Fast-sampler Correlator Configuration Comparison

Kevin P. Rauch (UMD)

Dec 2, 2011

ABSTRACT

This document provides a high-level comparison between several correlator board configuration options related to the fast-sampler hardware upgrade, with emphasis on the implications for spectral line observations. Possible upgrades to the correlation logic to increase resolution are also discussed.

1. Fast Sampler Hardware

The main focus of the current correlator upgrade is to expand the maximum simultaneous bandwidth coverage to 8 GHz; a second priority is improved support for 23-station observing, including dual-pol and full-pol (i.e., full-Stokes). To satisfy the MRI proposal, the requirement is 8 GHz in 23-station single-pol mode; the goal (wish) expressed in the latest NSF URO proposal is for 8 GHz in 23-station dual-pol mode and 4 GHz in 23-station full-pol mode—double the required processing power (for equal channel resolution). Given a maximum bandwidth of 1 GHz per band, this corresponds to 8 or 16 bands of 23-station single-pol, respectively. At *fixed* frequency resolution, this equates to 4 or 8 dual-pol bands and 2 or 4 full-pol bands; in practice, resolution will be lowered to increase the number of bands available. A tertiary desire is to maintain simultaneous 15-station and 8-station observing. The simplest way to achieve that is by operating in normal 23-station mode and discarding the unwanted cross-products (at the expense of effective logic utilization); fully optimized 15+8 operating modes can be created with additional effort.

The basic assumption made here is that the sampler/bandformer output consists of 46-inputs x 8-bands of sample data, with a bandwidth of up to 1 GHz each. The 46 inputs nominally correspond to 23-stations x 2-polarizations, but may represent any combination of antennas/polarizations the switchyard can produce. In this context a 'band' is a contiguous segment of IF bandwidth within the 8 GHz of usable receiver output. Note that bands can overlap arbitrarily.

Connection to existing CARMA board hardware requires the use of fanout cards to obtain both the fanout formerly provided by the CARMA digitizers and signal conversion into LVDS format. In this case the functionality formerly provided by the CARMA digitizer boards is split among three boards: ADC boards (sample digitization), bandformer boards (delay tracking, digital input processing and statistics, and band definition) and fanout boards (output fanout, signaling protocol, supplemental cross-correlation calculation). Fan-out requirements depend on the CARMA board cabling topology (determined by the partitioning of baselines between boards).

Dual use of bandformer boards as correlator boards obviates the need for separate fanout boards in the system, as inter-board connections can now utilize native communication protocols. The need for fanout remains, of course, and must be satisfied by some combination of inter-bandformer, bandformer-to-correlator, and inter-correlator communication. This is discussed in detail in Dave Hawkin's "Full-Stokes Wideband Correlator" document; the basic conclusion is that fanout would be produced solely via inter-correlator communication.

Below I consider the major configuration alternatives for the upgraded correlator system, two based on CARMA boards and two employing bandformer boards as correlators. The four alternatives are not quite mutually exclusive, depending on the disposition of the current spectral correlator hardware.

2. Configuration 1: As-is Reuse of Spectral Correlator

This option entails reusing the current CARMA board hardware (i.e., spectral line correlator) as-is, either as an interim measure until spectral operation is enabled on a new (initially) wideband correlator, or even as a long-term companion to the latter to provide additional observing bands. This configuration—our current one—consists of 8 crates of 8 digitizer boards and 7 correlator boards, each crate capable of generating cross-correlation products for one 15-station single-pol band of up to 500 MHz bandwidth, resulting in a maximum simultaneous coverage of 4 GHz. Each pair of crates can be configured as a single 30-station (15-station full-pol) band, allowing up to 2 GHz total coverage. Support for 23-station modes (single-pol only) is implemented via the 30-station configurations by simply ignoring the final 7 inputs. The spectral resolutions of the current 15-input single-pol modes are given in Tables 1-2; 15-input full-pol and 23-input single-pol resolutions are both precisely half these figures (and only 2-bit sampling is possible at 500 MHz bandwidth). Resolution for 4-bit sampling is half the corresponding 3-bit value.

Note that the current configurations are optimized for 15-station single- or full-pol observations, making them rather inefficient for 23-station single-pol use; also there is no support for 23-station full-pol. To be considered a serious option, specialized 23-input configuration modes would need to be implemented to maximize resolution (for 15-station and 8-station observing the existing modes can be reused). The corresponding FPGA configurations would calculate 20 baselines per chip (as opposed to 32 per FPGA for the 30-input configurations), increasing channel resolution for the 23-station modes by 60% (Table 3). However, given the higher-priority effort required to enable spectral operation on the new hardware, this optimization would only be implemented were the old hardware to be maintained for the long-term, which is unlikely.

3. Configuration 2: Recycled Spectral Correlator

This option involves reusing the current CARMA digitizer boards as correlator boards. All digitization is performed by the new fast samplers and bandformer output is fed to the CARMA boards indirectly via independent fanout boards. This immediately provides 15 correlator cards per crate. However, the CARMA crates can accommodate up to 16 CARMA boards (there are 18 target slots, one used by the timing card and one unusable due to a PCI conflict), and there are sufficient spares (including the lab crate) to fully populate all 8 crates at Cedar Flat. This arrangement approximately doubles current correlation logic capacity.

The use of 16 CARMA boards per band allows a very simple and efficient 23-station baseline partition

Table 1. Configuration 1 spectral resolution [15-input single-pol, 2-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
500	97	16	1500	5.2	500
250	193	4	750	1.3	250
125	321	1.2	375	0.39	125
62	385	0.49	188	0.16	62.5
31	385	0.24	93.8	0.081	31.2
8	385	0.061	23.4	0.020	7.81
2	385	0.015	5.86	0.005	1.95

Table 2. Configuration 1 spectral resolution [15-input single-pol, 3-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
500	41	38	1500	12	500
250	81	9.4	750	3.1	250
125	161	2.3	375	0.78	125
62	257	0.73	188	0.24	62.5
31	321	0.29	93.8	0.10	31.2
8	321	0.073	23.4	0.024	7.81
2	321	0.018	5.86	0.006	1.95

Table 3. Configuration 1 spectral resolution [best 23-input single-pol, 2-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
500	81	19	1500	6.3	500
250	161	5	750	1.6	250
125	257	1.5	375	0.49	125
62	305	0.62	188	0.21	62.5
31	305	0.31	93.8	0.10	31.2
8	305	0.077	23.4	0.026	7.81
2	305	0.019	5.86	0.006	1.95

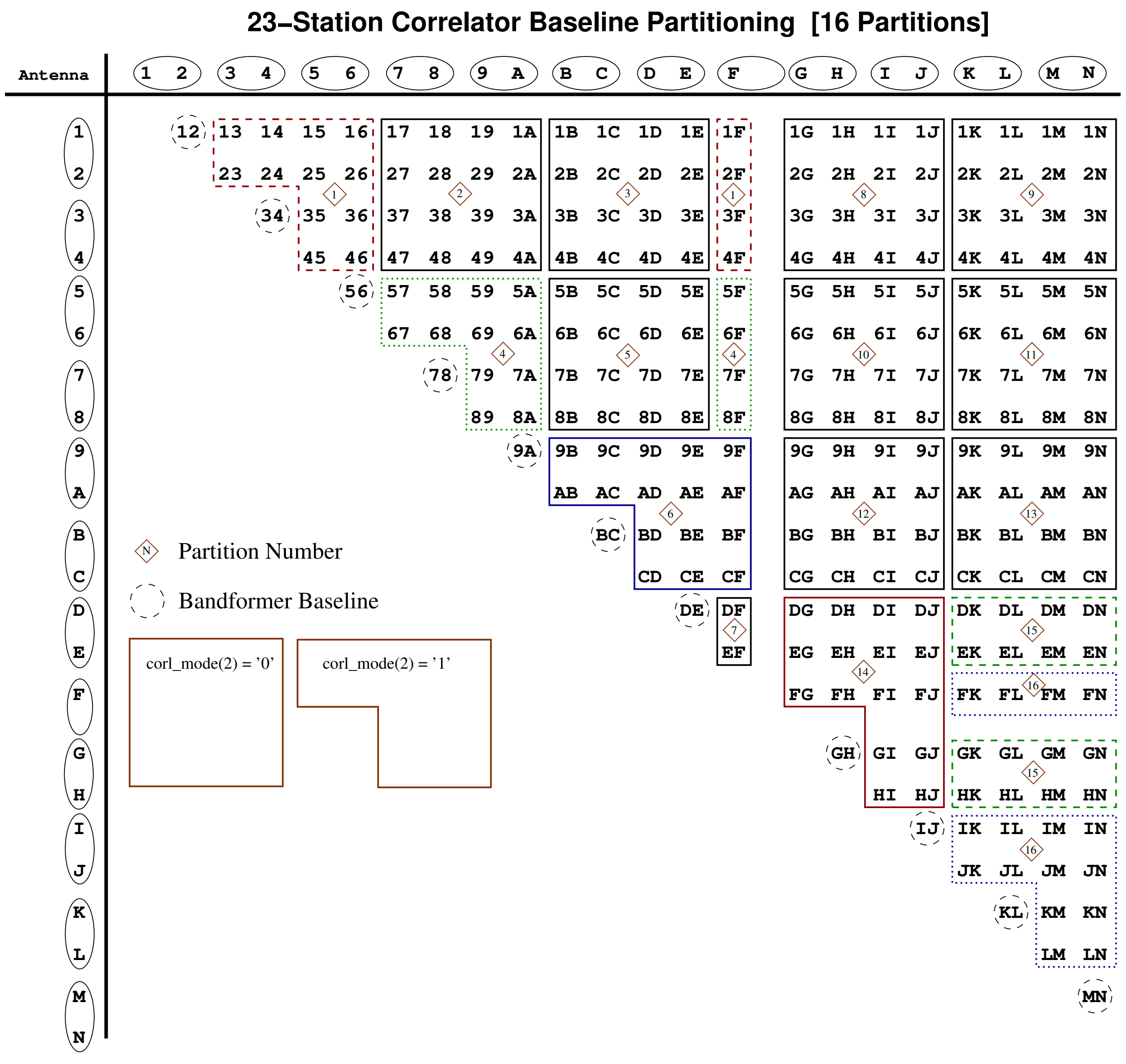

Fig. 1.— Single-polarization 23-station baseline partitioning for correlator configuration 2 (also 4b).

map to be employed, as shown in Fig. 1. The mapping is a superset of the current 15-station partitioning scheme. This makes it trivial to divide each crate/band into two banks, one consisting of the 7 original high-density (130K FPGA) correlator cards and the other of the remaining 9 low-density (90K FPGA) digitizer cards (possibly supplemented with correlators or über-digitizers to fill the ranks). Each high-density bank is a symmetric 15-input x 15-input correlator providing channel resolution identical to the current CARMA correlator. The low-density bank is a 23-input x 8-input correlator that can be used either independently or in conjunction with the other bank to form a symmetric 23-input correlator, with channel resolution limited by the low-density FPGAs. The estimated channel resolutions for 23-input single-pol mode are given in Tables 4-5; 4-bit resolutions are half the 3-bit values.

A major advantage of this setup is that the existing FPGA configurations can be reused as-is in all operating modes. The only additional requirements are (1) implementation of a 1 GHz mode; and (2) generation of reduced-resolution versions of all configurations, compatible with the low-density bank. The former is quite straightforward and the latter involves almost no (human) effort. Implementation of a 1 GHz full-pol mode does however demand use of 500 Mbps LVDS cable data, unless 1-bit correlations are used (not worthwhile given the efficiency loss). An intermediate contingency would be to use 3-level instead of 2-bit sampling, equivalent (in terms of continuum sensitivity) to reducing the bandwidth to $\approx$ 850 MHz.

Similar to the current situation for 15-input single/dual/full-pol, for 23-input dual-pol frequency resolution is identical to single-pol, but the number of available bands is halved (to four); for 23-input full-pol both the resolution and number of bands is halved. These capabilities more than satisfy the MRI requirements. With modest additional effort the option exists to support 23-input dual-pol modes with the same number of bands but half the resolution of single-pol, allowing 8 GHz total coverage. This would involve a small emendation of the full-pol FPGA configurations to select which correlations in the 2 x 2 cross-pol matrix to compute; it can be implemented by adding a single control bit (a separate set of FPGA configurations is not needed).

In this correlator configuration each front-panel cable continues to carry two (for single/dual-pol) or four (for full-pol) inputs per cable. Each input requires a fanout of 5. Twelve fanout boards are needed per band, each receiving two inputs with two polarizations each. The bandformer boards are also responsible for calculating a small number of cross-correlations—those now residing in the digitizers themselves.

4. Configuration 3: Minimum-cost Bandformer-based Correlator

4.1. Configuration 3a: Proof-of-Concept Design

This configuration concept consists of ten standard-width, single-FPGA bandformer boards per correlator band (single- or dual-pol). In split 15-input + 8-input operation, eight boards are allocated to the 15-input correlator and two to the 8-input section. Full-Stokes observing divides baselines between two bands, as for the current spectral correlator. Each bandformer board contains FPGA logic resources comparable to a (high-density) CARMA correlator board. The total logic available per band is somewhat less than for configuration 2, described in the previous section. The maximum number of cross-correlations computed per FPGA are 32 (single-pol) or 64 (dual-pol and full-pol). Estimated channel resolutions for 23-input single-pol are given in Tables 6-7. Resolution in each corresponding dual-pol or full-pol mode is one half that for single-pol; in addition, the number of available bands in full-pol mode is reduced by a factor of two (to four bands). As presented this arrangement only supports 2-bit sampling in 1 GHz dual-pol or full-pol mode.

Table 4. Configurations 2/3b spectral resolution [23-input single-pol, 2-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
1000	41	75	3000	25	1000
500	81	19	1500	6.3	500
250	129	6	750	2.0	250
125	225	1.7	375	0.56	125
62	257	0.73	188	0.16	62.5
31	257	0.37	93.8	0.12	31.2
8	257	0.092	23.4	0.031	7.81
2	257	0.023	5.86	0.008	1.95

Table 5. Configurations 2/3b spectral resolution [23-input single-pol, 3-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
1000	17	188	3000	63	1000
500	33	47	1500	16	500
250	65	12	750	3.9	250
125	129	2.9	375	0.98	125
62	177	1.1	188	0.36	62.5
31	225	0.42	93.8	0.14	31.2
8	225	0.10	23.4	0.035	7.81
2	225	0.026	5.86	0.009	1.95

Table 6. Configuration 3a spectral resolution [23-input single-pol, 2-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
1000	25	125	3000	42	1000
500	49	31	1500	10	500
250	97	7.8	750	2.6	250
125	161	2.3	375	0.78	125
62	193	0.98	188	0.33	62.5
31	193	0.49	93.8	0.16	31.2
8	193	0.12	23.4	0.041	7.81
2	193	0.031	5.86	0.010	1.95

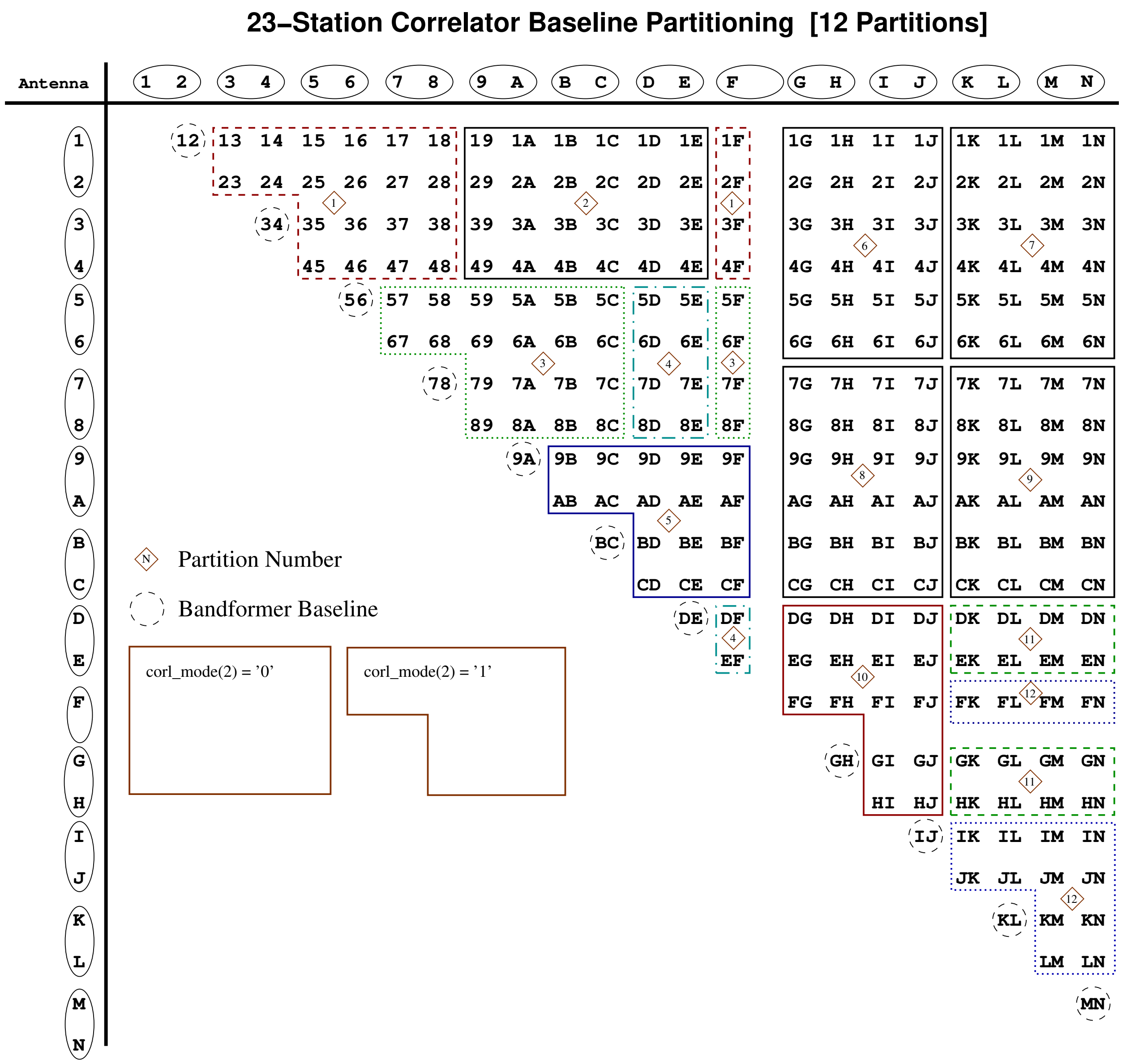

Fig. 2.— Single-polarization 23-station baseline partitioning for correlator configuration 3b.

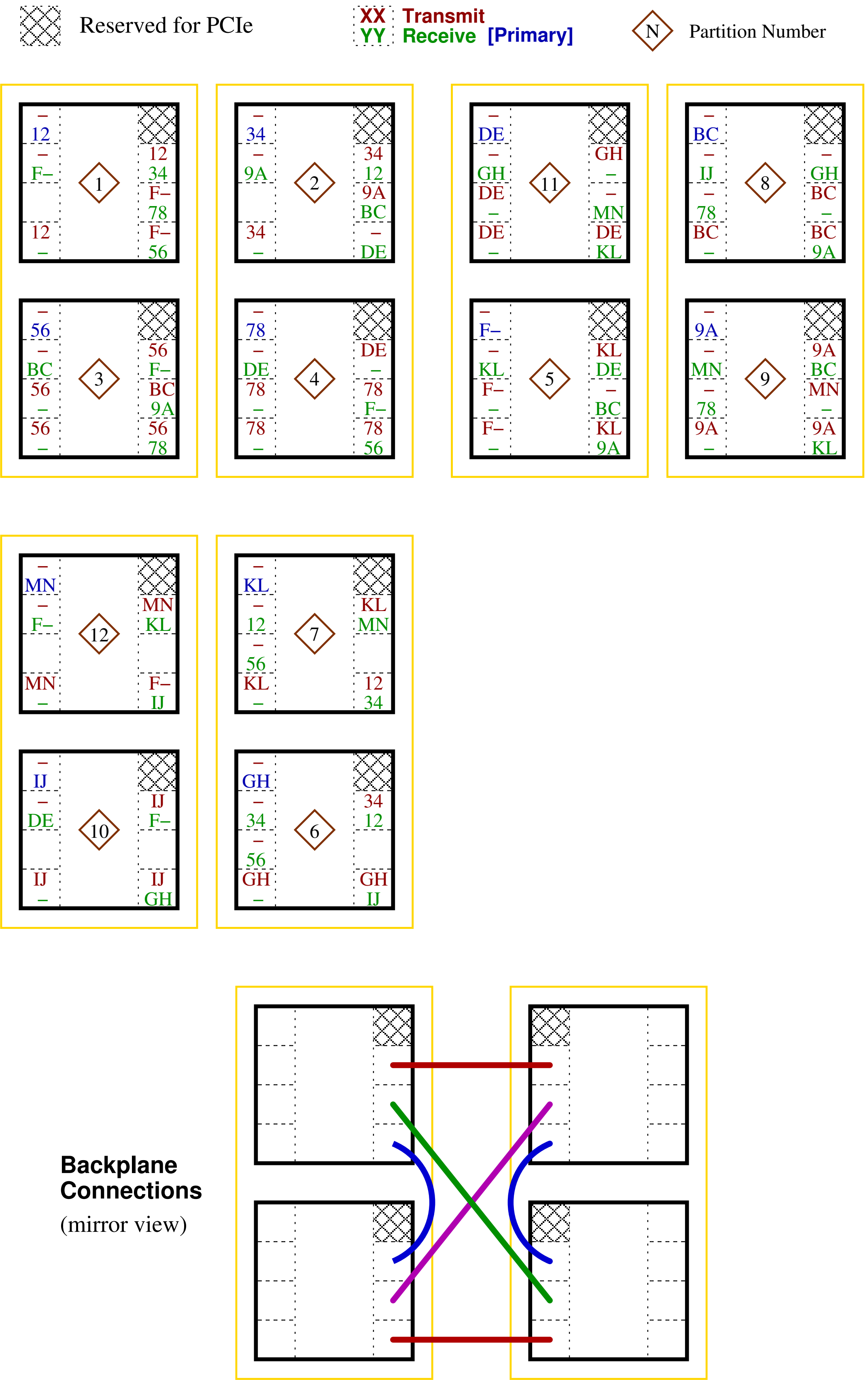

Fig. 3.— Signal fanout map for configuration 3b. Primary inputs are received from bandformer boards.

4.2. Configuration 3b: Optimized 23-Station Design

This configuration consists of 12 single-FPGA boards per correlator band. This number is special in that it matches the number of bandformer output cables per band, leading to a very regular input cable map. The baseline partition map for this configuration is shown in Figure 2; the corresponding signal fanout map is displayed in Figure 3. The maximum number of cross-correlations computed per FPGA are 24 (single-pol) or 48 (dual-pol and full-pol). Full-Stokes observing divides baselines between two bands. The use of 4-bit sampling is possible in all modes, including 1 GHz full-pol mode (albeit resolution is too low in this case for it to be useful). The total logic available per band is nearly identical to configuration 2 (Section 3), and estimated channel resolutions for 23-input single-pol are the same as given in Tables 4-5. Resolution in each corresponding dual-pol or full-pol mode is one half that for single-pol; in addition, the number of available bands in full-pol mode is reduced by a factor of two (to four bands).

5. Configuration 4: Extended Bandformer-based Correlator

5.1. Configuration 4a: Proof-of-Concept Design

This configuration concept consists of 18 bandformer boards per correlator band. In split 15-input + 8-input operation, 16 boards are allocated to the 15-input correlator and two to the 8-input one. All polarization modes utilize a single band of hardware, allowing a full 8 GHz bandwidth even in 23-input full-pol mode. The maximum number of cross-correlations computed per FPGA are 16 (single-pol), 32 (dual-pol), or 64 (full-pol). Estimated channel resolutions for 23-input single-pol are given in Tables 8-9; resolution is half these figures in dual-pol mode and one quarter of them in full-pol mode. Eight bands are always available. However, achieving this requires three independent sets of FPGA configurations (one each for single-pol, dual-pol, and full-pol), and resolution in full-pol modes is relatively low, limiting its appeal. Reusing the dual-pol configurations for full-pol (doubling resolution but halving the number of bands), as done by all other configuration options, would provide a better balance.

5.2. Configuration 4b: Optimized 23-Station Design

This configuration consists of 16 single-FPGA boards per correlator band. This is the same board count per band as configuration 2, and allows a very regular partition map to be employed (Figure 1). The number of baselines per FPGA is 16 (single-pol) or 32 (dual-pol and full-pol); full-Stokes observing divides baselines between two bands. The associated signal fanout map is shown in Figure 4; note that 4-bit sampling can be used in all modes, including 1 GHz full-pol mode. Channel resolutions are the same as for configuration 4a (Tables 8-9), due mainly to transfer of a subset of baselines to the bandformers.

Resolution figures assume the use of 530K (Stratix IV) FPGAs. One cost-saving option would be to load lower-density 360K FPGAs instead; this would give each correlator board logic equivalent to a CARMA digitizer board (using 530K FPGAs, logic is similar to a CARMA correlator board); resolutions in this case would match configuration 2 (and 3b; Tables 4-5). The advantage over configuration 3b is reduced implementation effort, as baseline partitioning matches that used by the current spectral correlator. Also the correlator board PCB could still be made pin-compatible with 530K (or even 290K) devices, allowing upgraded (or downgraded) logic capacity in future applications with no additional effort or risk required. On the other hand, option 3b (with 530K FPGAs) requires 25% fewer racks/chassis/boards than 4b (with 360K FPGAs); which one has the lowest overall hardware cost, however, depends strongly on the price differential between 360K and 530K FPGAs.

Table 7. Configuration 3a spectral resolution [23-input single-pol, 3-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
1000	11	300	3000	100	1000
500	21	75	1500	25	500
250	41	19	750	6.3	250
125	81	4.7	375	1.6	125
62	129	1.5	188	0.49	62.5
31	161	0.59	93.8	0.20	31.2
8	161	0.15	23.4	0.049	7.81
2	161	0.037	5.86	0.012	1.95

Table 8. Configurations 4a/4b spectral resolution [23-input single-pol, 2-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
1000	49	63	3000	21	1000
500	97	16	1500	5.2	500
250	193	4	750	1.3	250
125	321	1.2	375	0.39	125
62	385	0.49	188	0.16	62.5
31	385	0.24	93.8	0.081	31.2
8	385	0.061	23.4	0.020	7.81
2	385	0.015	5.86	0.005	1.95

Table 9. Configurations 4a/4b spectral resolution [23-input single-pol, 3-bit samples]

Bandwidth (MHz)	Channels (per sideband)	δV[3 mm] (km/s)	$V_{\rm tot}$[3 mm] (km/s)	δV[1 mm] (km/s)	$V_{\rm tot}$[1 mm] (km/s)
1000	21	150	3000	50	1000
500	41	38	1500	12	500
250	81	9.4	750	3.1	250
125	161	2.3	375	0.78	125
62	257	0.73	188	0.24	62.5
31	321	0.29	93.8	0.10	31.2
8	321	0.073	23.4	0.024	7.81
2	321	0.018	5.86	0.006	1.95

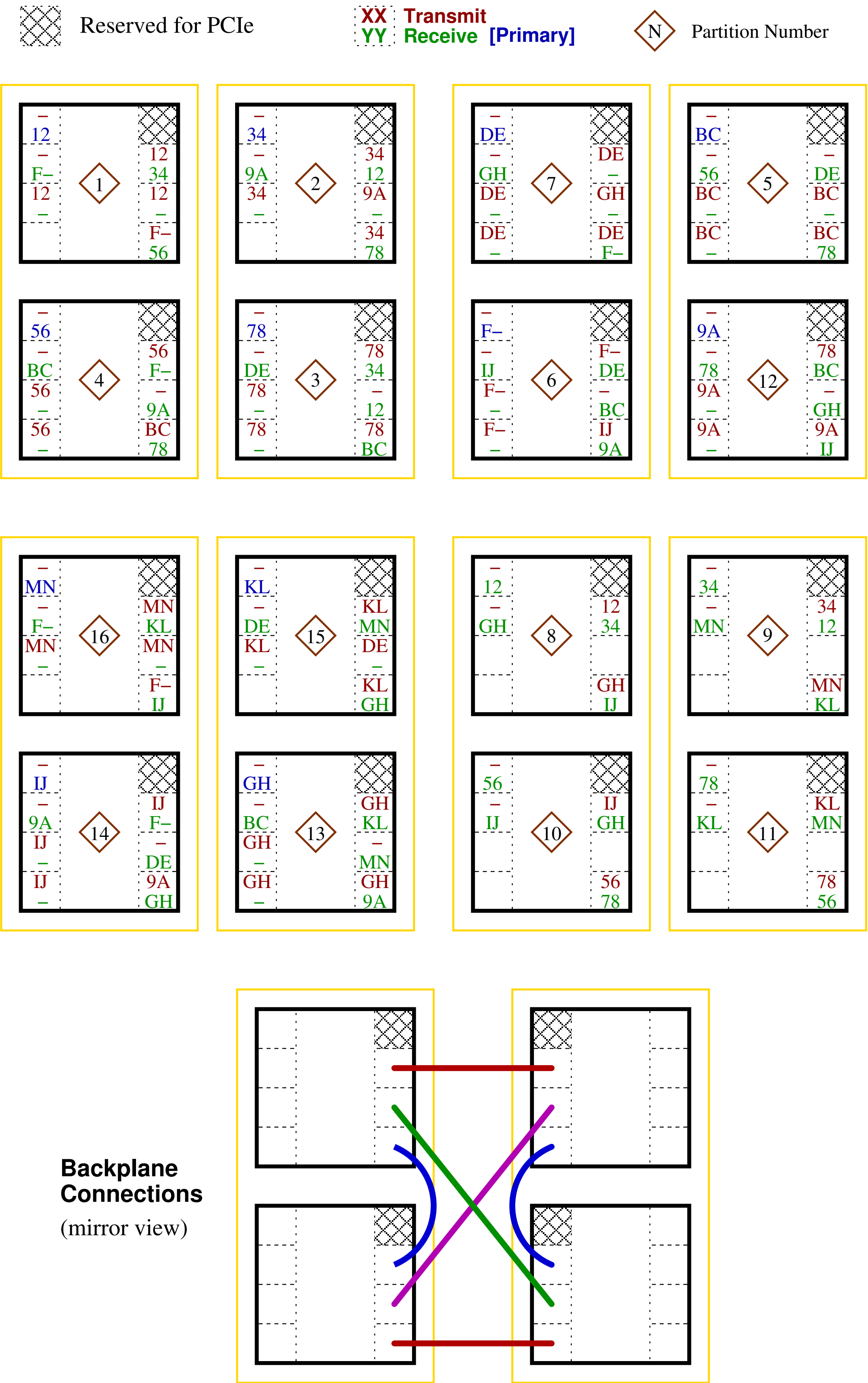

Fig. 4.— Signal fanout map for configuration 4b. Primary inputs are received from bandformer boards.

6. Correlation Logic Enhancements

In 23-station dual-pol and full-pol observing modes the estimated frequency resolution for several of the preceding configurations is low enough to impact scientific return. In particular, the optimal velocity resolution for galactic line observations is ~ 0.1 km/s over a full range of ~ 30 km/s (~ 10 MHz bandwidth for 3mm projects). In most cases only the single-pol modes satisfy this criterion. This motivates consideration into ways to improve the current correlation logic and augment lag counts without increasing logic usage. Doing so would be most useful for the narrowband modes (line observations), but would also benefit wideband modes (continuum observations) where coarse resolution can noticeably reduce efficiency due to the need to discard the end channels. The latter fact renders the 4-bit 1 GHz mode useless in *all* proposed configurations, for instance.

For wide bandwidth modes—125 MHz and above—correlation logic usage is dominated by the multiply-adder logic, due to the parallelism required to process the high incoming data rates (the correlation logic currently operates at a clock rate of 125 MHz). In this case the obvious way to increase lag counts is to increase the clock frequency and thus reduce the parallel logic. The limits to this are routability and $f_{\rm max}$ of the FPGA configurations and increased heat generation. For the existing CARMA boards this approach is not practical as our current 500 MHz wideband mode is already limited by heat, not logic. For the new boards the limits will depend on the family, density, and core voltage of the FPGAs used for correlator boards, as well as chassis cooling performance.

For narrowband modes—31 MHz and below—the effective input demux is less than one and the correlation logic is underutilized. In principle this can be exploited to increase resolution by buffering the input in RAM blocks and reading it out multiple times (with different delays) to compute additional lag segments. To satisfy real-time requirements the maximum number of readout passes (from RAM) is the inverse of the demux. Lags need to be dumped to RAM after each pass, and input must be double-buffered to accumulate new samples while the previous set is correlated. Absolute synchronization of each readout between all boards and FPGAs is required. Lag readout also needs to be double-buffered so that the CPU can retrieve one set while the latest is being calculated. Data latency increases by one frame (16 ms) and must be accounted for in the timestamp. Overall, implementing this scheme entails invasive updates to numerous components. Although the gain in resolution can be large (e.g., a factor of eight in the 8 MHz mode), implementation would require significant effort beyond the scope of the MRI.

A different underutilization exists in all modes that can be more easily exploited, namely the low rate of carry-in to the lag counters. When the carry-in rate is much less than one per cycle, blocks of lag counters can be stored in RAM and retrieved and incremented infrequently. If the maximum carry-in rate per lag is 1/N, groups of N lag counters can share the same increment register, reducing total logic usage. The increment logic needs to account for multiple carry-ins at the same time on different counters in the group, but this is straightforward. The modifications in this case are limited to the low-level `lag_counter` component, perhaps with tweaks to the lag dump state machine (to account for changes in readout latency). The applicability of this approach is enhanced by the fact that the current correlation logic divides the lag counter into two segments, each of which can be allocated bits arbitrarily. For the CARMA boards the allocation was tuned to maximize routability (and hence lag count) of the FPGA configurations. That tuning could be altered here to maximize the size of the RAM-based lag counter segments. Although applicable to all modes (both wideband and narrowband), in practice it is only useful for the low-demux narrowband modes. Rough estimates suggest $\sim 50\%$ improvement in resolution is possible for the latter.

Another avenue for logic optimization is to eliminate redundancy when a single input is used in multiple baselines within a single FPGA. Currently each baseline is treated independently, but in most cases a single input appears in several baselines. As a result the delay line for each input is instantiated more than once per FPGA, wasting logic. Restructuring the logic to remove these and any other duplicate logic would require modest effort and provide modest return ($\sim 10\%$ savings in narrowband modes).