CARMA Memorandum Series #57

Blank Sky Analysis and Statistics

Kevin P. Rauch and Demerese M. Salter (UMD)

March 20, 2012

ABSTRACT

This memo gives an overview of the blank sky task, data analysis details, and performance expectations. Details of the digital quantization schemes implemented by the CARMA correlator, which determine the expected noise levels, are also provided. We note that previous attempts to use noise source spectra for real-time bandpass calibration may have been biased by failure to correct for on-sky thresholding variations, which we find is necessary to accurately reproduce predicted noise statistics. Including these corrections, blank sky results generally agree with expectations to within $\sim 1\%$.

Change Record

Revision	Date	Author	Sections/Pages Affected
	Remarks		
0.1	2012-January-31	Kevin Rauch	
	Partial draft.		
0.2	2012-March-16	Demerese Salter	
	Added ‘Observation and Analysis Procedure’ section and appendix on MIRIAD analysis.		
1.0	2012-March-20	Kevin Rauch	
	Initial release.		

1. Blank Sky Observations

The CARMA blank sky task (project `ct004`) is one of several tasks run regularly to help assess array performance. It is used to exercise (over time) all possible combinations of correlator bandwidth (from 500 MHz to 2 MHz) and sample quantization (2-bit, 3-bit, or 4-bit) and collect on-sky, off-source (i.e., noise) visibility data for each. The primary uses of these data sets are to verify that RMS noise levels agree with expectations in general and to check for the presence of birdies or other anomalies in particular. This memo discusses the blank sky work flow, theoretical noise expectations, and the results of lab measurements and sample on-sky data sets.

2. Observation and Analysis Procedure

The CARMA wiki at `http://cedarflat.mmarray.org/twiki/index.php/Blanksky` maintains instructions for the observers on how to run the 'blanksky' array health task at the telescope, reduce data on-site, analyze the statistical results, and recognize problems in the noise statistics. To run a blank sky observation, type `run('ct004_blanksky', bw='BW31', bit='3', endtrack="12:00")` in the `sci1` SAC. Similarly, for `sci2`, the script name is `ct004_blanksky_sci2`, and since the correlator is limited to a bandwidth of 500 MHz and 2-bit quantization, only the endtrack keyword needs to be specified.

When project `ct004` is executed at the telescope, the script (which is located in the cedarflat directory `/array/rt/scripts/arrayHealth/`) searches the blanksky catalog for a bright calibrator above the horizon. This catalog currently consists of 4 calibrators (3C84, 3C273, 3C345, and 3C454.3) and their corresponding on-sky, off-source positions (3C84OFF, 3C273OFF, 3C345OFF, and 3C454.3OFF). The script integrates first on the noise source for 30 s, then on the calibrator for 3 min, and finally, on the nearby off-source, blank sky position for a period of 20 min, before repeating the entire cycle. For reliable statistics, at least one complete cycle should be completed during a blank sky observation. For blank sky observations, the correlator is tuned to a rest frequency in the USB of 95.0 GHz for `sci1`, 35.938 GHz for `sci2`, or 85.8286 GHz for CARMA-23. In addition, flux calibration, tsys, data blanking, and automatic flagging are turned off in the RTS pipeline. For best results, the MIRIAD data set should be filled using floats (as of March 2012 the default is to use scaled integers). This can be done by manually refilling the data with `/opt/rt/bin/sdpFiller` and specifying option `float=true`.

Analysis of the blank sky data and the overall health of the correlators is typically accomplished using the MATLAB software and a suite of scripts to plot the statistics of the recorded on-sky, off-source visibility data, including: minimum and maximum values, signal-to-noise (S/N) statistics, the standard deviation (RMS) of the amplitudes, and the mean values of the MIRIAD visibility data. These visibility statistics are calculated for all baselines, all channels, and separately for both the real and imaginary components of the data. The most recent version of the MATLAB scripts to perform these analyses are located in `/array/rt/blanksky/` on the cedarflat machines. In that directory, the analysis has been automated, such that the user can call the appropriate `blanksky_sci1` or `blanksky_sci2` shell script, followed by the name of the data set (which should be copied into the same blanksky directory) and the name of the off position (e.g. 3C84OFF). Once the script has completed, the relevant plots will be accessible on the web at `http://cedarflat.mmarray.org/blanksky/` where subdirectories are created according to each observation date.

In most cases, the observers will only need to concern themselves with the statistical analysis tools and plots provided automatically by the MATLAB code, in order to monitor the ongoing health of the correlator FPGA cards and to check for the presence of birdies or other anomalies. Thus, the CARMA wiki provides the best resource for the general user of blank sky data. However, in order to understand the theoretical noise expectations (provided as reference lines in the MATLAB plots), as well as the performance efficiency of the RTS data pipeline, we must obtain laboratory measurements of the correlator output before the sample blank sky data is passed through the pipeline and converted into a MIRIAD data file. Then, the statistics of the cross-correlation spectra from the correlator can be compared directly to those of the end-user's output MIRIAD visibility (noise) data.

In the absence of simultaneous laboratory data collection, the advanced blank sky user can perform a quick analysis of the blank sky pipeline-processed data correlation noise within MIRIAD (see Sect. 4 and Appendix A), and compare the results to those of an earlier laboratory and on-sky test (e.g., Table 3), where the results have already been shown to be consistent to within 1% of the expected values. The following sections describe the procedure and results of a simultaneous lab and sky test performed in Nov 2011.

3. Theoretical Noise Characteristics

The correlator outputs cross-correlation spectra (visibility data) at a fixed rate of 2 Hz. When the inputs sample uncorrelated noise, each channel of each baseline has an expected mean value of zero and an expected variance that depends on the number of bits used to quantize the digitized input samples from each antenna. The wideband COBRA correlator (`sci2`) is limited to 2-bit quantization whereas the spectral CARMA correlator (`sci1`) can employ up to 4-bit sampling in most observing modes.

The expected correlation statistics for generic quantization schemes are described in detail by Hawkins (2011). The results depend sensitively on the product rule for multiplying two quantized samples together. For 2-bit sampling, for example, the correlator utilizes a "deleted inner-product" rule that defines 1*1 = 0 when computing the cross-correlation product of 2-bit quantized samples. This is done to reduce hardware requirements and ultimately increase the output spectral resolution, at the cost of slightly reduced detection efficiency compared to full 2-bit multiplication. The implemented 3-bit and 4-bit quantization schemes—unique to CARMA—have been similarly fine-tuned to balance resolution and detection efficiency (see Appendix B).

Like the cross-correlation detection efficiency (defined as the signal-to-noise ratio of correlation coefficients derived using the quantized input samples, relative to the SNR that would have been obtained with infinite-precision sampling), the expected noise variance of correlation spectra depends on the details of the quantization scheme. The absolute normalization applied by the correlator to all data must also be taken into account. The blank sky analysis scales results such that if the observed noise variance matches expectations, then the scaled noise statistic $\hat{\eta}$ will equal the theoretical weak-correlation detection efficiency $\eta(0)$ for the tested mode. Explicitly, $\eta(0) = 0.8724$ (2-bit sampling), 0.9626 (3-bit sampling), or 0.9836 (4-bit sampling) for the CARMA correlator.

Controlled lab experiments have been performed to measure $\hat{\eta}$ for all possible correlator modes. For this analysis, the 2 Hz integration data produced by the correlator while integrating fully uncorrelated thermal noise was captured and analyzed directly, ensuring maximum precision and isolation from the rest of the system. The data were analyzed as follows. First, for each channel of each baseline the real and imaginary parts of the complex visibilities (one for each 500 ms integration collected) were grouped into a single statistical sample, and the mean and (sample) variance computed. For each baseline the individual variance values were then averaged over all channels (excluding the half-width end channels present in the raw data) to produce an average value for each baseline. The net observed noise variance for the band, $\hat{\sigma}^2$, was then obtained by averaging these over all baselines.

Theoretically the real and imaginary samples will be Gaussian random variables with zero mean and fixed variance σ^2, which can be computed analytically given the details of the quantization scheme and data normalization. The latter requires special attention. Correlator data are normalized by a Van Vleck correction, C, whose purpose is to remove the bias in the observed, quantized correlation coefficients, $\hat{\rho}$, and translate them into true correlation coefficients, $\rho = C\hat{\rho}$. At the time of writing (Jan 2012), all correlator data are normalized by the Van Vleck correction $C(1)$ appropriate for 100% correlated signals (such as the correlated noise source used for phase flattening)—even when observing weakly-correlated (i.e., astronomical) sources. This is scheduled to change in the near future, at which time $C(0)$, the theoretical Van Vleck correction for uncorrelated inputs, will be applied to on-sky data instead. This affects the conversion from $\hat{\sigma}^2$ to $\hat{\eta}$ and hence it is important to know which normalization was in effect when the data was taken.

The generic relation is $\hat{\eta}(\rho) = [C/C(\rho)]/\sqrt{2B\tau\hat{\sigma}^2(\rho)}$, where C is the applied Van Vleck correction, B is the bandwidth per channel, τ is the time per integration (*not* the total integration time), and ρ is the true correlation coefficient of the input signals. The factor of 2 arises from the Nyquist sampling criterion (two samples per cycle). For blank sky and these lab measurements, $\rho = 0$. It is important to note that τ is not identical to the integration time requested in the observing script, due to a phase switch blanking interval inside the correlator that is required to allow the analog signal to settle after each phase switch (which occur at a rate of 1024 Hz). For example, a nominal 1 s integration contains only 0.9728 s of actual data. However, the latter is reported properly by the system (e.g., in MIRIAD output) and no explicit correction factor is required during analysis. The values of $C(0)$ and $C(1)$ for each quantization scheme can be found in Appendix B.

The measured values $\hat{\eta}$ and expected values $\eta(0)$ for each correlator mode are shown in Table 1. Agreement is excellent, with worst-case discrepancies of a few tenths of a percent (excluding 500 MHz modes, which are known to be biased). It must be understood however that $\hat{\eta}$ itself is **not** an independent measurement of the actual correlator detection efficiency, because the applied Van Vleck correction C is a theoretical value only, not a measured quantity. Agreement between $\hat{\eta}$ and $\eta(0)$ merely verifies that the measured *noise* level is in line with expectations; determining $\eta(0)$ experimentally requires measuring both signal and noise levels and lies outside the scope of blank sky observations.

For additional discussion on correlator efficiency see the wiki page:

`http://www.mmarray.org/twiki/index.php/Correlator_efficiency`

4. On-sky Performance Comparison

The preceding lab experiments demonstrate that the digital correlator hardware performs as expected in all available operating modes (as far as RMS noise levels are concerned), but do not exercise the complete end-to-end signal chain and processing pipeline utilized by on-sky observations. To provide a secure reference point for normal blank sky measurements, we have performed a special set of blank sky observations in which 2 Hz correlation data was captured from the correlator simultaneously with the production of normal

MIRIAD output by the standard RTS pipeline. For this observation all available `sci1` antennas were commanded to track a blank patch of sky (specifically, NORTHPOL). The observing script disabled Tsys and flux scaling in the pipeline (noise analysis requires totally "raw" data), and requested that 64 x 8 s integrations be collected—about 8 minutes total integration time. This test was repeated in all possible correlator bandwidth and sample quantization modes; due to network bandwidth constraints, however, comparison 2 Hz data (captured directly from the correlator) was obtained for only a few of these: 250 MHz/2-bit, 125 MHz/3-bit, 8 MHz/4-bit, and 2 MHz/3-bit.

The 2 Hz data was analyzed precisely as described previously; the MIRIAD output was analyzed in a similar (but not identical) manner by constructing a data cube and using the `imstat` command to compute the noise RMS for each channel averaged over all times and baselines (the command collapses two dimensions at once). The channel RMS values were then averaged into a single band RMS for comparison with the 2 Hz results.

Before presenting the results we discuss several possible sources of disparity between them and when compared to theoretical expectations. The largest stems from the way in which digitizer threshold optimization (level control) is performed. Threshold optimization, which occurs after each bandwidth reconfiguration, normalizes the signal level of the digitized input samples to a fixed reference point maximizing the expected detection efficiency $\eta(0)$. This is done using a correlated noise source which is injected directly into the IF band downstream of the receiver output, which provides reliable output to all correlator RF inputs even when particular antennas or receivers are unavailable or otherwise misbehaving. Technically speaking therefore, the thresholds are optimal only when observing the noise source, not astronomical sources. However, automatic (analog) gain control in the downconverters (which filter the full IF into individual 500 MHz bands) ensures that the two types of RF input signals reach the correlator with equal power, nominally making them interchangeable for thresholding purposes. Unfortunately, the gain control circuit in the downconverters has a non-trivial frequency response, resulting in slightly different total output power depending on the input passband shape. As a consequence the thresholds end up being slightly non-ideal for on-sky observations; typical variations are a few percent. Note however that the impact on $\eta(0)$ is much less because we are operating near the peak of the efficiency curve; the change in *absolute* noise (and signal) level, on the other hand, is of the same order.

To first order thresholding variations can be corrected by normalizing the cross-correlation spectra by the autocorrelation amplitudes of the corresponding inputs; e.g., by dividing cross-baseline cross[A-B] by auto[A]$\cdot$auto[B] on a channel-by-channel basis. In MIRIAD this gain correction can be done by specifying `options=fxcal` to `uvcal`. In Table 2, we give the results both with and without this correction.

Table 1. Lab Experimental Correlation Noise Comparison

Bits	500 MHz	250 MHz	125 MHz	62 MHz	31 MHz	8 MHz	2 MHz	Average†	Expected
2	0.8454	0.8710	0.8720	0.8721	0.8744	0.8739	0.8720	0.8726	0.8724
3	0.9422	0.9575	0.9593	0.9616	0.9639	0.9634	0.9611	0.9611	0.9626
4	0.9751	0.9874	0.9876	0.9839	0.9858	0.9831	0.9844	0.9854	0.9836

†500 MHz noise results, biased by in-band artifacts, are excluded from the average.

Other differences arise in that the data sets being compared are not identical, nor (as a matter of convenience) analyzed in a strictly identical manner. The 2 Hz data contained 1024 x 500 ms integrations as opposed to 64 x 8 s integrations for the MIRIAD data. Grouping the former into 8 s packets for analysis changed the results by $\sim 0.1\%$, a negligible amount; for comparison, even grouping the data into a *single* 512 s integration changed the results by only $\sim 1\%$. The standard blank sky script utilizes 30 s integrations.

The 2 Hz data contained a full complement of baselines (105 for 15 antennas), but due to maintenance activity during the test only 45 baselines were present in the MIRIAD output, the remainder being masked by the pipeline. This should not bias the results on average, but can be expected to generate random differences in the individual trials. The fact that `imstat` collapses two dimensions of the data cube instead of one to compute the first RMS will act similarly. Experiments with the 2 Hz reduction suggest random variations of a few tenths of a percent. Finally, the RMS calculated by `imstat` is a population standard deviation, not the proper, statistically-unbiased sample standard deviation used by the 2 Hz reduction procedure; but for a sample size of $N \sim 45 \cdot 64 = 2880$, the two measures will agree to better than 0.1%.

Results for the scaled noise $\hat{\eta}$ for each data set, both raw and gain-corrected, are shown in Table 2. Overall agreement between the 2 Hz and MIRIAD results are very good, with typical differences of $\sim 1\%$ for raw results and $\sim 0.5\%$ for the corrected results. As anticipated the raw results deviate significantly from the expected value—by over 5% in some cases—whereas the gain-corrected results match to within 1%. We conclude that the simplified MIRIAD analysis procedure is sufficiently reliable to use as a reference for blank sky analysis.

CARMA has previously considered applying a real-time bandpass calibration to all correlation data by default, with normalization provided by noise source spectra. It was never enabled due to unresolved concerns over possible phase-closure artifacts. We speculate that failure to correct for thresholding variations via the auto-correlations, as done here, may have been involved. It would be interesting to repeat the previous calibration experiments with and without gain correction to test this hypothesis. The results should indicate the optimal method of bandpass calibration via the noise source, whether the solution is applied on- or off-line.

A. MIRIAD Code to Analyze Blank Sky Correlation Noise

This appendix describes the analysis of the Nov 2011 blank sky observation data set using MIRIAD, resulting in reference Table 3.

Table 2. On-Sky Correlation Noise Comparison

	— Raw —		— Corrected —		
Mode	$\hat{\eta}$[2 Hz]	$\hat{\eta}$[MIRIAD]	$\hat{\eta}$[2 Hz]	$\hat{\eta}$[MIRIAD]	$\eta(0)$
250 MHz 2-bit	0.8139	0.8011	0.8810	0.8826	0.8724
125 MHz 3-bit	0.9828	0.9732	0.9596	0.9595	0.9626
8 MHz 4-bit	0.9177	0.9049	0.9883	0.9895	0.9836
2 MHz 3-bit	0.9292	0.9401	0.9652	0.9870	0.9626

In the code below, we reference the `ct004.xmas31x8.1.mir` data set. This observation is 0.14 hours long and only looks at the source 'NORTHPOL'. Bands 1-3 consist of 31 MHz windows in 2-bit, 3-bit, and 4-bit quantization modes, respectively. Bands 5-7 are 8 MHz windows, also in each of the different bit quantization modes. Bands 4 and 8 are both 500 MHz/2-bit spectral bands that are ignored in the following analysis. We start by separating the data out window by window.

```
Miriad% set file = ct004.xmas31x8.1.mir
Miriad% foreach band ( 1 2 3 5 6 7 )
Miriad% uvcat vis=$file select='win('$band')' out=win.$band
Miriad% end
```

Next, we use the MIRIAD task `uvcal` to normalize the cross-correlation spectra by the auto-correlations to correct for thresholding variations. This step can be skipped in order to compare these 'gain-corrected' data with the 'raw' noise statistics.

```
Miriad% uvcal vis=win.1 options=fxcal out=UVCAL_31_2
Miriad% uvcal vis=win.2 options=fxcal out=UVCAL_31_3
Miriad% uvcal vis=win.3 options=fxcal out=UVCAL_31_4
Miriad% uvcal vis=win.5 options=fxcal out=UVCAL_8_2
Miriad% uvcal vis=win.6 options=fxcal out=UVCAL_8_3
Miriad% uvcal vis=win.7 options=fxcal out=UVCAL_8_4
```

We then use `uvimage` to convert the raw data into an image data cube of the uv data in a 'Time-Baseline-Channel' order. This will allow us to take advantage of several `imstat` utilities.

```
Miriad% foreach bw ( 8 31 )
Miriad% foreach bit ( 2 3 4 )
Miriad% uvimage vis=UVCAL_$bw\_$bit out=NORTHPOL_$bw\_$bit\_real.out
  view=real mode=3 line=chan,0,1,1,1 select="source(NORTHPOL),-auto"
Miriad% uvimage vis=UVCAL_$bw\_$bit out=NORTHPOL_$bw\_$bit\_imag.out
  view=imag mode=3 line=chan,0,1,1,1 select="source(NORTHPOL),-auto"
Miriad% imcat in=NORTHPOL_$bw\_$bit\_imag.out,NORTHPOL_$bw\_$bit\_real.out
  out=NORTHPOL_$bw\_$bit\_comb.out
Miriad% end
Miriad% end
```

We have created 3 separate data files: one with the real visibilities only, one with the imaginary visibilities only, and a data set that contains the real and imaginary data together. Now we can use `imstat` to calculate the statistics of the uv data 'image'. This includes the sum, mean, RMS, maximum, and minimum values of a region. We can save the output to a log file, or a plot.

```
Miriad% imstat in=NORTHPOL_$bw\_$bit\_comb.out plot=rms axes=x,y device=/xs
```

To convert these RMS values to the numbers listed in Tables 2 and 3, we need a little more information about the observation. We need to know the channel resolution of the data, which can be found using `uvlist` on the data set with `options=spec`. Just divide the bandwidth column by the number of channels. For the 31 MHz/2-bit data, this value is 81.3812 kHz. We also need to know the time integration steps for a record:

```
Miriad% uvio vis=$file | grep inttime
```

This should return just one number, unless the integration time has changed throughout the observation. In this example, the integration time is 7.782398701 s. Recall that the observations were set up for 8 s integrations, but that a settling time is required in the correlator after each phase switch (refer to Sect. 3). The output MIRIAD value is reported correctly. The system only integrates for 950 μs per phase switch (which occur at a rate of 1024 Hz); to get the expected integration time for your data, you must multiply $(950 \times 10^{-6}) \times 1024 \times t_{\mathrm{int}}$. Now we have all of the information to solve for the quantity $\hat{\eta} = [C(1)/C(0)]/[\hat{\sigma}\sqrt{2B\tau}]$.

The normalization factor $C(1)/C(0)$ is applied to the results to account for the correlation data scaling convention. It represents the Van Vleck correction appropriate for the data, $C(0)$, relative to the one actually applied, $C(1)$. In this case, just divide any 2-bit data by 0.9732, any 3-bit data by 0.9533, and any 4-bit data by 0.9256 to compare with the values listed in Table 3. Your numbers for a similar blank sky observation should be within $\sim 1\%$.

B. Correlator Quantization Schemes

This appendix provides details of the quantization schemes employed by the CARMA correlator system, two of which—the 3-bit and 4-bit variants—are unique to CARMA (as far as is known by their author [K. Rauch]). For general background on signal quantization in radio science applications, see Cooper (1970), Hagen & Farley (1973), and Thompson, Moran, & Swenson (2001).

An ideal digital correlator would digitize input analog signals to unlimited precision, compute cross-correlation coefficients exactly, and on average obtain ideal (100%) detection efficiency in the process: by definition, the cross-correlation detection efficiency $\eta(\rho)$ for signals with true cross-correlation coefficient ρ is the average signal-to-noise ratio of the computed correlation coefficient, $\hat{\rho}$, relative to the described ideal system.

Practical digital systems must quantize their inputs to a finite, typically small number of bits. For a digital lag correlator such as the CARMA correlator, hardware requirements (i.e., cost) for a given total bandwidth scale approximately $\propto N_{\mathrm{chan}} N_{\mathrm{level}}^2 N_{\mathrm{input}}^2$, where N_{chan} is the number of channels in the output spectra, N_{level} is the number of possible quantization levels (states) per sample, and N_{input} is the number of inputs (e.g., antennas) processed. The number of bits per sample is $\mathrm{ceil}[\log_2(N_{\mathrm{level}})]$. The quadratic dependence on N_{level} corresponds to the number of entries present in the associated quantized multiplication table. Increasing

Table 3. MIRIAD Observation Correlation Noise Comparison

Data	Bits	500 MHz	250 MHz	125 MHz	62 MHz	31 MHz	8 MHz	2 MHz
Raw	2	0.8584	0.8011	0.8766	0.7949	0.8553	0.7761	0.8643
Raw	3	0.9274	0.9522	0.9732	0.9041	0.9372	0.8505	0.9401
Raw	4	0.9450	0.9986	0.9948	0.9772	0.9438	0.9049	0.9325
Corrected	2	0.8358	0.8826	0.8675	0.8880	0.8759	0.8952	0.8987
Corrected	3	0.9461	0.9607	0.9595	0.9670	0.9627	0.9715	0.9870
Corrected	4	0.9804	0.9794	0.9794	0.9833	0.9854	0.9895	1.0109

$N_{\rm level}$ is desirable to maximize detection efficiency ($\eta(\rho) \to 1$ as $N_{\rm level} \to \infty$), but can only come at the expense of spectral resolution given $N_{\rm input}$ is normally a fixed design parameter.

Science requirements determine the minimum acceptable $N_{\rm chan}$ for a given total bandwidth and budget limitations determine maximum hardware capacity. As inputs from astronomical observations are nearly always weakly correlated, an optimal quantization scheme is one which maximizes the efficiency $\eta(0)$ within the imposed constraints. The value of $\eta(0)$ is set by three things which together define the quantization scheme: the value of $N_{\rm level}$, the spacing between levels (the precise analog values—here, voltages—defining the transitions between quantization states), and the quantized multiplication table (which need not follow simple arithmetic rules).

As part of development for the CARMA correlator a program (`multadd`) was written to search for optimal quantization schemes, with a focus on finding hardware-efficient multiplication rules. It can be found in the `carmacorl` CVS tree under share/fpga/test. The program is also able to calculate many ancillary properties of quantization schemes in the weak-correlation limit $\rho \to 0$, such as the detection efficiency $\eta(0)$ and Van Vleck correction $C(0)$. The CARMA correlator offers three sampling options: 2-bit (4-level), 3-bit (8-level), and 4-bit (15-level). The 2-bit scheme utilizes the popular deleted inner-product rule described by Cooper (1970). The 3-bit and 4-bit variants were derived with the aid of `multadd`.

For reference the characteristics of each sampling mode are provided in Table 4 (as direct output from `multadd`). The stated uncorrelated lag variance (for N samples) assumes data normalized by $C(1)$. Note that none of the listed product rules are commutative with normal multiplication; hence in particular $(\mathbf{X}\cdot\mathbf{Y})^2 \neq (\mathbf{X}\cdot\mathbf{X})(\mathbf{Y}\cdot\mathbf{Y})$, where $\mathbf{X}$ and $\mathbf{Y}$ are quantized values and '$\cdot$' is the quantized product operator. This has some important implications, such as in the behavior of the Van Vleck correction $C(\rho)$. By definition $C(\rho) = \rho/\hat{\rho}$, where $\hat{\rho} = \langle \mathbf{X}\cdot\mathbf{Y}\rangle$ is the quantized correlation estimate, and $\eta(\rho) = [\hat{\rho}/\hat{\sigma}(\rho)]/[\rho/\sigma(\rho)] = [\sigma(\rho)/\hat{\sigma}(\rho)]/C(\rho)$, where $\hat{\sigma}^2(\rho) = \langle(\mathbf{X}\cdot\mathbf{Y})^2\rangle - \langle(\mathbf{X}\cdot\mathbf{Y})\rangle^2$ and it can be shown $\sigma(\rho) = 1+\rho^2$. If $\rho = 1$ then $\mathbf{Y} = \mathbf{X}$ and hence $C(1) = 1/\langle\mathbf{X}\cdot\mathbf{X}\rangle$. In the limit $\rho \to 0$, $\hat{\rho} \to 0$ and $\hat{\sigma}^2(0) = \langle(\mathbf{X}\cdot\mathbf{Y})^2\rangle$. For any commutative/associative product rule (e.g., full multiplication) this reduces to $\hat{\sigma}(0) = \langle\mathbf{X}\cdot\mathbf{X}\rangle = 1/C(1)$, which implies $\eta(0) = C(1)/C(0)$ and an expected lag variance of $1/N$ (where N is the number of accumulated samples) for uncorrelated data normalized by $C(1)$. As seen below neither of these relations hold for the implemented sampling schemes. For the 4-bit scheme, it also happens that $C(\rho)$ is not monotonic (invertible) in the range $\rho \in [0.93, 1.00]$, meaning that $\rho(\hat{\rho})$ is double-valued there (where $\hat{\rho} \in [1.00, 1.03]$, for data normalized by $C(1)$). This does not impact its use with CARMA, which references only $C(0)$ and $C(1)$, nor any use whatsoever with sources less than 97% correlated (essentially all astronomical objects).

The optimized product rules all reduce the number of bits in the product compared to full multiplication with only a small drop in $\eta(0)$. For 2-bit sampling this is accomplished by zeroing low-level products and rescaling the results, shaving two bits off the product at the loss of $\sim 1\%$ efficiency (from 0.8812 to 0.8724). For 3-bit and 4-bit sampling, optimization is based mainly on constraining outer products. The 3-bit rule also removes two bits from the product but with negligible efficiency loss, $\sim 0.0001\%$, achieved in part through the use of non-uniform level spacing. Finally, the 4-bit (15-level) scheme removes two bits and loses 0.355% efficiency.

Table 4: CARMA Correlator Quantization Scheme Properties

```
i       quantized sample index           (zero/positive weights only)
w       quantized sample weight
x       quantized sample threshold range  (in units of v0)
prod    quantized sample product table    (zero/positive quadrant only)
hist    quantized sample histogram        (zero/positive weights only)
bias    product offset to ensure unsigned integer representation
eta0    weakly-correlated signal to noise relative to ideal
lag0    expected auto-correlation zero lag
v0      threshold scale  (assuming unit input variance)
Ex var  expected quantized sample variance
Ex kurt expected quantized sample kurtosis
VV C(0) Van Vleck correction for   0% correlation
VV C(1) Van Vleck correction for 100% correlation

Four level quantization with 3-bit (deleted inner) products:
i = 0:  w =  0.5  x = [0.0000, 1.0000]  prod = 0 1
i = 1:  w =  1.5  x = [1.0000,    Inf]  prod = 1 3
hist = {0.31763, 0.18237}
bias =  3, eta0 = 0.872446, lag0 = 1.0942, v0 = 0.906369 (1.3596 @ full value)
Expected variance = 0.97948, kurtosis = -1.03394
Expected uncorrelated lag variance = 1.38704/N
Van Vleck C(0) = 0.88943, C(1) = 0.91389, C(0)/C(1) = 0.97323

Eight level quantization with 5-bit (hand optimized) products:
i = 0:  w =  0.5  x = [0.0000, 1.0000]  prod =  0  1  1  2
i = 1:  w =  1.5  x = [1.0000, 2.0056]  prod =  1  2  4  6
i = 2:  w =  2.5  x = [2.0056, 3.1914]  prod =  1  4  6 10
i = 3:  w =  3.5  x = [3.1914,    Inf]  prod =  2  6 10 15
hist = {0.20156, 0.15404, 0.09869, 0.04572}
bias = 15, eta0 = 0.962559, lag0 = 3.1719, v0 = 0.528884 (1.8511 @ full value)
Expected variance = 3.14763, kurtosis = -0.67693
Expected uncorrelated lag variance = 1.18763/N
Van Vleck C(0) = 0.30054, C(1) = 0.31526, C(0)/C(1) = 0.95330

Fifteen (fixed) level quantization with 5-bit (hand optimized) products:
i = 0:  w =  0.0  x = [0.0000, 0.5000]  prod =  0  0  0  0  0  0  0  0
i = 1:  w =  1.0  x = [0.5000, 1.5000]  prod =  0  0  1  1  1  2  2  2
i = 2:  w =  2.0  x = [1.5000, 2.5000]  prod =  0  1  1  2  3  3  4  5
i = 3:  w =  3.0  x = [2.5000, 3.5000]  prod =  0  1  2  3  4  5  6  7
i = 4:  w =  4.0  x = [3.5000, 4.5000]  prod =  0  1  3  4  5  7  8  9
i = 5:  w =  5.0  x = [4.5000, 5.5000]  prod =  0  2  3  5  7  8 10 12
i = 6:  w =  6.0  x = [5.5000, 6.5000]  prod =  0  2  4  6  8 10 12 14
i = 7:  w =  7.0  x = [6.5000,    Inf]  prod =  0  2  5  7  9 12 14 15
hist = {0.13462, 0.12717, 0.10720, 0.08065, 0.05414, 0.03243, 0.01734, 0.01377}
bias = 15, eta0 = 0.983561, lag0 = 2.5876, v0 = 0.339063 (2.3734 @ full value)
Expected variance = 8.51468, kurtosis = -0.29598
Expected uncorrelated lag variance = 1.20653/N
Van Vleck C(0) = 0.35772, C(1) = 0.38646, C(0)/C(1) = 0.92561
```

REFERENCES

Cooper, B.F.C. 1970, Correlators With Two-Bit Quantization, Aust J Phys, 23, 521

Hagen, J.B., & Farley, D.T. 1973, Digital Correlation Techniques in Radio Science, Radio Sci, 8, 775

Hawkins, D.W. 2011, Digitizer Quantization and Correlator Efficiency Analysis, `http://www.mmarray.org/~dwh/wbsddc/correlator_efficiency.pdf`

Rauch, K.P. 2008, Revised CARMA Correlator FPGA Configurations, CARMA Memo 46

Thompson, A.R., Moran, J.M., & Swenson, G.W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd Ed (New York: Wiley), Chapter 8