Signal Processing for the WASP Analog Autocorrelation Spectrometers

A.I. Harris
Department of Astronomy, University of Maryland
September 2001

1 Introduction

Rather than sampling a signal's spectrum directly at a number of frequencies, an autocorrelation spectrometer measures the signal's autocorrelation function at a number of time lags τ. The spectral density $S(f)$ and the autocorrelation function $R(\tau)$ are linked by Fourier transforms:

$$S(f) = \int_{-\infty}^{\infty} R(\tau)\, e^{-i2\pi\tau f}\, d\tau \tag{1}$$

and

$$R(\tau) = \int_{-\infty}^{\infty} S(f)\, e^{i2\pi\tau f}\, df\ . \tag{2}$$

From the hardware perspective it is best to think in terms of the quantity that the device actually measures. The correlator samples the time-average autocorrelation function over an integration time T with one or more time delays and multipliers,

$$R(\tau) = \frac{1}{2T} \int_{-T}^{T} v_{in}(t) \times v_{in}(t+\tau)\, dt\ , \tag{3}$$

an averaged product of the input signal's voltage v_{in} at the two times t and $t+\tau$. In the important case that the spectrum has a well defined shape that is independent of time, then the autocorrelation function also has a well defined shape. A well-behaved instrument measures the average of the autocorrelation function over a finite time, which should be a good estimate of the true shape. Since the autocorrelation of a real function is symmetrical, only the even part of the transform kernel is important and measurement of only the positive (or negative) signal lags is sufficient to recover the spectrum:

$$S(f) = 2 \int_{0}^{\infty} R(\tau)\, \cos(2\pi\tau f)\, d\tau\ . \tag{4}$$

It is common to extend the range of lags slightly into the negative (or positive) delays to insure an accurate measurement of the entire zero-lag fringe.

Analog autocorrelators take their name from the fact that the delay is provided by signal propagation along delay lines and the multiplication is produced in an analog transistor circuit (Figure 1). The voltages v_{in} in equation (3) are voltages of the input signal itself, which has frequency components ranging from near zero to over 4 GHz for the WASP spectrometers. A digital correlator follows the same mathematical scheme but digitizes the analog signals before multiplication in digital hardware. To match WASP's full 4 GHz bandwidth, a digital system would have to digitize the input signal at at least 8 GHz, which is not possible at present. Practical wideband digital systems subdivide the band into narrower subbands with microwave circuits, find the spectra of the subbands, and then stitch the subbands back together to cover the whole input band. In addition to requiring a considerable amount of microwave and high-speed digital circuitry, digital correlators must eliminate slight mismatches in the spectral baseline between subbands. Analog correlators are well suited to spectroscopy with moderate resolution over a single wide band.

The analog part of the spectrometer electronics ends after some signal conditioning at the multipliers' outputs, and the remainder of the processing, most importantly accumulation and transformation into the spectral domain, is done digitally. Digital computers intrinsically process discrete samples, so both the lag τ and frequency f take on discrete values whether the hardware samples discretely or continuously. In the computations, sums replace integrals and the expression for the autocorrelation function becomes:

$$r_m = s_0 + 2 \sum_{n=1}^{N-1} s_n \cos(2\pi\, m\Delta\tau\, n\Delta f)\ . \tag{5}$$

where $\Delta\tau$ is the step size in lag time and Δf is the width of a spectral bin within the spectrum. The index m runs over the M multipliers or independent measurements of $R(\tau)$, with the index n denoting the number of the spectral channel.

For WASP, the separation between frequencies is established during the phase calibration, and the spacing between individual multipliers sets the discrete time delays $\Delta\tau$. Since the distance between delays is the finest sampling of the waves along the delay lines, this also sets the spectrometer's bandwidth. Equation (2) shows that a monochromatic signal with spectrum $S(f) = S_0\delta(f - f_{sig})$ has an autocorrelation function $R(\tau) = 2S_0 \cos(2\pi\tau f_{sig})$. Two samples per wavelength is the minimum needed to unambiguously measure a cosine's wavelength, so the lag spacing is related to the maximum frequency by $f_{max} = 1/(2\Delta\tau)$. WASP's structure is a ladder of lines, with signals suffering a propagation delay in both directions (Fig. 1), so the "corner" frequency beyond which aliasing may occur is

$$f_c = \frac{1}{4\Delta\tau}\ , \tag{6}$$

where $\Delta\tau$ is the delay between multipliers along one line, or half of the total delay. For WASP1's multiplier separation of 10.16 mm, $\Delta\tau$ = 62.5 ps and the corner frequency is 4.0 GHz; for WASP2's 9.65 mm, $\Delta\tau$ = 59.52 ps and f_c = 4.2 GHz. Any signals with frequencies above f_c will alias back into the fundamental band, from high frequencies to low for $f_c < f < 2f_c$, then from low to high for $2f_c < f < 3f_c$, and so on. WASP contains an anti-aliasing filter which cuts off power well above f_c. As a compromise between band flatness near the filter's edge and aliasing, a small band ($\sim$ 60 MHz) of high frequencies is allowed to alias into the highest one or two spectral channels. A small amount of aliasing here is not a practical problem since the highest few channels usually contain some edge effects and are discarded.

2 Signal Processing

It is convenient to treat the spectrometer data as vectors, or ordered sets of numbers. The advantage to this approach is that each operation on the data is then a matrix multiplication, so it is easy to understand the processing. This approach is also useful for writing simulation software since inverses of the matrices allow inverse processes. In practice, it may be simpler or significantly more efficient to perform the computations in other ways; many of the operations involve matrices that are diagonal or very sparse.

Noting that the multiplier output voltages are proportional to input power (v_{in}^2), but otherwise neglecting constant unit conversion factors, the elements of the average autocorrelation

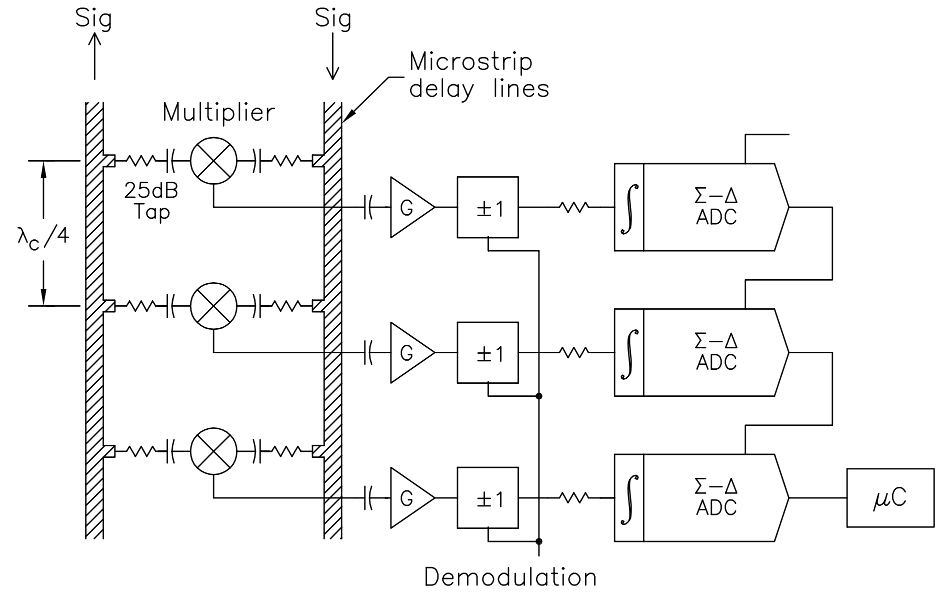

Figure 1: Schematic view of the WASP correlator lags. Sections of microwave stripline provide propagation-time delays between fast transistor multipliers.

function row vector $\mathbf{r}$ are the M output voltages from WASP's analog-to-digital converters, $\mathbf{r} = (v_0, v_1, \ldots, v_{(M-1)})$. The elements of the spectrum row vector $\mathbf{s}$ are the average power levels in N frequency bins across the spectrum, $\mathbf{s} = (p_0, p_1, \ldots, p_{(N-1)})$. The two are connected by the expression

$$\mathbf{s} = \mathbf{rXTCB}G_\alpha \, . \tag{7}$$

In successive operations on $\mathbf{r}$ (right to left with the choice of $\mathbf{r}$ and $\mathbf{s}$ as row vectors), $\mathbf{X}$ reshuffles the data from the correlator into a time-ordered series, $\mathbf{T}$ transforms between time lag and frequency spaces, $\mathbf{C}$ corrects for the passband gain and calibrates the spectrum in units of antenna temperature, and $\mathbf{B}$ bins or smooths the spectrum to its final spectral resolution. A scalar factor G_α is the gain of the attenuator that sets the spectrometer's input power level. WASP's software carries the data through all operations but the binning or smoothing operation $\mathbf{B}$ in equation (7); this last step is done interactively in the data analysis program CLASS. It is straightforward to add other operations to equation (7). As an example, a frequency-dependent gain factor maps only a single channel onto itself, so it is a pure diagonal $N \times N$ matrix $\mathbf{G}$ with the gain factors running along the diagonal.

2.1 Readout Order, $\mathbf{X}$

The sequence of the serial data stream from WASP2's analog-to-digital converters (ADCs) is a shuffled version of the delay-ordered data from the multipliers. It is useful to save the output from the spectrometer in delay-ordered form for diagnostic purposes; this is the purpose of the reordering operator $\mathbf{X}$ in equation (7). In order from index 0 to 15, the correspondence is:

ADC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Lag	1	3	5	7	9	11	13	15	14	12	10	8	6	4	2	0

In other words, the first datum is from the second multiplier (index number 1 in WASP's zero-base counting system), then the fourth, sixth, and so on to the last multiplier; then backwards from the fifteenth, thirteenth, and so on until the first multiplier. This pattern repeats for eight boards, with each board's data in the stream starting with its second ADC. This is equivalent to adding a board index b to the top and bottom rows of the table above, with $b = 0, 15, 31, \ldots, 16 \times l - 1, \ldots, 111$ for eight 16-lag boards.

As a matrix for a single 16-lag board, $\mathbf{X}$ would be 16×16 with elements $x_{1,0}$, $x_{3,1}$, $x_{5,2}, \ldots$, $x_{0,15} = 1$ ($x_{Lag,ADC}$ from the table above) and all other elements equal to zero. A 128-lag system would have a 128×128 matrix with sub-matrices of the single-board 16×16 matrices forming the diagonal, or nonzero elements $x_{1+b,0+b}$, $x_{3+b,1+b}$, $x_{5+b,2+b}, \ldots, x_{0+b,15+b} = 1$ where $b = 0, 15, 31, \ldots, 111$.

For reference, the lag-to-ADC correspondence is:

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ADC	15	0	14	1	13	2	12	3	11	4	10	5	9	6	8	7

In matrix form, this table would be replaced by the matrix $\mathbf{X}^{-1}$.

2.2 *Frequency Calibration and Spectrum Recovery,* T

Generalizing equation (2), a spectral density $S(f)$ produces an averaged multiplier output voltage $r(\tau)$ at lag τ of

$$R(\tau) = 2 \int_0^\infty S(f)\, k(\tau, f)\, df\,, \tag{8}$$

where $k(\tau, f)$ is the transform kernel. The working units of $R(\tau)$ are counts from the analog-to-digital converters connected to the multiplier outputs, which are proportional to power in the multiplier's input signal. In matrix notation for sets of discrete frequencies and time lags equation (8) becomes

$$\mathbf{r} = \mathbf{s}\mathbf{K}\,. \tag{9}$$

We use a set of basis vectors established by measurements of monochromatic signals instead of computing them from the theoretical Fourier transform kernels. Figures 2 and 3 show ideal and measured kernels for comparison. Although the two methods may seem unrelated, both rely on the basic principle of Fourier transforms and series: any signal may be decomposed into a family of appropriate monochromatic signals.

The autocorrelation function vector $\mathbf{r}$ has M elements, one for each lag; the elements of the spectral vector $\mathbf{s}$ represent power at N discrete frequencies; and the transformation matrix $\mathbf{K}$ is $N \times M$. We calibrate WASP with a family of continuous wave signals from a microwave synthesizer. A family of monochromatic signals with unit power is represented by the family of vectors containing delta functions in frequency: $\mathbf{s}_1 = (1,0,0,\ldots,0)$, $\mathbf{s}_2 = (0,1,0,\ldots,0)$, and so forth. Unit magnitudes and this ordering of frequencies is not fundamental but will prove convenient. Measurements of N of these row vectors can be easily handled by combining them in an $N \times N$ array:

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_0 \\ \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{(N-1)} \end{bmatrix} = \begin{bmatrix} 1\,0\,0\,\ldots\,0 \\ 0\,1\,0\,\ldots\,0 \\ \vdots \\ 0\,0\,0\,\ldots\,1 \end{bmatrix}. \tag{10}$$

The set of output voltages is then another array of vectors with M lag elements at N frequencies,

$$\begin{bmatrix} \mathbf{r}_0 \\ \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_{(N-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_0 \\ \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{(N-1)} \end{bmatrix} \mathbf{K} \qquad \text{or} \qquad \mathbf{R} = \mathbf{S}\mathbf{K}\,. \tag{11}$$

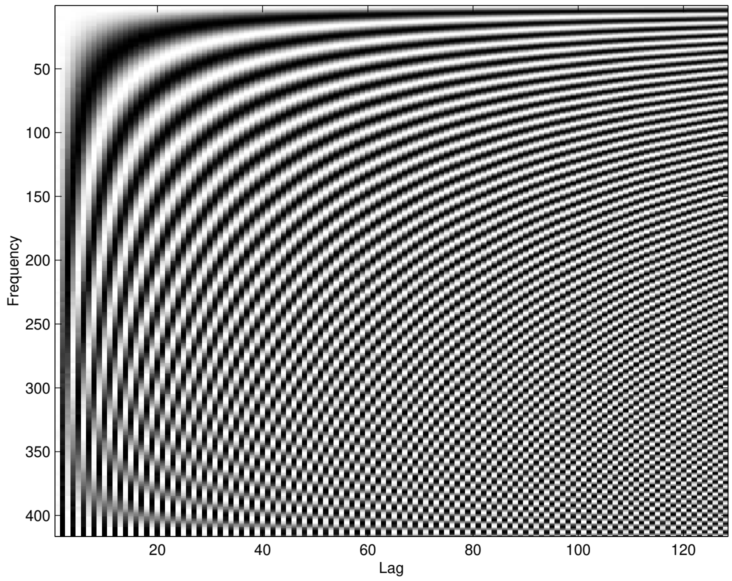

Figure 2: Image of an ideal Fourier transform kernel matrix **K**. The x-axis denotes lag number 0–127; the y-axis indicates samples at different frequencies. The wavelength of the autocorrelation function decreases from zero at the top of the plot to alternating peaks and valleys at the highest frequency at the bottom of the plot. The zero time lag falls in lag 0 in this ideal case.

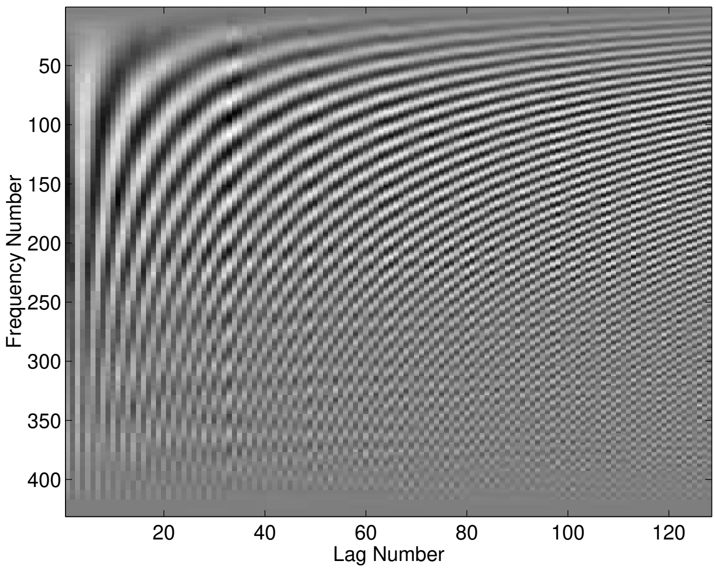

Figure 3: Image of a kernel matrix **K** from the WASP2 spectrometer. The y-axis shows steps in calibration frequency, with 431 samples from 100 MHz (long wavelength correlation functions) at the top to 4400 MHz (short wavelengths) at the bottom. The nearly vertical light fringe near lag 3 is the zero time lag position. Vertical structure near lag 35 shows multipliers with slightly higer responsivity than the others. Phase discontinuities between the eight 16-lag boards would appear as vertical structures at lags near multiples of 16 but are not present. Passband gain changes cause the intensity modulation from top to bottom.

Given a known set of $\mathbf{S}$ and measured voltages $\mathbf{R}$, it is simple to solve for the transform, $\mathbf{K}$:

$$\mathbf{K} = \mathbf{S}^{-1}\mathbf{R}\,. \tag{12}$$

It is easy to invert any $\mathbf{S}$ of individual nonrepeated monochromatic frequencies, but a simplification is possible by choosing unit magnitudes and the ordering of frequencies sequentially, as above. In this case $\mathbf{S} = \mathbf{I}$, the unit matrix, which is its own inverse and drops out of the right hand side of equation (12), leaving $\mathbf{K}$ equal to the measured outputs $\mathbf{R}$ without further calculation.

Once $\mathbf{K}$ has been obtained, it is possible to recover the unknown spectrum from the observed voltages at each lag:

$$\mathbf{s} = \mathbf{r}\mathbf{K}^{-1}\,. \tag{13}$$

Comparing this result with equation (7) shows that the transform $\mathbf{T}$ is equal to $\mathbf{K}^{-1}$, the inverse of the measured kernel matrix.

In a real measurement, the power from the calibration source may vary somewhat across the band, so the input matrix $\mathbf{S}$ is diagonal but is not equal to uniformly scaled unit matrix $\mathbf{I}$. The transform matrix is then $\mathbf{T} = (\mathbf{S}^{-1}\mathbf{R})^{-1} = \mathbf{R}^{-1}\mathbf{S}$. Calibrating the spectral amplitude separately retains the spectrometer phase from the monochromatic signal measurements but eliminates the dependence on the amplitudes within $\mathbf{S}$. As long as the calibration source's output power is roughly constant, an image of the measured matrix $\mathbf{R}$, such as the example in Figure 3, is quite a useful diagnostic tool.

In practice, we oversample in frequency during the spectrometer's calibration, covering a very wide band with frequency spacing a few times finer than the spectrometer's resolution. The main reason for this is to make calibration measurements at all frequencies where the spectrometer electronics have any significant response. Otherwise, the spectral recovery will alias signals that are outside the calibration band but are still within the spectrometer's response band: it can only assign components of the autocorrelation function to the measured frequencies in the calibration. The spectra produced with this scheme are as overresolved as the calibration, so only spectra smoothed or binned back to at least the frequency resolution set by the sampling theorem have a unique meaning.

A consequence of the oversampling is that the basis functions are overconstrained by the measurements, or $N > M$ in $\mathbf{K}$. A formal inversion of $\mathbf{K}$ is possible but is not particularly useful. If it were not for measurement noise, an oversampled measurement would contain redundant information and $\mathbf{K}$ would be singular. With noise, the matrix is merely nearly singular, with its inverse composed of a delicate combination of differences of large numbers. We invert $\mathbf{K}$ using singular value decomposition; the *Numerical Recipes* books contain a clear description of the method, and we use their algorithms. As part of the inversion, we set a magnitude cutoff that removes large and unphysical vectors in the reconstruction by setting them equal to zero within the range of the inversion. In spectra, these "missing" vectors correspond to frequencies where the spectrometer's power response is low to start with, so little information is lost in the spectrometer band (Figure 4). We find that keeping about 85% of the vectors in the inversion is best for the WASP1 spectrometer. Decreasing the number of discarded vectors increases the amount of power scattered across the baseline by residual numerical and phase errors. Increasing the number begins to noticeably eliminate signals that appear in low-gain parts of the main spectrometer band.

Inverting with singular value decomposition changes the information in the matrix once one or more vectors are discarded. Once a vector has been dropped, the computed inverse of the kernel matrix, $\mathbf{T}$, is not exactly equal to the inverse of the measured kernel $\mathbf{K}$: $\mathbf{TK}$ is not

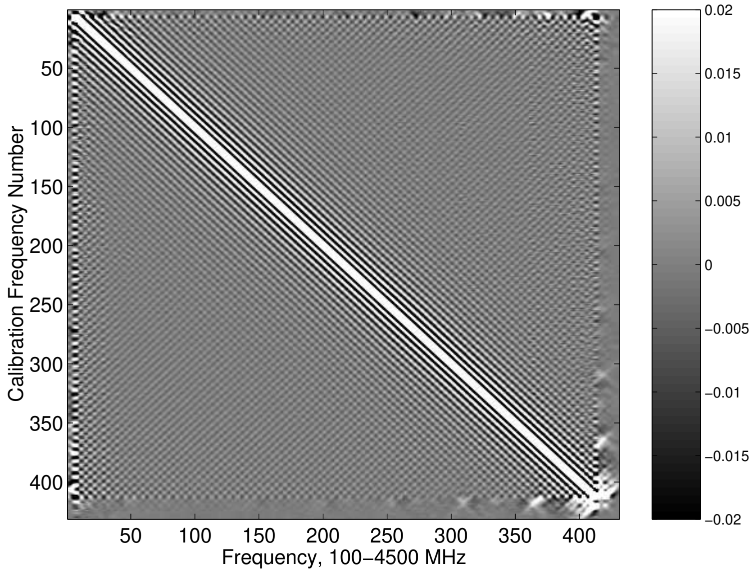

Figure 4: Image of the monochromatic signals recovered from the spectrometer, clipped about zero at ±2% of the peak value to show low-level structure. Each spectrum runs along a row in this figure, stacked from top to bottom in increasing frequency. The diagonal band through the figure is the peak of each spectrum, with the sidelobes from the spectrometer's finite length appearing as fainter stripes parallel with the main diagonal. Each spectrum has been smoothed to resolution slightly coarser than the spectrometer's highest possible resolution. The inversion of **K** retained 106 of 128 possible vectors, which suppressed the areas with low response near the band edges (see Figure 3).

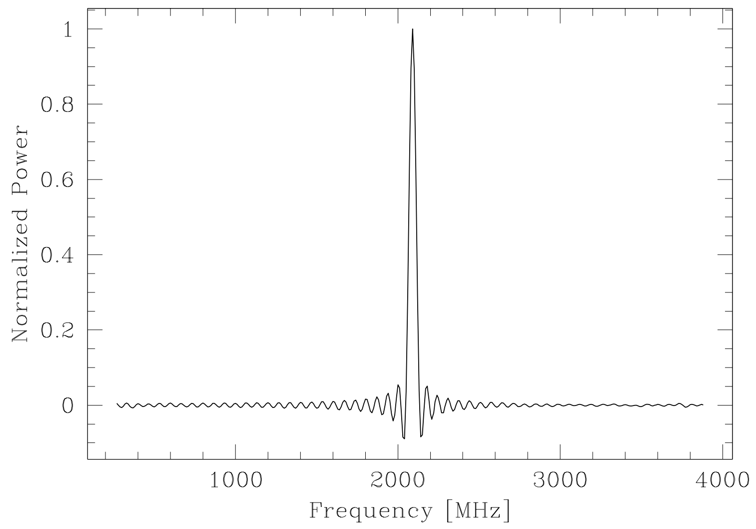

Figure 5: A single spectrum of a monochromatic signal from Figure 4. Smoothing or binning (apodization) will reduce the ringing at the expense of spectral resolution.

exactly equal to the unit matrix **I**, but it is as close as it can be, in a least-square sense, given the data and noise. It is this property that makes the SVD inversion robust. Figure 4 is the image of the calibration data times its own singular value decomposition inverse, smoothed to a resolution slightly coarser than the maximum. Information is missing from the lowest and highest frequencies, where the spectrometer has low response. Other than that, the matrix is very close to ideal. Ringing in the spectra, visible as bands parallel to the diagonal and shown in a sample spectrum in Fig. 5, is the consequence of truncation in the autocorrelation function measurement and occurs at the same levels as in an ideal transform. The ringing is less pronounced for spectra with broad features, which have little power at long lags.

2.3 Astronomical Signal Amplitude Calibration, **C**

At the Caltech Submillimeter Observatory we calibrate the astronomical data intensities with the chopper-blade method. This method assumes that the atmosphere, telescope, and spillover are at the same temperature but have different emissivities (the sky, for instance, emits as a low emissivity greybody with temperature equal to the telescope's). The advantage to this scheme is that it is simple, requiring measurements of only the sky and an ambient temperature load (the "blade") to make an estimate of the signal's strength at the top of the Earth's atmosphere T_A^*:

$$(T_A^*)_n = \frac{p_n(sig)}{p_n(blade) - p_n(sky)} T_{amb} , \tag{14}$$

where $p_n(sig)$ is the power in spectral bin n for the chopped and nodded spectrum, $p_n(blade)$ is the power from the spectrum of the blade, and $p_n(sky)$ is from a spectrum of the sky in the source's direction. It is prudent to switch off source in azimuth for the sky measurement if the source is bright. T_{amb} is the physical temperature of the blade, which we hope is the same as the ambient temperature provided by the CSO "scoreboard" program. With the same formalism, the system temperature referred to the same plane above the atmosphere is:

$$(T_{sys}^*)_n = \frac{p_n(sky)}{p_n(blade) - p_n(sky)} T_{amb} . \tag{15}$$

Chopped or position-switched measurements subtract the sky's temperature, leaving the difference in temperature between the two beam positions. Both the antenna and system temperatures affect one channel alone, or a pure diagonal matrix. This is an example of a correction where the operation is simple enough that a matrix formulation may be more effort than it is worth.

The full expression for the chopper blade calibration includes terms for atmospheric transmission toward the source, t_a, the ratio of the atmosphere's physical temperature to ambient temperature, $a = T_{atm}/T_{amb}$, and the cosmic background radiation temperature:

$$(T_A^*)_n = \frac{p_n(blade)}{p_n(blade) - p_n(sky)} \frac{T_{amb}(1 - (1 - t_a)a)}{t_a} - \frac{p_n(blade)}{p_n(sky)} 2.7\,\mathrm{K} . \tag{16}$$

Chopped or position-switched measurements subtract the 2.7 K term in addition to the sky's temperature. The usual expression for chopper blade calibration incorporates the assumption that $a = 1$, although $a = 0.9$–0.95 is more realistic. This correction is important at submillimeter wavelengths, where t_a is not close to unity. Taking these into account typically improves the calibration by 15–20%.

2.4 Spectral Smoothing or Binning, **B**

We leave the spectral binning to processing with the CLASS interactive data analysis program to keep flexibility in the frequency resolution of the final spectrum, but it is easy to construct matrices for smoothing and binning. These matrices combine information from neighboring frequency bins with weighting reflected in the magnitudes of the elements to either side of the diagonal.

2.5 Attenuator Gain, G_α

An attenuator at the spectrometer's input sets the overall power level for the input signal. Most attenuators are calibrated on a logarithmic scale of decibels α, which is related to the linear gain factor G_α by:

$$G_\alpha = 10^{-\alpha/10}\,. \tag{17}$$

The minus sign in the equation above is necessary because attenuation is a loss, or 1/gain. A 3 dB attenuator, for instance, has $\alpha = 3$ and $G_\alpha = 0.501$, a gain of almost exactly 1/2.