

The PSD function determines the power spectrum and divides the amplitudes by the interval of the frequency lines. The objective of this calculation is an improved comparability of spectra with evenly distributed frequency parts.

PSD is not suitable for signals with defined sine parts as was the case in the previous example for Autospectra.

PSD is always used for comparing measurements of even spectra and while preventing deviations through differing sampling rates and measurement periods.



This figure shows the averaged FFT results of an even noise signal with a lowpass of 8000 Hz. FFT calculations with 1024 values (blue) are contrasted with FFT calculations with 4096 values (red).

The power spectra in the top axis system show different amplitudes. The sum of all amplitudes is always the same; therefore the individual amplitudes must be smaller in the high resolution calculation. The blue curve combines, for example, always four frequencies of the red curve and therefore the amplitudes are four times as high.

PSD divides the amplitudes by the interval of the frequency lines. Both PSD curves have approximately the same dimension, because the interval of the blue curve is about four times as large.

PSD spectra are not summed but integrated, in order to retain the cumulative curve. The surface below the curve always corresponds with the total level regardless of the number of interpolation points.

FFT with Two Time Signals

In the previous examples, the FFT was inspected by a single independent time signal. An oscillation was monitored without considering what caused this oscillation.

In many technical areas though, for example, for mechanical vibrations, it is more informative to determine the response behavior to a known stimulation by observing cause and effect.

In this context the transfer function which determines the relation of output to input is an important analysis function.

To put it simply, a vibration is inserted into one side of the system to be analyzed, and the vibration result at the other side is then measured. The relation of output to input should always be the same, even if the input vibration is not always the same.

The transfer function provides exact information on which frequencies decrease the stimulation and which increase the stimulation. Resonance effects often produce especially strong vibrations out of small stimulations.

Transfer Function

Two signals are measured for transfer functions. These two signals are the stimulation and the response of the system under test. Both signals must be acquired simultaneously.

Oscillations usually do not occur on their own accord. Oscillations must be triggered by a stimulation. For exact analyses one must know what triggered the oscillation.

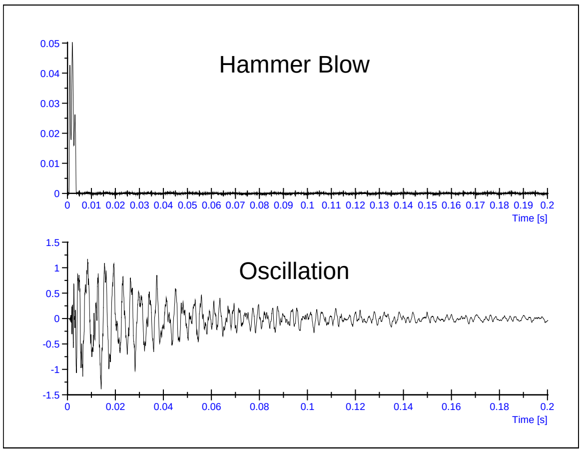
The stimulation must be suited for the analysis. Stimulations should trigger a system evenly in the analysis of the frequency domain.

Sinusoidal stimulations are usually not suited because they only contain one frequency. If a mechanical structure is stimulated with a single frequency, the structure will vibrate with exactly this one frequency.

There are two typical types of stimulations.

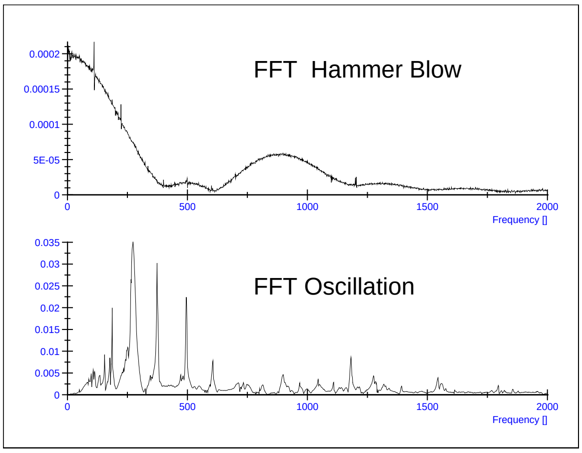
Noise type stimulations contain all frequencies nearly equal. The additional advantage is that the frequencies remain equal for a longer period of time. However, the technical preparations for noise stimulations are very complex.

Abrupt stimulations also contain all frequencies. In practice, for example, a hammer executes a short, dry blow on a structure and the vibration that is triggered by the blow is measured. The acceleration on the hammer and on significant spots on the mechanical structure is measured. The size of the hammer must correspond with the unit under test in order to execute normal vibration measurements without disturbances.



Stimulations with a hammer blow are easy to execute. After the blow though the decaying oscillation can only be measured for a short period of time. The time span should be sufficient for all relevant oscillations.

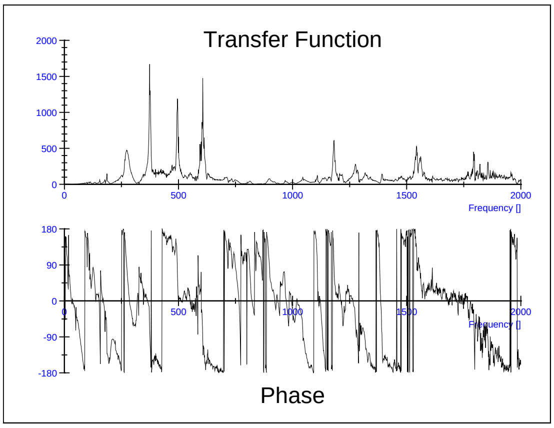
The following figure shows the individual FFT analyses for the two signals.



An ideal blow with a hammer would have stimulated all frequencies completely evenly. In practice this does not succeed. However, the figure illustrates that all frequencies occur in a wide area and that the frequencies are also contained in the wave signal.

The FFT of the oscillation shows at which frequencies the structure oscillates more and at which less. The greatest amplitude at approximately 260 Hz is only that large because the stimulation there is so strong. The resonance frequencies with a weak stimulation, for example 600 Hz, would be considerably higher if the stimulation were ideal.

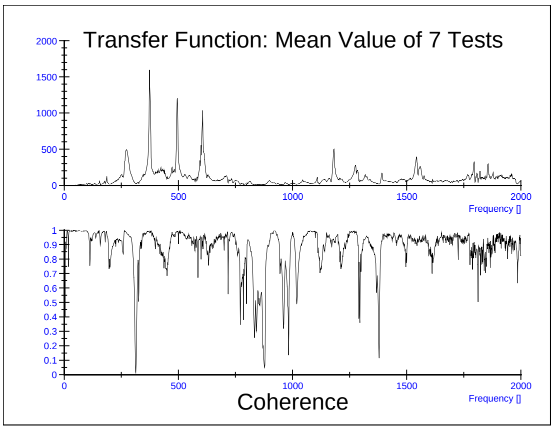
The transfer function calculates amplitudes in proportion to the respective stimulation. Phase shifting can interpret results further.



For the results shown here it not yet clear whether the measured vibration was caused by the stimulation. If the stimulation is a hammer blow, it is easy to evaluate due to the time signals. If the stimulation is a noise stimulation, it is not so easy. In this case the coherent function helps.

The coherent function requires several time signals that are measured consecutively. A hammer blow triggers the measurement (analog trigger). If the stimulation is triggered by noise, the measurement can start at any time.

If you compare the measurement, you receive different results. If, for one frequency, all amplitudes and phases of the transfer function correspond exactly, the coherence is one. If the concordance is very bad, the coherence is approximately zero. Therefore the coherence function shows very well which frequency domains of the transfer function show usable results. It only makes sense to evaluate frequencies that provide reproducible transfer values.



This figure shows the averaged results from seven tests. The coherence shows very different qualities in the different frequencies.

If the frequencies have high amplitudes and thereby a strong transfer rate, the coherence is usually very good. Frequencies with a low transfer rate very often do not have a good coherence, because random disturbances prevail. If nothing is transferred, the frequency and the phase cannot coincide.

If a bad coherence occurs at frequency peaks, it might be that a vibration is being measured that was triggered by disturbances and not by the hammer blow. Therefore vibrations should be measured in idle mode and not with a running engine.

Acquisition of Oscillation Signals

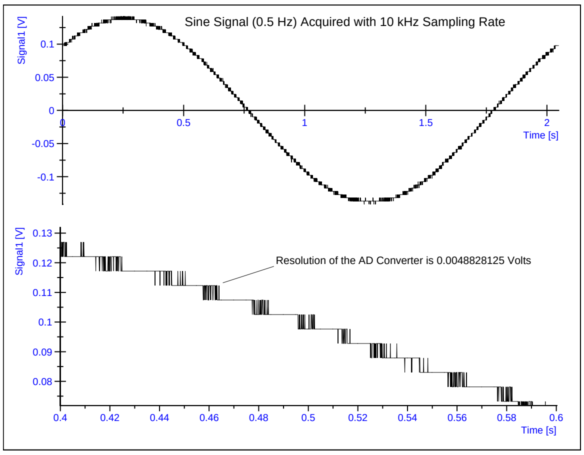
The acquisition of oscillation signals determines the quality of the frequency analysis. Important factors in the acquisition are the precision of the measurement equipment (recorder, measurement amplifier, AD convertor), the sampling rate, and the duration of the measurement.

General Information on the Acquisition of Measurement Data

Precision of the AD convertor

Every AD convertor has a defined precision which is determined by the voltage range and the resolution. Often AD convertors have a resolution of 12 bit and a voltage range of -10 to +10V. If the resolution is 12 bit, the voltage range is divided into 4096 steps of each 0.0048828125V. Usually the range does not include the +10V value.

Measured signals that do not use the complete voltage range are not as precise. The following example shows what results can look like if the voltage range is not well used.



The vibration ranges from approximately -0.15 to +0.15 volt. Only about 60 (approximately 1.5 percent) of the 4096 different voltages that the AD convertor can measure are used.

Warning:
The steps always exist for A/D convertor data but are not always so obvious.

The voltage steps are especially conspicuous due to the high sampling rate. With a lower sampling rate (for example 10 Hz) the steps would no longer be visible but the inaccuracy would still exist. If the sampling rate is low, the AD convertor does not specify the mean value of a specific period of time, but picks out the current value from a signal at a point of time. In the transition range between two steps it is thus coincidence which step is measured, regardless of the precision of the rest of the measurement chain.

Further errors occur on the analog side between the acquisition of the signal and the actual digitalization. Thus it is not guaranteed that the AD convertor always determines the "correct" step.

The measurement result provides further information:

⇨ In the transition range the values jump between two steps back and forth. Therefore it is possible to use the smoothing function to calculate a curve that seems to have a higher resolution. This only makes sense if the measurement error is small enough.

⇨ The high sampling rate prevents a meaningful calculation of the signal increase. In the top signal the increase is either zero, one step up, and one step down. To calculate increases accurately, the values must be compared in intervals that cover many steps. Therefore the increase in measurement data converted by the AD convertor is only rarely possible.

⇨ The precision of the AD convertor can be substantially improved if a measurement amplifier or an AD convertor with amplification is used. The signal can now be adjusted to the AD convertor and the measurement range can be utilized better.

⇨ AD convertors with a 16 bit resolution can divide the measurement range into 65,536 steps (16 times more than the 12 bit convertor).

Special measurement devices often can measure substantially more precisely. The precision is then specified in decimal places and not in bit resolution. 12 bit resolution corresponds to 3½ decimal places, 16 bit resolution corresponds to 4½ decimal places. Some devices can measure, for example, 8½ decimal places precisely. To prevent disturbing influences from the computer, these AD convertors are often located in a separate chassis. Often a time period is integrated and not just instantaneously measured. Often the integration lasts exactly 0.02 seconds (a 50 Hz net cycle) to average out net disturbances. Higher sampling rates are then no longer possible.

The technical data sheet of measurement devices list the expected measurement precision and the possible errors that can occur during measurements.

Calibration of the Measurement Signals

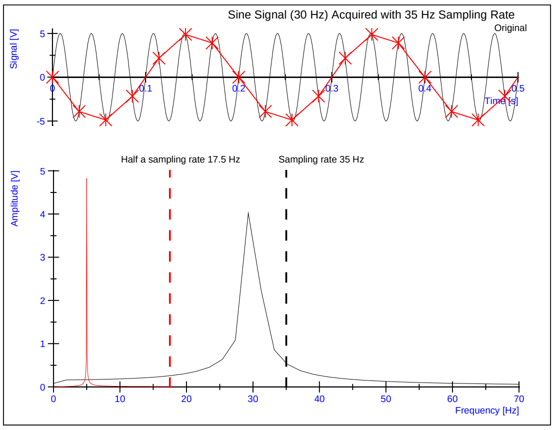
Due to technical reasons, you cannot prevent the AD convertor from being imprecise. Therefore it is important to estimate the possible errors and to ensure that the precision meets the requirements.

First, you should specify the precision you need for your configuration. The specification determines how precise the AD convertor must measure and which measurement amplification is necessary.

After this, a test measurement of known values can determine whether the precision was really achieved.

The Aliasing Effect

The aliasing effect poses a big problem when acquiring oscillation signals. The aliasing effect is sign of erroneous measurements due to too low sampling rates in the digitization. The effect occurs when a vibration, which has a frequency greater than half of the sampling rate, is contained in the signal. The following example measures the 30 Hz vibrations from previous sections in this article with a sampling rate of only 35 Hz. The red line in the top axis system shows the vibration which is simulated digitally by the low sampling rate. The frequency of the vibration is obviously incorrect.



The FFT of both measurements illustrates that the actual frequency is shifted due to the incorrect sampling rate. If the sampling rate is 35 Hz, only frequencies up to 17.5 Hz can be displayed correctly. The signal frequencies above this limit do not simply disappear, but emerge somewhere in the range between 0 and 17.5 Hz. A simple rule is: Higher frequencies are mirrored at the area boundary.

The signal frequency (30 Hz) is 12.5 Hz above the boundary of 17.5 Hz and therefore appears 12.5 Hz below the boundary at 5 Hz. If the signal frequency is exactly at the sampling rate, the measurement values appear as with a constant (0 Hz). Even higher frequencies (35 Hz to 52.5 Hz) rise to 17.5 Hz an. This convolution can be continued infinitely. If the recorder (for example, a microphone) is capable to do so, the signal can contain a vibration of 3502 Hz which reappears at 2 Hz.

After digitization it is impossible to distinguish real frequencies from the wrong frequencies that were generated by the aliasing effect. Therefore one must ensure before digitization that no frequencies above half the sampling rate are contained in the analog signal.

If an FFT determines large amplitudes before reaching half the sampling rate, signals must be just above this frequency. This means that you do not know whether the measured amplitudes are right or wrong. Even if the analog filters are very efficient the top quarter of the FFT should be used with care.

There are several possibilities to prevent aliasing effects. These options do not include digital filters because digital filters are not applied before the digitization. By then the aliasing effect is irretrievably contained in the signal.

A popular disturbance frequency is the frequency 50 Hz or 60 Hz of the power network (wall socket and/or fluorescent lights), which influences the measured signal in different ways. If the sampling rates are low (nearly always a fraction of 50 or 60), a random offset or a very low beat frequency is measured in the signal. If the sampling rate is 100 Hz, 2 values per 50 Hz or 60 Hz vibration are measured. Depending on the start of the measurement, the values are at the zero crossing and do not create a disturbance. However, if the values are at the antinode, a noisy effect is the result. If one measures with 100 Hz or 120 Hz and observes differing degrees of noise at different starts of the measurement, one should inspect this with a higher sampling rate. The power network emerges as an FFT peak at exactly 50 Hz or 60 Hz.

Lowpass Filter

To eliminate higher frequencies directly in the analog signal, lowpass filters are used. These filters are therefore called anti-aliasing filters. Thereby all frequency components above a certain limit frequency are filtered out quite well. This method can also cause problems if different measurement tasks must be constantly solved with different frequency ranges. Lowpass filters which are freely programmable are quite expensive and not always available. Therefore one has to resort to a number of fixed filters and the related sampling rates.

It is also possible to use signal recorders that have measurement amplifiers with integrated lowpass filters. The device or the respective manual usually indicate the frequency domain to use for the calculation. Often several frequency domains can be selected. If a vibration recorder has, for example, a frequency domain of 1 Hz to 1 KHz, higher frequencies cannot be measured. However, a sampling rate of 2 KHz does not incur an aliasing effect.

Note finally that If the sampling rate is high enough, aliasing effects are prevented. It is worth scrutinizing the frequency domains of recorders and filters.

Warning:

There is a typical error that should be avoided. If the recorder provides frequencies up to 1000 Hz, it is wrong to measure with a 1000 Hz sampling rate. The rate must be at least 2500 Hz.

Oversampling

Modern measurement devices, AD convertors, and computers have so much capacity that the optimal sampling rates and exact lowpass filters can be disregarded. To measure a vibration at approximately 1 Hz with a transducer that has a frequency domain of up to 100 Hz the sampling rate can be 250 Hz. The signal does not contain aliasing effects anymore but a lot of data.

A data reduction leads to the required result. To do so, one can use a digital filter. However one can also use a FFT over all data to extract the required area from the frequency domain (0–100 Hz).

Determining the Required Sampling Rate

The necessary sampling rate must be 2.5 times the highest important frequency. The highest important frequency might be the monitored vibration or a frequency in the signal that might be causing disturbances. If necessary, one should test whether a problem has occurred with 50 Hz disturbances (power network).

Determining Special Sampling Rates

A large number of all measurements is carried out with integer sampling rates (100, 200, 500 1000, ...). As the number of values for a FFT must be a power of 2, the measurement duration is often 'warped'. The FFT frequencies are then also at the 'warped' values because the frequency interval is calculated with one divided by the measurement duration.

For many evaluations it would be better to determine integer frequencies through the FFT. A search function such as the peak search in DIAdem, which can search for frequencies at the greatest amplitude, then provides, for example, 5 Hz and not 0.46875 Hz. A sampling rate with a power of 2 achieves this. If the sampling rate is 512 Hz and the measurement duration is precisely one second, the FFT after the frequencies is 0 to 255 in steps of exactly one Hz.

Now the special sampling rates are determined with the required frequency resolution. One divided by frequency resolution is the required measurement duration. The measurement duration must now be divided by a power of 2 samples. This produces the sampling rate which must correspond with the minimum required sampling rate.

Example:

An FFT result with a resolution of 0.1 Hz and a maximum signal frequency of 250 Hz is required. The signal frequency produces a required sampling rate of at least 625 Hz (2.5 times signal frequency).

The measurement duration is measured with 1/0.1 and must be precisely 10 seconds.

The correct sampling rate must be determined in order to measure for 10 seconds at a minimum of 625 Hz. 4096 measurement values divided 4096/10 produce a sampling rate of 409.6 Hz. This is not fast enough. Therefore one must acquire 8192 values in 10 seconds which corresponds to a sampling rate of 819.2 Hz.

Unfortunately, such sampling rates are not possible on every hardware.

Several Channels Simultaneously

When several channels are measured, it is not always guaranteed that the values of the various channels are acquired simultaneously.

Depending on the hardware there are substantial differences.

Simple Multiplexer

Many multi-functional boards contain an A/D convertor that can convert only one channel at a time. A multiplexer charges the A/D convertor consecutively with the voltages of the required channels that are converted consecutively.

For example, to measure 4 channels with 1000 Hz each, the A/D convertor uses a total sampling rate of 4000 Hz. The channels are sampled with an interval of 0.25 milliseconds. This means that the fourth channel is sampled with a delay of 0.75 milliseconds.

In frequency tests such as transfer functions such a delay leads to errors in the phase calculation.

Scan Frequency and Sample Frequency

Some multi-function boards provide different timers for the scan frequency and the sample frequency. The scan, which is the list of the required channels, is acquired with the specified sampling rate. Within the scan the multiplexer changes with the maximum rate from channel to channel. The time difference between the channels is thus minimized.

For example, a board with a maximum sampling rate of 100 KHz would enable an interval of 0.01 milliseconds from channel to channel. In this case the values are still not measured simultaneously, but nearly.

For high sampling rates which nearly reach the maximum board rate, this procedure is not a solution.

Sample/hold

To measure all channels simultaneously, boards with a Sample&Hold function are sometimes used.

Each board has a special measurement amplifier integrated before the multiplexer. Before the scan acquisition, these measurement amplifiers receive simultaneously a signal that switches the acquisition to the track mode. This mode 'freezes' the current voltage. Then the amplifiers are switched to the hold mode and consecutively sampled by the multiplexer.

In the sample&hold mode the channels are consecutively converted from analog to digital, however the voltage is the same voltage that was acquired and held in the track mode at the beginning of the scan.

A/D convertor boards with a sample&hold function can execute simultaneous measurements of several channels without phase shifting. This function is elaborate and therefore the boards are expensive.

Several A/D Converters

A further possibility to guarantee complete simultaneity is the usage of boards or devices with several A/D convertors.

Devices with one A/D convertor per channel produce simultaneously measured channels.

Some devices have two or four A/D convertors which measure several channels simultaneously. In practice this is often sufficient.

Comparing Results with FFT Analyzers

To ensure that the measurement results are correct and reproducible, one can compare different acquisition systems with each other. To do so, one does not use a random signal but, for example, a sine generator that can provide a signal which can be repeated any number of times. The results of computer-based measurement acquisition such as DIAdem and the results of an FFT analyzer often differ.

The reasons for the deviations between the various systems are not so much the different quality of A/D convertors or analog technology, but some of the already described details. Especially with clear sine signals of signal generators the A/D convertors and filters do not have to be one hundred percent precise. It is more important that the sampling rates and the measurement duration in both systems are simultaneous. Additionally the window function and the correction parameters should be the same.

The internal operation mode in FFT analyzers is not discernible at first sight. To reproduce the results one must read the documentation or study the display in detail.

FFT analyzers often have integer frequencies that were acquired with 'warped' sampling rates. Analyzers also often provide a number of even-numbered frequencies (for example, 400 not 512 lines). This is not achieved with a different FFT algorithm. Instead the top 112 calculated frequencies are not displayed. This makes sense, because the precision for these frequencies is low. It is also possible that the FFT analyzers only display the lower half of the calculated frequencies because they operate internally with twice the sampling rate. The advantage is that the analog lowpass filters have a large scope.

As a rule, analyzer with multiple channels measure inputs simultaneously without phase shifting.

In comparison to normal computer data acquisition technology, the FFT analyzers comprise programmable analog lowpass filters which usually prevent aliasing effects automatically. If one compares random distributed vibration signals and not pure sine signals, one must prevent the aliasing effects manually when one uses normal computer data acquisition technology.

External Clocks in Rotation Speed Dependent Vibrations

The FFT only provides precise results when the time signal is monitored for a longer period of time. In this period of time the signal must stay constant. If the vibrations are rotation speed dependent, one must monitor a larger number of revolutions to precisely analyze the frequencies and amplitudes that are produced by unbalance.

In this case the results are more correct if the data is measured over the angle of rotation and not over time. For example, if a quadrature encoder with 32 pulses per revolution is mounted on a rotating system, the measurement can run with an external clock. Now, precisely 32 values per revolution are acquired regardless of the rpm (revs per minute, revolutions per minute). All rotation speed dependent vibrations are contained in the input data for the FFT as integer numbers (for example 16 in 512 values). As shown above, vibrations that are contained as integer numbers in the signal are represented exactly by the FFT. The external clock with a pulse number of a power of 2 can analyze rotation speed dependent vibrations precisely even if the rotational speed changes.

In this case the sampling rate changes proportional to the rotational speed. After the FFT, one therefore receives amplitudes over the order not over time. In this case, the vibrations that correspond with the rotational speed are at the first order; vibrations that correspond with twice the rotational speed are at the second order and so on.

In machines, vibrations often occur that correspond with exactly the rotational speed or a multiple of the rotational speed. These vibrations are especially strong when they meet the resonance frequencies of the machine.

Determining the Parameters for External Clocks

The relationship between revolutions, resolutions and achieved order areas is analogous to the described relation between time-related data.

The order range stretches from 0 to half the number of measurement values per revolution.

The resolution of the order channel corresponds to the reciprocal number of the evaluated revolutions.

Example:

In a measurement with 64 measurement values per revolution and 1024 measurement values, one receives results up to order 16 which are divided into steps of 0.03125 orders. To evaluate higher orders, more measurement values per revolution must be measured. The number of measured revolutions affects the resolution of the orders.

In machines and motors the first order lies in the range of approximately 3,000 rpm or 50 Hz. Higher orders are necessary because many resonance, vibration, or noise problems have considerably higher frequencies.

In this context the order analysis is an additional topic. The order analysis tests rotational speed dependent vibrations. The aim is to analyze according to orders machine vibrations in different states (to power up, to shut down, different operational states).

Special Topic: Acoustic Measurement

Acoustic measurements determine the volume and the frequencies contained in noise signal. The noise range audible for humans is of special interest. Acoustic measurements have some special features that can pose unusual problems compared to other typical measurements.

- ⇨ The audible frequency range of sound lies between 16 and 20,000 Hz.
- ⇨ The audible volume (acoustic pressure) lies between 10^{-4} and 10^{2} Pascal (N/m^2).
- ⇨ To evaluate sound sources one needs a reproducible absolute precision.

The enormous measurement ranges in acoustic measurements require special measures that are not necessary for other physical dimensions to this extent. The frequency range requires high sampling rates and a large number of data to receive a sufficient resolution in the lower frequency domains. The measurement range with the factor 10^6 between extremely loud and barely audible cannot be handled with normal A/D convertors. Therefore one needs a special measurement amplifier that amplifies the microphone signal so that it can resolve with a normal A/D convertor.

The high precision in the measurement of sound amplitudes is also unusual. Normally in vibrations, only the relative differences of the amplitudes between the various vibration signals and the different frequencies are important. In acoustic measurements slight changes to the volume are of crucial importance (for example when implementing noise protection measures).

Sophisticated measurement tasks require sophisticated measurement technology. In the field of acoustic measurements several companies specializing in high-precision sound measurement technology have established themselves.

- ⇨ Microphones with linear frequency responses at different volumes and clearly defined recording angle.
- ⇨ Measurement amplifiers with a large measurement range (for example, amplification 8x10 dB)
- ⇨ Measurement amplifiers with evaluation filters important for sound evaluation (for example, A-weighting)
- ⇨ Measurement amplifiers with direct measurement of the volume in dB or dB(A)
- ⇨ Recorder with high playing precision (even speed and volume)
- ⇨ Calibration devices to test the measurement chain

Exact sound measurements can only be carried out with high-quality measurement technology. Precise quantitative statements cannot be derived for sound measurements executed with normal microphones and sound boards on the computer. The acoustic signal used in the example above was measured with sound board and can only be used for a qualitative evaluation of the frequencies but not for precise statements on the volume.

Acoustic Signals

Usually acoustic signals are measured with a microphone as acoustic pressure in Pascal (N/m^2). The acoustic signal is an oscillation that overlays the normal air pressure. Therefore it does not make sense to specify the acoustic pressure at a specific time. The frequency and the oscillation amplitudes are of interest.

A relatively simple definition of the acoustic signal is possible without FFT. The calculation of the root-mean-square of the acoustic pressure provides the first characteristic value of the volume of the signal.

First all individual acoustic pressures are squared in order to determine the root-mean-square. The square root is extracted from the mean value of all squares (sum divided by the number). The abbreviation RMS stands for root-mean-square. In a pure sine vibration the rms value is equal to the amplitude of the vibration divided by the square root of 2.

It is impractical to specify the acoustic pressure in N/m^2 because the acoustic pressures occur in different dimensions depending on the situation. Therefore a logarithmic scale is used to specify the acoustic pressure. To receive manageable dimensions the acoustic pressure is referred to a base size. This base size is an acoustic pressure of $2*10^{-5}$ N/m^2 for airborne noise which is a barely audible acoustic signal. The calculation is:

$$L = 20 * \log \frac{\text{acoustic pressure}}{base\ size}$$

with *L*=Noise pressure level

The value defined as noise pressure level is specified in the unit dB (decibel).

Acoustic signals with an equal acoustic pressure but different frequencies are received differently by the human ear. A sine tone of 200 Hz, for example, is not considered as loud as a sine tone of 2000 Hz with the same noise level. Several methods can adjust the rms values to those heard by the human ear. The most common method is the A-weighting. The various frequencies have correction values in dB, which are added or subtracted from the respective rms values that are measured. For example, 10.9 dB are subtracted from the sine at 200 Hz and 1.2 dB are added at the sine at 2000 Hz. The noise pressure levels that are evaluated with the A-weighting are then specified in dB(A).

At 1000 Hz the correction value is exactly zero. Therefore a signal with 1000 Hz and an acoustic pressure (RMS) of N/m^2 is usually used for calibration measurements. The acoustic pressure corresponds with 93.9794 dB or dB(A).

The signal can be filtered with a combination of highpass and lowpass filters for the A-weighting. Such analog filters are often integrated in the measurement amplifiers. It is also possible to use digital filters after measuring data.

A further option is to separate the signal into individual frequencies with the FFT. The frequencies can use smoothing functions with the respective correction values. The conversion is often executed after third and octave analyses because tables for the evaluation are standardized for thirds and octaves.