WHAT IS AN FFT SPECTRUM ANALYZER?

The SR760 FFT Spectrum Analyzer takes a time varying input signal, like you would see on an oscilloscope trace, and computes its frequency spectrum.

Fourier's basic theorem states that any waveform in the time domain can be represented by the weighted sum of pure sine waves of all frequencies. If the signal in the time domain (as viewed on an oscilloscope) is periodic, then its spectrum is probably dominated by a single frequency component. What the spectrum analyzer does is represent the time domain signal by its component frequencies.

Why look at a signal's spectrum?
For one thing, some measurements which are very hard in the time domain are very easy in the frequency domain. Take harmonic distortion. It's hard to quantify the distortion by looking at a good sine wave output from a function generator on an oscilloscope. When the same signal is displayed on a spectrum analyzer, the harmonic frequencies and amplitudes are displayed with amazing clarity. Another example is noise analysis. Looking at an amplifier's output noise on an oscilloscope basically measures just the total noise amplitude. On a spectrum analyzer, the noise as a function of frequency is displayed. It may be that the amplifier has a problem only over certain frequency ranges. In the time domain it would be very hard to tell.

Many of these types of measurements used to be done using analog spectrum analyzers. In simple terms, an analog filter was used to isolate frequencies of interest. The remaining signal power was measured to determine the signal strength in certain frequency bands. By tuning the filters and repeating the measurements, a reasonable spectrum could be obtained.

The FFT Analyzer
An FFT spectrum analyzer works in an entirely different way. The input signal is digitized at a high sampling rate, similar to a digitizing oscilloscope. Nyquist's theorem says that as long as the sampling rate is greater than twice the highest frequency component of the signal, then the sampled data will accurately represent the input signal. In the SR760, sampling occurs at 256 kHz.

To make sure that Nyquist's theorem is satisfied, the input signal passes through an analog filter which attenuates all frequency components above128 kHz by 90 dB. This is the anti-aliasing filter. The resulting digital time record is then mathematically transformed into a frequency spectrum using an algorithm known as the Fast Fourier Transform or FFT. The FFT is simply a clever set of operations which implements Fourier's basic theorem. The resulting spectrum shows the frequency components of the input signal.

Now here's the interesting part. The original digital time record comes from discrete samples taken at the sampling rate. The corresponding FFT yields a spectrum with discrete frequency samples. In fact, the spectrum has half as many frequency points as there are time points. (Remember Nyquist's theorem). Suppose that you take 1024 samples at 256 kHz. It takes 4 ms to take this time record. The FFT of this record yields 512 frequency points, but over what frequency range? The highest frequency will be determined by the period of 2 time samples or 128 kHz. The lowest frequency is just the period of the entire record or 1/(4 ms) or 250 Hz. Everything below 250 Hz is considered to be dc. The output spectrum thus represents the frequency range from dc to 128 kHz with points every 250 Hz.

Advantages and limitations
The advantage of this technique is its speed. The entire spectrum takes only 4 ms to measure. The limitation of this measurement is its resolution. Because the time record is only 4 ms long, the frequency resolution is only 250 Hz. Suppose the signal has a frequency component at 260 Hz. The FFT spectrum will detect this signal but place part of it in the 250 Hz point and part in the 500 Hz point. One way to measure this signal accurately is to take a time record that is 1/260 or 3.846 ms long with 1024 evenly spaced samples. Then the signal would land all in one frequency bin. But this would require changing the sampling rate based upon the signal (which you haven't measured yet). Not a good solution. In fact, the way to measure the signal accurately is to lengthen the time record and change the span of the spectrum.

FREQUENCY SPANS

Before we continue, let's clarify a couple of points about our frequency span. We just described how we arrived at a dc to 128 kHz frequency span using a 4 ms time record. Because the signal passes through an anti-aliasing filter at the input, the entire frequency span is not useable. The filter has a flat response from dc to 100 kHz and then rolls off steeply from 100 kHz to 128 kHz. No filter can make a 90 dB transition instantly. The range between 100 kHz and 128 kHz is therefore not useable and the actual displayed frequency span stops at 100 kHz. There is also a frequency bin labelled 0 Hz (or dc). This bin actually covers the range from 0 Hz to 250 Hz (the lowest measurable frequency) and contains the signal components whose period is longer than the time record (not only dc). So our final displayed spectrum contains 400 frequency bins. The first covers 0 - 250 Hz, the second 250 - 500 Hz, and the 400th covers 99.75 - 100.0 kHz.

Spans less than 100 kHz

So the length of the time record determines the frequency span and resolution of our spectrum. What happens if we make the time record 8 ms or twice as long? Well we ought to get 2048 time points (sampling at 256 kHz) yielding a spectrum from dc to 100 kHz with 125 Hz resolution containing 800 points. But the SR760 places some limitations on this. One is memory. If we keep increasing the time record, then we would need to store more and more points. Another limitation is processing time. The time it takes to calculate an FFT with more points increases more than linearly. The net result is that the SR760 always takes 1024 point FFT's to yield 400 point spectra.

Here's how it's done. The analyzer digitally filters the incoming data samples (at 256 kHz) to limit the bandwidth. This is similar to the anti-aliasing filter at the input except the digital filter's cutoff frequency can be changed. In the case of the 8 ms record, the filter reduces the bandwidth to 64 kHz with a filter cutoff of 50 kHz (the filter rolls off between 50 and 64 kHz). Remember that Nyquist only requires samples at twice the frequency of the highest signal frequency. Thus the digital filter only has to output points at 128 kHz or half of the input rate (256 kHz). The net result is the digital filter outputs a time record of 1024 points effectively sampled at 128 kHz to make up an 8 ms record. The FFT processor operates on a constant number of points and the resulting FFT will yield 400 points from dc to 50 kHz. The resolution or linewidth is 125 Hz.

This process of doubling the time record and halving the span can be repeated by using multiple stages of digital filtering. The SR760 can process spectra with a span of only 191 mHz with a time record of 2098 seconds if you have the patience. However, this filtering process only yields baseband measurements (frequency spans which start at dc).

Starting the span somewhere other than dc

Besides being able to choose the span and resolution of the spectrum, we would also like the span to be able to start at frequencies other than dc. It would be nice to center a narrow span around any frequency below 100 kHz. Using digital filtering alone requires that every span start at dc. What is needed is heterodyning. Heterodyning is the process of multiplying the incoming signal by a sine wave. The resulting spectrum is shifted by the frequency of the sine wave. If we incorporate heterodyning with our digital filtering, we can shift any frequency span so that it starts at dc. The resulting FFT yields a spectrum offset by the heterodyne frequency. When this spectrum is displayed, the frequencies of the X axis are the frequencies of the actual signal, not the heterodyned frequencies.

Heterodyning allows the analyzer to compute zoomed spectra (spans which start at frequencies other than dc). The digital filter processor can filter and heterodyne the input in real time to provide the appropriate filtered time record at all spans and center frequencies. Because the digital signal processors in the SR760 are so fast, you won't notice any calculation time while taking spectra. The longest it can take to acquire a spectrum is the length of the time record itself. But more about that later.

THE TIME RECORD

Now that we've described the process in simple terms, let's complicate it a little bit. The SR760 actually uses 512 point complex time records. Each point is a complex value (with real and imaginary parts) so the record actually has 1024 data points in it. But how does a real point get to be complex?

As we described in the previous section, the input samples are digitally filtered and heterodyned to produce a time record with the appropriate bandwidth and a constant number of samples. What we need to add to this is that the heterodyning is a complex operation. This means that the input points are multiplied by both sine and cosine to yield a real and imaginary part.

So instead of using 1024 real points, we use 512 complex points. The time records have the same duration so the complex record has half the sampling rate of the real record. Thus at full span, the real points would occur at 256 kHz and the complex points at 128 kHz. You can think of the complex record as two separate records, one real and one imaginary, each with 64 kHz of bandwidth. (1/2 of the sample rate). One covers 0 to +64 kHz and the other covers -64 kHz to 0 for a total bandwidth of 128 kHz (the same bandwidth as the real record). What a negative frequency means is beyond this discussion but suffice to say it works the same.

The time record display

What do you see when you display the time record? Clearly the time record is not as simple as the raw digitized data points you would see if this were a digital oscilloscope.

The analyzer stores the 512 point complex time record described above. Because the display is designed for 400 point spectra, only the first 400 points of the time record are displayed. You can use the trigger delay to "translate" the time record to see the part not normally displayed.

The time record for every span has been digitally filtered and heterodyned into a complex record. You can display the magnitude, real or imaginary part as well as the phase. Normally, the easiest display to understand is Linear Magnitude. Remember that magnitudes are always positive. The negative parts of the waveform will be folded around zero so that they appear positive.

Because of the filtering and heterodyning, the time waveform may not closely resemble the input signal. For baseband measurements (when the start frequency of the span is 0.0 Hz) the waveform will resemble the signal waveform (with folding if magnitude is displayed). The bandwidth will be limited by the anti-alias filter and the digital filtering. For zoomed measurements (when the span start is not 0.0 Hz) the displayed waveform will not closely resemble the input signal because of the heterodyning.

Why use the time record?

The time display can be useful in determining whether the time record is triggered properly. If the analyzer is triggered, either internally by the signal or externally with another pulse, and the signal has a large component synchronous with the trigger, then the time record should appear stationary on the display. If the signal triggers randomly, then the time display will jitter back and forth.

Watch out for windowing!

The time display is not windowed. This means the time record which is displayed will be multiplied by the window function before the FFT is taken (see "Windowing" later in this section). Most window functions taper off to zero at the start and end of the time record. If the transient signal occurs at the start of the time record, the corresponding FFT may not show anything because the window function reduces the transient to zero.

Either use a Uniform window with transients, or use the trigger delay to position the transient at the center of the time record. (Remember that the display only shows the first 400 points of the record. The center is always at the 256th sample, which is not at the center of the display.)

To repeat, the time record is not a snapshot of the input signal. It is the output of the digital filter and the input to the FFT processor.

MEASUREMENT BASICS

Now that we know that the input to the FFT processor is a complex time record, it should be no surprise to find out that the resulting FFT spectrum is also a complex quantity. This is because each frequency component has a phase relative to the start of the time record. If there is no triggering, then the phase is random and we generally look at the magnitude of the spectrum. If we use a synchronous trigger then each frequency component has a well defined phase.

Spectrum

The spectrum is the basic measurement of an FFT analyzer. It is simply the complex FFT. Normally, the magnitude of the spectrum is displayed. The magnitude is the square root of the FFT times its complex conjugate. (Square root of the sum of the real part squared and the imaginary part squared). The magnitude is a real quantity and represents the total signal amplitude in each frequency bin, independent of phase.

If there is phase information in the spectrum, i.e. the time record is triggered in phase with some component of the signal, then the real or imaginary part or the phase may be displayed. Remember, the phase is simply the arctangent of the ratio of the imaginary and real parts of each frequency component. The phase is always relative to the start of the triggered time record.

Power Spectral Density or PSD

The PSD is simply the magnitude of the spectrum normalized to a 1 Hz bandwidth. This measurement approximates what the spectrum would look like if each frequency component were really a 1 Hz wide piece of the spectrum at each frequency bin.

What good is this? When measuring broadband signals such as noise, the amplitude of the spectrum changes with the frequency span. This is because the linewidth changes so the frequency bins have a different noise bandwidth. The PSD, on the other hand, normalizes all measurements to a 1 Hz bandwidth and the noise spectrum becomes independent of the span. This allows measurements with different spans to be compared. If the noise is Gaussian in nature, then the amount of noise amplitude in other bandwidths may be approximated by scaling the PSD measurement by the square root of the bandwidth. Thus the PSD is displayed in units of $V/\sqrt{Hz}$ or $dBV/\sqrt{Hz}$.

Since the PSD uses the magnitude of the spectrum, the PSD is a real quantity. There is no real or imaginary part or phase.

Octave Analysis

The magnitude of the normal spectrum measures the amplitudes within equally divided frequency bins. Octave analysis computes the spectral amplitude within 1/3 octave bands. The start and stop frequencies of each frequency bin are in the ratio of 1/3 of an octave ($2^{1/3}$). The octave analysis spectra will closely resemble data taken with older analog type equipment commonly used in acoustics and sound measurement.

To compute the amplitude of each band, the normal FFT is taken. Those bins which fall within a single band are rms summed together (square root of the sum of the squared magnitudes). The resulting amplitudes are real quantities and have no phase information. They represent total signal amplitude within each band.

We will have more about octave analysis later.

DISPLAY TYPES

Spectrum
The most common measurement is the spectrum and the most useful display is the Log Magnitude. The Log Mag display graphs the magnitude of the spectrum on a logarithmic scale using dBV as units.

Why is the Log Mag display useful? Remember that the SR760 has a dynamic range of 90 dB and a display resolution of -114 dB below full scale. Imagine what something 0.01% of full scale would look like on a linear scale. If we wanted it to be 1 centimeter high on the graph, the top of the graph would be 100 meters above the bottom. It turns out that the log display is both easy to understand and shows features which have very different amplitudes clearly.

Of course the analyzer is capable of showing the magnitude on a linear scale if you wish.

The real and imaginary parts are always displayed on a linear scale. This avoids the problem of taking the log of negative voltages.

The PSD and Octave analysis are real quantities and thus may only be displayed as magnitudes. In addition, the Octave analysis requires the display to be Log Magnitude.

Phase
In general, phase measurements are only used when the analyzer is triggered. The phase is relative to the start of the time record.

The phase is displayed in degrees or radians on a linear scale from -180 ($-\pi$) to +180 ($+\pi$) degrees (rads). There is no phase "unwrap".

The phase of a particular frequency bin is set to zero if neither the real nor imaginary part of the FFT is greater than 0.012% of full scale (-78 dB below f.s.). This avoids the messy phase display associated with the noise floor. (Remember, even if a signal is small, its phase extends over the full 360 degrees.)

Watch Out For Phase Errors
The FFT can be thought of as 400 bandpass filters, each centered on a frequency bin. The signal within each filter shows up as the amplitude of each bin. If a signal's frequency is between bins, the filters act to attenuate the signal a little bit. This results in a small amplitude error. The phase error, on the other hand, can be quite large. Because these filters are very steep and selective, they introduce very large phase shifts for signals not exactly on a frequency bin.

On full span, this is generally not a problem. The bins are 250 Hz apart and most synthesized sources have no problem generating a signal right on a frequency bin. But when the span is narrowed, the bins move much closer together and it becomes very hard to place a signal exactly on a frequency bin.

WINDOWING

What is windowing? Let's go back to the time record. What happens if a signal is not exactly periodic within the time record? We said that its amplitude is divided into multiple adjacent frequency bins. This is true but it's actually a bit worse than that. If the time record does not start and stop with the same data value, the signal can actually smear across the entire spectrum. This smearing will also change wildly between records because the amount of mismatch between the starting value and ending value changes with each record.

Windows are functions defined across the time record which are periodic in the time record. They start and stop at zero and are smooth functions in between. When the time record is windowed, its points are multiplied by the window function, time bin by time bin, and the resulting time record is by definition periodic. It may not be identical from record to record, but it will be periodic (zero at each end).

In the frequency domain

In the frequency domain, a window acts like a filter. The amplitude of each frequency bin is determined by centering this filter on each bin and measuring how much of the signal falls within the filter. If the filter is narrow, then only frequencies near the bin will contribute to the bin. A narrow filter is called a selective window - it selects a small range of frequencies around each bin. However, since the filter is narrow, it falls off from center rapidly. This means that even frequencies close to the bin may be attenuated somewhat. If the filter is wide, then frequencies far from the bin will contribute to the bin amplitude but those close by will probably not be attenuated much.

The net result of windowing is to reduce the amount of smearing in the spectrum from signals not exactly periodic with the time record. The different types of windows trade off selectivity, amplitude accuracy, and noise floor.

The SR760 offers four types of window functions - Uniform (none), Flattop, Hanning and Blackman-Harris (BMH).

Uniform

The uniform window is actually no window at all. The time record is used with no weighting. A signal will appear as narrow as a single bin if its frequency is exactly equal to a frequency bin. (It is exactly periodic within the time record). If its frequency is between bins, it will affect every bin of the spectrum. These two cases also have a great deal of amplitude variation between them (up to 4 dB).

In general, this window is only useful when looking at transients which do not fill the entire time record.

Hanning

The Hanning window is the most commonly used window. It has an amplitude variation of about 1.5 dB (for signals between bins) and provides reasonable selectivity. Its filter rolloff is not particularly steep. As a result, the Hanning window can limit the performance of the analyzer when looking at signals close together in frequency and very different in amplitude.

Flattop

The Flattop window improves on the amplitude accuracy of the Hanning window. Its between-bin amplitude variation is about .02 dB. However, the selectivity is a little worse. Unlike the Hanning, the Flattop window has a wide pass band and very steep rolloff on either side. Thus, signals appear wide but do not leak across the whole spectrum.

BMH

The BMH window is a very good window to use with this analyzer. It has better amplitude accuracy (about 0.7 dB) than the Hanning, very good selectivity and the fastest filter rolloff. The filter is steep and narrow and reaches a lower attenuation than the other windows. This allows signals close together in frequency to be distinguished, even when their amplitudes are very different.

If a measurement requires the full dynamic range of the analyzer, then the BMH window is probably the best one to use.

AVERAGING

The SR760 analyzer supports several types of averaging. In general, averaging many spectra together improves the accuracy and repeatability of measurements.

RMS Averaging
RMS averaging computes the weighted mean of the sum of the squared magnitudes (FFT times its complex conjugate). The weighting is either linear or exponential.

RMS averaging reduces fluctuations in the data but does not reduce the actual noise floor. With a sufficient number of averages, a very good approximation of the actual random noise floor can be displayed.

Since RMS averaging involves magnitudes only, displaying the real or imaginary part or phase of an RMS average has no meaning. The RMS average has no complex information.

Vector Averaging
Vector averaging averages the complex FFT spectrum. (The real part is averaged separately from the imaginary part.) This can reduce the noise floor for random signals since they are not phase coherent from time record to time record.

Vector averaging requires a trigger. The signal of interest must be both periodic and phase synchronous with the trigger. Otherwise, the real and imaginary parts of the signal will not add in phase and instead will cancel randomly.

With vector averaging, the real and imaginary parts as well as phase displays are correctly averaged and displayed. This is because the complex information is preserved.

Peak Hold
Peak Hold is not really averaging, rather the new spectral magnitudes are compared to the previous data, and if the new data is larger, then the new data is stored. This is done on a frequency bin by bin basis. The resulting display shows the peak magnitudes which occurred in the previous group of spectra.

Peak Hold detects the peaks in the spectral magnitudes and only applies to Spectrum, PSD, and Octave Analysis measurements. However, the peak magnitude values are stored in the original complex form. If the real or imaginary part or phase is being displayed for spectrum measurements, the display shows the real or imaginary part or phase of the complex peak value.

Linear Averaging
Linear averaging combines N (number of averages) spectra with equal weighting in either RMS, Vector or Peak Hold fashion. When the number of averages has been completed, the analyzer stops and a beep is sounded. When linear averaging is in progress, the number of averages completed is continuously displayed below the Averaging indicator at the bottom of the screen.

Auto ranging is temporarily disabled when a linear average is in progress. Be sure that you don't change the input range manually either. Changing the range during a linear average invalidates the results.

Exponential Averaging
Exponential averaging weights new data more than old data. Averaging takes place according to the formula,

$$\text{Average}_N = (\text{New Spectrum} \cdot 1/N) + (\text{Average}_{N-1}) \cdot (N-1)/N$$

where N is the number of averages.

Exponential averages "grow" for approximately the first 5N spectra until the steady state values are reached. Once in steady state, further changes in the spectra are detected only if they last sufficiently long. Make sure that the number of averages is not so large as to eliminate the changes in the data that might be important.

REAL TIME BANDWIDTH AND OVERLAP PROCESSING

What is real time bandwidth? Simply stated, it is the frequency span whose corresponding time record exceeds the time it takes to compute the spectrum. At this span and below, it is possible to compute the spectra for every time record with no loss of data. The spectra are computed in "real time". At larger spans, some data samples will be lost while the FFT computations are in progress.

For all frequency spans, the SR760 can compute the FFT in less time than it takes to acquire the time record. Thus, the real time bandwidth of the SR760 is 100 kHz. This includes the real time digital filtering and heterodyning, the FFT processing, and averaging calculations. The SR760 employs two digital signal processors to accomplish this. The first collects the input samples, filters and heterodynes them, and stores a time record. The second computes the FFT and averages the spectra. Since both processors are working simultaneously, no data is ever lost.

Averaging speed

How can you take advantage of this?Consider averaging. Other analyzers typically have a real time bandwidth of around 4 kHz. This means that even though the time record at 100 kHz span is only 4 ms, the "effective" time record is 25 times longer due to processing overhead. An analyzer with 4 kHz of real time bandwidth can only process about 10 spectra a second. When averaging is on, this usually slows down to about 5 spectra per second. At this rate it's going to take a couple of minutes to do 500 averages.

The SR760, on the other hand, has a real time bandwidth of 100 kHz. At a 100 kHz span, the analyzer is capable of processing 250 spectra per second. In fact, this is so fast, that the display can not be updated for each new spectra. The display only updates about 6 times a second. However, when averaging is on, all of the computed spectra will contribute to the average. The time it takes to complete 500 averages is only a few seconds. (Instead of a few minutes!)

Overlap

What about narrow spans where the time record is long compared to the processing time? The analyzer computes one FFT per time record and can wait until the next time record is complete before computing the next FFT. The update rate would be no faster than one spectra per time record. With narrow spans, this could be quite slow.

And what is the processor doing while it waits? Nothing. With overlap processing, the analyzer does not wait for the next complete time record before computing the next FFT. Instead it uses data from the previous time record as well as data from the current time record to compute the next FFT. This speeds up the processing rate. Remember, most window functions are zero at the start and end of the time record. Thus, the points at the ends of the time record do not contribute much to the FFT. With overlap, these points are "re-used" and appear as middle points in other time records. This is why overlap effectively speeds up averaging and smoothes out window variations.

Typically, time records with 50% overlap provide almost as much noise reduction as non-overlapping time records when RMS averaging is used. When RMS averaging narrow spans, this can reduce the measurement time by 2.

Overlap percentage

The amount of overlap is specified as a percentage of the time record. 0% is no overlap and 99.8% is the maximum (511 out of 512 samples re-used). The maximum overlap is determined by the amount of time it takes to calculate an FFT and the length of the time record and thus varies according to the span.

The SR760 always tries to use the maximum amount of overlap possible. This keeps the display updating as fast as possible. Whenever a new frequency span is selected, the overlap is set to the maximum possible value for that span. If less overlap is desired, then use the Average menu to enter a smaller value. On the widest spans (25, 50 and 100 kHz), no overlap is allowed.

Triggering

If the measurement is triggered, then overlap is ignored. Time records start with the trigger. The analyzer must be in continuous trigger mode to use overlap processing.

INPUT RANGE

The input range on the SR760 varies from a maximum of 34 dBV full scale to a minimum of -60 dBV full scale. A signal which exceeds the current input range will cause the OvrLoad message to appear at the bottom of the screen. A signal which exceeds the maximum safe range will turn on the HI V indicator.

The input range is displayed in dBV. The maximum and minimum range equivalents are tabulated below.

Max	34 dBVpk
	31 dBVrms
	50.1 Vpk
	35.4 Vrms
Min	-60 dBVpk
	-63 dBVrms
	1.0 mVpk
	0.7 mVrms

Manual Range
The input range can be specified in the Input menu to be fixed at a certain value. Signals that exceed the range will overload and become distorted. Signals which fall to a small percentage of the range will become hard to see.

Auto Range
The input range can be set to automatically correct for signal overloads. When autoranging is on and an overload occurs, the input range is adjusted so that the signal no longer overloads. If the signal decreases, the input range is not adjusted. You must take care to ensure that the signal does not fall dramatically after pushing the input range to a very insensitive setting.

While the analyzer is performing linear averaging, the input range is NOT changed even if the signal overloads. The overload indicator will still light to indicate an over range condition. Changing the range during a linear average invalidates the average.