

ALMA Memo 452:
Passband Shape Deviation Limits

Larry R. D’Addario
2003 April 09

Abstract. Beginning with the ideal passband, which is constant from $Nf_s/2$ to $(N+1)f_s/2$ and zero elsewhere, where $N = 0, 1, 2, \dots$ is the Nyquist interval number for sampling rate f_s , we examine the effect of deviations from that form on radio telescope performance. The implications for both continuum and spectroscopic observations are considered. Among the deviations of interest are: finite band edge slope, leading to non-zero response outside the Nyquist interval (aliasing); gain slope across the passband; gain ripple across the passband; and mismatch of the responses among antennas in an array. When analog filters are involved, the variation of response with temperature is also a consideration. Most of this report is applicable to radio telescopes in general, but numerical examples are selected for relevance to the ALMA telescope and specific recommendations for ALMA are given.

Introduction

The minimum detectable signal (sensitivity) of one frequency channel of a radiometer or interferometer observing white Gaussian noise is proportional to $B_{\text{eff}}^{-1/2}$, where B_{eff} is the “effective” bandwidth of that channel. For an ideal channel bandpass shape where the gain is constant over a range of frequencies B and zero elsewhere, $B_{\text{eff}} = B$. This is true for continuous-time (analog) detectors and correlators; for discrete-time processing, where only samples of the signal are used, it is also true provided that the frequencies in B fall within one Nyquist interval ($Nf_s/2, (N+1)f_s/2$) for some integer N , where f_s is the sampling rate. Here we will be concerned primarily with discrete-time systems.

When the passband shape is not ideal, the effective bandwidth for continuous-time processing is [1]

$$B_{\text{eff},c} = \frac{[\int_0^\infty G(f)df]^2}{\int_0^\infty G(f)^2df} \quad (1)$$

where $G(f)$ is the system power gain function of frequency from input to detector; and for discrete-time processing it is

$$B_{\text{eff},d} = (f_s/2) \frac{R(0)^2}{\sum_{k=-\infty}^\infty R(k/f_s)^2} \quad (2)$$

where $R(\tau)$ is the autocorrelation function of the signal at the detector, given by

$$R(\tau) = 2 \int_0^\infty G(f) \cos(2\pi f\tau)df. \quad (3)$$

The discrete time result (2) is derived in Appendix A. The maximum possible value of $B_{\text{eff},d}$ is $f_s/2$, as might be expected from the sampling theorem. This is illustrated in Figure 1, which shows a plot of $B_{\text{eff},d}$ vs. B for a rectangular lowpass function, normalized by the sampling rate. When nearly all of the power is inside a Nyquist interval, $B_{\text{eff},c}$ and $B_{\text{eff},d}$ are nearly the same. Equations (1) and (2) apply to total power detection, and they also apply to interferometric detection (cross-correlation) provided that the transfer functions of the two antennas are identical; departure from identical responses is treated later in this report.

For *continuum* observations, meaning that the signal is assumed to have no interesting variation with frequency over the passband, it is usually our objective to maximize the effective bandwidth. For *spectroscopic* observations, where variations with frequency are interesting and must be accurately measured, we have the additional objective of establishing a well-defined passband, so that signals outside it are adequately rejected. There are two cases of this rejection, depending on whether the channel currently being considered is one of many that will be separately detected and used collectively to determine the frequency variation (sometimes called a “spectral channel”), or whether it is defining a passband that will be further analyzed later. The second case is of interest for ALMA, where we are concerned with defining a broad passband before sampling and quantization. The broad channel will be analyzed into many narrower channels using digital processing, but that will not be discussed here.

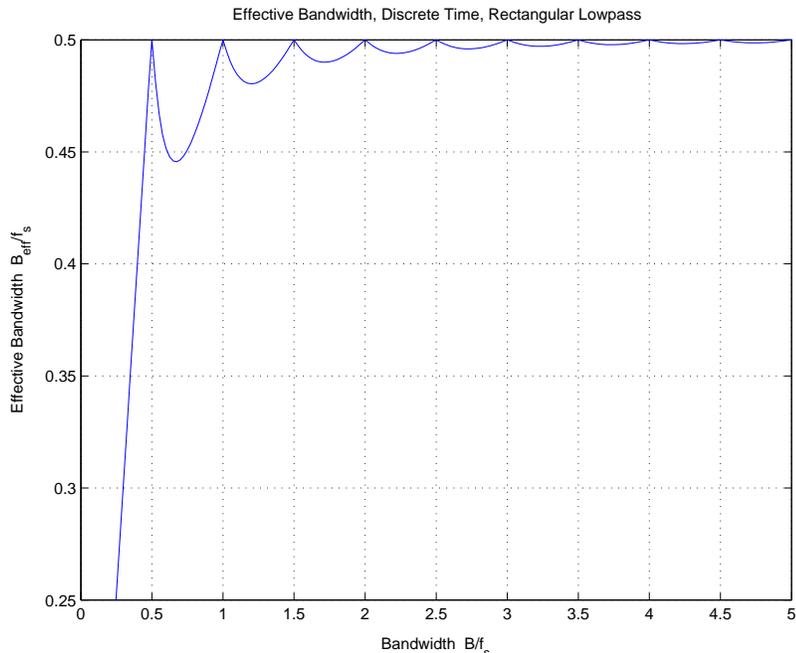


Figure 1 Effective bandwidth for sampled data with a rectangular lowpass gain function.

Effective Bandwidth vs. Alias Rejection

Let us now consider practical passband shapes as determined by analog bandpass filters, for the sampled data case. Results are given below for the $N = 1$ Nyquist interval ($f_s/2, f_s$) and for filters with equal ripple in their passbands (Chebyshev type I); these are useful cases, but results for other intervals and filter types could just as easily be calculated.

Figure 2 shows the effective bandwidths and 20 dB suppression bandwidths for filtered white noise with a sequence of different filters. The results are shown for $f_s = 4.0$ GHz and thus a Nyquist interval of 2 to 4 GHz, but they can be scaled to any sampling frequency. The nominal bandwidth of the filters varies from 1.6 to 2.2 GHz, centered so that the gains at 2 GHz and 4 GHz are equal, and their order (complexity) varies from 6 to 9. Maximally flat (zero-ripple or Butterworth) filters and 0.5 dB equal-ripple (Chebychev) filters are plotted separately. The 20 dB suppression bandwidth is defined as the range of frequencies f over which the gain is at least 20 dB greater than at the nearest aliased frequency, $kf_s - f$ for integer k . After sampling, a signal at an aliased frequency is indistinguishable from one at f . Assuming that the digitized signal will be analyzed into many smaller subchannels, almost any desired suppression factor a can be achieved by ignoring enough of the outer subchannels. The usable bandwidth is then the a -suppression bandwidth, B_a . A compromise must be made between maximizing B_a for spectroscopy and maximizing B_{eff} for continuum. Clearly a more complex filter or one where larger ripple is allowed produces larger usable bandwidth at a given effective bandwidth due to its sharper cutoff, but such filters are more expensive, have larger temperature coefficients, and are more difficult to match. Generally we choose the sharpest cutoff filter for which the temperature sensitivity, matching tolerances, and cost constraints can be met. Temperature sensitivity and matching are considered later in this report.

The compromise choice is somewhat subjective, since it depends on the relative importance of continuum and spectroscopic observations as well as on the desired suppression factor a . It can be seen from Figure 2 that for a given filter type and order there is a choice of nominal bandwidth that maximizes B_a and another (larger) that maximizes B_{eff} . A reasonable choice is a bandwidth somewhere between these. For example, if $a = -20$ dB, then using a 6th order filter with 0.5 passband ripple we can obtain maximum $B_a = 1.81$ GHz (90.5%) with $B_{\text{eff}} = 1.86$ GHz (93%); or we can obtain maximum $B_{\text{eff}} = 1.99$ GHz (99.5%) with $B_a = 1.65$ GHz (82.5%).

All these results are based the calculated responses of ideal lumped-element filters of the specified type, which means that the elements are lossless and perfectly accurate. To achieve comparable responses with

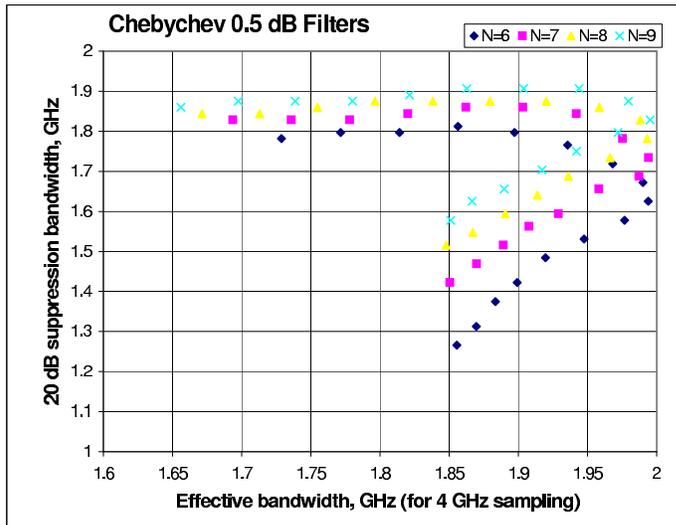
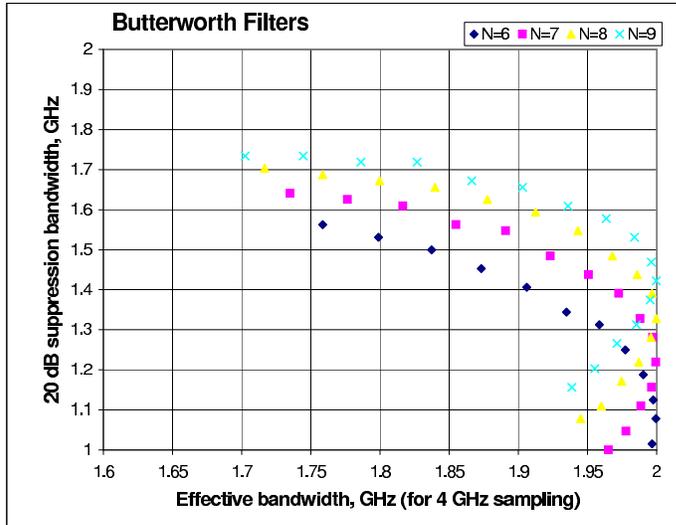


Figure 2 Effective bandwidth vs. 20 dB suppression bandwidth for bandpasses produced by a range of theoretical filter responses, computed for $f_s = 4$ GHz in the $N = 1$ Nyquist interval. The nominal bandwidth of each filter is varied from 1.6 to 2.2 GHz (starting at the upper left in steps of .04 GHz), and the filter order is varied from 6 to 9. Top: Butterworth (maximally flat) filters. Bottom: Chebychev filters with 0.5 dB ripple in passband.

practical filters, the order is normally increased by 1 or 2 so as to compensate for losses and errors. Filters based on transmission lines rather than lumped elements will have slightly different responses, but not by enough to affect these results.

Gain Slope and Ripple

The filters considered in the preceding section provide nearly flat gain over most of the active bandwidth. Other signal processing components may impose additional deviations from the ideal response, especially when the bandwidth is large. Of particular practical importance are transmission lines whose length is large compared with v/B_{eff} , where v is the propagation speed in the line. First, the attenuation in the line is normally larger at the high frequency than the low frequency end of the band, producing gain slope. Second, reflections at the line ends produce standing waves and sinusoidal variation of gain with frequency (ripple).

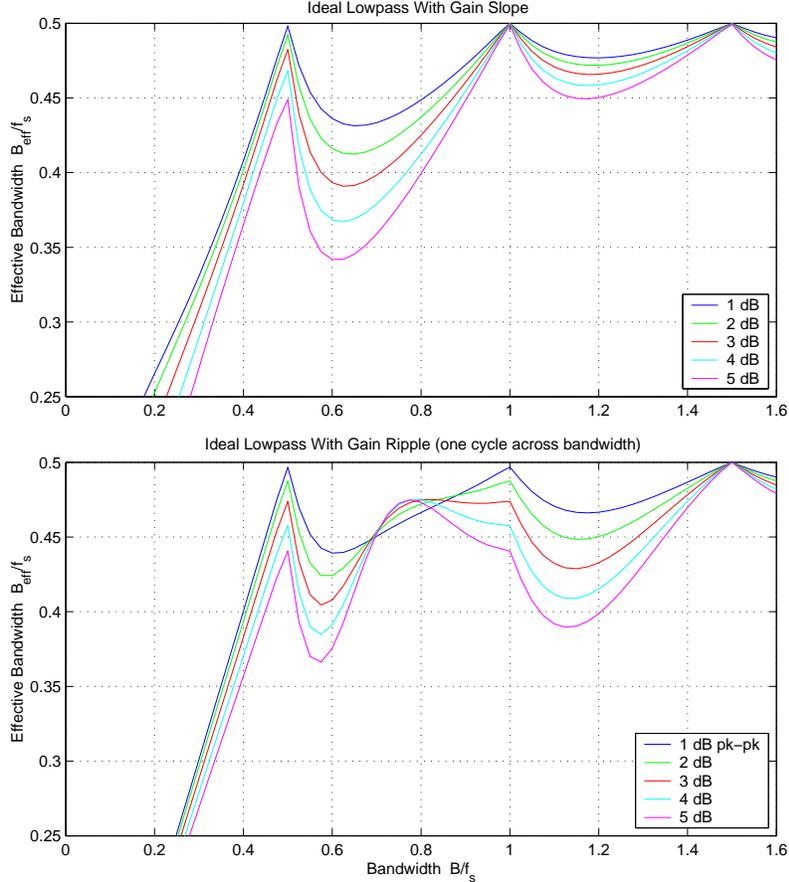


Figure 3 Effective bandwidth for sampled data with a non-ideal gain functions, normalized by sampling rate f_s . Top: Lowpass spectrum of width B and an exponential gain slope across the band of 1, 2, ..., 5 dB. Bottom: Lowpass spectrum of width B with sinusoidal ripple of various amplitudes, with one cycle of ripple across B ; the result is the same for any integer number of cycles and slightly different for a non-integer number. At $B/f_s = 0.5$ (Nyquist sampling), all results are identical to those for continuous time given in [2] and [3].

The effects of gain slope and ripple on sensitivity were studied in [2] and [3], essentially by evaluating (1). It was found, for example, that a 3.5 dB exponential (linear in dB) gain change across the band or a 2.9 dB peak-to-peak sinusoidal ripple each reduces B_{eff} by a factor of 0.95, producing an increase in minimum detectable signal of 2.5%.

The situation for sampled data can be determined by evaluating (2) under the same circumstances. The results are plotted in Figure 3, where the passbands are rectangular lowpass ($N = 0$) except for the slope or ripple. At the Nyquist bandwidth, the results match those of the continuous-time case.

Bandpass Matching

For an interferometer in which the transfer functions of the two antennas are not identical, the effective bandwidth is still given by (1) or (2) provided that squares are replaced by magnitudes squared, $R(\tau)$ is the cross-correlation function, and $G(f)$ is the complex gain of the antenna pair, which is

$$G_{nm}(f) = H_n(f)H_m(f)^* \quad (4)$$

where $H_i(f)$ is the transfer function of antenna i . The effective bandwidth for the interferometer will generally be less than the smaller of its single-antenna effective bandwidths unless $H_n(f) = hH_m(f)$ for constant h ; that is, mismatch reduces sensitivity. However, a more important effect of mismatch is on calibration, as explained in [2]. A calibration error occurs when continuum observations of known sources are used to

Table 1: Gain Residuals Due To Bandpass Mismatches

Type of Mismatch	1% Maximum Gain Error	0.2% Maximum Gain Error
Amplitude slope (edge to edge)	2.7 dB	1.2 dB
Sinusoidal ripple (peak to peak)	1.5 dB	0.75 dB
Center frequency displacement	.007 <i>B</i>	.001 <i>B</i>
Phase variation (random, rms)	9.1°	4.0°

determine the total complex gain on each baseline, and then this is approximated as the product of complex antenna “gains” by

$$\int_0^\infty G_{mn}(f)df = g_n g_m^* + \epsilon_{mn}. \quad (5)$$

where the g_i are the antenna gains and ϵ_{mn} is an (unknown) error. In this way the $A(A - 1)/2$ baseline gains are factored into only A antenna gains. This can be done without error ($\epsilon_{mn} = 0 \forall m, n$) only in the special case that all the complex transfer functions are identical except for a frequency-independent factor:

$$H_i(f) = h_i H(f) \quad \forall i. \quad (6)$$

When this is not the case, (5) can be solved for $\{g_i\}$ such that the mean-squared error is minimized. This was studied in [2][3] for several types of mismatch, including unequal exponential slopes, sinusoidal variations of unequal ripple phase, unequal phase variations of the transfer functions with flat magnitudes, and offsets of center frequencies. The results depend on the distribution of the mismatches among the antennas, and whereas there are many possible cases it is difficult to draw general conclusions. Attempts were made to identify the worst cases in order to provide some quantitative basis for design. Examples that produced peak residuals of 1% were found in [2] and those results are also given in Table 7.2 of [3]; see the references for descriptions of those examples.

Here we have repeated those calculations with a wider range of parameters. The results are given Figure 4 and Table 1. For each type of deviation, the relative error in amplitude has been calculated for each baseline of a 6-antenna array. In this array, two antennas have ideal rectangular passbands; two have the specified deviation; and two have the specified deviation in an opposite sense. In spite of the small number of antennas, this model is reasonably representative of a large array. For the passband ripple case, the calculation is done for 1 cycle of ripple across the passband, the worst case. With more ripples, the error stays about the same for the same peak-to-peak gain variation.

In practice, calibration errors of the type considered here can be mitigated by data processing techniques. First, determining the antenna gains can be avoided entirely if a sufficiently strong calibration source can be used to measure the gain of each baseline. However, the antenna gain estimates are less noisy than the baseline gain estimates by about $\sqrt{A - 1}$ for A antennas due to noise averaging, so stronger calibrators are needed to achieve the same accuracy. Also, the assumption that baseline gains can be factored into antenna gains is fundamental to the “self calibration” technique [4], which is important in many circumstances. For these reasons, baseline-by-baseline calibration is effective only in special cases. A second approach to mitigation of the errors depends on the availability of spectroscopic measurements of both calibration sources and target sources. Since modern correlators analyze the complex cross-power of each baseline into many separate “spectral channels,” it is in principle possible to solve for the complex antenna gain for each such channel separately. Over the narrow bandwidth of these channels, the passband shapes are determined mainly by the digital cross-correlation process and are thus very well matched among antennas. However, the narrow bandwidth also leads to poor SNR on most continuum calibrators, so this technique usually cannot be used directly. If all spectral channels are averaged to improve the SNR, then this is equivalent to having a continuum-only correlator. By selective averaging of channels, a good compromise between SNR improvement and mismatch error reduction can often be achieved. For example, if most of the bandpass mismatch is at the band edges, then the outer channels can simply be omitted from the average (and also ignored in observations of target sources); or the bandwidth can be partitioned into two or more sections, within each of which the channels are averaged and for which antenna gains are separately determined.

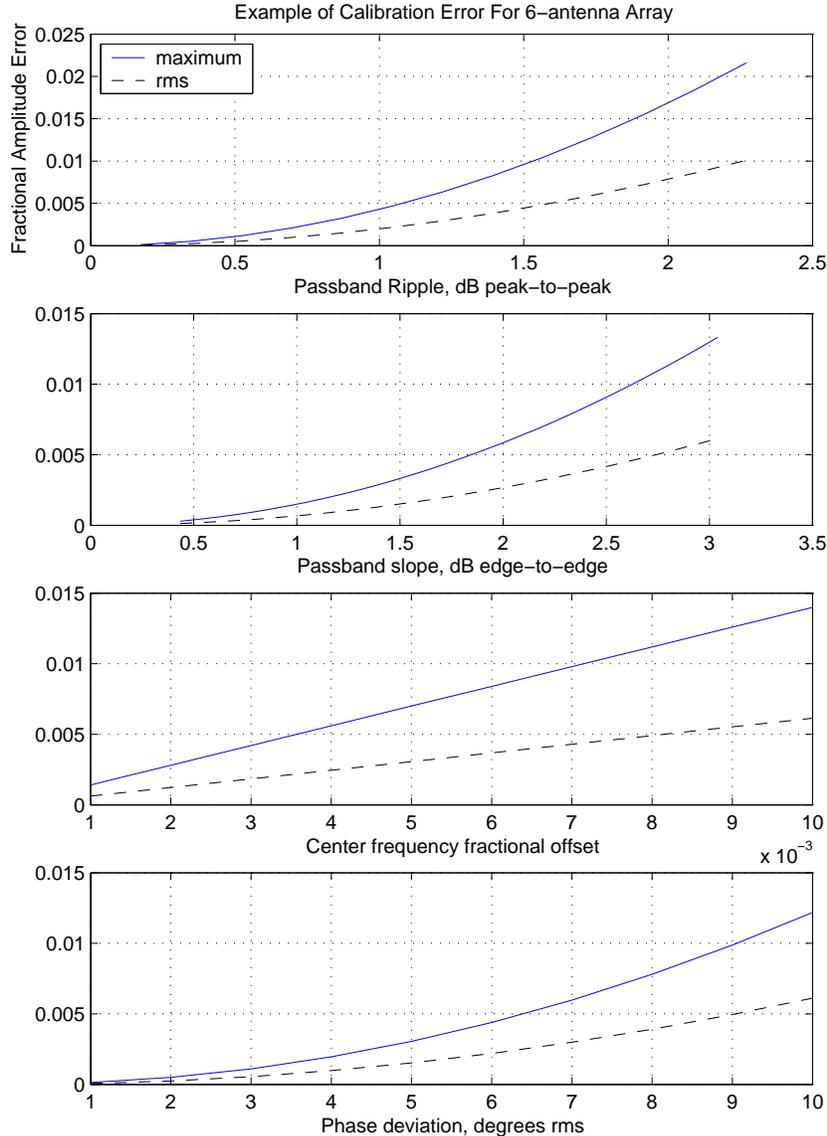


Figure 4 Calibration errors for a model array of 6 antennas with various deviations from the ideal passband.

Finally, a third approach also depends on having spectroscopic data and furthermore on the assumption that each antenna's gain can be described by

$$H_i(f, t) = h_i(t)b_i(f) \quad (7)$$

where $h_i(t)$ is independent of frequency and $b_i(f)$ is stable in time. Then separate calibration observations can be used to determine the two factors. First, a very strong calibration source is used to solve for the channel-by-channel gains at one epoch t_0 . These solutions should have residuals limited only by SNR and not by bandpass mismatches. Then subsequent observations of weaker calibrators, made more often and interspersed with observations of target sources, are first divided by $H_n(f_k, t_0)H_m(f_k, t_0)^*$ for baseline m, n and spectral channel k , thus removing the effect of mismatches over the full bandwidth, before averaging of the spectral channels to obtain adequate SNR. The averages are then used to solve for the antenna gains, which should now be free of residuals from bandpass mismatches. This yields essentially the time-dependent part of the gains, $h_n(t)$. The second and third techniques described here can be combined.

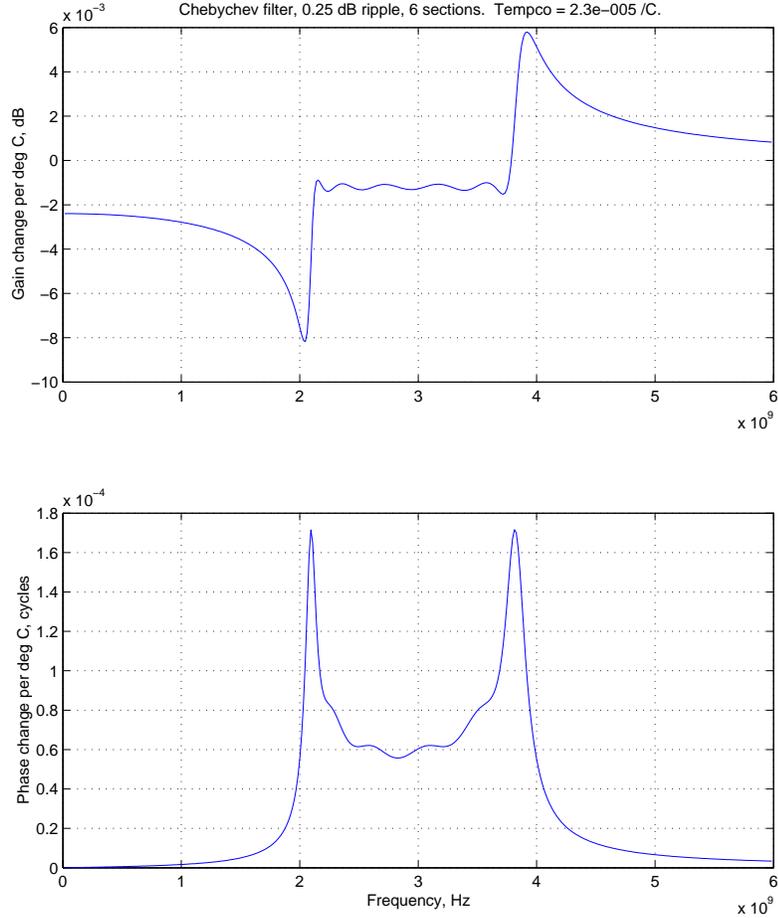


Figure 5 Calculated temperature sensitivity of a 6th-order, 0.25 dB ripple octave-bandwidth filter made of aluminum. The cutoff frequencies are 2.1105 and 3.7905; the results are independent of frequency scaling.

In this section we have considered only continuous-time signal processing. Since our examples have all had strictly zero gain outside the Nyquist interval, the results for sampled data will be identical provided that $f_s \geq 2B$. For practical systems with finite gain slope at the band edges, the results cannot be significantly different if the energy outside the Nyquist interval is small.

Temperature Sensitivity

The temperature coefficients of analog electronic devices are generally larger when they have higher Q or smaller fractional bandwidth. Partly for that reason, modern radio telescopes are designed with very broad band devices whenever possible. Filters used to define passbands can be exceptions to this. For those that select a Nyquist interval just before sampling, it is best to use the lowest numbered interval N , and to operate at baseband ($N = 0$) if possible.

If a filter is made entirely out of a single material, the temperature coefficient of its electrical properties can be predicted from the coefficient of expansion of the material. Otherwise, with mixtures of materials, the behavior is hard to predict. Near room temperature, the expansion coefficient of aluminum is 2.3×10^{-5} and for copper it is 1.7×10^{-5} . In Figure 5, the temperature coefficients of gain and phase are calculated for a 6th order Chebychev 0.25 dB filter for the 2 to 4 GHz interval ($N = 1$), assuming that it is made of pure aluminum. Due to the large fractional bandwidth, the coefficients are very small and should not be a significant design consideration. This calculation is done by assuming that the frequencies of all poles and zeros of the theoretical filter response scale inversely with the filter dimensions, which vary linearly with temperature. The responses are calculated at two different temperatures and subtracted.

Table 2: Recommendations for ALMA

Final filter sharpness	-20 dB/-3 dB width ≤ 1.21 (est. order: 7) $B_{-20\text{dB}} = 1.78\text{GHz} = 89\%$, $B_{\text{eff}} = 1.86\text{GHz} = 93\%$	
Tolerance on -3 dB frequencies	.001 B (1 MHz)	< 0.2% maximum gain error
Amplitude Matching	± 0.5 dB deviation from average after removing exponential slope	0.2% gain error
Phase Matching	$\pm 4^\circ$ rms deviation from average after removing linear slope	0.2% gain error
Gain Slope	< 3 dB edge to edge < 1.2 dB diff. between antennas	< 2% SNR loss 0.2% gain error
Ripple	< 2.0 dB peak-peak	< 2% SNR loss

Recommendations For ALMA

ALMA operates with 8 channels per antenna, each occupying the first Nyquist interval at 2 to 4 GHz. Recommended overall specifications for each such channel and for the matching of channels among antennas are given in Table 2.

It is assumed here that the passband shape is mostly determined by the final filter, since everything ahead of it has larger bandwidth. Therefore, all of the above recommendations except the slope and ripple can be taken to be specifications for the final filter.

The filter’s sharpness of cutoff is limited so that it can be realized in practice by a Chebychev I design of 8th order or less. This is primarily because the matching of higher-order filters is expected to be very difficult, and our matching requirements are fairly stringent. Cost also increases rapidly with filter order due to the labor required in final adjustments. A theoretical passband ripple of 0.5 dB, as used to calculate Figure 2, is believed to be too large a contribution to the overall system ripple; yet to achieve sufficient sharpness, a maximally flat (Butterworth) design would require too high an order, so something in between should be selected. As illustrated by Figure 5, the temperature coefficient should not be a significant consideration if the filter is made primarily of one metal.

Figure 6 shows the response of an ideal filter (lossless and perfectly accurate) that meets these requirements. It is a 6th order 0.25 dB ripple Chebychev I type. The -3 dB and -20 dB bandwidths are 1.785 GHz and 2.15 GHz, respectively, and the gain at the 2 GHz and 4 GHz edges of the Nyquist interval is -13 dB. The effective bandwidth and the 10, 20, and 30 dB suppression bandwidths, as fractions of the 2 GHz Nyquist bandwidth, are 93%, 95%, 89%, and 77%, respectively. A practical filter of higher order will be needed to achieve the same sharpness.

The matching requirements are chosen to keep the calibration errors less than about 0.2%. This is already difficult. To achieve the higher accuracy that might be desired for high dynamic range imaging, we must count on using spectroscopic calibrator observations and channel-by-channel bandpass corrections, as described earlier under “Bandpass Matching.” For the final filters, adequate matching of amplitude and phase should be achievable by a tight tolerance on just the -3 dB frequencies, assuming that all filters are of the same type. Figure 6 shows that the phase and delay of the theoretical filter vary substantially over the passband; the matching of these is expected to be similar to that of the cutoff frequencies, which should be adequate.

Phase and amplitude mismatches caused by components other than the final filters are problematical. Table 2 gives limits on total slope and ripple for any one channel that are chosen only to avoid degradation of the effective bandwidth, and are thus allowed to be rather large. The differences among antennas must be controlled more tightly, and this is included in the recommended limits on amplitude and phase matching.

Gain mismatches that vary rapidly across the bandwidth are the most difficult to mitigate in data processing. For this reason, mismatched ripples with multiple cycles across the band may require tighter limits than those given for amplitude and phase in Table 2, and the limits for ripples with 1 to 2 cycles may be slightly relaxed.

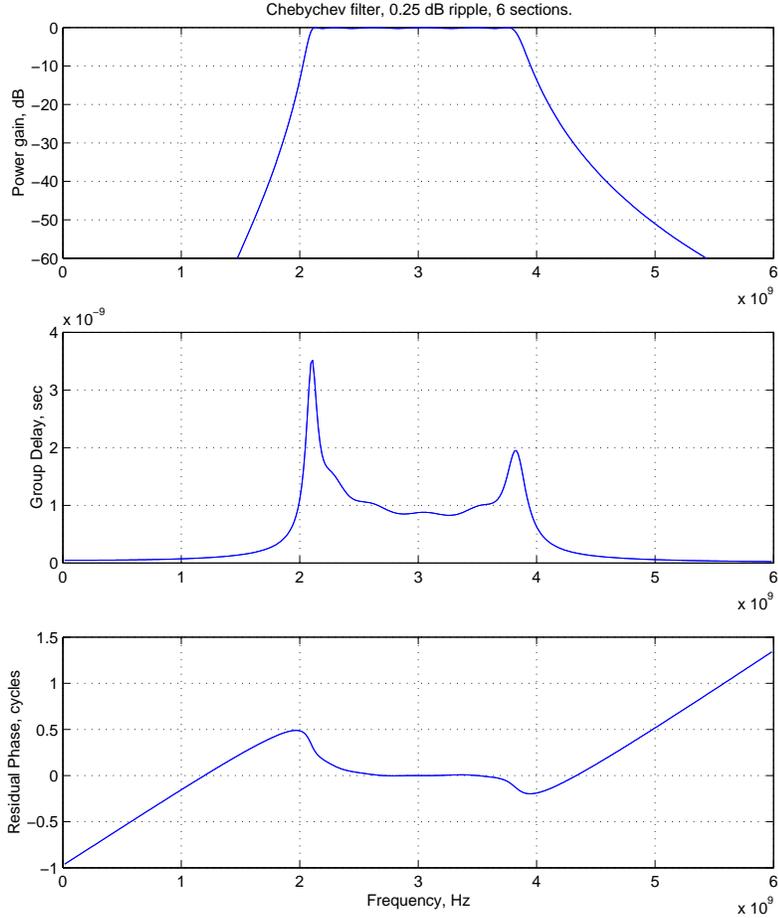


Figure 6 Transfer function of a 6th-order, 0.25 dB ripple octave-bandwidth filter. (Same filter as Figure 5.) The residual phase was obtained by subtracting the group delay at band center from the phase of the transfer function.

Acknowledgement

Calculations for this report were done in MATLAB 6.1 using MATLAB functions written by the author. Copies of the source code are available on request.

REFERENCES

- [1] J. Kraus, *Radio Astronomy*, 2nd ed., Cygnus-Quasar Books, Powell, OH, 1986. Eqn (7-14), quoting M. Tiuri, "Radio astronomy receivers," *IEEE Trans. Ant. Prop.* **AP-12**, 930–938, 1964.
- [2] A. Thompson, J. Moran, and G. Swenson Jr., *Interferometry and Synthesis in Radio Astronomy*, 2nd ed., Wiley:New York, 2001. Section 7.3.
- [3] A. Thompson and L. D'Addario, "Frequency response of a synthesis array: performance limitations and design tolerances." *Radio Sci.*, **17**, 357–369, 1982.
- [4] A. Thompson *et al.*, *op cit.*, section 11.4 of ref [2].

Appendix A: Derivation of Discrete-Time Effective Bandwidth

Let $n(t)$ be a real, zero-mean, Gaussian-distributed, stationary random function of time with variance σ^2 , and let it be regularly sampled at rate f_s producing samples $n_i = n(i/f_s)$. After observing a finite

number K of these samples, the power in $n(t)$ can be estimated by

$$\hat{P} = \frac{1}{K} \sum_{i=1}^K n_i^2. \quad (\text{A1})$$

This estimate is unbiased since its expected value is the actual power: $\langle \hat{P} \rangle = P = \sigma^2$. The variance of the estimate is then

$$\begin{aligned} \text{var } \hat{P} &= \langle (\hat{P} - P)^2 \rangle \\ &= \langle \hat{P}^2 \rangle - P^2 \\ &= \left\langle \left(\frac{1}{K} \sum_i n_i^2 \right)^2 \right\rangle - \sigma^2 \\ &= \frac{1}{K^2} \sum_i \sum_j \langle n_i^2 n_j^2 \rangle - \sigma^2 \\ &= \frac{1}{K^2} \sum_i \sum_j \langle n_i^2 \rangle \langle n_j^2 \rangle + 2 \langle n_i n_j \rangle \langle n_j n_i \rangle - \sigma^2 \\ &= \frac{2}{K^2} \sum_i \sum_j R_{i-j}^2. \end{aligned} \quad (\text{A2})$$

Here we have used the theorem

$$\langle abcd \rangle = \langle ab \rangle \langle cd \rangle + \langle ac \rangle \langle bd \rangle + \langle ad \rangle \langle bc \rangle$$

valid for jointly Gaussian zero-mean random variables [A1], and we define

$$R_{i-j} = \langle n_i n_j \rangle = \langle n(i/f_s) n(j/f_s) \rangle = R(i/f_s - j/f_s) \quad (\text{A3})$$

where $R(\tau)$ is the autocorrelation function of $n(t)$. Since $n(t)$ is stationary, R depends only on the time difference τ and R_k depends only on the sample number difference $k = i - j$. We now assume that the spectrum of $n(t)$ is of non-zero and compact support, which means that it is confined to a limited region of frequency space and has non-zero bandwidth. Then $R(\tau)$ decays at large $|\tau|$, as does R_k at large $|k|$. For sufficiently large K , the inner sum in (A2) can then be extended over all j , so that the result reduces to

$$\text{var } \hat{P} = \frac{2}{K^2} \sum_{k=-\infty}^{\infty} R_k^2. \quad (\text{A4})$$

The relative standard deviation of the power estimate is then

$$\frac{\Delta P}{P} = \frac{\sqrt{\text{var } \hat{P}}}{\sigma^2} = \frac{1}{R_0} \sqrt{\frac{2}{K} \sum_k R_k^2} \quad (\text{A5})$$

where we have used $R_0 = \sigma^2$.

To relate this to the effective bandwidth, compare (A5) with the well known similar result for estimation of the power in $n(t)$ using continuous time processing when $n(t)$ has a flat rectangular spectrum of bandwidth B . In that case,

$$\left(\frac{\Delta P}{P} \right)_{\text{continuous}} = \sqrt{\frac{1}{BT}} \quad (\text{A6})$$

for an observation interval of length T . By replacing B with $B_{\text{eff,d}}$ and T with K/f_s and equating (A6) with (A5), we find that

$$B_{\text{eff,d}} = (f_s/2) \frac{R_0^2}{\sum_k R_k^2} \quad (\text{A7})$$

as claimed in (2).

REFERENCE

[A1] J. Wozencraft and I. Jacobs, *Principles of coimunication Engineering*, Wiley:New York, 1965, p. 205; and earlier works.