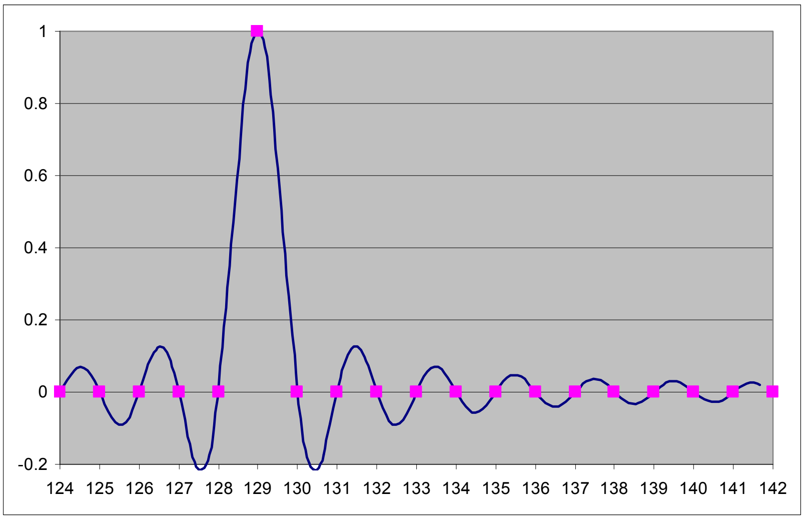
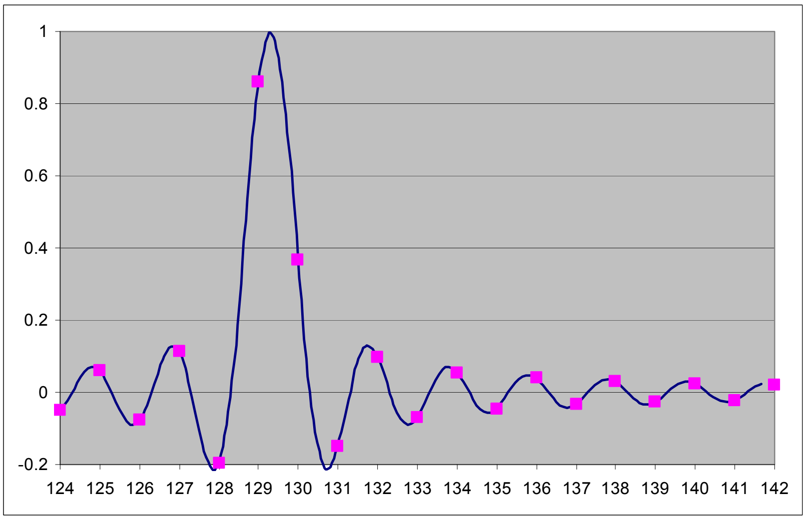
Note on Spectral Response

Like all spectroscopic instruments, ALMA can only measure the spectrum of the signals received with a finite degree of fidelity. To interpret the data we need to understand the relationship between the output spectrum produced by the telescope and the true spectrum of the astronomical source. We expect that this relationship will be linear (although there will of course be added noise, which we mostly ignore here). We also expect that, to a good approximation, the output spectrum will be simply the convolution of the input spectrum with a known "Spectral Response Function". Furthermore this response function should not vary significantly in different parts of the spectrum. The form of the response function is however determined by some choices that are made in processing the data. These are under the control of the user and are described here.

The correlator operates by forming the auto-correlation and cross-correlation functions of the incoming signals and then taking the Fourier transform to generate the spectrum. The correlation functions are only calculated out to a certain maximum value of the time difference (the "lag"), which is set by the amount of hardware available. This maximum lag, τ_{max} sets the spectral resolution, $\Delta\nu \sim 1/\tau_{max}$. The truncation of the correlation function at this maximum lag corresponds to multiplying the true correlation function with a "top hat" function. This means that, when we take the transform, the Spectral Response will be a sinc function, $\sin(\nu)/\nu$, which looks like this:

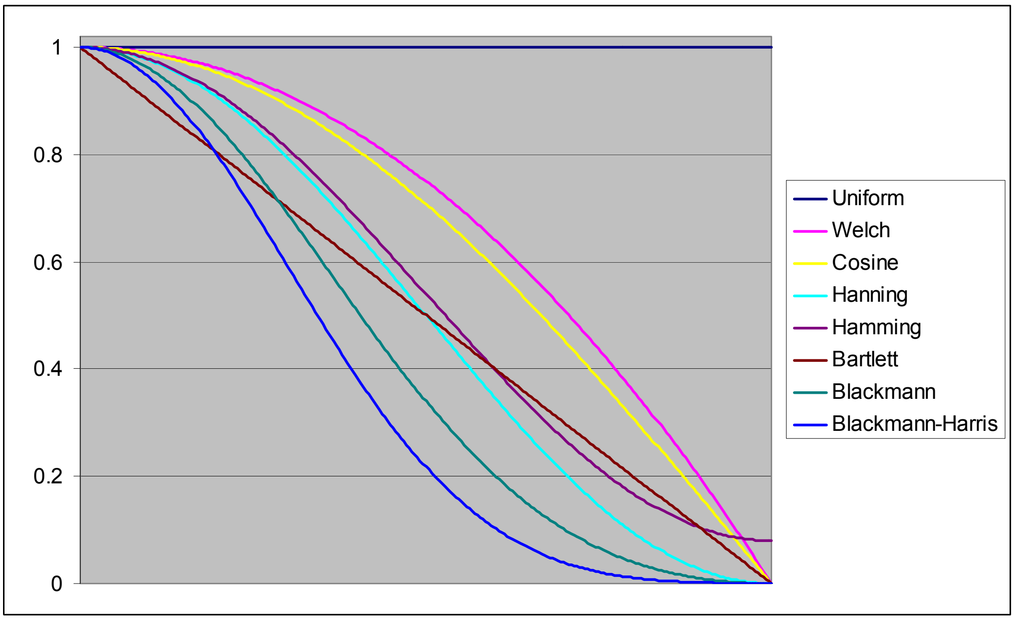


This can be thought of as the response of the instrument to a pure sine wave, i.e. an infinitely narrow feature in the input spectrum, which we have arbitrarily chosen here to fall at channel 129 of the output. In practice we do not calculate all the intermediate points in the output spectrum but only sample it at a finite number of points (pink in the plot above) separated by the "channel spacing", $\Delta\nu_{ch}$, which we set to be the Nyquist interval, $1/2\tau_{max}$. When the input frequency exactly matches that of one of the output channels, then the response is ideal: unity for that channel and zero for all the others. As soon as the input frequency differs from an integer times the channel spacing, however, the "ringing" implicit in the sinc function shows up:

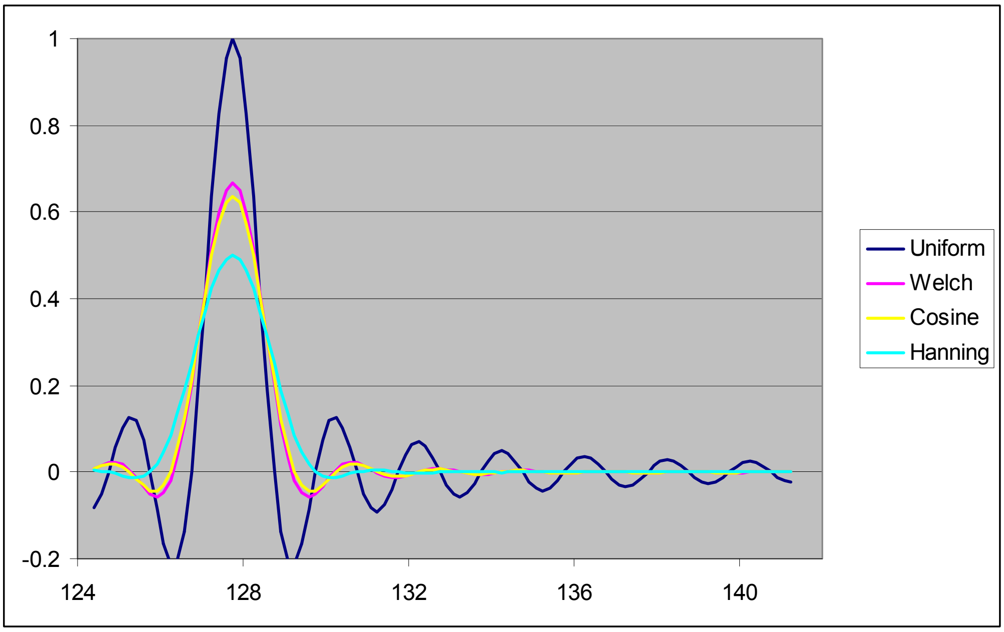


This is generally felt to be an undesirable feature of this form of spectrometer. (Note that a sinc-function spectral response is also produced by an optical/IR Fourier-transform spectrometer, whereas a diffraction grating with a uniform illumination produces a sinc-*squared* response, i.e. one which falls off much faster in the sidelobes far away from the peak response.) The solution is to apply a weighting to the correlation functions before transforming them. This is equivalent to tapering or "apodizing" an aperture in order to reduce the sidelobes on the image. The spectral case has been thoroughly studied: see in particular the famous book by Blackman and Tukey[1] and the Wikipedia article http://en.wikipedia.org/wiki/Window_function.

A total of 8 weighting or "window" functions are implemented in the software of the ALMA baseline correlator[2]. Here they are plotted in the lag domain:



Here are the spectral response functions corresponding to the four "softer" weighting functions:



Welch (parabolic) and Cosine weighting are quite similar. Note that Cosine weighting in the lag domain corresponds to averaging two channels in the spectral domain, i.e. convolving with the

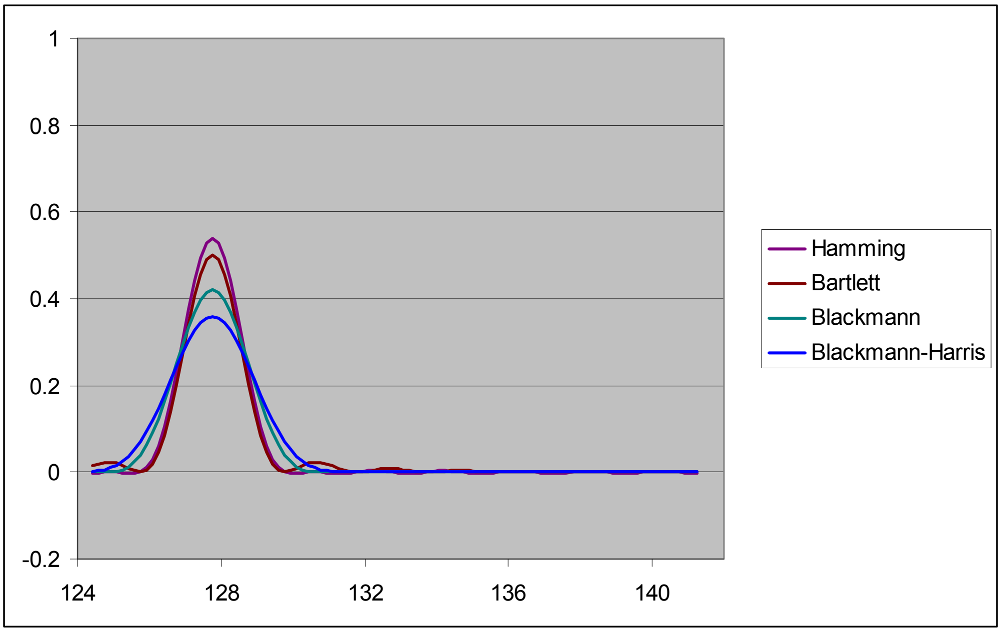
---

[1] "The Measurement of Power Spectra", R.B.Blackman & J.W.Tukey, Dover edition 1959, 486-60507-8.

[2] So far this description only covers the 64-input ("Baseline") Correlator: the ACA Correlator should be added.

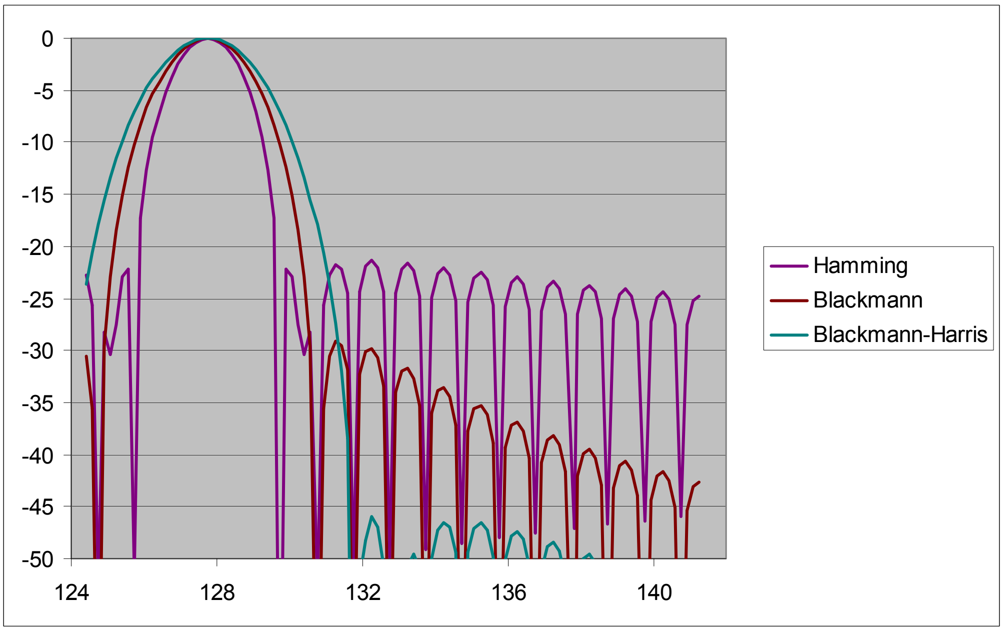
function 0.5 x [$\delta(-½)$ + $\delta(½)$]. Similarly Hanning weighting, which is cosine-squared in the lag domain, corresponds to averaging by two channels twice, which is equivalent to convolving with the function 0.25 x $\delta(-1)$ + 0.5 x $\delta(0)$ + 0.25 x $\delta(1)$.

Here are the spectral response functions for the other four cases:



The Bartlett function is a simple linear taper, giving rise to a sinc-squared spectral response.

As can be seen from the plots the effects of the weighting are to reduce the height of the peak (recall that this is the response to an infinitely narrow spectral line), reduce the sidelobes, and increase the width of the response function, i.e. reduce the spectral resolution. The remaining three functions, Hamming, Blackmann[3] and Blackmann-Harris, have been designed to maximize the suppression of the sidelobes (the spurious responses away from the peak) while minimizing the loss of resolution. This is more clearly seen in a log plot. (Note that these have been normalized with respect to the peak value of the response function. The vertical scale is in dBc.)



It can be seen that Hamming smoothing suppresses the first sidelobe rather well, but the fall-off after that is rather slow. Blackmann has a much better fall-off but produces some loss of

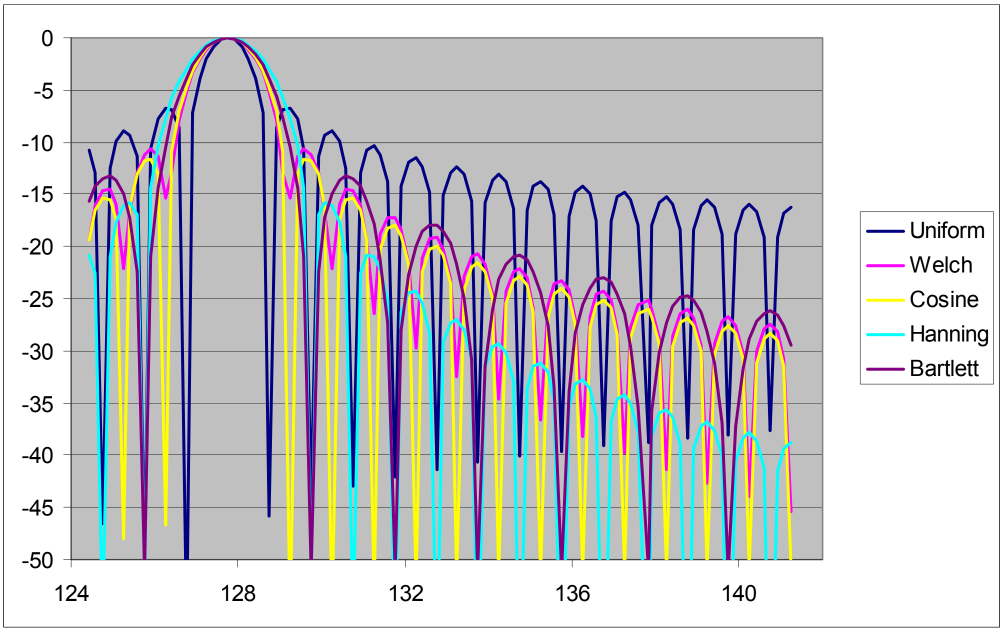
---

[3] Note that the name Blackman seems to have been wrongly spelt in all the ALMA documents. I believe that the double "n" is spurious but I have kept it for consistency with the ALMA literature.

resolution and Blackmann-Harris manages to suppress all sidelobes to below -45dBc, with only a modest additional loss of resolution.

It is worth pointing out that ALMA does have a specification of -40dBc for "spectral dynamic range" – the detection of weak spectral lines in the presence of strong ones. The specification does not give any number for the separation between the strong and the weak features so it is not clear how this relates to the topic of this note, but it is good that the Blackmann-Harris does give the required performance, at least in principle.

For completeness here is the log plot of the other five functions.



The key parameters that the user needs to know is are the effective resolution, which is most simply described by the Full-Width-Half-Maximum (FWHM) of the Spectral Response Function, and Effective Bandwidth, which determines the rms noise fluctuations in the spectrum. The latter is the value that one needs to use in the term 1 / square root (bandwidth x integration time) in the standard expressions for the noise fluctuations[4]. One of the effects of applying the weighting is of course to reduce the noise on the spectrum.

These are as follows:

	FWHM	Effect BW
Uniform	1.207	1.000
Welch	1.590	1.875
Cosine	1.639	2.000
Hanning	2.000	2.667
Hamming	1.815	2.516
Bartlett	1.772	3.000
Blackmann	2.299	3.283
Blackmann-Harris	2.666	3.877

The values are all expressed in units of the nominal channel spacing in the output spectrum, $\Delta\nu_{ch}$.

It can be seen that reduction in the noise, which is proportional the square root of the effective bandwidth, is quite significant. It should however be remembered that the data points in the

[4] Note that the "effective bandwidth" we have used here is not the same as the "noise equivalent bandwidth" used in the engineering literature. The reason for this is the most common engineering application is finding unresolved (e.g. coherent) signals in the presence of noise, whereas in astronomy we normally use enough resolution to resolve the lines we are observing quite well. We therefore calibrate our data so that the peak value of a very wide spectral feature will be independent of the smoothing applied, whereas the engineers calibrate theirs by injecting an unresolved feature and keeping the measured value the same.

resulting spectrum are not independent and the correlation between them increases as we use a stronger taper. This means in particular that averaging across N channels will NOT cause the noise to go down by a factor of root(N). Instead, for N >> 1, the noise on the averaged spectrum will be approximately 1/root(N) times the noise expected with no averaging and uniform weighting, i.e. effective BW = N, independent of what weighting function was originally applied to the data.

We will return to the question of averaging over channels below, but there is one final point to make on the subject of weighting functions. This is that, in principle, the process is (almost) reversible: since we have simply applied weights to the correlation coefficients before doing the Fourier transform, we could, if we wished, do the inverse transform, divide by the original weights, apply a new set of weights and then transform back to get a new spectrum. In fact we cannot do this exactly for two reasons: 1) the weight applied to the longest lag is zero (except in the case of the Hamming window), so this is lost, 2) the transforms and the data storage are performed with finite numerical precision so, if the weights are small, as is the case for the stronger taper functions at the larger lags, this re-weighting process will add noise due to round-off errors. It is also true that CASA does not presently provide us with the capability of doing this, although it does enable us to average channels together.

The advice at this point therefore is that, unless one is doing something very specialized, the best thing to do is to apply a relatively light taper and then, if the data analysis indicates that smoothing is needed, one can average over channels at that point. One solution would be simply to use the Uniform weighting but, as noted, this can lead to very messy spectra as a result of the ringing from a strong interference spike, e.g. at the edge of the spectrum. My personal preference is to use the Welch weighting[5] but Cosine gives very similar results. The present default is Hanning, which does suppress the ringing quite well (e.g. about -26dBc at 5 channels from the peak, compared to ~ -19 and -20dBc for Welch and Cosine) but the spectral response functions is really quite wide with the FWHM equal to twice the channel spacing.

Channel Averaging

The two basic modes of operation of the ALMA baseline correlator are Time Division Mode (TDM) and Frequency Division (FDM). From the point of view of the users the most important difference is that they get 32 times more channels in the output spectra in FDM than in TDM. The totals are 1024 channels in TDM mode and 32,768 in FDM. In the most typical operation these are divided between the two polarizations and four basebands so that one gets either 128 or 4096 channels[6] per polarization per baseband. This means that in practice most continuum observations will be made in TDM mode whereas most spectral line work will use FDM. The snag is that in FDM mode the data rate is high and the resulting data sets will be large, increasingly so as the number of antennas in the array increases. As a result it will become difficult to transport the data from spectral line observations and, in particular, the reduction will become slow and laborious. If the observations really require all the spectral information – e.g. ~15,000 channels per polarization – then there is of course no way round this. It is however likely that many observations require a far smaller number of spectral channels and in that case we are simply making things difficult for ourselves by storing and processing the whole set.

The hardware does not allow any intermediate cases between TDM and FDM, so the software feature “Channel Averaging” has been introduce to enable the user to control the number of output channels so that it matches the scientific needs better. As implied by the name, the output values are formed by averaging N neighbouring channels together, where N is 2, 4 or 8. *(Why is 16 not allowed?)* The channel spacing in the output spectra becomes N times larger than that of the underlying hardware mode and the number of output channels is N times smaller.

As far as the Spectral Response Function is concerned, the result is that the functions described above are replaced by versions where the original has been convolved with 2, 4 or 8 delta-functions. As one would expect, this means that as N increases the shapes tend towards top-hat functions and the FWHM and the Effective Bandwidth both tend towards N times the underlying channel spacing, $\Delta\nu_{ch}$, where N is the number of channels averaged.

---

[5] Jack Welch was my Ph.D, supervisor, so I may be biased.

[6] This ignores the fact that ~1/32nd of the channels are dropped at each edge of the band to avoid aliasing.

Here are the actual values for the first four weighting functions, again in units of $\Delta\nu_{ch}$. The first table gives of the Full Width Half Maximum. (In fact I have taken the width at half the central value here since the central value is not always the maximum.)

N	1	2	4	8
Uniform	1.207	1.639	4.063	8.033
Welch	1.590	1.952	4.007	8.001
Cosine	1.639	2.000	4.000	8.000
Hanning	2.000	2.312	3.970	7.996

Here are the values of the Effective Bandwidth.

N	1	2	4	8
Uniform	1.000	2.000	4.000	8.000
Welch	1.875	2.565	4.499	8.470
Cosine	2.000	2.667	4.571	8.533
Hanning	2.667	3.200	4.923	8.828

One can note various relationships here, which can be confirmed analytically, e.g. averaging two channels combined with uniform weighting is equivalent to cosine weighting, N = 2 plus cosine equals Hanning.

A final important but obvious comment on channel averaging is that it is NOT reversible. In particular the fact that we record a factor N fewer output points obviously means that information is lost. It is therefore important that the combination of the underlying spectral mode and the number of channels averaged together is correctly chosen to ensure that the resulting spectral resolution is sufficient for the scientific goals.

This table shows, for each value of N, the number of useful channels in the output spectrum per baseband for each polarization, assuming that two polarizations are used, together with the spectral resolution (FWHM) for the correlator modes that are available. Here I have assumed that Welch weighting is adopted.

Tot BW	Ch Spacing	Useful BW	1	2	4	8
Num Chans per pol per baseband			3840	1920	960	480
MHz	kHz	MHz	kHz	kHz	kHz	kHz
2000	488	1875	777	953	1957	3907
1000	244	938	388	477	978	1953
500	122	469	194	238	489	977
250	61	234	97	119	245	488
125	31	117	49	60	122	244
62.5	15	59	24	30	61	122
31.25	7.6	29.3	12.1	15	31	61

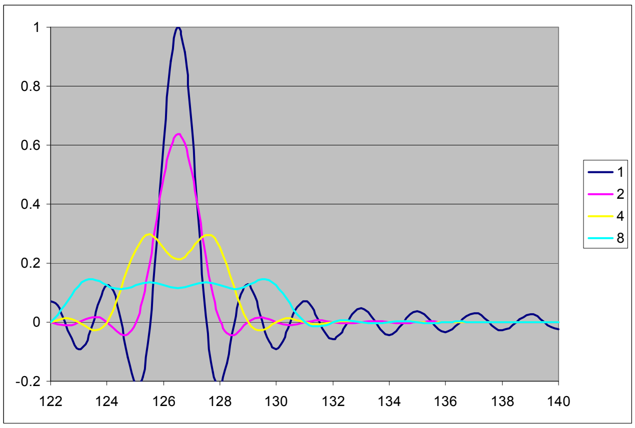
(The final line of this table requires double-Nyquist sampling which is not yet implemented.)

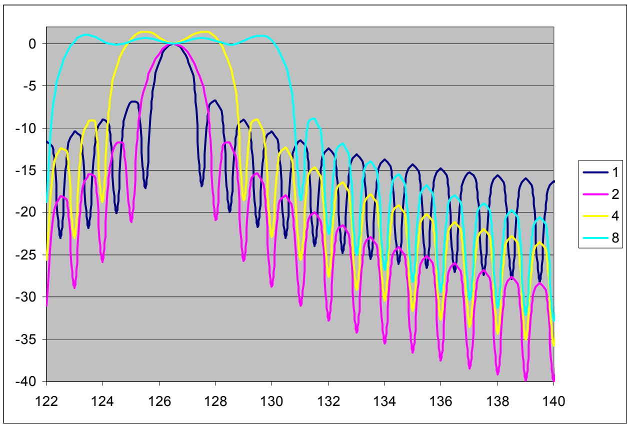
Plots of the resulting spectral response functions are given below for completeness. As before the linear plots of the spectral response function show the output value of a single channel as a very narrow spectral line is moved across the band. These are not normalized, so the lower value of the peak when weighting and averaging is applied reflects the resulting reduced response to an unresolved feature. The logarithmic plots are normalized by the central value to make it easier to see the shape of the function. Note that the varying depths of the nulls in the logarithmic plots are just a result of the finite sampling in the calculation: the functions do all contain zeros but I have avoided having the sampled points fall on those.

The x-axis on all these plots represents frequency in units of the underlying channel spacing, $\Delta\nu_{ch}$. Obviously the data points in the final output spectrum will lie 1, 2, 4 or 8 points apart in accord with the value of N that is chosen.

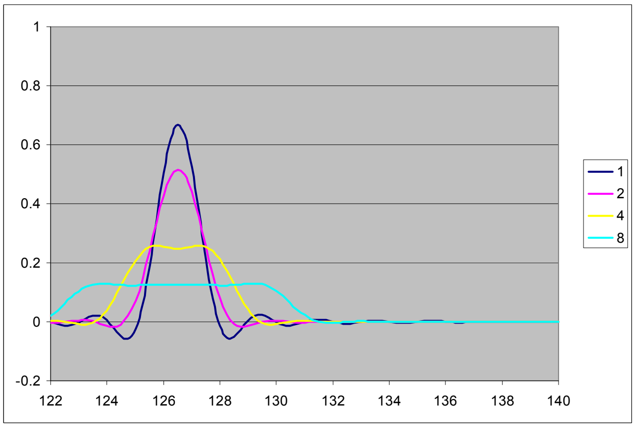
Richard Hills 8th April 2012

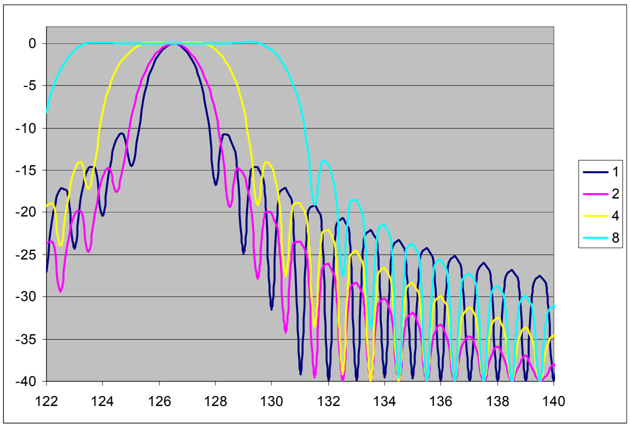
Spectral Response Functions with Averaging: Uniform weighting



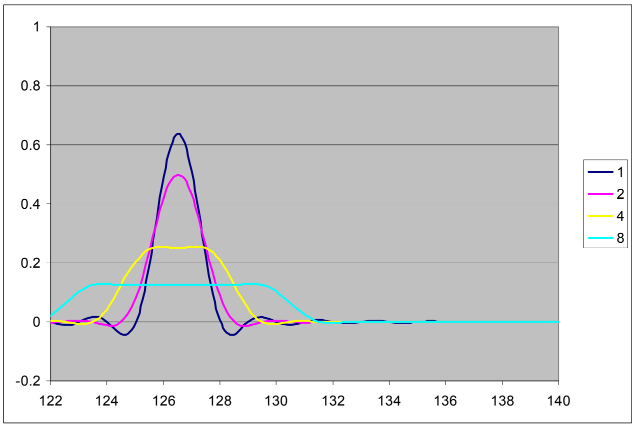


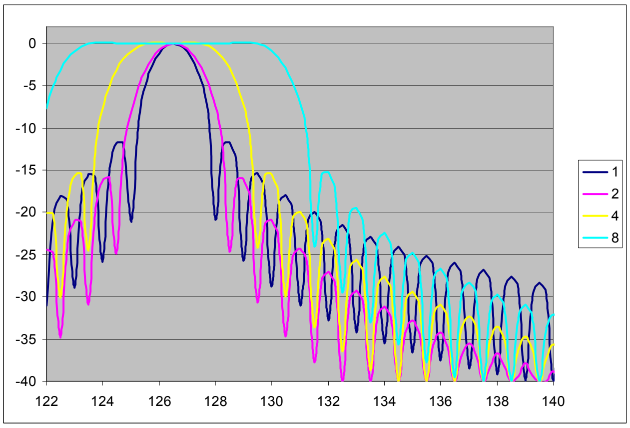
Welch weighting





Cosine weighting





Hanning weighting