Atacama
Large
Millimeter /
submillimeter
Array

Specifications for a Second-Generation ALMA Correlator

ALMA-05.00.00.00-0049-A-SPE

2021-January-08

Prepared by:	**Role, Organization:**	**Date and Signature:**
Alain Baudry	ALMA Digital Front-End Group, University of Bordeaux, L.A.B. / EU	
Crystal Brogan	NA Program Scientist, NRAO / NA, *Chair*	
Takeshi Kamazaki	ACA Correlator Scientist, NAOJ / EA	
Jongsoo Kim	PI ACA Spectrometer Project, KASI / EA	
Neil Phillips	ALMA Instrument Scientist, ESO / EU	
Michael Rupen	Director, Dominion Radio Astrophysical Observatory, NRC Canada / NA	
Alejandro Saez	ALMA Correlator and DTS Technical lead / JAO	
Baltasar Vila-Vilaro	ALMA System Astronomer / JAO	
Approved by:	**Organization Role:**	**Date and Signature:**

Change Record

Version	Date	Affected Section(s)	Author	Reason/Initiation/Remarks
V.0	2020-July-06	All	All	Initial draft
V.1	2020-Oct-23	All	All	First Draft to AMT
V.2	2021-Jan-07	Many	All	- Addressed comments from AMT reviewers - Significant improvements to S6.4 (Subarrays) - Significant improvements to S6.5 (VLBI) - Reorganization and improvements to S6.8 (Deployment); creation of Appendix A.4

Table of Contents

1 Executive Summary

The ALMA Development Roadmap [AD04] recommends a long-term development strategy for ALMA with near-, mid-, and longer-term goals, in aggregate these improvements will lead to the "ALMA2030 System". The highest priority near-term goal is to increase the correlated bandwidth of ALMA by at least a factor of two. A key component of such an upgrade will be a new ALMA Correlator that can meet the ambitious science goals of the next decade and beyond.

In order to define the requirements for a new ALMA Correlator, the ALMA Management Team convened the 2nd Generation ALMA Correlator Working Group (CorrWG) on January 22, 2020, the Charges for which are summarized in S2.1. In order to solicit initial feedback from the community a technical workshop was held in Charlottesville, VA Feb. 11-13, 2020 entitled "*The ALMA2030 Vision: Design Considerations for the Next ALMA Correlator*" the results of which are described in S4.1 and Appendix A.1. Key outcomes of the workshop was a consensus that the new correlator should be of FX (or FFX) design, and that if at all possible the new correlator should be located at the Operations Support Facility rather than the high-site.

One impediment to a full understanding of the requirements for the 2nd Generation ALMA Correlator is that other working groups are similarly working to define the needs of the other key upstream subsystems (Front-End, Digitization, Data Transport) in parallel. Therefore, in order to make progress, we first present a series of assumptions about the ALMA2030 System in S5.1, incorporating the recommendations of the Frontend & Digitizer Requirement Update Working Group. For example, we assumed that the minimum requirement for the maximum E-W baseline length is 35 km (S5.1.5). Based on these assumptions, detailed requirements for the 2nd Generation ALMA Correlator are presented in S6, as a series of minimum requirements and stretch goals. Included in S5.2 are questions that need to be addressed by other subsystems (or across subsystems), in order to finalize the correlator requirements.

A few of the most impactful correlator requirements are summarized in the Table below:

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.1.1 Number of antenna inputs	66 (all 12m + 7m) in subarrays and all together, *but expandable*	>=80 total antennas in subarrays and all together, *but expandable*	Main Array: 50x12m + ACA: with 12x7m and 4x12m = 66; Stretch: Recover original design, 64x12m in Main Array
6.1.2 Maximum bandwidth that can be correlated	8 GHz per pol per sideband (2x current) ***but expandable.***	Max Digitizer Design: present projection 16 GHz per pol per sideband (4x current)	**Stretch goal strongly preferred.** Allowed observing setups will be needed, to impose data rate limitations.
6.1.4 Correlator efficiency (minimum number of bits for multiplication + DSP losses)	97.5% (4-bit multiplication + DSP losses)	98.5% (6-bit multiplication + DSP losses)	Assumes system wide efficiency goal of >96%, and upstream digital efficiency of ~98.5% (ADC > =5-ENOB+losses). DSP=digital signal processing

6.2.1 Coarsest channel resolution	6.6 MHz (5% loss at 35 GHz, 0.2PB for 35km baseline)	2.9 MHz (1% loss at 35 GHz, 0.2PB for 35km baseline)	Avoid BW smearing at lowest freq/max baseline for largest correlated BW mode; 0.2PB = 20% level of the primary beam response. Used for calibration, continuum, and high-z spectral scans
6.2.2 Finest resolution at max correlated bandwidth	23.35 kHz (0.2 km/s at 35 GHz)	11.67 kHz (0.1 km/s at 35 GHz)	Spectral Scans of line-rich Galactic sources, as well as high-z absorption-line studies
6.2.3 Finest possible channel resolution	2.33 kHz (20 m/s at 35 GHz)	1.17 kHz (10 m/s at 35 GHz)	Very cold clouds, protoplanetary disk structure, small-body rotation, interplanetary probes
6.2.4 Bandwidth at finest possible channel resolution	1.6 GHz	1.6 GHz	Derived from max channels required for 6.2.2 minimum goal
6.5.1 Elements in VLBI (Tied-Array) Sums	One to all antennas	One to all antennas	It shall be possible to form a VLBI Tied-Array (aka Phased-Array) subarray from any number of available antennas, including those in the Main 12m-array and the ACA.

In order to begin the discussion of how the new correlator and its prodigious data rates will impact downstream subsystems, we provide a brief analysis of the peak output data rates in S7.

In S8, we describe options for deploying the new ALMA 2030 System, and find that a parallel deployment strategy is likely the only viable option to avoid significant delays and science downtime. The results of some initial fact-finding related to this option are also discussed (also see A.4). We recommend that the HILSE eventually be upgraded to serve as a testbed for the 2nd Generation Correlator.

Appendix A.2 describes our concerns related to doing a first coarse frequency channelization (FFT) *at the antennas*, and why the CorrWG prefers that the correlator receive the data in time-series format. Finally, Appendix A.3 describes options for overcoming the challenges presented by doing an accurate delay correction, to inform future efforts to define an ALMA 2030 System-wide delay correction strategy.

2 Introduction

2.1 Goal of this Document and Charges

The primary goal and charges for the 2nd Generation ALMA Correlator Working Group, as assigned by the ALMA Management Team, and addressed in this document are described below, see [AD06] for more details:

The primary goal of the CorrWG shall be to develop specifications for the 2nd Generation ALMA Correlator that should be deployed by 2030. Specifications are taken to mean both the detailed technical requirements of the correlator itself, as well as, describing the key prerequisites that will define the ultimate correlator design.

- In deriving the technical requirements CorrWG shall consult with digital correlator experts world-wide, and ensure equal representation of ideas and views across the ALMA partnership.
    - As part of this activity, members of the CorrWG will participate in, and receive feedback from the "The ALMA2030 Vision: Design Considerations for the Next ALMA Correlator" Workshop to be held on Feb. 11-13 in Charlottesville, VA. By agreement within the working group, a representative subset of the CorrWG will also attend the "The ALMA 2030 Vision: Design considerations for Digitizers, Backend and Data Transmission System" Workshop to be held on March 11-13 in Mitaka, Japan, with the intent of sharing information and resolving issues that cross multiple systems.
        - We note that the Mitaka workshop was delayed until October 14-16, 2020, and virtual due to the Covid-19 pandemic. Much of the CorrWG did attend, but there has been little time to synthesize or incorporate the results.
    - The CorrWG shall summarize the correlator workshop for the AMT.
- The CorrWG will also consult with the ALMA Integrated Science Team (IST) and ALMA Science Advisory Committee (ASAC), as well as, regional SACs for advice on the scientific requirements.
- The CorrWG shall present the detailed requirements and functionality in the form of both the minimum goals that achieve the science priorities of the ALMA2030 Roadmap and "stretch" goals that anticipate the forefront of technology (though likely production-ready) at the start of the next decade.
    - The ultimate meaning of minimum and stretch goals shall be understood as the minimum for which a particular correlator design is considered to "meet spec", while compliance with stretch goals may be used to distinguish between competing designs.
    - Stretch goals, in particular, should have both technical and scientific justification.
    - Areas of particular technical uncertainty should be clearly highlighted for future followup.
    - Areas of consensus should also be clearly denoted.
- The CorrWG shall describe the likely pros and cons of different correlator architecture as they apply to ALMA Key science goals.
- The CorrWG shall propose prototyping for the Second-Generation ALMA Correlator, testing, and a deployment framework that minimizes ALMA science observing downtime, while not unduly prolonging the time to deliver new capabilities. The primary goal (for now) is to identify how various testing/deployment options could affect the ultimate correlator design/requirements.

2.2 Applicable and Related documents

2.2.1 Applicable Documents

The following documents are part of this document to the extent specified herein. If not explicitly stated otherwise, the latest issue of the document is valid.

Appl.	Document Title	ALMA Document Number
[AD01]	ALMA Product Assurance Requirements	ALMA-80.11.00.00-001-D-GEN
[AD02]	ALMA Safety Manual	ALMA-10.08.00.00-011-D-MAN
[AD03]	ALMA System Technical Requirements, rev C	ALMA-80.04.00.00-005-C-SPE
[AD04]	The ALMA Development Roadmap	Roadmap_public_20180415.pdf
[AD05]	ALMA Scientific Specifications and Requirements, rev. A	ALMA-90.00.00.00-001-A-SPE
[AD06]	Charge: Specifications for a Second Generation ALMA Correlator	ALMA-60.00.00.00-0173-A-SPE

2.2.2 Related Documents

The following documents are referenced in this document, a number of other references are given in the text as footnotes.

Appl.	Document Title	ALMA Document Number
[RD01]	Front End and Digitizer Requirements	ALMA-40.00.00.00-1612-A-SPE
[RD02]	64 Antenna Correlator Specifications and Requirements (Baseline Correlator)	2008-08-07-ALMA-60.00.00.00-001-C-SPE
[RD03]	ACA Correlator Technical Specifications and Requirements	ALMA-62.00.00.00-001-A-SPE
[RD04]	The Atacama Millimeter Array Implications of Potential Descope	https://www.nap.edu/download/11326
[RD05]	Specifications and Clarifications of ALMA Correlator Details	COMP-70.40.00.00-0007-A-MEM / ScottCorrelNormalization.pdf
[RD06]	NRC Talong Frequency Slice Architecture Correlator / Beamformer (AT.CBF) for ALMA	ALMA_Talon_correlator_study_report_vRELEASE_2020-09-21.pdf
[RD07]	Data Rate Impact of the ALMA Correlator Upgrade Project	ALMA-60.00.00.00-0148-B-MEM Correlator Upgrade Data Rate Memo.pdf
[RD08]	ACA Spectrometer Technical Specifications and Requirements	ALMA-64.00.00.00-0005-B-SPE
[RD09]	ALMA Back End Electronics Design Description	BEND-50.00.00.00-077-B-DSN
[RD10]	ALMA Memo 561 (Delay Errors In Single- and Double-Sideband Interferometer Systems)	https://library.nrao.edu/public/memos/alma/main/memo561.pdf
[RD11]	ALMA Technical Handbook Cycle 8	https://almascience.eso.org/documents-and-tools/cycle8/alma-technical-handbook

[RD12]	ALMA Memo 607 (Digital Correlator and Phased Array Architectures for Upgrading ALMA)	https://library.nrao.edu/public/memos/alma/main/memo607.pdf
[RD13]	Implementing the concurrent operation of sub-arrays in the ALMA correlator	https://ui.adsabs.harvard.edu/abs/2016SPIE.9913E..3KA
[RD14]	The ALMA Phasing System: A Beamforming Capability for Ultra-high-resolution Science at (Sub)Millimeter Wavelengths	https://ui.adsabs.harvard.edu/abs/2018PASP..130a5002M
[RD15]	Phasing ALMA with the 64-antenna correlator	https://ui.adsabs.harvard.edu/abs/2012evn..confE..54B

2.3 Acronyms

A complete set of acronyms and abbreviations used in ALMA is maintained at the acronym list web page, (also see the external website https://www.almaobservatory.org/en/siglas/). In addition, throughout this report we will abbreviate "Second Generation ALMA Correlator" as 2nd Gen Correlator.

3 Key Science Drivers

Below we list the three new key science drivers identified by the ALMA2030 Roadmap [AD04] for the next decade, followed by the type of improvements required to realize them. The term "spectral grasp" is shorthand for the ability to tune to a wide range of diagnostic spectral lines within a single receiver band in a single observation, while "spectral range" indicates the need to access the full (sub)millimeter frequency range visible from the ground, such as to accommodate a particular source's redshifted spectral emission or the rest frequency of a unique diagnostic line transition.

1. **Origins of Galaxies:** Trace the cosmic evolution of key elements from the first galaxies ($z>10$) through the peak of star formation ($z=2$-4) by imaging their cooling lines, both atomic ([CII], [OIII]) and molecular (CO), and dust continuum, at a rate of 1-2 galaxies per hour.
    + Spectral line sensitivity
    + Spectral grasp
    + Spectral range
    + Continuum sensitivity
2. **Origins of Chemical Complexity:** Trace the evolution from simple to complex organic molecules through the process of star and planet formation down to solar system scales (~10-100 au) by performing full scans of a whole frequency band at a rate of 2-4 protostars per day.
    + Spectral line sensitivity
    + Spectral grasp
    + Spectral range

+ Finer spectral resolution
+ Increased angular resolution

3. **Origins of Planets:** Image protoplanetary disks in nearby (150 pc) star formation regions to resolve the Earth-forming zone (~1 au) in the dust continuum at wavelengths shorter than1 mm, enabling detection of the tidal gaps and inner holes created by nascent planets.
   + Continuum sensitivity
   + Increased angular resolution

These ambitious goals, while far-reaching from a science perspective, are unified by a relatively small number of required upgrades: improved spectral line and continuum sensitivity, increased spectral grasp and spectral range, and enhanced spectral and angular resolution. As proven by the original Level One science goals, the capabilities driven by these new key science goals will enable a vast range of new discovery space across the whole spectrum of ALMA science.

With the exception of spectral range (which is driven by the capabilities of the receivers), all of the improvements needed to enable these ALMA2030 key science goals require a 2nd Gen ALMA Correlator that can handle a wider instantaneous bandwidth (spectral grasp and continuum sensitivity), correlate a higher number of bits (spectral and continuum sensitivity), accurately correct for larger delay rates (higher angular resolution), produce more channels for a given bandwidth (spectral grasp, higher spectral resolution), and allow the placement of numerous independent spectral windows (spectral grasp).

4 Community Feedback

4.1 ALMA Correlator Workshop

The North American ALMA Science Center hosted an ALMA correlator workshop in Charlottesville, VA on Feb. 11-13, 2020 entitled "*The ALMA2030 Vision: Design Considerations for the Next ALMA Correlator*". Hereafter, we simply refer to the "2020 Correlator Workshop"[1]. There were about 70 participants from around the globe, including correlator representatives from ALMA (BLC, ACAC, TPS), JIVE, LOFAR, MeerKAT, ngVLA, NOEMA, SKA1-Mid, SKA1-Low, SMA, VLBA, and VLITE. Presentations were given on the ALMA Development Program and the ALMA 2030 Roadmap, "Lessons Learned" from recent correlator projects, and "Emerging Technologies" from on-going efforts. The presentations can be viewed at NextALMACorrelator_Presentations. Additionally, there were three discussion/feedback sessions specifically geared toward the 2ndGen ALMA Correlator on (1) open questions solicited from the participants, (2) optimal correlator architecture, and (3) potential deployment options. Outcomes from the first two sessions are described below, while the third is presented in S7. Additional feedback is provided in Appendix A.1.

4.1.1 Outcome from Discussion of Key Questions Picked by Participants

1. **Science benefits of XF or FX architecture? Do we need a first "F"? If so, where?**

---

[1] http://go.nrao.edu/NextALMACorrelator

Correlator architectures are traditionally divided between "XF" (in which cross-multiplication in the time domain is followed by Fourier transformation into the frequency domain) and "FX" (in which Fourier transformation is done first, followed by cross-correlation in the frequency domain).
Complete Consensus: Many of the previous advantages of XF (with ASICs for the X-engine) can now be surpassed by FX. There was agreement across all three Groups that FX offers significant advantages:

- Upgrade flexibility; Easier to expand # of antennas
- Better isolation of channels (sinc^2)
- Better for large numbers of channels and easy to define windows of desired width and channelization
- Easier to do sideband rejection like 90deg Walsh function switching and fringe rotation
- Take advantage of the N x log(N) complexity the FFT algorithm can afford

Near Consensus that a first "F" could be necessary (i.e., FFX architecture)
FFX (Coarse Channelization => Fourier transform=> Complex Multiplication)

- An initial coarse channelization step may be required to reach the finest spectral resolution target of 1 kHz.

Majority of discussion on first "F" concerned where: antenna, or centralized with correlator

- Consensus: Final data transport requirements should drive the answer
- Multiple 400Gb/s transmissions on a single fiber are feasible (up to 25.6Tb/s), so it may not be necessary to reduce data volumes for transmission
- Concern was expressed about adding complexity to the antennas; however most felt that the savings in electronics complexity from an overall new design would still yield a net improvement
- There was concern that putting the first "F" at the antenna may result in efficiency loss (depending on the # of bits which are preserved)

2. **Are there aspects of data processing which we should pull forward to the correlator? (gridding, weighting, WVR/Tsys application, baseline-dependent averaging,, ...)**

Consensus (with caveats) – No strong driver for adding more capability in this area for near-term ALMA 2030 goals, though this could change in the future if the number of antennas increases significantly

- Removes ability to mitigate problems in data (as corrections are "locked-in"), so likely not desirable for science data (but see caveats)
- Takes large investment in commissioning at the telescope, and inhibits development of improvements in offline processing
- Caveats:
    - It is possible to do WVR correction now (allowing further time averaging on-line) but this is not being used except for VLBI. This option should be preserved. Research into whether WVR correction can be improved online or offline desirable – offline can more easily model changing corrections both forwards and backwards in time.
    - TelCal will likely need to apply Tsys and bandpass corrections on-line to flatten wide bandwidth before making channel-averages for QA0 assessment. However, the uncorrected channelized science data should be stored for offline correction.

In the future:

- If RFI becomes significant, we may want to perform RFI excision in the correlator
- GPU processing could be interesting in the future, especially if the number of antennas increases significantly

3. **What is the best way to correct for the instrumental delay?**

Complete Consensus – Need more centralized method for delay application
Current issues with the ALMA delay mechanisms:

- Delay corrections are derived and applied in several different stages (based on centralized delay server) -- Bulk delay (Correlator Station Card, 250ps steps), Fine delay (TFB phase shift used for VLBI implementation of baseband delays normally done in the CDP), Ultra-fine delay (Digitizer clock-phase, 15.625ps steps), and Residual delay (CDPs, <15.625ps) [values taken from RD11]
- Communication traffic is high, especially on long baselines (with faster-changing delays). The situation will become critical when we implement longer baselines for ALMA2030. We must then either replace the CAN bus, or transmit only polynomial coefficients (rather than actual delays) and compute delays locally
- The delay server uses weather parameters which themselves change as a function of time
- VLBI: Doing ultra-fine delay correction at the antenna limits the accuracy of those corrections

Some fundamentals to consider:

  - Must ensure return to phase when switching frequency and sources (when the phases of time-series data are rotated)
  - Better done at correlator after FFT, so that no loss due to decorrelation occurs (this is optimal for narrow channels, but less effective the larger the channel size)
  - Fine delay tracking after the FFT filter bank may use phase delays while integer delay steps occurring before the FFT can generate delay/phase discontinuities which can propagate through the poly-phase FFT filter bank and cause significant decoherence at the longer baselines. Further investigation for ALMA is needed.
  - Cable training=cabling delays within the correlator that are predictable/stable; once accurately measured, they can be easily accounted for in "instrumental delay". However, if data are transmitted with timestamps as is currently under discussion, cable training will no longer be necessary

4.1.2 Additional Consensus Feedback

- The number one comment was that ALMA must produce a system-wide plan as soon as possible to get the wider-bandwidth initiative going
- The following should be considered for design trade-offs:
  - Total power consumption (hardware, computing, & cooling)
  - Required down-time of science observations to deploy/commission
  - Operational cost
  - Need for new fiber
- The 2ndGen Correlator should be located at the OSF (also see RD12]
- Build with future maintenance and commissioning in mind – standardize interfaces
- Liquid cooling for major correlator components or LRUs deserves a look
- A CPU-based correlator is not practical for ALMA
- Also deserving a closer look: a few talks suggested that we may want to offer multiple baseband data access points -- “spigots” for future “experiment” type user-developed back-ends. Currently we just have one that is used for phased-array mode observations.
  - For operations, allocating a baseband data access spigot for a limited (# of antennas, channelization, for example) CPU-based software correlator could offer a useful platform for debugging and commissioning

4.2 Feedback from ALMA Science Advisory Committees

So far draft versions of the science drivers and requirements have been presented at:

- the ALMA Science Advisory Committee face-to-face meeting March 3, 2020
- the ALMA North American Science Advisory Committee face-to-face meeting (held virtually), May 28, 2020

Both groups expressed support for the on-going process, with no major additional suggestions. Future community outreach activities will include: similar presentations to the East Asian and European SACs, participation / discussion at the upcoming "*ALMA 2030 Vision: Design considerations for Digitizers, Backend and Data Transmission System"* in October 2020**"**; and webinars for registered ALMA users, towards the end of the process.

5 Key Assumptions and Critical Design Decisions

The second generation correlator design will depend upon a number of design decisions for the upstream system (e.g., [RD01]), as well as potential updates to the System Level Requirements [AD03] to reflect the ALMA2030 goals. In particular, decisions regarding the digitization and data transport designs will have a critical impact.

5.1 Key Assumptions

The assumptions described in this section, in effect form part of the 2nd Generation ALMA Correlator requirements. If any of them are later proven to be unfounded, then all the correlator requirements described in S6 will need to be reviewed for potential impact.

5.1.1 Receiver Design and Digitization Goals

Below we briefly summarize those aspects of the Front-End/Digitizer WG recommendations [RD01] that directly affect the correlator subsystem.

- Instantaneous Bandwidth to be Digitized: ≥ 8 GHz per polarization per sideband, 16 GHz per sideband highly desirable
- Bandpass Shape (within digitizer baseband width): flat to within < 5.4 dB peak-to-peak
- Effective Number of Bits ≥ 5
- Digitizer sampling speed: > 40 GSamples/s
    - Implied Number of Basebands: 1 per sideband assuming 16 GHz per sideband IF

As in [RD01], for this document we assume that future ALMA receivers are likely to all be 2SB with the exception of Band 1, which is expected to retain its current SSB design within the scope of likely operation of the 2nd Generation ALMA Correlator. It is notable that there are still some open questions regarding optimal design for Bands 9 and 10 depending on correlated bandwidth[2], but for the sake of simplicity we follow [RD01] in assuming that these will be 2SB eventually. *As a result of this assumption, bandwidths are written as* ***per sideband*** *quantities.* Requirements essential to the future operation of the current Band 9 & 10 receivers are, however, included for completeness (i.e. ability to perform 90° Walsh switching).

For the purposes of this document, *we assume that the [RD01] digitization goal of ENOB ≥ 5 (99.6505% efficient[3]), together with any required digital signal processing and losses due to clock jitter will yield an ADC/BE sensitivity of ≥98.5%.* Additionally, there is a strong desire in the astronomical community to avoid any gaps in the usable bandwidth (if there is more than one baseband per sideband), or any significant difference between the correlated bandwidth and the usable bandwidth. Toward this aim, *we assume the digitized bandwidth will exceed the required correlated bandwidth by a sufficient amount to prevent the edge roll-off from the digitizer anti-aliasing filter from impacting the usable correlated bandwidth (also see [RD01]).* It is also

---

[2] Using a DSB receiver with sideband separation in the correlator (e.g., 90° Walsh switching) results in double the effective correlated bandwidth. This gives a sensitivity gain for continuum and spectral survey cases which is comparable to that gained from the reduced T_{sky} contribution to T_{sys} for a sideband-separating receiver. Achieving very wideband DSB receivers for Bands 9 and 10 may be considerably easier (and cheaper) than achieving a comparable bandwidth sideband-separating design; a sideband-separating receiver design may only be competitive if the entire bandwidth in both sidebands can be correlated, so that there is no loss of bandwidth compared to the simpler DSB option.

[3] Number of bits to efficiency values taken from Thompson, Emerson, and Schwab, Radio Science, Vol. 42 2007, https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2006RS003585

worth noting that while [RD01] does not include a specific recommendation for the ALMA 2030 IF range (2–18 and 4–20 GHz are given as possible options, with the final choice dependent on performance), provided that twice the low end of the range is ≤ the maximum correlated bandwidth per sideband, it will be possible to avoid the inter-tuning "gaps" that presently impede efficient observation of spectral scans.

5.1.2 Correlator Design

Based on the unanimous feedback received from the 2020 Correlator Workshop (S4.1.1), *we assume that the 2nd Generation ALMA Correlator design will be FX:* Fourier Transform first, converting input signal samples in the time domain to the (complex) frequency domain; followed by complex multiplication and accumulation, taking each piece of the input bandwidth and multiplying the data flows for all antenna pairs in the array. The current technologies for correlator implementation impose practical limits on the instantaneous bandwidth of signal that can be processed. Thus, it is likely that the correlator design will actually need to be FFX, with an initial "F" stage that channelizes the wideband signal into a moderate number of "chunks" (of order 10-100), followed by a second "F" stage that produces the finer spectral sampling needed to meet the science requirements. The first "F" stage then delivers a continuous time series of spectral ''chunks'' that are presented to the second "F" stage. (The first "F" stage could be either at the antenna, but see SA.2 for difficulties, or in the main Correlator room.) The time-to-frequency transpose and the frequency-to-antenna transpose operations needed before correlation are performed after the second "F" stage.

5.1.3 Correlator Subsystem Definition

The correlator subsystem processes all basebands (as defined in S5.1.4) delivered by each antenna in the array to form all cross- and auto-correlation functions required for further astronomical data analysis. ALMA consists of two arrays of antennas. We assume that the 2nd Gen Correlator will at minimum be able to process the current 50x12m antennas of the Main Array, as well as the 12x7m antennas + 4x12m Total Power antennas of the Atacama Compact Array (ACA, also known as the Morita Array). Such observations shall be possible via independent subarrays, or with all 66 antennas combined into a single array. Presently, the data from these two arrays are typically correlated by the Baseline Correlator (BLC, [RD02]) and the ACA Correlator (ACAC, [RD03]), respectively, though ACA data can optionally be processed by the BLC.

Presently Total Power data is typically processed by the ACAC, but in the near future, such data will be processed by the new ACA Spectrometer (ACAS, [RD08]), targeted to be commissioned for general use in Cycle 9. While we assume that the 2nd Gen Correlator must be able to correctly process ALMA Total Power data, in addition to interferometric data, we note that it may be operationally advantageous to upgrade and optimize the ACAS for spectral compatibility to the 2nd Generation Correlator for Total Power observations. *We assume that even if this path is chosen, that it will still be possible to use the 4x12m Total Power antennas that are co-located with the compact 7m-array interferometrically with the 2nd Gen Correlator.* Especially for high frequency observing, the addition of the co-located 12m antennas to the 7m-array can be critical for accurate calibration.

For the current BLC, the Correlator Data Processing nodes (CDPs) are sometimes considered separate from the Correlator, though many of the important "correlator" functions are performed there, including the FFTs (the 'F' in "XF"), sub-band stitching, Van Vleck corrections, normalization by the autocorrelations, and spectral averaging to name a few. It is also notable that the correlator is presently responsible for performing the instrumental delay tracking (the station cards apply the bulk delay, while the baseband and residual delays are applied in the CDPs). At this time no detailed system-wide delay tracking plan is available for ALMA2030. *In this document we specify requirements for the "combined correlator subsystem" because some functions that are currently handled in software by the CDPs could be done in hardware/firmware in a future design.* The expected correlator input is described in S5.1.7 and the output is described in S6.7 of the requirements.

5.1.4 Baseband Definition

Our definition of one baseband channel is closely linked to the electrical intermediate frequency (IF) bandwidth delivered by a Front-End receiver and to the sampler actual performance. One *baseband channel bandwidth* (BB) is defined here as the unit of digitized electrical bandwidth presented to the correlator. This bandwidth may differ from the sampler input bandwidth, depending on the digitizer module design and performance (in particular, the adopted exact sampler clock rate and anti-aliasing filter properties). The correlator must process the entire BB bandwidth presented to the correlator, although it need only correlate the scientifically useful part of this band. For a 2SB receiver (S5.1.2), there will be a minimum of four BBs, two for each polarization per sideband; more BBs may be required and presented to the correlator if a single digitizer (per sideband and polarization) cannot cover the IF bandwidth of the receiver.

Broader BBs allow the processing of very broad spectral lines (as needed, e.g., for galaxy clusters or Solar observations) as well as collections of individually-narrow spectral lines or related isotopic molecular species distributed over a wide instantaneous bandwidth. This suggests that very broad BBs are desirable scientifically, especially if they provide a high signal-to-noise ratio performance.

Smaller increments within a BB are called *spectral channels* and *spectral windows*. The smallest frequency increment produced at the output of the correlator subsystem is called a *spectral channel*. The width of such a channel is set by the actual correlator design, or by internal averaging required to reduce the output data rate to a reasonable level. A *spectral window*[4] (sometimes called a sub-band) is a collection of contiguous spectral channels which share the same frequency widths and spectral responses. Spectral windows may overlap in frequency. The bandwidth of a spectral window is at most the bandwidth of an individual BB.

Many science cases may require independently configuring multiple spectral windows within a single BB, e.g., to allow observers to "zoom in" on interesting parts of the spectrum with different spectral resolutions, while also maximising continuum bandwidth (for calibration at least). Thus, each spectral window must have adjustable frequency centers, bandwidths, and spectral channel widths (S6.8).

5.1.5 Maximum Baseline Lengths

The maximum delay rate that has to be taken into account in the correlator corresponds to the longest east/west extension of the Array, while the maximum total delay to be buffered depends on the longest baseline in any direction (in addition to the maximum fiber length difference). Since the 2nd Generation Correlator may operate for a significant fraction of the active life of ALMA, the possibility that additional pads are added to increase the length of the maximum baselines has to be taken into account.

---

[4] In the current ALMA terminology a spectral window may include all polarization products thus requiring correlation of a BB pair.

The SE arm of the Array could be extended further along the same direction, since after about 12 km from the center of the Array, it could merge with the gorge along Route 27. This could continue for another 10 km until the National Flamingo Preserve near La Pacana. At that point the geography becomes significantly more complex, and a more or less linear extension becomes complicated (road access would be challenging), until possibly the border into Argentina has been crossed. The NE arm can be extended into Bolivia for another ~10 km into the Laguna Verde/Laguna Blanca area before encountering a more complex landscape. The EW arm can be extended easily across the Atacama Valley for about 30 km before reaching the Valle La Luna area. The extension in that direction, however, comes with a significant reduction in altitude (~2000m lower than the Chajnantor Plateau), which would significantly affect the performance of the Array at high frequencies. All these extensions may be hampered by external geopolitical factors, such as the need to get the pertinent allocation of the land to set the pads, difficulties in placing pads outside the territory of Chile, etc. Taking into account all the above considerations, *we assume that the maximum East-West extension during the operation phase of the 2nd Generation Correlator (i.e., Stretch Goal) will be* ***=<60 km***. The maximum North-South extension would be no greater than this.

A more realistic value for the maximum extension in the mid-term future comes from the maximum extent of the ALMA concession and surrounding Atacama Astronomical Park, and by the Paso Jama road in the NE, within which permission to locate stations may be most feasible. In this case, the extension would be largest in the Westerly direction, along the access road from AOS to OSF. The OSF site with multiple existing stations at 2900m altitude is a likely end-point, with PWV already typically 3 times higher than at AOS (going down to the gatehouse would mean higher still PWV). The EW distance from OSF to the most likely Easterly location (on an SE extension of the S branch) is about 32 km (see Figure 1). *To allow some flexibility in final station locations, we will assume that the system must support* ***35 km East-West baselines***. The North-South extent would be less than this.

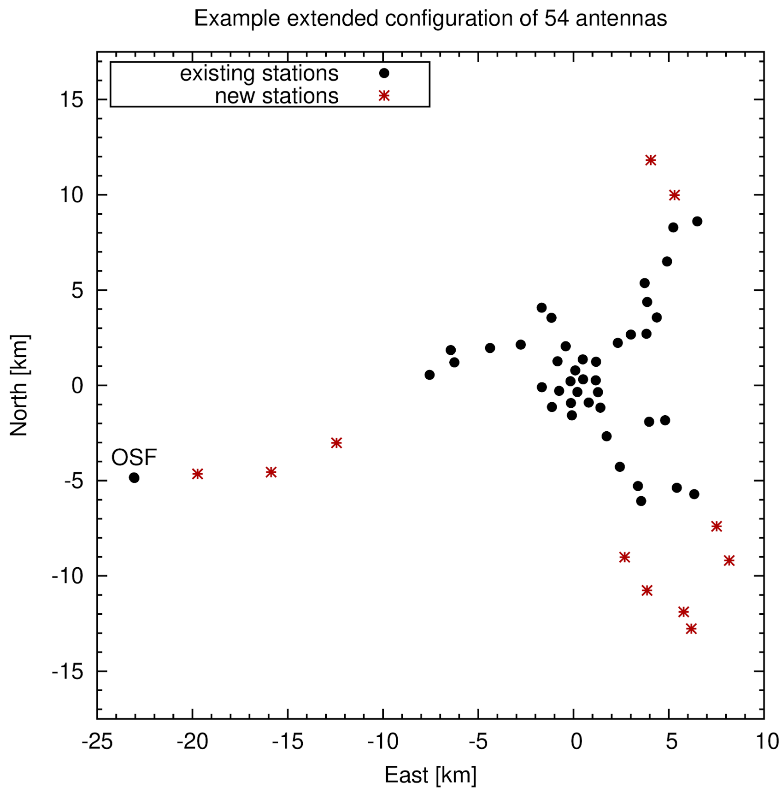

Figure 1: *Representative extended configuration with stations added to the West (OSF-AOS access road, including existing OSF-TF stations), North (within the ALMA concession, as limited by the Paso Jama road) and South (within the Atacama Astronomical Park). Note that obtaining N-S extent is crucial for complete uv coverage, especially for sources near the equator. The coordinates are relative to the array centre currently used for computing delays online for existing array configurations.*

Feasibility testing so far has demonstrated that the existing LO and DTS systems can be used between OSF and AOS. However, the usability of the LLCs is so far untested, and the expectation is that the LLCs will not be usable in their normal way. Thus we assume that a more extended array will likely operate with the LLCs either operating differently, or not being used at all, on the assumption that atmospheric phase fluctuations dominate over fiber path length variations on the relevant temporal and spatial scales of astronomical phase referencing. This issue is, however, not relevant to the correlator subsystem.

5.1.6 Time Synchronization

The meaning of time synchronization in this section is: the mechanism for synchronizing the correlator operations to the ALMA telescope timing references.

The CLOA (**C**entral **L**ocal **O**scillator **A**rticle) is the ALMA telescope entity responsible to generate the timing reference signals to be distributed to the ALMA telescope components by means of the Back End infrastructure. No significant changes to the CLOA are expected during the lifetime of the 2nd Generation Correlator, and its resources can therefore be used for synchronization purposes. Two sets of timing Reference signals are generated by the CLOA:

- Low frequency
    - 125[MHz]
    - TE → 48[ms] period square wave signal
    - 2[GHz]
- High Frequency
    - produced by the Photonic References and intended to synchronize the First LO (see [RD09], sections 3.6.3 and 3.6.4)

The correlator gets two low-frequency signals from the CLOA (see [RD09], section 3.6.1)

- 125[MHz] sine wave, 8[dBm], 50[Ohms]
- TE LVDS

The 125[MHz] is derived from an ultra-stable 5[MHz] oscillator. The TE signal is made using as time reference the 125[MHz], and its phase is synchronized to a 1PPS signal output of a GPS[5]. The usage of the previously listed signal by the correlator is not mandatory, but, those are the only available signals which provide a time reference to the correlator. *The correlator must process the samples from the antenna according to the current telescope conditions, and the time reference signals provide a time-base for this. We further assume that scans start and stop on TE boundaries.*

---

[5] A 1 PPS signal is also emitted by the Central LO. Together with the GPS and TE signals it is used by the BLC to provide timing sanity checks for Phased-array observing.

5.1.7 Metadata that the correlator expects

In addition to the actual digitized data, the correlator requires certain other information to be delivered in real-time. The details depend on the system-level implementation of basic functions such as digitization, timing, application of delays, and the like. While this is known for the current system (e.g., S5.1.6), there is not yet a design for the ALMA2030 system (see S5.2). It is clear however that the correlator will require the following:

- Time stamps and synchronization patterns, to align the data from the different antennas;
- Verification information, to monitor missing samples and communications problems (e.g., check sums, bit error rates, etc.);
- Subscan sequence specifications, both to label the data properly and potentially to allow application of the correct delay models;
- Sub-integration flags[6] (if any) and excision controls, to pass along flags upstream of the correlator system (if any) and to control any automatic flagging done by the correlator itself (e.g., based on high power levels);
- Delay information (delay events and/or delay models [including validity intervals]);
- WVR data for real-time phased-sum path length correction (if desired);
- Total power information, it there is digital processing upstream of the correlator system (if not, the correlator can calculate power levels itself), together with corrections if this information comes from a separate detector rather than from digitiser statistics;
- Any other calibrations which must be applied in real-time (e.g., complex gains from TelCal);
- Information related to tied-array beams (e.g., complex antenna weights to be used in the tied-array sum, including delay offsets from the subarray delay centres and sensitivity weighting).

5.2 Critical Design Decisions for ALMA2030

In this section, we describe critical decisions that are required to facilitate an actual 2nd Generation ALMA Correlator design. Each open issue is followed by a brief explanation from the point of view of the Correlator Subsystem/CorrWG:

Digitization

1. How many basebands of data will be delivered to the correlator?
    a. Although the correlator subsystem needs a decision on the number of basebands (as defined in S5.1.4) to be made early in the process, it is largely agnostic about the actual number. This is because the correlator will need to divide data into suitably sized frequency "chunks" for optimal processing that are almost certainly smaller than the adopted baseband width in any case.
    b. The Front-End/Digitizer WG report recommends avoiding a second down-conversion thereby producing only one baseband signal per sideband, and the CorrWG has no concern with this suggestion (assuming that it is feasible to achieve it on a reasonable development timescale).
    c. For ease of testing and implementation the correlator will likely be built, and at least initially deployed and tested, one baseband at a time.
2. What will the usable bandwidth of each baseband be? It is highly desirable that the usable bandwidth delivered to correlator be >= the maximum correlated bandwidth.
3. How many bits will the transported data have?
    a. The Front-End/Digitization WG recommends that the effective number of bits be ≥ 5, and we concur with this goal in terms of

---

[6] Flags which apply on timescales as long or longer than an output integration need not be applied in the correlator, while sub-integration flags must be applied during correlation.

the expected digitization efficiency. However, the correlator needs to know how many physical bits (which will likely range from 6-8) will be transmitted, in order to design the correlator, as well as the requisite requantization correction.

   b. The physical number of bits will also determine the requisite data rate to be transported.

Data Transport

4. It remains an open question whether the data rate of the full time series data produced by the ALMA2030 Front-ends / digitization can feasibly be transported to the correlator without incurring infeasible cost. A potential remedy would be to perform a first Fourier Transform with relatively coarse spectral resolution at the antennas.
   a. Because it will fundamentally dictate the nature of the data that is transmitted to the correlator (time-series vs spectral) and how it must be processed, and corrected, the Correlator Subsystem is keenly interested in this issue, and requests that addressing this issue be of highest priority for moving forward with the ALMA2030 Roadmap.
   b. After weighing the pros and cons, the CorrWG would strongly prefer that all post-digitization data manipulation occur within the correlator subsystem, i.e. no first “F” at the antennas unless data transport is otherwise infeasible. A complete description of our concerns are described in A2.1.
   c. Beyond the raw data itself, a complete list of all the information that needs to be transmitted needs to be developed and included in the data transport question. The additional information required by the correlator is described in S5.1.7.
   d. How exactly will the data be packetized?
      i. In particular will timestamps be used and exactly how will that be implemented? What will the timestamp accuracy be?
   e. What will the data transport protocol be?
5. What will the maximum instantaneous data rate of the transported data be?
6. Can the data be transmitted to the OSF without any significant loss in data integrity?
   a. At the 2020 Correlator Workshop it was unanimously agreed that placing the 2nd Gen Correlator at the OSF would convey many advantages, see below for two key examples, however we need to know if this is feasible ASAP.
      i. The HVAC needs of the 2nd Generation ALMA Correlator will be considerably reduced if it can be located at the OSF rather than AOS (as well as its maintenance). For the HVAC, the savings will be of order a factor of 2.
      ii. The complexity of deployment and commissioning, as well as future maintenance and trouble-shooting of the 2nd Gen Correlator will be significantly reduced if it can be located at the OSF.

Power Infrastructure

7. After construction, the power consumption of a correlator (and its HVAC needs) represents a significant recurring operational cost. Presently there are no concrete system-wide constraints on the ALMA2030 power needs overall or on the correlator power budget specifically. We urgently need to know the upper limit on correlator power consumption.
8. Will the 2nd Gen Correlator be backed-up with UPS power? If not additional protections will be needed in the final design.
   a. The UPS could be in the form of short-term protection of the whole subsystem against brief glitches, or longer term protection for just key hardware.

Downstream Capabilities

9. What is the likely maximum sustainable instantaneous dump rate from correlator to storage in the ALMA2030 era? Data Processing Constraints? etc
    a. It would be a waste to build a correlator that greatly exceeds (by orders of magnitude) the foreseeable ability of the downstream subsystems to handle the data rate. We need near, mid, and far-time look aheads that span at least the next decade to understand what is likely to be possible.

Other (Cross-Subsystem) Issues

10. We need an ALMA2030 System-wide efficiency goal in order to know the number of bits the correlator needs to process. A system-wide value of 96% is typically being assumed (i.e., the product of digitization and correlation losses).
11. We need an ALMA2030 System-wide specification on the maximum allowed power variation across each IF passband (the FE/Digitizer WG Report only recommends a constraint for the digitizer portion of 5.4 dB) [Note this item does not directly affect correlator design but was identified as a notable missing specification for the ALMA2030 wideband system, i.e., a wideband equivalent to [AD03] 7.2.21, Requirement 272].
12. We need a cross-subsystem strategy for how to implement delay tracking, time synchronization, and communications in the full ALMA2030 system.
13. How do we ensure return-to-phase for the system as a whole? Various subsystems, hardware and software, are inter-related.
14. We need a system-wide requirement and strategy for preserving the X-Y phase in the system over (up to) several hours in order to calibrate polarization calibrators over the required range of parallactic angle coverage (> 60 degrees).
15. Will WVR continue to be applied offline (online is a combination of TelCal and Correlator System)?

6 Correlator Requirements

The Second Generation ALMA Correlator requirements are presented as “minimum” and “stretch” goals. The “minimum” requirements are those required to meet the near-term goals laid out in the ALMA2030 Roadmap [AD04], specifically at least doubling the correlator bandwidth of the ALMA System. The “stretch” requirements anticipate what will likely be possible at the start of the next decade (considering likely technology readiness), and incorporate requirements for the ALMA2030 Roadmap longer-term goals such as extended baseline lengths, and additional collecting area. The requirements are broken into topical subsections, and within each subsection, the requirements are presented as both short summary table entries, as well as descriptive text, both denoted by subsubsection numbers -- in case of confusion, the text shall be considered as defining the full requirement.

6.1 Basic Properties

Parameter	Minimum Req’mt	Stretch Goal	Comment
6.1.1 Number of antenna inputs	66 (all 12m + 7m) in subarrays and all together, *but expandable*	>=80 total antennas in subarrays and all together, *but expandable*	Main Array: 50x12m + ACA: with 12x7m and 4x12m = 66; Stretch: Recover original design, 64x12m in Main Array

6.1.2 Maximum bandwidth that can be correlated	8 GHz per pol per sideband (2x current) ***but expandable.***	Max Digitizer Design: present projection 16 GHz per pol per sideband (4x current)	**Stretch goal strongly preferred.** Allowed observing setups will need to impose data rate limitations.
6.1.3 Correlation products	1, 2, or 4 (XX or YY; XX and YY; XX, YY, XY, YX)	1, 2, or 4 (XX or YY; XX and YY; XX, YY, XY, YX)	Science motivation for 1 pol weak after significant improvement in overall spectral resolution
6.1.4 Correlator efficiency (minimum number of bits for multiplication + DSP losses)	97.5% (4-bit multiplication + DSP losses)	98.5% (6-bit multiplication + DSP losses)	Assumes system wide efficiency goal of >96%, and upstream digital efficiency of ~98.5% (ADC > =5-ENOB+losses). DSP=digital signal processing
6.1.5 Max Delay compensation distance, buffer time range, and buffer capacity per station and baseband	400us (35 km East-West baseline + 56km fiber difference; 16Msample @ 40Gs/s, 96Mbit if 6-bit/sample)	1000us, same as BLC (40Msample @ 40Gs/s, 240Mbit if 6-bit/sample)	Note that the overall delay compensation strategy for ALMA2030 is not yet clear Buffer time is sum of atmosphere geometric delay difference + fiber (2e8m/s) delay difference
6.1.6 Power Consumption	TBD	=< now	Current (hardware + computing) BLC=165 kVA + 25.3 kVA + HVAC ACAC+ACAS=66.7 kVA + 20 kVA + HVAC

6.1.1 Number of antennas

The minimum requirement is to correlate 66 antennas, i.e., the current number of ALMA antennas. The Second Generation ALMA Correlator should have subarray modes which enable it to produce not only cross-correlations from 50 x 12m and 12 x 7m arrays but also auto-correlations from the 4 x TP array. All the products from these three arrays then have the same spectral response. In this way, we avoid the need for any frequency synthesis to match one correlator to another. The bottom line is that the Second Generation ALMA Correlator should produce cross-correlations from both 12m and 7m arrays and auto-correlations from the TP array, for a total of 66 antennas. Additionally, it shall be possible to correlate all the antennas together in a single observation.

The stretch goal for the number of antenna inputs is related to the longer term ALMA 2030 Roadmap [AD04] goal of increasing the main-array's collecting area, though there is no information at this stage as to how many new antennas this goal would entail. The original ALMA design envisioned 64 x 12m antennas in the main array, with the drop to the current 50 x 12m antennas a consequence of the 2005 ALMA Re-baselining Activity for cost-savings [RD04]. However, the improved image fidelity afforded by the full 64 antenna main-array configurations, especially for longer baselines, is well-documented [RD04], so this forms the lower bound of the stretch goal, with further expandability desirable. The stretch goal is thus a total of at least 80 antenna inputs (64 x 12m + 12 x 7m + 4 x 12m, TP).

6.1.2 Maximum bandwidth that can be correlated

The ALMA 2030 Roadmap [AD04] states that the most important near-term goal is to at least double the bandwidth of ALMA, including the 2nd Gen Correlator, in order to achieve the ambitious new key science goals described in S3. Thus, the smallest maximum correlated bandwidth (MCB) goal is 8 GHz per polarization, per sideband (2x current). However, since the Roadmap was begun, there has been significant progress in developing wider bandwidth receivers (e.g. ESO is building a Band 2 receiver with a likely IF bandwidth of >=16 GHz (required >=12GHz)[7]; an upgraded Band 6 being studied by NRAO will likely have an IF bandwidth of 12 GHz[8]; and NAOJ has demonstrated a Band 7/8 concept with 17 GHz IF bandwidth)[9]. Additionally, digitizers that can handle up to 40 Gsample/s have been demonstrated, enabling direct instantaneous sampling of an IF range of 0-20 GHz, yielding at least 16 GHz of *usable bandwidth* per pol, per baseband. **Therefore, the stretch goal is to allow for a maximum correlated bandwidth equivalent to the maximum that the upgraded digitization/DTS system can produce, which at this moment appears likely to be 16 GHz per pol per sideband (4x current). We further amend the minimum goal to explicitly require the ability to expand the bandwidth if it is initially limited to 8 GHz per pol, per sideband due to cost or scheduling constraints.** These requirements are consistent with those of the Front-End/Digitizer WG [RD01]. All other applicable requirements in this document must be fulfilled at the maximum correlated bandwidth.

From a technology point of view, a number of recent and planned correlator designs suggest that a correlated bandwidth of 16 GHz per pol, per sideband for 66 to 80 antennas (S6.1.1) is feasible in the ALMA 2030 timeframe as described at the 2020 Correlator Workshop. Examples include the SMA[10] (8 antennas, MCB = using two independently tunable, but overlapping RF receivers 12 GHz per pol (dual polarization), per sideband, or 24 GHz single pol, per sideband), NOEMA[11] (12 antennas, MCB = 8 GHz per pol, per sideband), and the upcoming SKA-mid Phase 1[12] (197 antennas, MCB = 5 GHz, per pol) correlators. Thus, these current and upcoming correlators are approaching ALMA's needs in #Antenna^2 x MCB space. Indeed, a recently completed Cycle 7 North American ALMA Development Study carried out by the NRC describes a proof of concept for adapting the extensively reviewed SKA-mid Phase 1 correlator design to ALMA with 8 GHz per pol, per sideband and 80 antennas, which is potentially expandable to 16 GHz per pol, per bandwidth [RD06].

It is notable that while the technology upstream of, and including the correlator is reaching a maturity consistent with the desired 16 GHz per pol, per sideband goal (4x current) with 66 to 80 antennas, the downstream subsystems (Telcal, Archive, Global Data Transport, and Data Processing) have not yet demonstrated that they would be able to handle the prodigious data rates (or volumes) that would be afforded by the recommended ALMA2030 system. A modern FX correlator will natively produce ~1 million of channels, as compared to the maximum of 30,720 available with the BLC or ACAC (though the ACAC itself produces many more channels natively that are averaged down using frequency

---

[7] https://www.aanda.org/articles/aa/full_html/2020/02/aa36777-19/aa36777-19.html

[8] https://zenodo.org/record/3240369

[9] https://www.aanda.org/articles/aa/full_html/2020/08/aa38713-20/aa38713-20.html

[10] SMA Presentation: https://osf.io/8dt2c/

[11] NOEMA Presentation: https://osf.io/mcwkf/

[12] SKA-mid Phase 1 Presentation: https://osf.io/k4jpt/

profile synthesis to match the BLC). However, the CorrWG recommends that these challenges should not inhibit the design of the 2nd Gen Correlator, because its output can be constrained to feasible observing modes that restrict the data rates to manageable levels, that can be gradually relaxed over time as the downstream subsystems become more capable. For more on observing modes and data rates see S6.8 and S7, respectively.

We would also like to reiterate from the assumptions described in S5.1.1 (also see [RD01]), that from an observing efficiency point of view it is essential to avoid the tuning “gaps” for spectral scans that currently arise from having the *usable* bandwidth in both TDM and FDM modes be less than the digitized bandwidth due to filter roll-off (i.e. 1875 GHz per sideband usable versus 2 GHz digitized). This issue will be overcome provided that the 2nd Gen Correlator is delivered a usable bandwidth consistent with the maximum correlated bandwidth requirement and if twice the low end of the IF range (which defines the size of the gap between the sidebands) is ≤ the maximum correlated bandwidth per sideband for the 2SB receivers.

6.1.3 Correlation products

The Second Generation ALMA Correlator shall generate products $X_iX_j^*$ or $Y_iY_j^*$ for observing modes of single polarization, $X_iX_j^*$ and $Y_iY_j^*$ for observing modes of dual polarization, and $X_iX_j^*$, $Y_iY_j^*$, $X_iY_j^*$, and $Y_iX_j^*$ for observing modes of full polarization, where X and Y represents two orthogonal polarization components, the superscript * is a notation for the complex-conjugation operation, and the subscript i or j index denotes an antenna in a given subarray. If i is equal to j, then auto-correlation products are generated. The main purpose of correlation products of single polarization is to improve spectral resolution. Since the Correlator shall have much improved spectral resolution, the observing modes of single polarization may not be supported. Dual polarization modes are likely to remain useful, to limit output date rates for experiments which do not require cross-polarization products.

6.1.4 Correlator Subsystem Sensitivity

As described in S5.2, a system-level sensitivity goal for ALMA 2030 has not yet been defined, although 96% seems a likely goal. The losses considered here are all multiplicative and result in a loss of signal-to-noise ratio. Hence, the overall digital system efficiency can be defined as

$$E_{DS} = E_{Digitizer}\ x\ E_{Back\text{-}End}\ x\ E_{Correlator}$$

If we assume a combined $E_{Digitizer}\ x\ E_{Back\text{-}End}$ efficiency of 98.5% (including the expected ENOB used for digitization & clock jitter, see S5.1.1), then *the Correlator Subsystem should achieve a sensitivity of ≥ 97.5%.* The efficiency of the Correlator Subsystem ($E_{Correlator}$) is defined by three major (multiplicative) components: (1) Efficiency loss due to the number of bits employed; (2) Losses due to imperfect delay correction; and (3) Losses due to blanking. The expectations for the performance of these components are described in sections 6.1.4.1-6.1.4.3.

6.1.4.1 Correlator efficiency (minimum number of bits for correlation)

The correlator efficiency, defined as the ratio of the actual SNR of the correlation products to the ideal SNR of the products that would be obtained with an infinite quantization resolution, is affected by a number of approximations including digital filtering losses, fixed bit multiplication and truncation errors, and fine delay inaccuracies across a coarse channel. The correlator efficiency will thus be known only when the detailed architecture (assumed FX) is frozen. Here we note two options for demonstration purposes:

1. Using: (a) 4-bit input to the multiplier in the X engine (giving 98.8457% efficiency); preceded by (b) several 6-bit equivalent processing stages before multiplication (up to 13 stages at 99.896% efficiency would be allowed, as .988457*.99896^13 = .9752), thus enabling digital filtering and re-quantization in the F engine, and a correlator efficiency of 97.5%.

2. Similar to (1) but using 6-bit correlation for the multiplication step (a), and still allowing for several 6-bit equivalent processing stages before multiplication (b), implies a correlator efficiency of 0.99896*0.98659 = 0.98556 ~ 98.5%.

Option (1) uses up the entire assumed Correlator Subsystem allowance to meet the system level goals (if our assumptions for the System and Digitization/BE goals are valid). Option (2) would provide an extra 1% of "cushion" to account for imperfect delay correction (S6.1.4.2), and/or poorer performance than hoped in the upstream subsystems (S5.1.1). *Given that current estimates for other losses may be optimistic, the stretch goal of 6-bit correlation is likely to be preferred.*

6.1.4.2 Sensitivity loss due to imperfect delay correction

The amount of delay-related sensitivity loss depends on how often the delay correction can be performed in the correlator. The maximum sensitivity loss is roughly estimated by the following formula[13],

$$1-\frac{sin(2\pi\Delta\nu\Delta\tau)}{2\pi\Delta\nu\Delta\tau}$$

where $\Delta\nu$ is the maximum frequency offset from the phase rotation center and $\Delta\tau$ is the maximum uncorrected delay change due to Earth rotation and/or source motion. Note that this estimate assumes that the residual delays, i.e. the component shorter than the shortest quantized correction applied in hardware, will not contribute any additional sensitivity loss because they shall be compensated by applying a phase slope in the frequency domain (in other words, phase rotation of the FFT products), a technique known as the "residual delay correction" (S4.1.1.3).

For demonstration purposes, if we assume a sensitivity loss goal of 1% [i.e., the same as current ALMA performance, 0.33-1.3%, RD10] and take the minimum requirements for the maximum baseline distance of 35 km (S5.1.5, S6.1.5), and the maximum delay rate 8.49 [ns/s] (S6.3.4, for objects moving at the sidereal rate), as well as a sampling speed of 40 GSa/s (i.e. the goal digitizer sampling speed, S5.1.1, which will be the same as the correlator sampling speed if there is no downconversion), **the delay correction must be applied every 0.230 [ms], as a minimum goal.** The time drops to **0.134 [ms] for a max baseline of 60 km as the stretch goal.**

However, further considering the implications for the FFT segment length (in time) under these assumptions, along with the finest spectral resolution required (min = 2.33 kHz, stretch = 1.17 kHz: S6.2.3), it is clear that the segment times (0.839 ms and 1.049 ms, for a maximum baseline of 35 km) are too long to be updated at the requisite delay correction rate. Ideally of course the sensitivity loss

---

[13] Kamazaki et al., 2012, PASJ, 64, 29

due to imperfect delay correction would be a factor of a few smaller than 1%, making this problem even worse. Moreover, the data rate for such a mode would be extreme. For example, data taken at 35 GHz and a spectral resolution of 20 m/s (2.33 kHz) would produce 6.9 to 13.7 *million* channels (per pol), for a max bandwidth of 8 or 16 GHz per pol per baseband. The long baselines also require very short dumps to avoid time-averaging smearing, putting the output data rates at 10s to 100s of GB/s, compared to the current limit of 70 MB/s (see S7).

Thus, it is clear that accurate delay correction in a single stage, adhering to the most challenging requirements, is infeasible. For this reason, we have also supplied additional requirements for the finest resolution at the maximum correlated bandwidth (S6.2.2), and the bandwidth required at the finest possible spectral resolution (S6.2.4), that provide scientifically-motivated options for mitigation of the largest required FFTs. In SA.3 we discuss additional considerations for the ALMA 2030 delay tracking, and associated requirements, though it should be emphasized that better than 1% efficiency is certainly feasible.

6.1.4.3 Blanking loss

Due to implementation details, it may be that a fraction of input samples are not processed (skipped) or are replaced with zeroes (blanked) to avoid incorrect correlations in an overlap period. This may for example happen at the beginning or end of FFT segments for an FX correlator, or time segments in the current BLC's TDM modes. Mismatches between segment lengths and other timing boundaries, such as the 16ms switching period for 90^o Walsh switching, can lead to such a scenario arising, and this could be particularly acute for long FFT segments to achieve large numbers of spectral channels. *Any losses from blanking (or skipping or incorrect segment overlaps etc.) should be included in the overall correlator sensitivity used to judge this requirement, but we believe the contribution will be small (<0.1%).*

6.1.5 Maximum Delay compensation distance, buffer time range, and buffer capacity

The bulk of the delay compensation (also known as instrumental delay) is implemented using a circular buffer, with an accuracy (precision) of one sample. Delays for a time difference less than 1 sample are implemented in a different way and are beyond the scope of this section. The aim of this section is to consider the minimum buffer size needed to implement the integer delay compensation.

The buffer size primarily depends on the maximum baseline length and the maximum difference in fiber delays between stations. At any moment in time the delay of the station needing the least delay can be subtracted from the computed delay of all stations, and indeed this is what ALMA currently does. The current BLC provides a buffer of 1ms. The maximal distance between stations is limited by the operation range of the LLC (**L**ine **L**ength **C**orrector; the LLC ensures a constant phase for the timing reference signals). This limitation plus other infrastructure and logistic limitations allow us to assume the maximal baseline length will be 35 [km] (see also S5.1.5).

Taking:

- Fiber length difference: 30[km]
- Index of Refraction (IoR): 1.4682 (Fiber: Corning SMF-28E)
- Speed of Light (SoL): 299792458 m/s
- Theoretical travel time in fiber (IoR/Sol) = 4.897 ns/m
- longest baseline: 35[km],

the time difference due to fiber length differences is: 30[km] x 4.897 ns/m = 147[uS].

The time difference due to the array geometry is:

- minimum requirement: 35[km] x cos(20[degrees])/Sol = 110[uS]
- stretch goal: 60[km] x cos(20[degrees])/Sol = 188.1[uS].

Therefore the maximal delay compensation must ensure to be able to compensate a time differences equal to:

- minimum requirement: 147[uS] + 110[uS] = **257[uS]**
- stretch goal: 147[uS] + 188.1[uS] = **335.1[uS].**

Converting the previous result to memory size gives (for one station):
Buffer Size [bits] = Sampling Rate * samples' bit width * polarizations * basebands pairs * time difference.

For example, assuming the minimal requirements for the baseline length and sampling rate: Sampling rate: 16[GS/s], samples' bit width: 6-bits, polarizations: 2, Basebands: 2, and time difference: 257[uS], we find

Buffer Size [bits] = 98.688e6 [bits].

The above results only show the required memory size for compensating the maximal delay in one station. Aspects like buffer word size and a mechanism for accessing the buffer are not covered and strongly depend on how the data is processed within the correlator and what is the DeMux factor for dealing with the incoming samples.

6.1.6 Power Consumption

ALMA currently operates two independent correlators: the BLC and the ACAC, and soon there will be a third, the ACA Spectrometer (ACAS) for total power observations. The power used by these correlators/spectrometer in the current system at the AOS (i.e. 5000m site) are:

- BLC = 165 kVA (hardware) + 25.3 kVA (computing) + HVAC
- ACAC + ACAS = 66.7 kVA (hardware) + 20 kVA (computing) + HVAC

The HVAC (heating, ventilation, and cooling) requirements are served by three air handler units (AHU-1, AHU-2, both for the BLC correlator room, and AHU-6 that handles the ACAC+ACAS correlator room), that together use 62.5 kW, as well as, the power used by the AOS technical building heat exchanger / chiller (note this serves the whole building, not just the correlator rooms).

We expect that the 2nd Generation Correlator will take over the interferometric observations of the BLC and ACAC, but it remains TBD whether the ACAS will be upgraded and continue to be used for Total Power observations in the future (See S5.1.3).

Presently the power consumption constraints for an ALMA2030 correlator are unknown. However, we do know that the single 2nd Gen Correlator will be required to correlate ~100x more channels (covering 2-4x more bandwidth), and all ALMA antennas (66 up to 80; both as one interferometric array and using subarrays). Therefore, we adopt as a stretch goal, that the new correlator will not require more power than the current correlators/spectrometer combined (including hardware, software, and HVAC). Considering advances in FPGA and GPU technology (favored over ASICs by the 2020 Correlator Workshop), and that the 2nd Gen Correlator will hopefully be placed at the OSF (3000m site, significantly reducing the HVAC needs), this requirement seems feasible.

A more practical requirement may be that optimally the 2nd Gen correlator (+ACAS if it also remains in operation) should not *routinely* require the operation of more than one turbine, as this represents a significant break-point in cost. However, we also note that given the preferred parallel deployment scenario described in S8, it will be necessary to power both the old and new correlators for the period of deployment/commissioning, because while we do not envision simultaneous observing with the old and new correlatores, unnecessary power-cycling is to be avoided.

More detailed considerations of the 2nd Gen Correlator power consumption requirements will need to await a detailed assessment of the ALMA 2030 power budget, and confirmation of the viability of placing the correlator at the OSF.

6.2 Spectral Domain

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.2.1 Coarsest channel resolution	6.6 MHz (5% loss at 35 GHz, 0.2PB for 35km baseline)	2.9 MHz (1% loss at 35 GHz, 0.2PB for 35km baseline)	Avoid BW smearing at lowest freq/max baseline for largest correlated BW mode; 0.2PB = 20% level of the primary beam response. Used for continuum and high-z spectral scans.
6.2.2 Finest resolution at max correlated bandwidth	23.35 kHz (0.2 km/s at 35 GHz)	11.67 kHz (0.1 km/s at 35 GHz)	Spectral Scans of line rich Galactic sources, as well as high-Z absorption-line studies
6.2.3 Finest possible channel resolution	2.33 kHz (20 m/s at 35 GHz)	1.17 kHz (10 m/s at 35 GHz)	Very cold clouds, protoplanetary disk structure, small-body rotation, interplanetary probes
6.2.4 Bandwidth at finest possible channel resolution	1.6 GHz	1.6 GHz	Derived from max channels required for S6.2.2 minimum goal
6.2.5 Linearity as a function of input power	14 dB	17 dB	Considerations: power level changes during observations of bright continuum and maser sources
6.2.6 Spectral dynamic range	10000:1	15000:1	Considerations: of absorption against strong continuum sources and maser emission; satellite constellation downlinks in Bands 1--3.
6.2.7 Bandpass flatness	0.2 dB peak-to-peak (passband filter prior to correlation)	0.1 dB peak-to-peak	Post-digitization ripple added by digital filtering prior to correlation (single channel or channel average)
6.2.8 Normalization by autocorrelations	Default mode but must be reversible offline per integration	Default mode but must be reversible offline per integration	Likely needed for Telcal and QA0 to ensure spectral flatness if channel averaged assessments used
6.2.9 Spectral channel independence & leakage	-60 dB	-80 dB	Power at least 60 dB down at filter stop band edge defined by filter transition band

6.2.10 Consistency of overlapping frequencies	Overlapping spectral windows supported.	Overlapping spectral windows supported.	Identical spectral responses for overlapping or discontinuous SPWs
6.2.11 RFI Flagging	None, but RFI monitoring should be implemented.	RFI flagging synced with input data stream if found to be necessary	Flagging on short timescales by the correlator could be useful if new RFI sources in the ALMA bands are intermittent, however this is highly uncertain at present

6.2.1 Coarsest channel resolution

An easily configurable correlator observing mode, with relatively coarse spectral resolution that minimizes losses, including from bandwidth smearing at the longest baseline lengths, is essential for ALMA operations. Such a “low” data rate mode would be analogous to the usage of the current TDM (Time Division Mode of the BLC, with a dual polarization channel width of 15.625 MHz[14]), and would routinely be used for many calibration purposes, as well as science observations of continuum sources and high redshift spectral scans (depending on spectral resolution needs). For a modern FX correlator design, it is likely that the “native” spectral resolution will be significantly higher than any “coarsest” spectral resolution constraint, so this requirement may turn out to simply be a requirement on the coarsest resolution observing mode (using post-correlation channel averaging, S6.7.2, S6.8.4). However, it is possible that such a mode could be generated internally by, for example, doing some or all of the channel averaging before multiplication and then using a higher number of bits for improved correlator efficiency. At minimum, the output spectral resolution employed for the “coarsest” resolution mode must avoid significant bandwidth-smearing effects which we evaluate below.

As described in S5.1.5, we expect that in the ALMA 2030 timeframe, the maximum baseline lengths are unlikely to grow beyond about 35 km, with a maximum of 60 km imposed by the geography of the surrounding region. The lowest operating frequency of ALMA will be 35 GHz once the new Band 1 receivers are operational in the next few years. To assess the effects of bandwidth smearing on an image, it is necessary to define the power level of the primary beam that will be imaged. In the current ALMA pipeline (Cycle 7, and indeed since its inception in Cycle 4), by default the image products are created with a field of view equivalent to the 0.2 level of the primary beam (PB) response for both single field and mosaic imaging (though a reduction in field of view to image can be triggered if the image products would be excessively large). Using the formalism described in *Synthesis Imaging in Radio Astronomy 2, Bridle & Schwab 1999,* the coarsest resolution that could be used with <1% loss at the 0.2PB level at 35 GHz, and max_baseline=35 km is 2.9 MHz (the coarsest resolution limit would be roughly half this for a 60 km max_baseline). For this calculation we have employed the “square-shaped” rather than “Gaussian” channel response since we expect the 2nd Gen Correlator to produce channels that have a high degree of independence (see S6.2.6).

---

[14] It is notable that using the methodology described here, the TDM dual polarization mode of the BLC, accounting for the automatic online Hanning smoothing to an effective resolution of 31.25 MHz, will produce a bandwidth smearing loss of ~20% at the 0.2PB level for a 16.2 km max_baseline at 35 GHz, i.e. FDM modes will be required for all wide-field long-baseline observations at both Bands 1 and 2.

The value of 2.9 MHz is then the maximum (coarsest) spectral resolution that will avoid significant bandwidth smearing at the 0.2PB level at the lowest possible observing frequency, at the ALMA 2030 goal max_baseline length of 35 km, and represents the stretch goal. It may be desirable to relax this requirement if the relatively high resulting data rate compared to now for TDM (and associated data volume) for this commonly used mode is deemed problematic for ALMA operations (see S7). For example, using the same assumptions as before, but only imposing a <5% loss at the 0.2PB level at 35 GHz and 35 km max_baseline implies a coarsest resolution of 6.6 MHz (the corresponding loss at the upper end of Band 1 would then be ~2.5%), and represents the minimum requirement goal. If the coarsest resolution constraint is imposed post-correlation, then it could also be increased as a function of observing band and/or baseline length, ideally with <1% loss at 0.2PB through channel averaging (S6.8.4).

A coarsest resolution of 6.6 MHz would conveniently provide < 10 km/s velocity resolution at 200 GHz and above. In Bands 6-10, such a resolution will comfortably support high-redshift spectral line surveys and line imaging, which are essential techniques in support of one of the new key science drivers for ALMA2030 - "*The Origins of Galaxies*" (S3). For example, the recent ALMA detection of the 88 micron [O III] line in a z=8.312 Lyman break galaxy at 364 GHz with a fitted linewidth of 140 km/s was performed with 7.8125 MHz channels (6.4 km/s resolution) via an FDM setup with 16x online channel averaging[15] (Figure 2). Also, because it shall be possible to configure the new correlator to frequency resolutions at many steps in between the coarsest and the finest values, a matching velocity resolution will be selectable in the lower bands (1-5) as well.

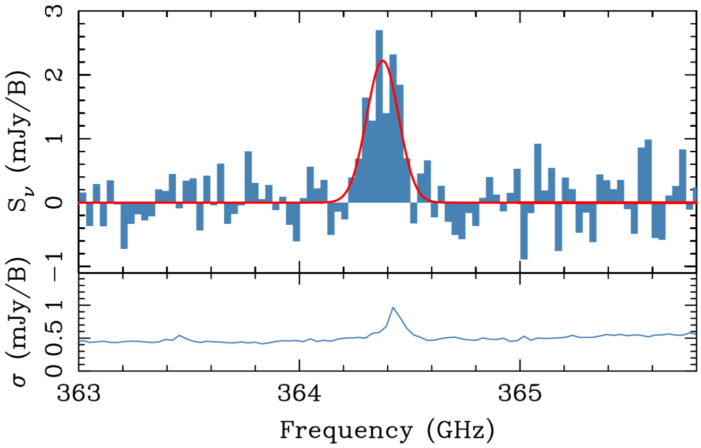

Figure 2: *ALMA detection of [O III] at z=8.312 from Tamura et al. 2019.*

6.2.2 Finest resolution at maximum correlated bandwidth

One of the most demanding modes of the 2nd Gen Correlator will be to correlate data at the maximum possible bandwidth with the finest spectral resolution possible at that bandwidth. On the other hand, such a mode is also one of the most highly anticipated capabilities for ALMA2030 by the scientific community -- finally having the ability to do spectral scans of line rich sources with both high spectral resolution (to capture fine kinematic details) with wide bandwidth (to simultaneously capture many diagnostic lines efficiently for accurate chemical and physical modeling). Indeed, spectral scans of protoplanetary disks are one of the new key science drivers for ALMA2030 - "*The Origins of Chemical Complexity*" (S3). By its nature, achieving this science goal requires the comparison of data from numerous protostars across the full range of protoplanetary formation environments and evolutionary states, implying that the capability must be not only possible, but it must be efficient[16]. It is important to note that wide bandwidth at high spectral resolution affords further scientific benefits than merely more efficient observing. A major source of uncertainty with such datasets arises from

---

[15] https://ui.adsabs.harvard.edu/abs/2019ApJ...874...27T

[16] https://ui.adsabs.harvard.edu/abs/2019BAAS...51c..81C/

the natural differences in calibration (absolute flux, phase transfer, antennas present, uv-coverage etc.) that result from data taken at different times. Obtaining the required data with the fewest possible tunings minimizes these added uncertainties, maximizing the science return.

One of the most impactful high spectral resolution spectral scans undertaken with ALMA to date is the *Protostellar Interferometric Line Survey* (PILS) survey toward one of the most nearby astrochemical templates, the Class 0 protostellar binary IRAS 16293-2422 (Jorgensen et al. 2016[17], see Figure 3 below). To date, the PILS survey data has generated more than 22 refereed publications and is the subject of several hundred citations. Unfortunately, despite their scientific impact, such surveys with ALMA are currently time-consuming due to the limited ability of the current correlators to achieve high spectral resolution at wide bandwidth. For example, the Band 7 PILS survey (329 to 363 GHz) could only obtain data in increments of 1.78 GHz per tuning in order to achieve a spectral resolution of ~0.23 km/s (18 tunings in total!). Figure 3 below demonstrates the spectral coverage that will be possible with the 2nd Generation ALMA Correlator with at least bandwidth doubling and an order of magnitude more spectral channels, covering most of the band in only 2 tunings. Presently, it is unknown how representative the PILS results from IRAS 16293-2422 are of the earliest phases of star and planet formation, a question that requires ALMA2030 performance to answer (i.e. spectral scans that are efficient in observing time).

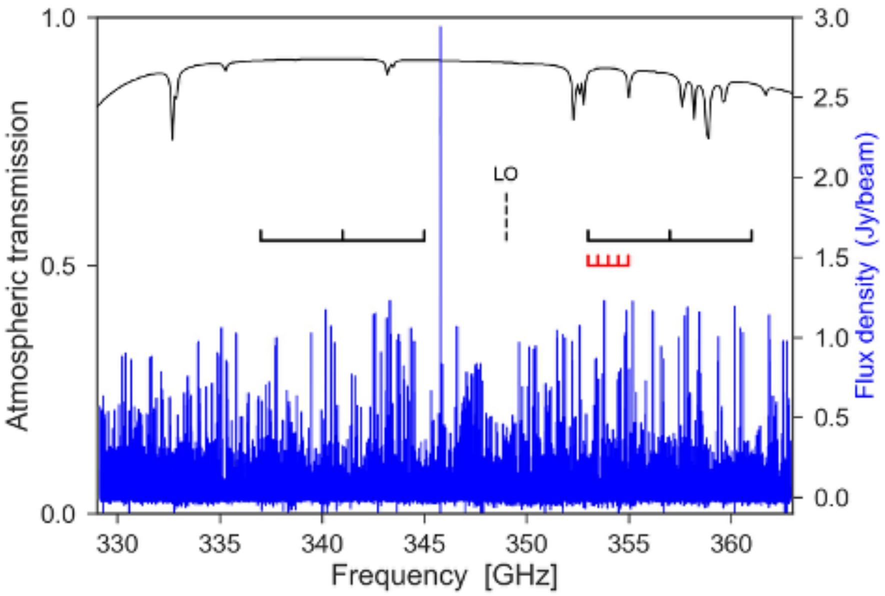

Figure 3: *ALMA PILS Band 7 spectral scan survey data toward IRAS 16293-2422, the red line segments show the spectral range that can be obtained in one tuning with the BLC at ~0.23 km/s spectral resolution. The black horizontal line segments show the frequency coverage that will be possible with the 2nd Generation Correlator with 8 GHz per polarization per sideband.*

Though spectacular, it is noteworthy that the PILS survey data cannot be substantially improved because it already achieved a sensitivity equivalent to the spectral confusion limit at Band 7[18]. The ability to find new molecular species (which must be weak to have eluded discovery thus far), including those that may be biologically significant, requires similar surveys to be carried out at lower frequencies where the line confusion (spectral density) is significantly lower, but where the requisite spectral resolution requires proportionally narrower frequency channels. Additionally, many biologically interesting molecules are complex, containing > 6 atoms (i.e. COMs, complex organic molecules). The detailed assessment

---

[17] https://ui.adsabs.harvard.edu/abs/2016A%26A...595A.117J/

[18] https://ui.adsabs.harvard.edu/abs/2020arXiv200607071J/

of the detectability of such molecules is challenging, but a general rule of thumb is that spectral complexity (density of line transitions) scales with the number of atoms. Higher spectral complexity generally equates to weaker intensity for each individual transition, and for molecules with similar elemental composition, the larger molecules will have their rotational transitions, including the ground state, shifted to lower frequencies[19]. Therefore, to reduce line confusion and target large biologically interesting species, high spectral resolution at wide bandwidth at the lower frequencies of ALMA, coupled with ALMA's exquisite sensitivity, will offer the most *new* spectral scan discovery space.

Thus, we must set the requirements for the finest resolution at maximum correlator bandwidth based on the **lowest** ALMA frequencies. For example, with the current correlators, at the lowest frequency of Band 2: 67 GHz, in order to achieve a velocity resolution ≤ 0.2 km/s (0.14 km/s), the associated aggregate bandwidth would only be 0.234 GHz, so it would take ~209 tunings to cover the full Band 2+3 (67 to 116 GHz)! **We take as the minimum goal for the finest resolution at maximum correlator bandwidth (S6.7.2) a spectral resolution of 0.2 km/s at the lowest RF of Band 1: 35 GHz, which requires a frequency resolution of 23.35 kHz. For the stretch goal, 0.1 km/s should be possible at the lowest RF of Band 1: 35 GHz, which requires a frequency resolution of 11.67 kHz for the maximum correlated bandwidth.** Thus, the minimum goal will enable spectral scans at the lowest frequencies of ALMA, at the same spectral resolution as the PILS survey (i.e. 5 channels across a 1 km/s spectral feature), but at a frequency favorable to the discovery of new molecules. As described in S7, the requisite data rates will be very challenging to handle and process such data downstream, but the data mitigation options in S6.7.2 can be employed until such modes are feasible. Additionally, it should be noted that via channel averaging, it should be possible to select a wide range of spectral resolutions between the minimum value and the coarsest possible resolution (S6.2.1) -- at minimum in powers of 2 x minimum possible channel width.

6.2.3 Finest possible channel resolution

The current ALMA correlators have a finest dual polarization spectral resolution of 92 m/s at a frequency of 100 GHz (taking into account the need for Hanning smoothing), a factor of nine times poorer than the original ALMA science specification of 10 m/s at 100 GHz [AD05] -- indeed this is one of the few high level ALMA science requirements that has not yet been met. The need for such high spectral resolution includes (1) spectrally resolving molecular emission from very cold clouds and comets, and non-thermal maser profiles (2) probing the kinematics of protoplanetary disk structure, especially the cold gas in the mid-plane, and (3) measuring atmospheric winds in Solar System objects.

The narrowest spectral line emission that can arise from cold (thermally excited) molecular gas is >= to its Doppler line-width (FWHM) in km/s $\Delta Vthermal$= 0.214*sqrt(T/mA), where T is the temperature of the gas, and mA is the atomic mass of the molecule. The Doppler line width for heavy molecules at low temperature, say 10 K can be very small indeed, take for example CN $\Delta Vthermal$= 0.13 km/s. The observed linewidth may be increased by non-thermal turbulent or bulk motions along the line-of-sight, but these effects can be minimized by using an angular resolution that spatially resolves the kinematics. Especially, for science cases (1) and (2), it is essential to be able to spectrally resolve the emission at longer wavelengths since as described in S6.2.2, large complex organic molecules tend to have their brightest emission at lower frequencies, and it reduces the high

---

[19] https://ui.adsabs.harvard.edu/link_gateway/2018ApJS..239...17M

continuum opacity present at shorter wavelengths for protoplanetary disks. A few recent examples of thermal line science results that required “ultra-high” spectral resolution include:

1. **Dark cloud chemistry:** The GOTHAM project at GBT recently detected two new heavy organic molecules in TMC-1 via observations of numerous transitions in multiple bands including Ka band (26-40 GHz). Several tunings of the spectrometer, each covering 0.75 GHz of instantaneous bandwidth, were used to survey the predicted lines at a resolution of 1.4 kHz (0.012 km s^{-1} at 35 GHz), which was essential to resolve the hyperfine splitting of the line profiles and add confidence to the detections. Propargyl cyanide ($HCCCH_2CN$) and benzonitrile ($c-C_6H_5CN$) showed linewidths of 0.144 and 0.121 km/s, respectively[20].

2. **Protoplanetary disk kinematics:** The nearby disk TW Hya was observed by ALMA in the three lowest transitions of CN, including at the highest possible spectral resolution in Band 3 (31 kHz = 80 m/s at 113 GHz).[21] An excitation analysis showed consistency with formation models via vibrationally-excited H_2 in the upper layers of the disk atmosphere. In addition, the Band 7 observations of the N=3-2 transition, taken with only 215 m/s spectral resolution, revealed hyperfine splitting previously unresolved by laboratory spectroscopy.

3. **Atmospheres of Solar System Objects:** ALMA Band 7 spatially-resolved observations of Titan in nine molecular species at spectral resolutions as fine as 50 m/s provided direct measurements of the atmospheric winds via Doppler shifts in six molecules.[22] Analysis of the images revealed clear evidence of zonal winds including an equatorial jet that reaches into the thermosphere to a height previously unconstrained by the Huygens probe. The rotational velocity shift between the east and west limbs is only ~400 m/s (Figure 4).

---

[20] https://ui.adsabs.harvard.edu/abs/2020ApJ...900L..10M

[21] https://ui.adsabs.harvard.edu/abs/2020ApJ...899..157T

[22] https://ui.adsabs.harvard.edu/abs/2019NatAs...3..614L

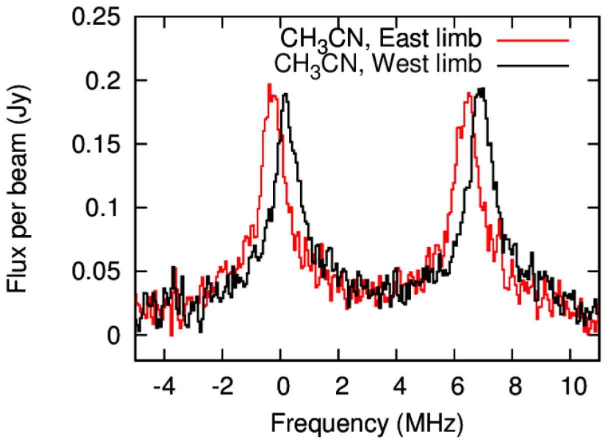

Figure 4:
Overlay of two spatially resolved ALMA spectra of Titan showing the two lowest K-components of the methyl cyanide CH_3CN J=18-17 line at 349 GHz (Lellouch et al. 2019). The spectral resolution is 61 kHz = 50 m/s and the on-source time is 2.5 hours. To achieve a comparable velocity resolution in Band 1 (such as on the J=2-1 line at 36 GHz) would require a frequency resolution of 6 kHz.

In the era of ALMA2030, the lowest ALMA observing frequency will be 35 GHz using Band 1. **Thus, the minimum (highest) spectral resolution required is 2.33 kHz, consistent with better than 20 m/s velocity resolution at all ALMA observing frequencies, including 35 GHz. This goal will allow for at least 5 channels across the narrowest currently known thermal lines, and make it possible to observe at 100 GHz with a spectral resolution of 7 m/s, and finally meet the original ALMA science requirement.** This spectral resolution will also be sufficient for non-thermal maser emission which can also have line feature widths as narrow as ~0.1 km/s, especially during the exponential growth phase when their widths can be narrower than pure thermal[23]. For example, when observed at 0.05 km/s resolution at the VLA to measure the Zeeman effect, the 36 and 44 GHz methanol maser lines in DR21W show typical linewidths of 0.2-0.6 km/s[24]. Similarly, the SiO 43 GHz J=1-0, v=1 and v=2 maser features toward evolved stars are often much broader than 0.2 km/s, which is the typical resolution they are observed with by numerous authors[25,26]. Note: if a window function is required to damp spectral ringing effects, or the channel independence is worse than expected, then the finest channel width will need to be commensurately smaller to achieve the required spectral resolution. The stretch goal is 1 kHz, consistent with better than 10 m/s spectral resolution at all ALMA frequencies (> 35 GHz), though we have not found a concrete science goal that requires such an extreme spectral resolution at 35 GHz -- such a signal would likely need to be man-made such as RFI or an interplanetary beacon. The bandwidth at which such fine spectral resolution is required is covered in the next section.

6.2.4 Bandwidth at finest channel resolution

This requirement is set by two orthogonal facts: (1) the velocity extent of spectral lines that one might wish to observe with ultra-high spectral resolution is limited, implying that a relatively narrow bandwidth is required per transition; and (2) it is most scientifically interesting and effective to observe such a line in combination with other diagnostic transitions of the same or different molecules (at ultra-high spectral resolution or not, as appropriate), together with a decent amount of continuum bandwidth. In the case of (1), it is important to resolve the emission at a particular spatial location which may be very spectrally narrow, but also to cover the whole kinematic structure of the object which could be up to tens of km/s wide in a single spectral window (protoplanetary disks, protostars,

---

[23] http://articles.adsabs.harvard.edu/pdf/1991ApJ...367L..63N

[24] https://ui.adsabs.harvard.edu/abs/2019ApJ...872...12M/

[25] https://ui.adsabs.harvard.edu/abs/2020A%26A...638A..17O

[26] https://arxiv.org/pdf/2009.09771.pdf

AGB star stellar winds, atmospheric features of Solar System objects); note we do not consider outflows as prime targets for ultra-high spectral resolution. For (2) the scientific possibilities are essentially unbounded. Therefore, we use values from other requirements to set an upper bound on the total number of channels that can be used. For example, the minimum goal of S6.2.2: 0.2 km/s resolution at the minimum requirement for the maximum correlated bandwidth of 8 GHz per pol per sideband at 35 GHz implies the need for at least 685,257 channels per polarization. If this value was to be considered the largest number of channels (per polarization) available to use at ultra-high resolution (20 m/s), and 32 spectral windows per baseband are available for placement (S6.8.2), then each window could use up to 428 km/s of bandwidth, i.e. 50 MHz per spw, for an aggregate bandwidth of 1.6 GHz. This maximum bandwidth is likely to be sufficient as the stretch goal as well.

6.2.5 Linearity as a function of input power

It is expected that the correlator will act as a linear device within the range of power expected during normal observations. By this, it is understood that at any stage of the processing, the output of that stage has to be strictly proportional, for the whole spectral range involved, to the corresponding input at that stage. There are several steps where this linearity can be affected, i.e., frequency windowing/conversion, digitization and data transport, multiplication and data accumulation, where clipping can occur. The probability of clipping needs to be minimized at the engineering design stage, such as ensuring that the proper number of bits and requantization are utilized at each stage.

The input of the correlator at the start of an observation is set to the average power coming from the IF on an OFF position (i.e., including the Tsys but not the target itself), and the linearity must therefore be preserved for departures from that level whenever a target is included in the beam. It is expected that the largest signal deviations from the Tsys levels will occur whenever the Tsys values are at their lowest, and the brightest targets on the sky/Hot Loads are included. The lowest Tsys are expected at the lowest frequency range of the ALMA telescope, where Tsys is dominated by the instrumental components of the ALMA system and the atmosphere contributes very little. As an example in the continuum mode, an observation of the center of the Moon would result in a 12x increase in power, that is 10.8 dB (for Tsys ~ 30K and Moon brightness temperature at center ~330 K), and similarly, a measurement of the Hot Load in the ACD would produce a 13.3x increase in power, that is 11.2 dB. The most powerful spectral-line emission sources are the cosmic masers, which may exhibit very narrow lines with flux densities as high as a few times 10000 Jy in some exceptional cases. For the same values of Tsys (equivalent to ~1260 Jy) and a putative maser peak emission of ~30000 Jy, this implies an increase of power by a factor of 23.8, or 13.8 dB. In terms of total power increase, absorption-line projects can be accounted for with the expected continuum level of the background astronomical source, and therefore, are already covered by the above estimates. Assuming that the RFI is not significant during normal observations (see section 6.2.11) **it is therefore required that the correlator can linearly handle at least differences of power input of about 14 dB without any changes in internal set-ups/configurations during a given observation.** Given the expected performance enhancement in the low frequency bands, **a factor of 2 buffer over the minimum requirement is what is considered the stretch goal for the specification, i.e., 17 dB.**

6.2.6 Spectral dynamic range

Spectral dynamic range represents a different kind of requirement. In this case, the correlator is expected to behave linearly despite a wide range of power levels within the same spectrum. For pure continuum observations at (sub)millimeter wavelengths, astronomical sources tend to exhibit smooth, linear or gently curved continuum variations as a function of frequency (for the bandwidths of the

ALMA receivers), and therefore, the expected variations are quite modest (~ less than 3 dB as specified by the FE in an auto-correlation, and significantly smaller for cross-correlations). The stringent limits come, therefore, for hybrid projects (continuum + absorption lines), for maser observations (with weak or no continuum emission), or for non-astronomical sources. In such cases, and assuming that the cross-correlations eliminate completely the uncorrelated signals between any pair of antennas, the dynamic range within a given spectrum can be in some cases as high as ~ a few 10000:1. Only the brightest sources will achieve this dynamic range in timescales of ~ 1sec for any 2-antenna baseline, but the requirement should drive the specification in this case. **Therefore, as a minimum the correlator should be linear in the representation of spectra with a dynamic range of ~ 10000:1**.

A possible stretch goal would be to try to keep the linearity of the spectral dynamic range also for the extreme situations of strong RFI/pick-up of satellite signals in the far sidelobes of the antennas. Estimates of the possible noise contribution from satellites, Gie Han Tan (ESO) has provided us with an estimate of 500K excess in the Tsys for Bands 1-3 (that is 10dB excess). However this will result in a quite weaker contribution to the cross-correlated flux of any observation as the delay/phase errors will vary quickly between antennas. If a satellite ended up near the primary beam of the antennas, it is expected that the power would increase by several tens of dB and will quite probably saturate analog saturation at the FEs, and therefore will not be considered in these requirements (see 6.2.11). Assuming that a 500K narrow-band signal from a satellite is included in the spectra, **the stretch goal for the correlator would be a linear representation of spectra with a dynamic range of 15000:1**.

6.2.7 Bandpass stability and correlator passband flatness

Gain variations over a baseband can be assigned to the Front-End subsystem and to the analog part of the Back-End including the analog input stage of the analog-to-digital converter (ADC) device. In the current Front-End subsystem, less than 5 dB peak-to-peak (p-p) is allocated over the entire IF. Assuming gain variations allocated to Back-End are < 3 dB p-p, the ADC input adapter/amplifier must contribute well below 3 dB p-p if an IF down-conversion stage is required after the Front-End. Bandpass shape and time stability of the analogue part of the ALMA system is calibrated on sky thanks to interferometric bandpass calibration techniques. Stability impacts the spectral dynamic range.

An additional weak ripple could be added by the fully digital filtering subsystem of the Correlator. This ripple does not vary with time and thus does not contribute to any bandpass changes with time. The current 2-stage FIR filter (TFB) achieves a passband ripple better than 0.3 dB p-p over more than 93% of the 62.5 MHz nominal passband. A polyphase filter coupled to an *n*-point FFT could achieve 0.1 dB p-p passband ripple and 60 dB stopband attenuation. 0.1 dB ripple keeps the SNR efficiency loss below 0.1% with respect to the ideal flat case and enables averaging several individual channels to provide broader channels with ripple within 0.2 dB, still within 0.1% SNR degradation. To avoid remaining spectral platforming effects, fine channel-to-channel gain power alignment shall be supported.

6.2.8 Normalization by autocorrelations

In the current ALMA correlators, the cross-correlation products are normalized by the autocorrelations, so that the output is fairly flat across the width of the spectral range selected in the cross-correlation [RD05]. There are, however, situations in which this normalization is not optimal.

In particular, for channels at the band edges and for observations dealing with bright sources (relative to the T_{sys})[27].

For the latter, whenever the flux of the source is a significant fraction of the equivalent flux of the system temperature, since the autocorrelations also include the source emission, they will alter the flux scales of the final products, by a factor of 1/(1+Tsource/Tsys), in a way that is not easily recoverable. The only options for proper data reduction of those datasets is either undoing the normalization or devising complicated atmospheric calibration methods including observations on source and off source. Given the normal slow cadence of atmospheric calibrations, corrections that use them must assume that in-between calibrations the behavior of the possible variations have been linear in time, which may not be an adequate assumption for 5-15 minute timescales. **It is therefore desirable that the data that the correlator either allows the undoing of the normalization per integrated sample or that the option for un-normalized cross-correlation data products is offered.** Depending on which of the two options is offered, the offline data reduction would either undo the automatic normalization and then apply an improved normalization, or proceed directly to the computation of the correct normalization factors. In both cases, the observations for targets that require these offline corrections would have to include data outside the target area. Given that the number of projects that is expected to be affected by this problem is a small fraction of the total, the default output option should continue to be that the cross-correlations are normalized, and that the autocorrelation data is also saved with the same cadence as the cross-correlations (as is currently the case for the ALMA BLC). However, to tackle the issues outlined in this section, the correlator should **in addition** offer the possibility of un-normalized output **without** significant time loss to operations (perhaps through some parameter that is passed to it at the start of a specific observation). For such observations, the correlator should still output both the (un-normalized) cross-correlation and auto-correlation products for processing down the line. **The minimum requirement is therefore that the correlator offers the choice of normalized and un-normalized cross-correlations, and the stretch capability is that switching between these two possibilities can be done without significant interruption of operations (~ a few minutes).** The implications on processing by Telcal and Pipeline for an unnormalized data stream will need investigation.

6.2.9 Spectral channel independence & leakage

The measured cross-power spectrum of an FX correlator varies as a sinc squared function which has low sidelobe responses compared to the sinc function response of an XF correlator. The spectral purity of an FX correlator can further be improved by the use, for example, of a polyphase filter preceding the FFT so as to provide a steep channel-to-channel transition response. Digital filters can be designed with a low passband ripple (see S6.2.7) and a -60 dB stopband attenuation goal at the bottom of the filter transition width between passband and stopband seems achievable. We thus expect that, after correlation and for any channel, the spectral channel independence defined as the power leakage of the correlated products from the passband of an adjacent channel is well below -60 dB. We conservatively suggest **-60 dB as our power leakage goal**. We note that this is consistent with similar requirements for SKA1_Mid. **A stretch goal of -80 dB** seems achievable.

6.2.10 Consistency of overlapping frequencies

The bandwidth of the spectral window (SPW) used for synthesis imaging can be as large as the maximum bandwidth presented to the analog-to-digital converter (ADC) or narrower, depending on how the digital filter subsystem is configured prior to correlation. The same spectral channel

---

[27] https://help.almascience.org/index.php?/Knowledgebase/Article/View/419

independence, spectral channel leakage, dynamic range and passband flatness shall be possible whether processing a single SPW or overlapping SPWs. The SPWs must have adjustable frequency centers and widths (see also S5.1.4 and S6.8).

Calibration derived for one source from one spectral window with one spectral resolution must be transferable to other sources, spectral windows, and/or spectral resolutions, without the correlator system introducing additional errors (beyond those due to atmospheric dependence on frequency, position, and/or time) above 0.1% in amplitude and 0.001 radians in phase.

6.2.11 RFI Flagging

Various spurious signals may present themselves in the basebands processed by the correlator. Sources include external RFI from terrestrial transmitters and satellites, and internal RFI from instrumentation such as the YIG oscillators in the WCAs and their harmonics, the LO chain of the WVRs, a second LO if present (i.e. the current LO2s), the digitizer clock (which at 40 GHz will fall within Band 1 and have harmonics in higher bands), and digitizer interleaving artefacts (if interleaved samplers are used). Spurious external and internal signals entering the IF chain can be suppressed in cross-correlation by 180º Walsh switching and LO offsetting, and uncorrelated spurious signals between antennas are also largely suppressed in cross-correlations. However, all of them still impact autocorrelations which means they affect data calibration. Strong spurious signals relative to the system noise can also impact linearity and quantisation efficiency, or may even perturb the astronomical spectral analysis due to channel-to-channel finite spectral power leakage.

A particular concern for ALMA2030 are satellite internet mega-constellations, which have proposed downlink bands in ALMA Bands 1 and 2, and one uplink band extending even into ALMA Band 3. The details of these transmissions are as yet not well known. It is so far assumed that the RFI spectrum will be very wide-band (covering most of each ~4 GHz wide radio frequency spectrum allocation) and continuous in nature. Power levels are likely such that when a satellite passes near the primary beam of the ALMA antennas observations in the affected receiver bands will need to be completely flagged due to analogue saturation. For the majority of the time, when we have the situation of a superposition of many satellites in the far sidelobes of the ALMA antennas, the expected effect is an increased noise level in the transmit bands in the region of 10dB (with significant uncertainty). In either case, we don’t see yet a need for flagging or special treatment within the correlator for normal ALMA operations. The affected frequency ranges and transit time ranges can be flagged offline just as well (and this includes in processing of phased-array data). This assessment could change if the transmit duty cycle or instantaneous bandwidth utilization were so low that even with many satellites above the horizon there would be quiet periods or frequency ranges that could be momentarily exploited on sub-integration durations within the correlator.

We do not currently anticipate in the ALMA bands the sort of narrow-band intermittent RFI that is common at lower frequency radio bands (e.g. for pulsar low frequency observations) which could be usefully flagged by the correlator. Although external RFI in the IF frequency range (e.g. WiFi and cellular telephones) are intermittent, the expected impact, especially in cross-correlations, is negligible based on experience with the present system, which covers the problem frequency ranges in baseband (e.g. 2.4 GHz WiFi) or IF (e.g. 5GHz WiFi). However, this does depend on adequate shielding of the receivers and their interfaces.

As the future of external RFI in the ALMA bands is far from clear, it is plausible that high cadence RFI detection and flagging *may* be of use, so we specify this as a stretch goal. This flagging should apply both for correlated data and phased-sum output data. With Band 1 scheduled to begin deployment over the next year, it should be possible to achieve a greater understanding of the RFI implications on the ACAC/BLC before the 2nd Generation Correlator requirements are set in stone, and eventually decide if a dedicated RFI detection station would be of any use.

Internally generated spurious signals are generally continuous in nature, so also do not require flagging in realtime by the correlator, as the affected channel ranges can be equally well flagged offline. The correlator itself could potentially introduce spurious signals into the spectrum due to imperfections in processing stages. However, the level of these spurious signals can be monitored by implementing pseudo-random generators and built-in subsystem checks in the correlator cards. Spurious signals should be minimized to a level negligible in 100 hours of cross-correlation integration on each baseline.

The correlator subsystem can be configured as a multi-antenna phased-up VLBI station or, for pulsar observations, as a beam-forming engine providing several beams and pulse profiles over several frequency channels per beam (S6.5). RFI could be particularly detrimental to the data stream output from the spigot. This topic also requires additional investigation.

We recommend that an ALMA working group be formed to further monitor and analyze the likely need and best means of RFI excision (based on its nature) and specifically what “hooks’’ in the correlator may be needed.

6.3 Time Domain

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.3.1 Shortest integration duration (standard interferometry)	16 ms (cross) 1 ms (auto)	TBD	Needs science case investigation. Also: must be fraction of 48 ms timing event, and feasible to carry out calibrations, sideband separation (90° Walsh) etc.
6.3.2 Longest integration duration (standard interferometry)	Any: 10s	Any: 30 seconds	2048ms needed for 90° Walsh integrations; phase stability allows up to a few tens of seconds in some cases; long durations mostly just serve for data rate mitigation.
6.3.3 Correlator configuration time	< 0.96 s (20 TEs)	<=1 TE (48 ms)	See text for various additional caveats based on experience with the current correlator
6.3.4 Maximum delay rate	Sidereal at 35 km: 8.5 [ns/s]	4x sidereal at 35 km: 34 [ns/s]	Ability to map near-earth asteroids and comets. We do not find OTFI to be a driver for ALMA.
6.3.5 Minimum supported subscan duration	<= 0.96 s (20 TEs)	1 TE (48ms)	Subscan durations of ~2 second duration or less are needed for some routine calibrations and must be supported; minimum duration is also a multiple of the integration duration; 48ms is minimum possible subscan duration in the ALMA system; note there may be many such subscans in a sequence
6.3.6 Maximum supported subscan duration	>= 30 minutes	>= 30 minutes	Long scans without change of observing parameters should not need to be broken into multiple subscans (with inter-subscan overhead); bandpass scans can currently be up to 30 minutes duration

6.3.1 Shortest integration duration (standard interferometry)

This refers to the shortest integration produced by the correlator system for auto/cross-correlations (auto-correlation spectra and visibilities). As yet we have not identified a science motivation for shorter than the current 16ms/1ms numbers, but we note that one might come up in future, so we leave the stretch goal TBD.

Cross-correlation integration durations as short as 16ms are potentially of use to science observations of very rapidly varying phenomena on the Sun, and occultation experiments. The 16ms duration is very useful for commissioning tests and observations as it allows a time-resolved picture of many things happening in the system which would be averaged-out in common science observation integrations of ~1 second. It allows constraining timing issues to a particular TE edge or finer, and allows dumping individual 90° Walsh switch states. If the 90° Walsh switching demodulation by the correlator subsystem will be carried-out completely in software (as e.g. the ACAC does), then the correlator hardware must be able to produce 16ms (or 16ms/N, with N integer) dump durations to allow this.

To support continuum single-dish observing, it is required to support autocorrelation-only integrations at least as short as 1ms. A correlator/spectrometer is needed for continuum single-dish observing to allow separating spectral-line contributions from the real continuum, and allow spectral weighting to mitigate strong atmospheric spectral features in the expected wide basebands. Data rates can be made feasible for a wide bandwidth either by using very coarse channelisation, or by online averaging over channel-average regions within each spectral window and only storing the channel-average data at high rate.

6.3.2 Longest integration duration (standard interferometry)

From a scientific perspective, the choice of integration duration for a particular project is influenced by a number of factors including the maximum baseline length, atmospheric phase stability, and the brightness of the science target (which sets the feasible range of self-calibration solution intervals). While longer dump durations primarily serve to reduce the data rate out of the correlator hardware, longer integration durations at the correlator subsystem output serve to reduce archiving data rates and processing times in other subsystems. If hardware binning of 90° Walsh switching states (as opposed to a fully software implementation receiving 16ms dumps) is required by the implementation then a dump duration of 2048ms, equal to the Walsh switching cycle duration, must be supported (and those dumps would be of 3 bins). The maximum integration duration at the correlator subsystem output must be at least 2048ms regardless of implementation. In the most stable observing conditions, for short baselines, it is sometimes feasible to average over some tens of seconds in an integration as output from the correlator subsystem. Thus we recommend supporting a maximum integration duration of at least 10 seconds, and a stretch goal of 30s as the longest we could conceivably consider recommending averaging over online. This integration duration could be achieved in the software and the hardware dump duration could be much shorter (e.g., the ACAC uses only 16ms dump duration).

The longest dump time will drive the design of the accumulator, so the needed accumulator (X stage) must be able to store results according to the longest dump time and the Correlator output format (see S6.7.2, S6.7.3, and S6.8.5 includes constraints due to time-smearing).

6.3.3 Correlator configuration time

Prior to a subscan execution, and between subscans in a sequence, the correlator system may require some time for configuration and/or calibration. What actions are required is implementation dependent, and could include things such as: changes of hardware operating mode, measurement and/or upload of post-re quantization scaling factors/spectra. and initialising delay compensation. Updates to hardware common to multiple subarrays can introduce further complications. Measurement of any required calibrations can require interaction with other systems to ensure that the calibration is made when the input signals are as expected.

The setup time for subscans directly contributes to observation time overhead and science calibration cycle times when the time is greater than other overheads that can occur in parallel, or in general if actions cannot be parallelised (e.g., calibrations). Note that during ALMA observations correlator configurations/calibrations generally must be changed between types of scan (e.g., switching between a full-bandwidth short-integration mode for pointing/focus calibration scans, the chosen bandwidth/resolution mode for scientific scans, and a short-integration autocorrelation-only mode for atmospheric calibration scans). A worst-case is when “bandwidth-switching“ phase referencing is employed. This entails switching between the desired narrow-bandwidth (high spectral resolution) mode on the science target and a full bandwidth mode on the gain calibrator to achieve sufficient S/N. This mode requires “differential gain” calibration scans, which rapidly cycle the two modes on a single calibrator source in order to measure the instrumental differential gain between the two modes, with minimum atmospheric variation between mode switch cycles. This is a case for which correlator configuration time both determines directly the observing overhead, and the calibration cycle time (and thus accuracy).

To eliminate a correlator contribution to subscan setup overheads in general, the configuration/calibration durations would ideally always happen within a 1 TE (48ms), as this is the minimum non-zero time possible between subscans in ALMA. Thus this is what is specified as the stretch goal. This is likely to be difficult to achieve. In the current system, correlator configuration overheads are up to around 1.5 seconds, which has become reasonable with various optimisations in the correlator software. As an improvement is highly desirable, we set a compromise minimum requirement of <0.96 seconds (1 second rounded down to the nearest TE), but this must be subject to the following refinements:

- Configuration/calibration actions for one subarray should not be significantly blocked by other subarrays, i.e., configuration of one subarray should not have to wait for completion of an on-going subscan in another subarray (which may be many minutes in duration). If configuration/calibration actions of multiple subarrays would overlap in time, the total time should not exceed the sum of the two actions (i.e., no further added overhead), and ideally they should occur in parallel so that overheads in each subarray are not increased at all.
- The correlator software should minimise hardware configuration actions by inspecting the subscan specifications in each sequence and only perform configuration actions between subscans if something really changes (e.g., the correlator mode differs between subscans). When configuration actions between subscans are absolutely required they should ideally be scheduled to occur in parallel with the previous subscan as far as possible, to minimize actions in the inter-subscan period.

6.3.4 Maximum delay rate

Assuming the correlator system is responsible for tracking (at least part of) the instrumental delay, this section discusses the maximum rate at which that delay should be updated. That maximum rate is set primarily by changes in the geometric delay, due to the rotation of the Earth, the proper motion of the object being tracked, and any further (commanded) changes to the pointing of the antennas. The geometric delay, τ_g, is defined as the difference in the arrival times of the wave front at the different antennas, and thus depends directly on the maximum baseline length, D.

The minimum requirement then is that the correlator correctly account for changes in the delay due to the rotation of the earth, for the maximum east/west baseline - i.e., for sidereal tracking. The geometric delay is:

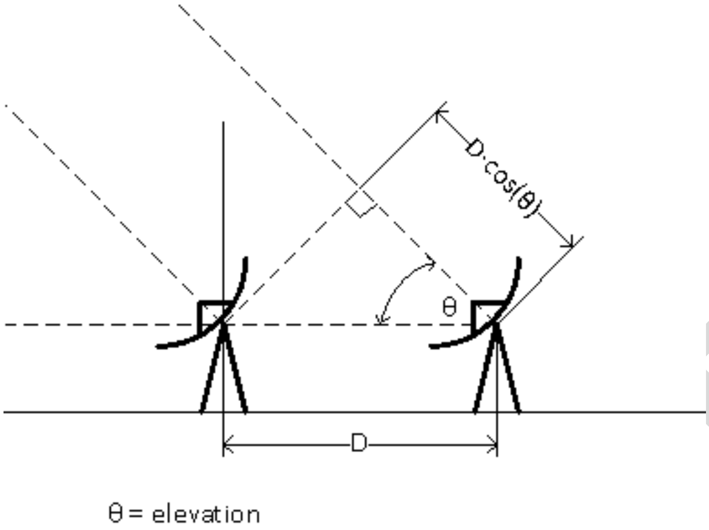

Figure 5: Diagram of geometric delay in a radio interferometer.

$$\tau_g = \frac{D \cdot cos(\theta)}{c}$$

where c is the speed of light (2.9979e8[m/s]), $\theta$is the elevation angle, and D is the east-west baseline projection (see Figure 5). The delay rate is first derivative of τ_g with respect to time:

$$\frac{\partial \tau_g}{\partial t} = \frac{D \cdot sin(\theta)}{c} \cdot \frac{\partial \theta(t)}{\partial t}$$

where the maximum $\frac{\partial \theta(t)}{\partial t}$for a source fixed on the celestial sphere is the sidereal motion, $\frac{2\pi}{Day} \simeq 7.27e - 5[rad/sec]$. The highest delay rate for such a source is at zenith,$\theta = \frac{\pi}{2}$, and the corresponding maximal delay rate is:

Maximal Delay rate = D[km] * 2 pi/day / 2.998e8[m/s] = 0.24 D [ns/s]

Taking D=35[km] as the required and 60[km] as the strength goal (see S5.1.5), this yields maximum delay rates for sidereal tracking of **8.5** and **14.6[ns/s]** respectively.

Near-earth asteroids and comets have their own substantial proper motions across the celestial sphere, requiring super-sidereal tracking. Apart from nearby airplanes, the most demanding case currently is that of Low-Earth-Orbit satellites (LEOs).[28] Such man-made sources have led the SKA to set their tracking requirement at 17x sidereal[29]. For ALMA a more appropriate goal might be that associated with near-earth asteroids and comets, for which a value of 4x sidereal may be sufficient[30]. *If* it is

---

[28] See https://en.wikipedia.org/wiki/Quasi-Zenith_Satellite_System, https://en.wikipedia.org/wiki/Tundra_orbit

[29] For SKA the intent is to track fast-moving RFI sources to facilitate the removal of their unwanted emissions.

[30] https://ictjira.alma.cl/browse/ICT-6008

important to map such objects at the maximum ALMA resolution, the corresponding maximum delay rates would be **34 (58)[ns/s]** for 35 (60)[km] baselines. Alternatively, the sidereal rates above would suffice for tracking these near-Earth objects on baselines as long as ~ **8.7 (15)[km]**.

Another potentially demanding case is *on-the-fly interferometry* (OTFI). In this mode data are correlated and written out as the antennas scan across the sky without stopping. This avoids the start-and-stop overheads of mosaicking observations, but requires dumping the correlator fast enough to track the motion of the primary beam. OTFI also leads to rapid changes in the array geometry as seen by the source, i.e., high delay rates.[31] OTFI is most useful for very shallow surveys of large areas of the sky, and it is not clear ALMA has the sensitivity to make this useful on its longer baselines. For instance, the 4x sidereal tracking rate mentioned above for 35[km]-baseline observations of near-earth asteroids would allow tracking 16x sidereal scanning for the 8.5[km] array, i.e., 16 * 15[arcsec/s]= 4[arcmin/sec]. ALMA's primary beam is ~3.4[arcmin] at 30 GHz, so scanning at this rate would give only ~3.4[arcmin]/4[arcmin/s]~ 0.9[s] effective integration time at each position; this integration time like the primary beam scales *down* linearly with frequency. The improvements embodied in the ALMA2030 program will lead to substantially higher sensitivity, but even so, it seems unlikely that faster scanning (and hence faster delay updates) will be required. The conclusion then is that, for ALMA, OTFI does not drive the desired maximum delay rate.

In sum, setting the requirements for maximum delay rate according to sidereal tracking on the longest baseline seems a reasonable approach.

6.3.5 Minimum subscan duration

Several routine calibration scans such as pointing, focus, and atmospheric calibrations employ subscans of a few seconds duration, and in some cases (e.g., very bright sources for pointing calibration, low spectral resolution atmosphere scans) would be more time efficient with even shorter subscans. Currently ALMA reliably achieves minimum subscan durations of about 2 seconds for a wide variety of setups. We recommend the minimum officially supported duration be 1 second, rounded down to the nearest TE, giving 0.96 seconds (as stated in S5.1.6, scans must start and stop on TE boundaries). With the same time also specified as minimum setup time, this means the worst-case time efficiency of subscan sequences due to the correlator subsystem will be 50%. The theoretical minimum subscan duration in the ALMA system is 1 TE, so we give 48ms as the stretch goal minimum duration. In all cases, the subscan duration should include an integer number of integrations.

6.3.6 Maximum subscan duration

Some use cases require staring continuously at a source without any change of parameters for many minutes. The most common example is a bandpass calibration scan, which for high spectral resolution, small numbers of antennas (e.g., 7m array), or low atmospheric transmission, can take up to 30 minutes. Breaking such scans into multiple subscans adds unnecessary inter-subscan overheads which reduce observing efficiency. We thus wish to be able to complete these scans as a single subscan. It is assumed that the output files from the correlator will be streamed in real time (e.g., after each integration is processed) so that output data buffering in the correlator subsystem is not a concern. We note that the maximum subscan duration may set requirements on the delay tracking if the implementation applies this in such a way that only a limited delay variation can be accommodated

---

[31] One can think of drift scans as a special case of OTFI moving east at the negative of the sidereal rate. Here the delay update rate is zero. But generally OTFI, to be useful, requires faster-than-sidereal mapping.

without data glitches (the TFB delay application in the BLC is the origin of this remark). We do not anticipate bandpass calibrations needing longer than 30 minutes in the upgraded system, so we do not give a larger stretch goal than the minimum.

6.4 Subarrays

There are many ways antennas can potentially be grouped into arrays and subarrays for achieving different science and calibration purposes and maximising use of the available hardware resources. These have varying levels of relevance to the correlator subsystem. For example, fully independent arrays of antennas performing different observations while sharing the same correlator are an important design consideration for the correlator subsystem (both hardware and software). This is what is most commonly meant by “concurrent subarrays” in the current ALMA system, with the BLC currently supporting 4 such independent subarrays (up to 16 in firmware) of arbitrary sets of antenna inputs, and the ACAC supporting 2 such subarrays of specific antenna inputs (0--3 for the TP array and 4--15 for the 7m array). At the other extreme, allowing subarrays of antennas within the same observing array to apply different pointing/focus (referred to as “pointing subarrays” in the ALMA control system, and routinely used for calibrations) doesn’t require anything particular from the correlator, as all baselines can usefully be correlated normally.

The Working Group has assessed the correlator requirements of the following potential subarray use cases for the next generation correlator (which is not likely exhaustive):

A. **Fully independent observations**
    - e.g. independent 12m science array + 7m science array + TP science array[32] + antenna integration + engineering tests + VLBI with one or a subset of antennas
    - possibly there could be cases for multiple independent 12m science observations in parallel if *uv* coverage is not critical and observing time is not driven by sensitivity (e.g. bright point source flux monitoring, polarization observations needing parallactic angle coverage but not sensitivity-limited)
    - each fully independent observation requires a separate photonic reference, so the maximum possible number is 6 (and there is no plan to change that number).

B. **Continuous phase calibration/monitoring (science source + calibrator with 2 sets of antennas)**
    - either using a single tuning or different tunings e.g. “band-to-band”
    - subscans in the participating subarrays should ideally be synchronous (same start times)
    - it is unclear if this is a strong use case as *uv* coverage and sensitivity are reduced for a given total number of antennas, and it imposes complicated constraints on array configurations in order to place monitoring and science antennas close together

C. **Simultaneous multiple delay centres for mosaic pointings or multiple sources**
    - subscans in the participating subarrays should ideally be synchronous (same start times)
    - could be useful for multiple bright simple sources (not limited by *uv* coverage or sensitivity)
    - the mosaic use case would need to consider variation of *uv* coverage between pointings but it may be interesting for rapidly varying bright sources like the Sun

D. **Simultaneous multi-band observations of one target (rapid time variability studies)**
    - subscans in the participating subarrays should ideally be synchronous (same start times)

---

[32] Although there may be a separate total power spectrometer (see S5.1.3), there is no reason why spectral line TP observations could not be performed with the next generation correlator given the requirements in this document (e.g. 6.3.1, 6.6.2)

  - different spectral window configurations should ideally be possible for each band e.g. to capture different available spectral lines, avoid atmospheric lines, and provide the most useful velocity resolutions (e.g. narrower channels for lower frequency bands)
  - there are potential use cases for rapidly varying sources e.g. Sun, peculiar stars, comets, occultations, GRBs
  - for each case it should be decided if this simultaneous approach or sequential band cycling is optimal e.g. considering uv coverage in addition to variability timescales
  - a single photonic reference may be used for all participating subarrays that use harmonically related LO1 frequencies

E. **Antenna pointing/focus offsets per set of antennas ("pointing subarrays")**
  - no correlator impact as a single correlator subarray is used
  - used generally for calibration purposes, e.g. different pointing/focus offsets for 7m and 12m antennas in the same array, or allowing reference (boresight) antennas with no offsets

F. **Interferometry+TP with same tuning (à la ACA "dynamic subarrays" concept)**
  - TP and interferometry scans need not be synchronous
  - need to switch between single and multiple subarrays on per-scan basis (calibration together, science scans apart)
  - would be especially useful if sources are significantly time variable
  - potentially interpolate TP antenna WVR data to 7m array (although this places an operational constraint on having 4 TP antennas available on the designated TP array pads)
  - a downside is that usually much more TP integration time is needed compared to 7m so simultaneous observing is not necessarily efficient

For VLBI another category of subarray use case arises: tied-array beams (see S6.5). This means independently forming multiple beams, i.e. tracking multiple delay directions, with a single set of antennas. This is used when a calibrator is available within the primary beam of the science target (e.g. using the core of a quasar as calibrator for an observation of part of a jet). As this case is quite distinct and is specific to VLBI, it is covered separately in the VLBI section (S6.5).

The requirements of these use cases relevant to the correlator subsystem are summarised in the following Table. These requirements have been distilled to whether or not subscans in multiple subarrays should be synchronous (i.e. start simultaneously), whether different correlator spectral configurations may be needed (spectral resolution, bandwidths, number of windows, location of windows within the IF/basebands, use of sideband separation), whether different integration durations may be needed, whether or not the subarray definitions may need to change during an observations, whether a common LO (photonic reference) could be used for the participating subarrays and whether baselines between antenna in the subarrays could be meaningful. Additionally for each use case it is noted whether the use case could be achieved without any correlator subarray support, i.e. including the different groups of antennas in a single correlator subarray with common correlator configuration, and whether the use can be achieved using the functionality required for performing fully independent observations in multiple subarrays.

Table: Summary of correlator related requirements (or desired features) of subarray use-cases

Use case	Synchronous subscans (same start times)	Different spectral window configuration needed	Different integration durations needed	Switch between single and multiple subarrays within observation	Common LO[5]	Inter-subarray baselines meaningful	Feasible without correlator hardware subarray support	Feasible with independent non-synchronous subarrays
Fully independent observations	no	yes	yes	no	no	no	no	yes
Continuous phase calibration / monitoring	desirable	desirable	no	yes	yes/no[4]	no	yes[1,2]	yes[3]
Simultaneous multiple delay centres for mosaic pointings or multiple sources	desirable	no	no	yes	yes	no	yes[1]	yes[3]
Simultaneous multi-band observations of one target	desirable	desirable	desirable	no	yes/no[4]	no	yes[1,2]	yes[3]
Antenna pointing/focus offsets per set of antennas	yes	yes	no	yes	yes	yes	yes	no
Interferometry+ TP with same tuning	no	no	yes	yes	yes	no	no	yes

[1] For these cases a single correlator subarray can be used for the observation, with the baselines between sets of antennas ignored.

[2] A caveat is that without correlator subarrays the spectral window configuration must be the same for all sets of antennas in the observations, which can be sub-optimal

[3] For these cases independent non-synchronous subarrays could be used, although synchronous subscan timing would be preferable

[4] Multiple photonic LOs would be needed in case of multiple bands (or tunings within one band) that cannot use harmonics of the same photonic LO frequency

[5] A "switch between single and multiple subarrays within observation" requires "common LO". Whilst not strictly impossible to swap photonic reference allocations to antennas during an observation, there are excessively long time overheads involved (up to 2 minutes).

From the point of view of the correlator subsystem, all the use cases can be covered by the requirements for fully independent subarrays, as long as

1. **Subscan times can be controlled such as to start at the same time in multiple subarrays (i.e. to allow enforcing synchronous subscan timing when desirable)**
2. **The time necessary to create and destroy subarrays, to allow joining and splitting groups of antennas within an observation, is no more than around 1 second so as to minimise overheads**

Additionally, to avoid entirely the overheads and control complexity of (1) and (2) for the use cases B and C, it is desirable to be able to define groups of antennas within a subarray for which cross-correlations between groups are excluded from the output data (or set to zero and flagged if sparse output data cannot be supported). These correlation groups should be selectable on a per-subscan basis to allow switching within a subscan sequence.

To handle all the possible correlator subarray use cases, and ensure operational flexibility, **it must be possible to select the antennas, i.e. correlator antenna inputs in each subarray with complete flexibility, and to create and destroy subarrays at any time without disturbing other active subarrays** (e.g. it must not be required to wait for subscans in other subarrays to complete).

Because it is envisioned that the 2nd Generation ALMA correlator will be responsible for correlating completely independent data from all three major components of ALMA: the 12m Main Array, 7m Atacama Compact Array, and the Total Power Array (a minimum of 3 subarrays), as well as allowing for other simultaneous activities (>=1 subarrays) concurrently, **it is essential that seamless subarray operations be fully implemented and commissioned prior to the correlator entering into service.** As documented in [RD13], the implementation of independent subarrays internal to the BLC (several years after its initial commissioning) proved time-consuming and difficult, and continues to present challenges to its seamless use even today (full commissioning of BLC "science subarrays" remains elusive).

The quantitative requirements for subarrays are given in the following table, followed by detailed explanations in the following subsections.

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.4.1 Independent (concurrent) subarrays	16	>24	ALMA currently has 5 production photonic references plus a pre-production unit; upgrades are not foreseen by ALMA2030 Roadmap. Maximum number of bands with common photonic reference frequency is 4 (Bands 3,5,7,10 at around 97GHz × 1,2,3,9). Minimum: 16 = 5 production photonic references times 3 bands (or other splitting reason)

			plus one subarray for non-science use. Maximum: 6 photonic references times 4 bands (or other splitting reason).
6.4.2 Correlation groups per subarray	1 (feature not supported)	$N_{ant\_in\_subarray}$ - 1	This feature is desirable to reduce data rate (or improve data compressibility if sparse data output is not supported) in some use cases. The maximum possible number of groups while still producing a cross-correlation is one less than the number of antennas in the correlator subarray (leaving a single baseline).

6.4.1 Independent (concurrent) subarrays

The correlator must support fully independent concurrent subarrays. This functionality is necessary to allow multiple independent observations with different sets of antennas to run in parallel using the same correlator. This functionality also supports the various use cases described previously where multiple subarrays would be used for a single observation.

For only fully independent observations the number of subarrays would be limited in practice by the number of photonic references and associated hardware in the ALMA CLO, i.e. to 6 subarrays. However, use cases for multiple subarrays within an observation can employ a single photonic reference for multiple correlator subarrays, either when all subarrays observe the same frequency or multiple receiver bands can use different harmonics of a single photonic reference frequency. The absolute maximum number of bands that can share a common photonic reference frequency is 4 (there is a small common range for Bands 3, 5, 7, and 10). We consider that a reasonable maximum number of subarrays to consider per photonic reference is 3, and we consider that the 6th photonic reference (a pre-production unit) will continue to only be used for simple observations requiring a single correlator subarray, so **the minimum requirement is to support 16 independent concurrent subarrays (5*3 + 1)**. This is the same number as implemented in the current BLC firmware. The stretch goal considers the potential for 4 correlator subarrays for all 6 photonic references for 24 in total. These numbers are also sufficient in practice to accommodate more fully independent observations if the CLO were to be expanded in future to include more photonic references, given that most observations will still use only a single photonic reference.

Independent concurrent subarrays must be implemented such that there is negligible impact on the performance of any subarray due to the execution of another subarray. For example, creating/destroying subarrays or setting up configurations for subscans should not induce pauses in execution of other subarrays. Subscan setup times should not depend significantly on the number of subarrays in operation (and if caching of configurations or calibrations in the firmware is used to reduce setup times, this caching should allow for multiple configurations/calibrations per subarray with the maximum number of subarrays). It must be possible to schedule subscans to start at the same time in different subarrays when needed (i.e. to allow enforcing synchronous subscan timing), in coordination with the ALMA control software.

Finally, the requirements for phased-array and other specialized observing modes that utilize the "spigot" access to baseband data are presented in S6.5, but we mention here explicitly that it must be possible for a subset of antennas (or a single antenna) to be used in a phased-array / VLBI mode in at least one subarray, concurrently with the other subarrays.

6.4.2 Correlation groups per subarray

To facilitate trivial subarray use cases where all groups of antennas should be using exactly the same configuration and timing, but baselines between groups are not meaningful (e.g. tracking different source directions on the sky), it is recommended to implement a mechanism to select which baselines within a correlator subarray are output. This will allow these use cases to be made with a single correlator subarray, reducing their complexity and minimising time overheads, while also minimising output data rate and potentially reducing correlator resource usage for the unnecessary baselines. If the output storage format will not be able to efficiently support a sparse set of baselines, there would still be benefit to compressibility by blanking the unneeded baselines, and there could still be correlator resource usage reduction.

As this feature would only be used by a relatively small subset of observations, and as the effect is mostly just to reduce archiving data rate, it is not considered as a minimum requirement. It is however a stretch goal, and the maximum possible number of groups one could consider per subarray is one less than the number of antennas (if one group per antenna were defined, the outcome would effectively be the same as specifying autocorrelation only). These correlation groups should be selectable on a per-subscan basis to allow switching within a subscan sequence.

6.5 VLBI

During the 2005 ALMA Re-baselining Activity for construction cost-savings [RD04], the ALMA implementation for VLBI/Phased-array observations was de-scoped except for the provision that the capability must not be designed-out. Subsequently, funding by the ALMA Development Program (led by NA) led to the successful completion of the ALMA Phasing Project 1 (APP1) [RD14], which implemented this capability through access to the baseband data stream of the BLC via a “spigot” and a special purpose BLC Phasing Card [RD15]. The APP1 project provided continuum phased array observation capabilities at Bands 3 and 6, as well as a precise hydrogen maser clock for the ALMA site. This development ultimately led to the remarkable image of the shadow of the supermassive blackhole in M87[33]. Phased array observations for pulsars was also enabled, and resulted in the detection of the Vela pulsar[34] during commissioning -- this mode will be offered to the community in Cycle 8. A funded follow-on NA Development project, called APP2, is currently underway to expand the capabilities to Band 7, allow for spectral line observing, and implement a number of other improvements for sensitivity.

A full system-wide set of requirements for VLBI/phased-array ALMA observing does not currently exist, largely because of the original de-scope and subsequent incremental implementation. We therefore provide here a proposed draft set of these requirements going into VLBI in some detail because it is important to build the basics into the correlator *design* from the start. We believe this can be done with modest additional expense. At the system level the major cost will likely be in the software and commissioning effort, which can be deferred and/or implemented in stages. The main VLBI cost driver *for the correlator* is the total instantaneous bandwidth to be phased, and the minimal requirements here correspond to what ALMA can currently produce; further, in many modern correlator designs, even the total phased bandwidth can be built up over time.

VLBI observations have a number of special requirements, with regards to (1) real-time calibration and flagging; (2) formation of tied-array (also known as phased-array) beams, i.e., the coherent sums of data from a number of antenna elements; and (3) data and metadata outputs. A *tied-array beam* is the coherent sum of data from antenna elements within a single subarray, applying independent, complex, possibly frequency-dependent weights to each antenna element in the sum. In the simplest case (as for ALMA currently) these weights are simple phase corrections. More generally they may include “delta” delays to shift the beam away from the subarray delay center; polarization corrections; phase and/or delay corrections derived and applied in real-time (calibration measurements); sensitivity-based weights to ensure maximum SNR for the sum; and zero weights for antennas which are not to be included (e.g., due to flagging). Historically tied-array beams have usually been referred to as *phased-array beams,* since phase corrections and simple flagging were in the past the only available real-time weights. Such phase-only corrections generally only suffice for narrow-band observations using arrays whose elements have uniform sensitivity. To capture the need for these additional corrections, we shall follow the SKA nomenclature and use the term *tied-array beam* rather than *phased-array beam*.

Due to their extremely high angular resolution, VLBI observations benefit greatly from the maximum possible sensitivity. This leads to a desire to maximize the effective collecting area and recorded bandwidth, and to minimize quantization losses when recording the data for subsequent VLBI

---

[33] https://www.almaobservatory.org/en/press-releases/astronomers-capture-first-image-of-a-black-hole/

[34] https://ui.adsabs.harvard.edu/abs/2019ApJ...885L..10L

correlation. This is reflected in the requirements given below. Note that the effective sensitivity of a given array is set largely by the accuracy of the antenna weights, both intrinsic (based on each element's relative sensitivity) and atmospheric (based on real-time atmospheric calibration). The need for accurate real-time calibration leads to system-level requirements on the turn-around time between observation of a calibrator[35] and application of the derived calibrations. There is also a strong interest in allowing the use of in-beam calibrators, where "in-beam" here means within the primary beam of the individual array elements. Further, it is very useful to record not only the tied-array sum, but also the ALMA visibilities and the weights used to derive those sums, to allow the VLBI observer to understand how effective the real-time calibration was, and to allow correction of any coherence losses in the corresponding tied-array sums. This is similar to the correction of long-baseline data in highly extended millimeter arrays, when the integration time has exceeded the coherence time for those baselines.

In these requirements, we assume that the correlator subsystem will be creating the tied-array sums, rather than making baseband data (spigots) available for the individual antennas. While a departure from current ALMA practice, this is the standard approach for arrays used more regularly for VLBI. Access to raw baseband data for backwards compatibility and other purposes is discussed in S6.5.19.

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.5.1 Elements in VLBI (Tied-Array) Sums	One to all antennas	One to all antennas	It shall be possible to form a VLBI Tied-Array (also known as Phased-Array) subarray from any number of available antennas, including those in the Main 12m-array and the ACA.
6.5.2 VLBI: Access to Receiver Bands	Bands 1-7	VLBI allowed using any observing Band	Upon completion of APP2, observing in Band 7 will be available in the current system, in addition to Bands 3 and 6.
6.5.3 VLBI: Maximum Tied-Array Baseline Length	3[km]	Max. baseline	Max. baseline length in tied-array sum
6.5.4 VLBI: Maximum Recorded Bandwidth per Tied-Array Beam	8 GHz (summed over polarizations and sidebands), for each tied-array beam	Full BW per pol per sideband, for each tied-array beam	"Full BW" means the maximum scientifically useful BW available to the correlator. Multiple BBs may be split into multiple VLBI outputs.
6.5.5 Number of Independent VLBI	1 (strong preference for at least 2)	6	*N* subarrays would allow simultaneous observations in *N* bands. 6 is set by (1) allowing

---

[35] As of December 2020 the ALMA system requires computation of the phasing solutions based on observations of the target itself, although the intent is to allow transfer of calibration between sources in the next Cycle (L. Matthews, priv.comm.).

(Tied-Array) Subarrays			source + calibrator for 3 observing bands; and (2) the number of independent LOs.
6.5.6 VLBI: Number of Tied-Array Beams per Subarray	1 per tied-array subarray	2-4 per simultaneous VLBI observing band	2 needed for in-beam calibration, 4 for 2D interpolation
6.5.7 VLBI: Output Data Format	VDIF; 2-bits	VDIF; selectable from among {2, 4, 8} bits	Requantization correction needed
6.5.8 VLBI: Output Sampling Rate	Selectable between Nyquist and x2 oversampling	Same	
6.5.9 VLBI: Antenna Weights	Ability to apply complex gains and delays per antenna, per pol	Same, but apply pol corrections before summing, and allow per-antenna weighting (including flagging) of spectral regions	Allow application of real-time gain corrections and weighting (based e.g. on Tsys & efficiency)
6.5.10 VLBI: Real-time Calibration	Ability to transfer gains between sources & frequencies in real time: 2sec turn-around; Ability to apply WVR-derived corrections within 1 second	same but 1sec turn-around for gain transfer between sources and frequencies	Turn-around for gain transfer: time between taking data on a different source/frequency, and using resulting the gains in calculating the tied-array sum
6.5.11 VLBI: Spectral Window Used for Real-time Calibration	Ability to derive real-time gains based on an arbitrary contiguous subset of the correlated frequencies	Same	Allows deriving gains from strong spectral lines. Must also be able to apply those gains to the entire spectral window. Similar use as defining channels for channel averaged data for Telcal use.
6.5.12 VLBI: Delay Center and Timestamping	VLBI data to have time traceable to a consistent delay center near the center of the array, to an accuracy of 1 microsec	Same, but accurate to 2 nsec	Ensure consistent time basis for all subarrays and over long periods (at least a few days, preferably years). The accuracy is set primarily by pulsar timing.

6.5.13 VLBI: Placement of Tied-array Beams	VLBI beams placed anywhere within the 50% response of the primary beam	VLBI beams placed anywhere within the 20% response of the primary beam	Rare use case for ALMA due to small PB
6.5.14 VLBI: SNR Losses	VLBI output signal-to-noise ratio >=90% of that of an ideal system presented with the same inputs	VLBI output signal-to-noise ratio >=95% of that of an ideal system presented with the same inputs	Assumes no RFI. Inputs include calibration information (data weights).
6.5.15 VLBI: Polarization Response	Summed array polarization leakage should be comparable to that from a single antenna	Same	
6.5.16 VLBI: Simultaneous Correlation	Correlator should produce standard correlated visibilities (possibly less BW) at the same time as the tied-array sums	Same but full BW and full range of spectral channelization etc., flexibly configured	Existing requirement
6.5.17 VLBI: Metadata	Per-antenna time-dependent weights to be archived for use in VLB processing; calibration & flagging information	Same	
6.5.18 VLBI: Configurability	Each beam fully and independently reconfigurable on scan boundaries, in <10 sec	Same, but <5 sec	Includes changes in input antennas, position on sky, centre frequency, etc. Individual beams are completely independent.
6.5.19 Spigot access to baseband data	yes	yes	This is how tied–array data are currently accessed, is essential for some important observing modes, and provides unique telescope diagnostics.

6.5.1 Elements in VLBI (Tied-Array) Sums

Subarrays may contain any number of antennas, from 1 up to the total number of available antennas. The requirement here is that tied-array sums may include any number of antennas, from a single antenna up to the number of antennas in the parent subarray. The maximum tied-array baseline length (S6.5.2) may restrict this, but that depends on the distribution of any new antennas; we ***assume*** that there will be some configurations at least in which *all* antennas are included within that maximum baseline limit.

Note that a given tied-array sum is associated with a *single* subarray - one may not have tied-array sums which includes antennas from multiple subarrays.

Antennas which are *not* in a tied-array subarray should be usable for normal (imaging) observations. Per S6.5.15, tied-array subarrays should also produce correlated visibilities, though there may be some restrictions e.g. on spectral resolution and number of channels.

Note that use of a tied-array subarray involving only 7m antennas *is* desirable, as is the ability to form tied-array sums involving both 7m and 12m antennas.

6.5.2 VLBI: Access to Receiver Bands

VLBI observations are currently restricted to Bands 3 and 6, but extension to cover Bands 1 to 6 should be easy. Upon completion of the NA Development Project APP2, Band 7 will also be implemented, so **the minimum goal is to allow VLBI observations in Bands 1-7**. The **stretch goal is to allow VLBI observations in any of the outfitted observing Bands**. While there are currently few other stations that could observe in the highest frequency Bands, this may change in future, and the sensitivity of ALMA will make it a key element in any such VLBI array.

6.5.3 VLBI: Maximum Tied-Array Baseline Length

The maximum baseline length for a tied-array sum sets various parameters such as the maximum (delta) delay rate for an offset tied-array beam. The minimum requirement is set by the desire to include most of the collecting area in the compact array configurations. The stretch goal is to add signals from all the antennas, regardless of baseline length, for weather conditions and observing bands which would allow this.

6.5.4 VLBI: Maximum Recorded Bandwidth per Tied-Array Beam

The current ALMA system allows recording up to **7.5 GHz of total bandwidth (summed over polarization and sidebands)**, so this seems a reasonable minimum requirement for the new correlator. For sensitivity one would like to record the **entire available bandwidth**, which sets the stretch goal for this parameter. Note that in each case this is the desired recorded bandwidth ***for each tied-array beam***.

In many cases one will not wish (or be able) to record the full bandwidth for every tied-array beam. For spectral line observations only a small frequency range is useful; while for strong, compact sources (such as calibrators) a narrower bandwidth may be all that is needed to reach the desired sensitivity. Further, it may simply not be possible to record the maximum bandwidth for every tied-array beam. The intent here is that such constraints be imposed by scientific need or practical limitations which could be lifted in the future, rather than being forced by the correlator subsystem itself.

6.5.5 Number of Independent VLBI (Tied-Array) Subarrays

The minimum requirement is to allow **a single independent tied-array subarray**. **Two such independent subarrays** using different subsets of antennas would allow simultaneous VLBI observations in two different bands, which would be a major benefit. VLBI necessarily involves scheduling simultaneous observations at multiple observatories spread across multiple countries and continents, which leads to special observing sessions scheduled without knowing the actual observing conditions (which in any case can vary wildly from site to site). Allowing simultaneous observations in multiple observing bands helps to mitigate the risks due to weather, while also taking advantage of any particularly favorable conditions. Scientifically there is also great interest in simultaneous observations over a broad frequency range, since sources can change dramatically on very short timescales, and even data taken near in time (but not simultaneously) may not reflect the same phenomena.

There are two plausible stretch goals. First, one might wish to observe in all ALMA bands simultaneously. This argues for one independent tied-array subarray per band, but the number of available independent LOs limits this to 6 at the moment[36]. Second, ideally, each tied-array beam (see S6.5.5) would be assigned to a different subarray. This allows simultaneous observations of calibrators and sources, when they are not close enough to each other on the sky to allow observations in the same primary beam; and also maximizes the number of sources which can be observed at the same time.[37] Overall, a stretch goal of **six tied-array subarays** seems reasonable, allowing source + calibrator observations simultaneously in each of 3 observing bands, or at 6 completely independent frequencies.

From the correlator perspective the number of allowed subarrays is less interesting than the number of independent tied-array beams, discussed in the next section.

6.5.6 VLBI: Number of Tied-Array Beams per Subarray

The number of tied-array beams per subarray is set by a combination of calibration requirements, possible interest in (relatively) wide-field imaging, and the desire for simultaneous observations in multiple bands (see S6.5.4).

- Simultaneous observation of calibrator(s) and source: The minimum requirement is to allow simultaneous observation of the source and the calibrator. If the source-of-interest is itself suitable for calibration only one tied-array beam is needed, but in general this requires 2 tied-array beams per observing band**[38]**. Observing multiple calibrators allows better spatial interpolation - e.g., three calibrators evenly distributed about the source allows planar fitting[39]. This suggests a stretch goal of 4 tied-array beams per subarray**.**
- (Relatively) wide-field imaging: If there is more than one potential VLBI source in the primary beam, extending over more than one synthesized beam, one would like to place multiple tied-array beams to “tile” the entirety of the interesting field-of-view. For instance, there may be multiple particle acceleration regions along a jet, or a merging galaxy with a double nucleus (cf. Arp 220). There are also cases where the position of an interesting transient source is not known to be better than a synthesized beam-width. In most such cases two tied-array beams per observing band would suffice, although more would occasionally be appreciated.

---

[36] Actually a single LO can be used for multiple observing bands, if the frequencies are harmonically related

[37] For some observations the VLBI uv-coverage is more important than the sensitivity, so it may be desirable to have more subarrays, even though each would then have fewer antennas.

[38] Here the number of observing bands refers to the number being used simultaneously for VLBI. Thus the absolute maximum is set by the number of independent LOs (six, with the current system).

[39] Note that here we are considering interpolation of the VLBI phases post-observation. Interpolation to derive better *ALMA* phases in real-time could be useful but does not necessarily require multiple tied-array beams.

All of these use cases are accommodated by the requirement that there be **2 tied-array beams per subarray**, with a stretch goal of **4 tied-array beams per subarray**.

Requiring multiple tied-array beams within the same subarray implies that at least one such beam will not be at the same delay centre as that used for the cross-correlated visibilities. For instance, if there is a suitable calibrator which can be observed within the same primary beam as the source of interest, it is highly beneficial to put one tied-array beam on the calibrator and one on the source, and record the data for both simultaneously. While this is not a very common case for an arbitrary position in the millimeter sky, VLBI observers often preferentially choose to observe sources which have very close calibrators. The case of multiple targets within the primary beam has already been mentioned.

Such offset beams require "delta-delay models" which transfer the calibration information to that secondary position. See S6.5.13 for proposed restrictions on the magnitude of allowed offsets, which sets the maximum update rate required for these delta-delay models. Note that, in addition to the phase and delay corrections needed to properly position the offset tied-array beam, significant offsets may require modifications of the antenna-based weights and polarization corrections (possibly as a function of frequency), if the offset is a significant fraction of the primary beam. This is a particular issue when combining 7m and 12m data in the same tied-array subarray.

The maximum *total* number of tied-array beams is the number of tied-array beams per subarray, times the number of tied-array subarrays.

6.5.7 VLBI: Output Data Format

The current standard for VLBI data recording is the VLBI Data Interface Format (VDIF; see https://vlbi.org/vlbi-standards/vdif/), and the ALMA VLBI output should adhere to this standard.

Currently ALMA outputs 2-bit data for VLBI, compatible with the rest of the VLBI network, and that is a reasonable minimum requirement, but the ability to choose between 2-, 4-, and 8-bit outputs would be useful for spectral line work, and hopefully, as recording rates increase, for the whole network in the future. The use of more bits is also an inexpensive way to achieve higher VLBI sensitivity.

Ideally this requantization would include provision of the information needed to apply requantization corrections for the VLBI data.

6.5.8 VLBI: Sampling Rate

While less effective than increasing the number of bits, sampling at rates faster than Nyquist improves the sensitivity, e.g., by ~7% for 2-bit data oversampled by a factor of 2 (e.g., Table 9.5 of Thompson, Moran, & Swenson 2017, *Interferometry and Synthesis in Radio Astronomy*). It seems reasonable to require that the new correlator support this recording mode.

6.5.9 VLBI: Antenna Weights

Tied-array beams represent the (complex) weighted sum of the individual antenna data, in which the weights should be chosen to optimize the sensitivity. In addition to "standard" calibration this should include per-antenna weighting by sensitivity, based for example on Tsys and efficiency. As discussed in the VLBI header section (S6.5), these weights must include per-antenna complex gains and delays, and polarization corrections[40]. At minimum the polarization (Jones matrix) corrections should be applied after summation, but it is preferable to apply them before summing, to ensure coherence of the polarization sums. A stretch goal would allow per-antenna calibration and weights to be frequency-dependent. As discussed above (S6.5.5) the antenna weights will be different for each tied-array beam, even within the same subarray.

This is really a system-level rather than a correlator requirement, as the correlator system as currently defined is responsible for applying but not for deriving these weights.

6.5.10 VLBI: Real-time Calibration

Forming a coherent sum requires deriving and applying antenna gains in real time, fast enough that the atmosphere has not changed significantly between the time the calibration data are taken, and the time the derived gains are applied. The ideal case would work just like off-line calibration, with the raw data stored while the gains are being derived, possibly with interpolation between adjacent calibrator scans. While possible in principle this may be difficult in practice, and for some antenna configurations in reasonable weather conditions is not actually required. A more reasonable requirement is that one be able to **apply antenna-based gains within 2 seconds** of taking the corresponding calibration data. **WVR-based corrections should be applied within 1 second**. The **stretch goal is to apply *all* gain corrections within 1 second.**

Again this is a system-level requirement, involving the M&C system (including TelCal) as well as the correlator.

6.5.11 VLBI: Spectral Window Used for Real-time Calibration

Strong spectral lines (e.g., masers) can be excellent calibrators at ALMA frequencies. While standard continuum calibrators benefit from averaging together the entirety of the correlated spectrum, the optimal use of line calibrators requires the ability to derive real-time gains from a user-supplied subset of the correlated frequencies. As this is already possible for "standard" TelCal it seems a reasonable requirement for real-time tied-array calibration as well.

6.5.12 VLBI: Delay Centre and Timestamping

VLBI data benefit from use of a well-defined and consistent delay center[41] (for all subarrays, and over long periods) and consistent time-stamping, primarily to allow accurate and coherent correlation with other VLBI stations. This is analogous to the need for a consistent time reference and known positions for all antennas within the ALMA array itself. For VLBI purposes it would suffice to keep time consistent to within a few microseconds between observing sessions (to avoid excessively painful fringe searches)[42], and this is the minimum requirement. The stretch goal is primarily set by pulsar timing, where one wishes to compare arrival times to this accuracy over years or even decades. This may not be a major science driver for ALMA however.

---

[40] It may also be useful to convert to a different polarization basis, e.g., from circular to linear polarization, as is currently done off-line for VLBI observations with ALMA (L. Matthews, priv.comm.).

[41] This is generally the array center position.

[42] Standard observing requires much more consistent timing *within* an observing session, to maintain the white light fringe and to allow reasonable calibration.

6.5.13 VLBI: Placement of Tied-array Beams

As discussed in S6.5.5, there are several reasons why one may wish to place multiple tied-array beams simultaneously within the same primary beam during VLBI observations. Different weights are required in such tied-array sums, both to shift the delay centres, and to account for different antenna responses at the two positions; for well-designed antennas with smooth and uniform primary beam responses the latter is important when the tied-array beams are separated by a significant fraction of the primary beam. Forming well-separated tied-array beams within a single subarray are thus more challenging in requiring first, more frequency delta-delay updates, and second, more substantial corrections for differing primary beam responses. The situation for ALMA is further complicated by the presence of multiple dish designs, and particularly the possibility of subarrays mixing the 7m and 12m dishes. Thus the allowed placement of tied-array beams within the primary beam is an interesting system-level requirement, affecting (at minimum) the correlator, M&C, and TelCal subsystems.

The main driver for this requirement is to maximize the possibility of in-beam (primary beam) calibration (see S6.5.5), since the likelihood of this goes as the square of the allowed distance from the source of interest. At the same time there is no point in forming a tied-array beam so far from the source that the effective gain of the telescope leads to miniscule apparent signal for either source or calibrator. A reasonable minimum requirement then is that **tied-array beams may be placed anywhere within the 50% response power of the primary beam of the smallest dish** (7m for ALMA), while a stretch requirement allows **placement anywhere within the 20% response.**[43]

6.5.14 VLBI: Sensitivity Losses

VLBI observations naturally benefit from the greatest possibility sensitivity, so an important system-level requirement is that the effective signal-to-noise ratio (SNR) of the VLBI data be the maximum allowed by the weather conditions. For the correlator this is a requirement on acceptable losses due primarily to finite-precision arithmetic and quantization losses, as well as any inaccuracies of application of the antenna-based weights in the sum. A reasonable minimum requirement is that **the signal-to-noise ratio of the VLBI output be no less than 90% of that of an ideal system presented with the same inputs, ignoring RFI**[44], The stretch goal sets this to **no less than 95%.**

Re-phrased as requirements comparing the actual response to that of an ideal system observing under the same (weather) conditions with the same hardware, these may be made into system-level requirements. This would add constraints on the accuracy and timeliness of the weights supplied to the correlator.

6.5.15 VLBI: Polarization Response

The coherent mechanisms that produce detectable VLBI emission generally produce highly polarized emission, so it is essential that these properties be preserved during VLBI observations. A reasonable requirement is that **the polarization leakage in each VLBI tied-array sum be comparable to that of a single ALMA antenna.** As with the sensitivity requirements this is really a requirement on the system as a whole, primarily the calibration and M&C subsystems, rather than on the correlator alone.

---

[43] More distant calibrators may be accommodated through by fast slews, or through the use of multiple subarrays. Dividing the antennas between multiple subarrays naturally reduces the sensitivity of each of the subarrays, and does not allow direct calibration of antennas in the subarray pointing at the source. See S6.5.4.

[44] Note that "the same inputs" implies the identical weather conditions, input weights, ADC quantization, etc. The requirement is that imperfections in the correlator subsystem decrease the resulting SNR by no more than 10%.

6.5.16 VLBI: Simultaneous Correlation

A primary requirement is that **VLBI observations in one subarray must not limit standard observations in other subarrays.**

The correlator must also perform "standard" correlations in each subarray that is producing VLBI data. Reasons include:

- Calibration: to derive real-time gain solutions, to check the effectiveness of the derived gains through after-the-fact imaging, and to provide accurate total flux densities to aid in calibration of the VLBI data
- Science: to track variability of the source of interest, and to allow deep imaging of the entire ALMA field-of-view[45] (which is generally also interesting for VLBI sources)

At minimum the correlated data should cover a substantial fraction of the VLBI bandwidth, and provide reasonable spectral resolution (within a factor of a few of that provided by "standard" broadband continuum observations) and integration times (within a factor of 5-10 of the minimum for "standard" observations). As a stretch goal these correlations should cover the same bandwidth and allow the same spectral channelization, dump times, etc. as non-VLBI correlations. This last is particularly important for spectral line VLBI, where "continuum" channelization is not really adequate, and where simultaneous ALMA line observations may be especially useful - particularly since VLBI observations may be many hours long.

6.5.17 VLBI: Metadata

In addition to the standard metadata ALMA should record the weights used in each tied-array sum, as a function of time. This is useful in comparing these weights with the best weights which can be derived in post-processing, and in determining the appropriate weighting of the VLBI correlations when analyzing the VLBI data. Tsys and phasing efficiency should also be reported, as should degrees-per-flux-unit (DPFU), antenna gains, and a flag table. The positions of the individual tied-array beams must be stored, along with the frequency coverage. The details of the VLBI recording should also be provided, e.g., setup (number of bits, oversampling rate, etc.), data rates, recording ON and OFF times, data flagging, and the like.

6.5.18 VLBI: Configurability

Configuration of tied-array sums include defining the antennas and (types of) antenna weights[46] used in those sums; the sky positions of the tied-array beams; the frequency ranges being used; and the details of the VLBI data recording setup. The basic requirement is that these be configurable independently for each tied-array beam, and be reconfigurable on scan boundaries. At minimum these can be changed in <10 seconds, to allow switching configurations rapidly e.g. between calibrator and source; the stretch goal is to do so more rapidly (<5 seconds).

6.5.19 Spigot Access to Baseband Data

"Spigot access" to baseband data from individual antennas is still required, despite the proposal for a full VLBI implementation in the correlator. Apart from backwards compatibility, such data are useful for special observing modes, such as pulsar observations and others requiring extremely high

---

[45] Generally set by the primary beam of the constituent antennas.

[46] The antenna weights themselves must change continuously for delta-delay and (potentially) atmospheric and sensitivity tracking. During configuration one sets the *types* of weights to be applied, e.g., whether Tsys should be used.

temporal and/or spectral resolution. Moreover, such direct access provides unique telescope-based diagnostics, which have proven quite useful.

6.6 Calibrations/Corrections

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.6.1 Sideband separation and rejection	Support LO frequency offsetting and 90° Walsh switching sideband separation	Support LO frequency offsetting and 90° Walsh switching sideband separation	Expect all receivers 2SB eventually but sideband rejection still needed with LO frequency offsetting; 90° phase Walsh switching desirable for any remaining DSB bands and for measuring receiver sideband gains.
6.6.2 Quantization Correction(s) [rate and accuracy]	Per dump duration in all correlator modes with 1% accuracy and 0.01% systematic error	Per dump duration in all correlator modes with 1% accuracy and 0.01% systematic error	Will depend on adopted architecture; will need ADC histograms and autocorrelation power

6.6.1 Sideband separation and rejection

The existing ALMA system employs two techniques for sideband rejection/separation in cross-correlation data: LO frequency offsetting for sideband rejection, which is generally used with 2SB/SSB receivers to improve rejection to levels that are not astronomically measureable (>>30dB image rejection); and 90° Walsh switching for sideband separation, which is generally used for DSB bands to allow recording spectra for both sidebands simultaneously. 90° Walsh can also be used for calibration purposes with 2SB/SSB receivers, e.g. when measuring the receiver sideband gains.

The LO offsetting implementation primarily uses the LO2 for applying the second frequency offset, which counters the one applied on LO1. The BLC also supports LO offsetting in the TFBs, which is optionally used in FDM modes (in this 3-LO offsetting case the offset applied to LO1 is doubled and half the offset is removed by LO2s and half by the TFBs). In the next generation system it is likely there will not be a second analogue downconversion, so **LO offsetting absolutely must be implemented by the correlator**. The frequency offsets currently implemented in ALMA are multiples of approximately 30.5kHz (exactly: 125 MHz / 2^{12}), so to support 80 antennas with unique non-zero offsets the maximum frequency offset to be applied by the correlator is 2.5 MHz. Due both to the uncertainty of future upgrades to Bands 9 and 10, and to other calibration uses of sideband separation, we recommend to also implement 90° Walsh switching. Assuming the FLOOG will not be upgraded, we recommend to support the existing ALMA implementations of LO offsetting (i.e. the frequency offsets as just described) and 90° Walsh switching (i.e. 128 sequence length with 2048ms sequence period). We note however that the effectiveness of 90° Walsh switching would improve if it is possible to increase the sequence length and reduce the sequence period, so as to allow maximally spacing assigned Walsh sequences between antennas and incurring less atmospheric phase variation during each cycle period. This would, however, be a system-wide change. The existing Walsh sequence implementation in ALMA already supports up to 127 antennas with unique non-constant sequences.

In summary both LO frequency offsetting and 90° Walsh must be available, as minimum requirements and stretch goals.

6.6.2 Quantization Correction(s) - rate and accuracy

Quantization with a modest number of bits causes nonlinearity between input and output signals, which can be seen as nonlinear power responses in auto-correlation and reduced correlation coefficients in cross-correlation. To compensate for the nonlinear relations, quantization correction is needed. Since the correction coefficients depend on the level histogram (or total power) of digitizer outputs, the histogram should be measured and provided. Furthermore, the histogram can vary within an integration duration (e.g. on-the-fly mapping and solar observations). The correction should be applied to each dump data output from correlator and spectrometer every dump duration. The correction accuracy is directly connected with amplitude accuracy of auto and cross correlation spectra. Thus, the correction should achieve 1% or better accuracy and less than 0.01% systematic error, which are required for spectral amplitude and spectral dynamic range in ALMA. The correlator implementation must ensure direct communication between the entity which measures the level histogram and the entity which calculates the quantization correction. The latency must not be larger than 16msec (it is assumed the shortest dump time is equal to 16msec, see S6.3.1, and a stationary process within a time slot of 32msec).

6.7 Correlator Output

Parameter	Minimum Req'mt	Stretch Goal	Comment
6.7.1 Products output from Correlator Subsystem	Cross-correlation and autocorrelation spectra, and correlator-based flags, +TBC	TBD	With all necessary correlator calibrations applied. See detailed comments.
6.7.2 Data rate mitigation options	It shall be possible to average data in both time and spectral space before data output	Same	All necessary correlator corrections should be applied before averaging.
6.7.3 Correlator output format	Options for 16 or 32-bit spectral output	TBD	Is there a need for additional flexibility?

6.7.1 Products output from Correlator Subsystem

At minimum, the cross-correlation and autocorrelation spectra, with any correlator-based corrections applied, will need to be output from the correlator subsystem, along with any flagging produced in/by the correlator (and their-integration based centroids). Channel-averaged data will also need to be produced. Presently the output from the CDPs is in ASDM Binary Data Format (BDF[47]), with one file per subscan for spectral and one for channel-averaged data (i.e. two files per subscan). The BDF files are sent incrementally per integration. The 4-channel WVR data, as well as, the derived correction coefficients from Telcal are delivered to the correlator to enable online WVR correction. The correlator software subsystem is currently responsible for writing the WVR data to a BDF file per subscan, although this task could be given to a dedicated component outside the correlator subsystem in future. Although online WVR correction is not presently being used, we do want to preserve the option, as well as, the current capability to output both WVR-corrected and uncorrected datastreams. All of these data then go straight to archive, while Data Capture handles assembly of the ALMA Science Data Model metadata. In Phased Array modes, the phased, summed data are delivered to the spigot output S6.5.4.

6.7.2 Data rate mitigation options

Options to mitigate the output data rate are required in both time and spectral space (also see S6.8), as mentioned in S5.1.2 these actions may be done in either hardware or software. Correlator corrections should be applied at the appropriate cadence before averaging is done. Additionally, it must be possible to apply different levels of spectral and/or temporal averaging for each output spectral window. In particular, the QA0 (quality assurance level 0) assessment presently runs on the

---

[47] BDF definition and ALMA-specific usage is documented in the ALMA software repository in ICD/HLA/ASDMBinaries/doc/bdf.pdf

channel-averaged data with a visibility integration time of 1.008s, regardless of the time averaging used for the science (channelized) spectral windows and this capability must be preserved. Allowed time and spectral averaging should be restricted to values that do not incur significant time or bandwidth smearing (S6.3.2, S6.2.1). It shall also be possible to only output selected polarization products (dual pol for example, even if full polarization correlation is always produced). Beyond the required calibration and metadata information (described above in S6.7.1), only the data defined by the user-specified spectral windows shall be output (see S5.1.4 and S6.8).

6.7.3 Correlator output format

The current system outputs BDF data from the CDPs with either 16 or 32-bit depth, according to the dynamic range of the spectra per integration [RD05]. It is notable, that when the ASDM is converted to a CASA measurement set (MS) for data processing, it is stored as 4-byte (single precision, signed) complex numbers, equivalent to 32-bits, regardless of input (double precision is used for internal calculations as necessary). So changes to the bit-depth will only affect the archival raw datasize, not the size of an MS unless changes are also made to CASA.

We could not find any information/science driver to support broadening the output bit-depth options. As noted on ICT-6114, the vast majority of ALMA science data does not presently trigger 32-bit (even when 32-bit is triggered, it is typically < 1-2% of the data for a single dataset).

6.8 Capabilities for User-facing Correlator Observing Modes

Until the correlator design is known, it will be impossible to define specific user-facing observing modes. Here we note the key correlator observing mode *capabilities* that will be required from the science point of view. Typically observing modes are considered merely "software", but user expectations can often interact with correlator design and the ease with which such modes can be implemented in software in subtle ways. These capabilities are understood to apply to the correlator subsystem output, which as described in S5.1.3, we define as including the CDPs. These requirements apply to the primary user facing output of the correlator known as spectral windows (spws), see S5.1.4.

Parameter	Minimum Req't	Stretch Goal	Comment
6.8.1 Placement of spectral windows (spws) within basebands	Full flexibility (including overlapping)	Full flexibility (including overlapping)	No restrictions beyond those imposed by the upstream subsystems, i.e. proximity to baseband boundaries etc.
6.8.2 Maximum number of spectral windows per baseband	32	64	Intersection of what is technically feasible (lots) and users ability to effectively set up very complex spectral setups.
6.8.3 Bandwidth per Spectral Window	Independent selection of bandwidth per spw	Independent selection of bandwidth per spw	Each spw can have different BW, provided each individual spw's BW, and the aggregate of all spws < max allowed
6.8.4 Spectral window channelization characteristics	Independent selection of spw channelization	Independent selection of spw channelization	Each spw can have different channelization (via spectral averaging), provided the aggregate #channels is < max allowed
6.8.5 Visibility integration time per spectral window	Independent selection of visibility integration time per spw	Independent selection of visibility integration time per spw	Each spw can have different visibility integration times (via time averaging, provided min dump time < Tint < max time smearing

6.8.1 Placement of spectral windows within basebands

It shall be possible to place spws within each baseband subject only to the restrictions imposed by the upstream subsystems (i.e., optimal receiver tuning range, baseband edges, etc.) and what is stated in section 6.2.10.

6.8.2 Maximum number of spectral windows per baseband

In the absence of a specific correlator design there are no obvious limits to the number of spws that could be placed per baseband, except that each must include at least one channel (that must have a width less than the bandwidth smearing limit if imaging the full primary beam) and there will be a maximum number of channels. However, there are practical limits on the number of spws that a user can feasibly be expected to individually setup, not to mention the difficulty on the operations side of producing a user-friendly observing tool to enable such complex setups along with associated restrictions (i.e., the OT and P2G process), as well as the requisite user training. It seems likely that 32 to 64 independent spws will be adequate, and represent the minimum and stretch goals, respectively.

6.8.3 Bandwidth per Spectral window

It shall be possible to independently select the bandwidth of each spw, provided that they meet the criteria for minimum and maximum spw widths, and the other requirements for max correlated bandwidth and channelization in aggregate.

6.8.4 Spectral window channelization characteristics

It shall be possible to select the number of channels for each spw (to be correlated) independently, as well as their output channelization after spectral averaging, providing that the limitations of other requirements are met, bandwidth smearing limits for example S6.2.1.

For each spectral window it will be possible to define a number (<=32) of channel-average regions: channel ranges which are spectrally averaged to produce the channel-average data that is written to the channel-average BDFs for online calibration and QA0 purposes. These data should have an independently configured integration duration (in units of the hardware dump duration) such that they can be shorter than the spectral integrations.

6.8.5 Visibility integration time per spectral window

As already mentioned in S6.7.2, it must be possible to independently select the visibility integration time (Tint) for each spectral window, up to the time-smearing limit. The longest integration duration out of the correlator is restricted by the desire to avoid significant time-smearing effects at the longest baselines. This is frequency independent, as the field of view decreases linearly with frequency while the phase variation at a given absolute distance from the phase centre increases linearly with frequency, cancelling out. Following similar logic to that derived for the Bandwidth Smearing constraints in S6.2.1, for a baseline of 35 km and at the 20% power point in the primary beam response, we define time-smearing limits of 2.55s for <5% loss and 1.14s for <1% loss. As for bandwidth smearing, these constraints can be relaxed for more compact configurations.

Different choices for Tint will be primarily driven by the need to optimize the calibration as a function of time, while minimizing the data rate. Thus, channel averaged data should be dumped at the cadence of the QA0 assessment of 1.008s (S6.7.2), while the science data (at least for lower frequency bands and smaller configurations) can generally withstand time averaging of several seconds. Though presently the science spectral windows must use the same uniform Tint, in the future one might want to mitigate the data rate by using more time averaging for high spectral resolution, that can be cross-calibrated with coarser resolution "continuum" windows with a finer Tint, to apply a decorrelation correction.

We note that the independent integration time per spectral window may have implications for the output file streaming (nominally done in chunks per integration) and there may be implications in file structures / schemas and in other subsystems. Currently the integration duration is a property of the APDM correlator specification as a whole. Currently each BDF file, one of which contains all spectral integrations, and another contains all channel-average integrations, only includes data with a single integration duration.

7 Brief Analysis of Output Data Rates

Here we provide a brief, representative analysis of the data rates that the 2nd Generation correlator will be capable of producing based on the requirements listed in S6. This analysis is only meant as a guide to aid a future more comprehensive analysis, as well as to hopefully foster discussion on how the downstream (of correlator) subsystems might be upgraded to deal with the onslaught in the ALMA2030 era. It is also notable that the "real" data rates will need to be reassessed once an actual 2nd Generation Correlator design is chosen. The current peak data rate allowed on the BLC is < 70 MB/s. We adopt the data rate calculation formalism developed in [RD07] for the now defunct Baseline Correlator Upgrade Project (CUP). As in that document, we define:

Output Data Rate = ((2Nbyte x Napc x Nant(Nant-1)/2 +4Nant) x Nchannels x Npols) / Tintegration
2Nbyte = 4 for cross correlation data (16bit real +16bit imaginary), while autocorrelations are 4 byte (32bit)
Napc = Number of atmospheric streams, currently WVR corrections are applied offline, 1 stream stored
Nant = Number of antennas
Nchannels = Total channels in all basebands per polarization
Npols = Number of polarizations
Tintegration = Visibility integration time

As described in S6.1.1, the 2nd Gen Correlator will need to correlate data from all three ALMA arrays (Main-12m, ACA-7m, and ACA-TP) together, or in independent subarrays. However, **for the purpose of this initial analysis we consider only a 47-antenna interferometric array, as well as visibility integration times that are typical of current long baseline (>= configuration 7) observing practices, and dual polarization.** The first two examples below use a fiducial Band 6 frequency of 220.398 GHz (i.e. ^{13}CO(2-1)) for a "coarse resolution" continuum mode, as well as, a "decent spectral resolution" spectral science mode with 0.17 km/s spectral resolution. In both cases, the results are compared with those of the closest matching BLC observing mode.

Coarse Spectral Resolution "Continuum Mode" Band 6 220.4 GHz	**BLC TDM mode**	**2ndGen 8 GHz per sideband, dual pol**	**2ndGen 16 GHz per sideband, dual pol**
Total Bandwidth per polarization (GHz)	7.5	16.0	32.0
Channel width (kHz) to achieve spectral resolution (2x for Hanning and 1.0x without)	15625	6,640	6,640
Total channels in all basebands per polarization	512	2,410	4,819
Visibility integration time (sec)	1.008	1.008	1.008
Peak Output Data Rate (MB/s)	**4.58**	**21.57**	**43.14**
Coarsest resolution chosen to prevent BW smearing at lowest freq/longest baseline, could be relaxed for higher freq/smaller configs	In TDM mode the spectral resolution is only 50 km/s!	The "coarse" resolution mode has 9 km/s velocity	The "coarse" resolution mode has 9 km/s velocity

		resolution	resolution

"Decent" Spectral Resolution Spectral Science Mode: 0.17 kms at 220.4 GHz	**BLC FDM mode**	**2ndGen 8 GHz per sideband, dual pol**	**2ndGen 16 GHz per sideband, dual pol**
Total Bandwidth per polarization (GHz)	0.9375	16.0	32.0
Channel width (kHz) to achieve spectral resolution (2x for Hanning and 1.0x without)	61.0352	125.0	125.0
Total channels in all basebands per polarization	15360	128,025	256,050
Visibility integration time	3.024	3.024	3.024
Peak Output Data Rate (MB/s)	**45.84**	**382.04**	**764.09**
NOTES: This table demonstrates the current issue that to get required spectral resolution, many projects must give up significant bandwidth.	Hanning + required spectral resolution limits the total BW by 8x	Without Hanning takes 8x the channels for 16x the bandwidth	Without Hanning takes 8x the channels for 16x the bandwidth

Finally, in the table below, we present the peak output data rate for the finest spectral resolution at maximum correlated bandwidth: 0.2 km/s at 35 GHz (S6.2.2). These data rates will be very challenging indeed to handle downstream, and it is likely that data rate mitigation options will need to be applied (limiting bandwidth, using channel averaging etc, S6.8). Even so, the 2nd Generation Correlator will be able to deliver vastly superior capabilities given that the current correlator would be restricted to a spectral resolution of 0.26 km/s and an aggregate bandwidth of only 234 MHz.

Finest Spectral Resolution at Max Correlated Bandwidth Spectral Scan Mode: 0.2 km/s at 35 GHz	**BLC FDM mode**	**2ndGen 8 GHz per sideband, dual pol**	**2ndGen 16 GHz per sideband, dual pol**
Total Bandwidth per polarization (GHz)	N/A	16.0	32.0
Channel width (kHz) to achieve spectral resolution (2x for Hanning and 1.0x without)	N/A	23.3	23.3
Total channels in all basebands per polarization	N/A	685,257	1,370,514
Visibility integration time	N/A	3.024	3.024
Peak Output Data Rate (MB/s)	N/A	**2044.89**	**4089.79**
NOTE: The current correlators cannot achieve better than 0.26 km/s at 35 GHz, and with that the aggregate bandwidth is only 234 MHz.			

Because the maximum bandwidth would be restricted by a factor of 10 (S6.2.4), the finest possible spectral resolution mode (20 m/s at 35 GHz) would have similar peak output data rates.

Using the equations above, these fiducial values can be easily adjusted to other combinations of observing parameters.

8 Deployment Options for a Second Generation ALMA Correlator

We were asked in our Charge [AD06] to consider the implications of potential Second Generation ALMA Correlator deployment options, with the aims of minimizing ALMA Science Operations downtime, while not unduly prolonging the time needed to deliver new capabilities. Three broad deployment options were considered and are briefly described below:

1. Deploy the new correlator first with backwards compatibility with upstream system
    a. It can replace the BLC as soon as its "backwards compatible" operation is commissioned, realizing significant science benefits immediately (similar to those that would have been afforded by the CUP Phase 1).
    b. It would need to be backwards compatible with the current 3-bit digitizer and tunable filter bank output until the upstream system is available
    c. It can operate with 4-bit correlation for a spectral sensitivity gain of 11.6%, and without the factor of two loss in continuum bandwidth and spectral resolution inherent in the BLC.
    d. Though the correlated bandwidth limitations (and baseband specifications) would remain, tuning within the full RF/IF range of new receiver bands would be possible.
    e. The available number of channels/spectral resolution options of the 2ndGen Correlator would be available (assuming the downstream system can handle it).
2. Deploy the new upstream digital system first with backwards compatibility with the BLC
    a. As originally conceived this option assumed that the new digitization hardware would need to ***replace*** the current hardware due to limited space in the antenna receiver cabin.
    b. The new digital hardware would have to be capable of mimicking the current output of the 3-bit digitizer / tunable filter bank signal, with the ability to later switch to its full capabilities (i.e. via a firmware update).
    c. There would be no interim science benefit.
3. Deploy the new ALMA 2030 digital system in parallel with existing hardware, with all new components deployed to different physical locations, so that the old system can continue Science Operations until the new system is commissioned.
    a. The old and new systems would NOT be required to work simultaneously (apart from the IF Switch or similar unit at which the IF signal is split to old and new systems), but a fairly easy method of switching between them must be possible.
    b. The bulk of commissioning of the new digital system could then be accomplished during periods of unfavorable / less favorable observing conditions (i.e., daytime), while Science Operations continue with the old system during optimal conditions.

The first two options were developed under the assumptions that (A) it would be challenging from a development/funding point of view to get the full ALMA2030 digital path updated on the same timeline in order to deploy it all at once; and (B) unlike the 2ndGen Correlator, the digitization components would have to be *replaced* in situ where the current hardware resides and thus could lead to significant downtime. The third option known as "parallel deployment" was conceived at the 2020 ALMA Correlator Workshop with the recognition that (B) may not be true. **The parallel deployment option has emerged as by far the favorite, since it minimizes downtime, and reduces the complexities that would be introduced by backwards compatibility requirements which would likely prolong the overall timeline considerably.**

We also find that there is no deployment path that can meet the goal of minimizing downtime, unless the 2ndGen Correlator is located at a different physical location than the current AOS correlator rooms for the BLC and ACAC (and ACAS). Moreover, an outcome of the 2020 ALMA Correlator Workshop was a consensus that placement at the OSF would be optimal, though there are a number of technical challenges that require further investigation. Also see ALMA memo 607 [RD12], which also considered the trade-offs for locating a new correlator at the OSF.

Our initial findings related to the Parallel Deployment path and locating the 2nd Gen Correlator at the OSF are described further in S8.1 and A.4. A potential commissioning testbed/prototype is described in S8.2.

8.1 Parallel Deployment

The CorrWG consulted with ALMA experts to *begin* the process of assessing the feasibility of the Parallel Deployment concept, specifically with the 2nd Gen Correlator located at the OSF with the goal to answer the question: are there any showstoppers? None have been identified so far based on preliminary discussion with ALMA engineers, but there are many open questions and concerns. Our initial findings and suggestions for future study are outlined in S8.1.1-8.1.2, and additional details are provided in A.4.

8.1.1 Location of new digitizers and IF Switches

Due to the rare congregation of a wide range of ALMA engineers from multiple subsystems at the 2020 ALMA Correlator Workshop, a significant conceptual breakthrough was achieved: It may well be possible to deploy new digitizers and IF switches in (or on) the Front-End Electronics Chassis (See Figures 6 & 7 in A.4), rather than *replacing* the existing digital components as had been heretofore assumed. This is a game-changer because it would permit the old digital system to continue to be used for Science Operations, while the new components are deployed, integrated, and commissioned.

We further note that there appears to be space for the new hardware needed to format the digital signal and put it onto the fiber on the existing Back-End Rack (there is presently an empty subrack). If, after study, co-locating the digitizers with the IF Switches on the Chassis is infeasible, there may also be room in the Back-End Rack to place the new digitizers (together mimicking the current DTX assembly). However, it was noted that the increased IF Bandwidth of the upgraded system may require significantly upgraded connectors/coaxial cables to transmit the wider analog signal from the Front-End Assembly to the Back-End Rack.

The above findings (also see A.4) suggest two possible actions:

- Initiate a cross-subsystem study of the physical requirements of a new "Digital Assembly Box" including needs of the hardware itself, as well as constraints from FE Assembly including removal from antenna, and FE Handling Vehicle.

- Consult the full Back-End Team to confirm the availability of space for the new Formatter, and if necessary new digitizers in the Back-End Rack. Additionally, if the digitizers are to be located in the Back-End Rack, assess the cabling implications given the IF Bandwidth goals.

8.1.2 Fibers from Antennas to 2ndGen Correlator

Each ALMA antenna pad has eight fibers that go to the AOS. Presently, one is used for the LO, two are used for antenna control, and one is used for digital transmission of data. Therefore, four fibers are, in principle, "available" to transmit data from the new digital system to the AOS. Though many details, especially of the digital transmission capacity required for the new system, remain undecided (e.g., is there a first F at the antennas which would decrease the transmission load?), it is possible that only one additional fiber from the antennas to the AOS is required to transmit the "new" data stream. From the AOS to the OSF there are 24 fibers, with eleven presently in use. There are an additional 24 fibers that run between the AOS and the Power Station at the OSF which could potentially add more capacity, but this needs investigation.

However, a few concerns were identified and are listed in A.4. We suggest that a detailed study of the data transport requirements, including availability for use from existing fibers be carried out.

8.2 Hardware in the Loop test environment

The "Hardware In the Loop Simulation Environment" (HILSE) is to be a mini-interferometer made up from current ALMA instruments for testing purposes that can operate in parallel with Science Operations. The HILSE is an ongoing JAO/NRAO development project that is currently undergoing its Critical Design and Manufacturing Review (CDMR). It builds on experience of the AOS Test Interferometer setup in which the ACA and production LO hardware were operated from an independent computing system (AOS2), and the fully independent OSF interferometer using a 2-antenna BLC-based correlator (TFINT), both of which have been used for testing in parallel to science operations. The HILSE will be able to process the data produced by 4 antennas with an instantaneous bandwidth of 2 GHz (¼ of the current ALMA capacity, one baseband). The objective of the HILSE project is to provide a representative infrastructure of the production environment for software and firmware testing while providing an environment to allow engineers to perform maintenance activities efficiently. The HILSE correlator will be located at the OSF, and at least initially will employ correlator hardware identical to that of the BLC. The HILSE will be connected "on demand" to up to four ALMA antennas located at the AOS. In order to enable the interferometer functionality to the HILSE, it is needed to bring from the AOS the:

- Time synchronization signals
- DTS signal

The needed software infrastructure for supporting the HILSE is based on the current ALMA software and it will only be customized for matching the number of stations and available bandwidth. The HILSE infrastructure and also the experience gained along the development can be largely reused in the development stages of the ALMA 2nd generation correlator. For example, these aspects include:

- Data transmission from the AOS to OSF
- Environmental requirements
- Time synchronization

In addition to the previously stated aspects where the HILSE project experience can be reused, it is also important to highlight the fact that the HILSE infrastructure can also be used as a test bench for evaluating the new correlator using the actual ALMA instrumentation. This will positively impact the commissioning process since it will be possible to test the new correlator without disturbing the ongoing ALMA operations. The technical room where the HILSE correlator will be deployed could also house a prototype of the 2nd generation correlator, which would provide many functional and infrastructure advantages including:

- Access to a replica of the Timing reference signal
- Access to data collected by up to 4 ALMA antennas
- HVAC capacity in place
- Fire protection

- Electrical power

8.2.1 Recommendations for the 2nd generation correlator project

To allow test and maintenance tasks to continue to be efficiently carried-out with minimal requirement of time on the production system, it is crucial than an independent test environment be available in the 2nd generation correlator era. *Therefore it is strongly recommended that the 2nd generation correlator project deliver not only the production correlator itself, but also a representative small version of it for use in such a system as the HILSE.* By “representative” it is meant that it should be able to run with the same software and firmware as the production correlator, it should support the same capabilities (subject to smaller number of antennas) including spectral setups, subarrays, VLBI, etc., and the output data should be identical to the production correlator for the same antenna inputs. If it is convenient for the 2nd generation correlator project, this small correlator could be installed first as a prototype test environment, as long as it is upgraded to production hardware/firmware/software by the end of the project.

It is recommended that the 2nd generation correlator project work with the HILSE team on how best to accommodate any prototype hardware and exploit the HILSE infrastructure for any on-site testing.

Acknowledgements

The Correlator Working Group would like to especially thank Todd Hunter for helpful comments and suggestions on many parts of the document, as well as double checking many of the most important calculations. We would also like to thank K. Saini, B. Shillue, C. Jacques, and G.H. Tan for helpful discussions that underpin much of the information provided in S8 (Deployment Options). Alain Baudry (AB) would like to thank the Digitization/DSP group at LAB (Univ. de Bordeaux) for several exchanges on possible ALMA digital system upgrades. AB also wishes to thank T. Cavalié (LAB) and N. Biver (Obs. de Paris) for preliminary discussions on future spectral and total bandwidth requirements from the planetary and comets communities.

Appendix

A.1 Additional feedback from Correlator Workshop held in February 2020

A.1.1 Additional Questions were addressed according to interest of each Break-out Group and time available so no judgement is made about level of consensus

1. What sets the practical limit on the IF bandwidth and max IF frequency? Is it digitizer performance or SIS mixer bandwidth?
   - IF limitations will be set by digitizers not by RF bandwidth
   - Rx will have different max IF – priority for upgrades means some of the more recently deployed Rx will not be upgraded for a long time (likely beyond 2030)
   - Atmospheric transmission windows in the mm/submm also play practical role
2. Are there extreme requirements for 10% of the science cases which drive the design, which might better be handled like VLBI (i.e., dump baseband data and process offline in software correlator)?
   - Needs of the many should outweigh the requirements of the few (especially where they add significant cost, or limit the more popular modes)
   - Examples included: super high spectral resolution, specific pulsar observations, etc.
   - Would limiting the spectral or special modes offered by the correlator make a new correlator significantly cheaper? – worth investigation
3. How do we reach a system-level timeline of phased development of all subsystems?
   - No group discussed this though it was a popular submitted question
4. Are there engineering/practical drivers for keeping the correlator at the high site?
   - No. Fiber length is not an issue but cost of laying fiber, new OSF infrastructure, etc. versus savings from maintenance, downtime, cooling, etc. should be investigated in detail. Also see [RD12].
5. Special requirements related to subarrays, fly's-eye use cases, paired array calibration (a la CARMA).
   - Require at least 3 subarrays for test/integration purposes +3 for science ops (12m, 7m, TP)
   - Support Main Array, 7m-array, and possibly TP array on same correlator for consistency
   - Consider ability to remove/add antennas while observing, and fast observing array creation should be a goal, although currently mostly determined by Central LO hardware and software
6. Are there possible trade-offs needed to accommodate scalability? (i.e., future expansion/upgrades)
   - Build bigger correlator room to start with, and enough cooling capacity for future expansion or upgrade
   - If the opportunity comes to increase the elements in the array, you probably want fresh design and parts rather than buying more obsolete components. So don't worry about scalability
   - Question might be better directed at surrounding equipment (fiber patch panel, number of fibers from each antenna, etc).
   - Can we minimize the software changes needed if we frequently replace the correlator? A correlator only costs about 2 antennas, so not a big value to preserve it? However, the time (and FTE) spent commissioning is large. How do we decrease this effort?
   - Things that might change on the timescale of replacing the correlator should not be put into the antenna, because it takes years to change out parts in an antenna – parts go obsolete much faster than in the past

A.1.2 Discussion on pros/cons of GPU, FPGA, and ASIC technologies for "X"

Overall choice depends on application, # calculations per byte, arithmetic intensity

- GPUs may be able to scale to large N (for cross-correlation)
  - They are a lot less power hungry than they used to be
  - New (since 2018) "tensor cores" can be exploited for significant speed improvement compared to previous generations
  - Still concerns about limitations on input bandwidth, though this may be solved in time (currently nVidia A100 GPUs support up to 600GB/s link bandwidth via custom interface)
- FPGAs are becoming more capable all the time and are still the most popular option
- ASICs were not viewed favorably as a design option for future correlators by the majority of attendees – summarized by one attendee as "not worth the trouble"
  - Always several generations behind in feature size and hence capability
  - Lots of cost/effort up front in design phase
  - However, still the lowest power option but not by as much as it used to be

A.2 CorrWG Comments and Concerns Regarding First "F" at Antenna

Assuming that the 2nd generation correlator architecture will be FX (S5.1.1) and because high spectral resolution will require one first stage channelizer it is useful to investigate if there are advantages or show-stoppers with a design including a first F stage in the antennas. We specifically refer to conversion of the time-series data to frequency-based data ahead of the correlator subsystem. The alternative would be, as usually, to implement this first F stage in the central correlator room. Several of our comments below are focused on a first channelization obtained by Fourier transforming the digitized astronomical data of each sideband of a receiver to provide digital sideband separation (DSBS) at a high rejection level. (The benefits of DSBS have been demonstrated for 2SB receivers.) We do not consider here any channelizer or specific digital processing scheme that would be required to format and transport the data captured by a new generation fast ADC and assume that the ADC data flow remains in the time-series domain.

There are two main **advantages** and significant **unknowns** with DSBS as described below:

1. A first F stage at the antennas would permit digital sideband separation with rejection levels around 30 dB (or above, depending on the ADC effective number of bits). DSBS requires the derivation of four calibration coefficients and to specify calibration channel widths. An unpublished study made by the Digital Front-End WG estimates channel widths in the range 20 to 300 MHz depending on the ALMA bands and on the achieved rejection level. Such channel widths could be useful for a first spectral channelization at antennas. However, there are also several caveats to this advantage:
   a. The total time required for DSBS calibration is uncertain, time stability of the coefficients is still unknown and coefficients need to be refreshed at each new LO tuning.
   b. We further note that a rejection level around 30 dB is already provided with interferometric LO offsetting and we are not aware of any science case requesting more than 30 dB. DSBS would be useful for single dish observations with DSB receivers or receivers with poor SB rejection. However, all future receivers will be 2SB (except for Band 1) and we are not aware of specific requests for SB rejection higher than what is planned or currently offered.

   c. It is also important to note that no current or upcoming interferometric facility uses or plans to use DSBS.
2. Data transmission requirements could be optimized because a first F processing at antennas could include re-quantization to best match the fiber links capability. However, the future high speed links do not seem to imply strong rate limitations, a point also made in [RD01].

There are also a number of **difficulties** with a first F-stage at antennas (the concerns listed below still need deeper investigation).

1. Because coarse delay corrections to compensate geometric delays must be: (i) applied in the time domain before F processing (but see architecture dependence below), and (ii) synchronized with the remaining fine delay corrections, we anticipate more complexity with first F windows derived at antennas.
   a. Synchronization constraints depend on the details of the F-engine implementation which could include for example a polyphase filter bank (PFB) followed by per-channel DFT or a tunable filter bank (TFB) similar to the current TFB followed by per-channel PFB.
   b. In all cases, synchronization of coarse delays with sub-sample delays at the fine channelization level (in e.g. complex gain modules also performing Walsh demodulation) is facilitated with all correlator functions managed in the same environment.
2. Although future ADCs will deliver many bits there will be science cases requiring application of non-linearity and quantization corrections. With a first F at antennas additional information such as the analog power levels per ‘frequency chunk’ may need to be communicated to the central correlator.
3. Re-quantization may be needed to lower the data rate with F antenna products at antennas. Then, controlling the signal levels at the re-quantizer input stage to optimize the sensitivity loss may also add complexity.
4. Providing the ‘return to phase’ capability across the full data path could be more difficult with first F at antennas. The control software of all correlator sub-systems and full performance check of the correlator could also be more complex.
5. Benefiting from technology advances or implementing further flexibility or new capabilities in future upgrades will probably be easier in a single correlator room compared with an antenna-distributed first F-stage (even though a first F design at antennas could be upgraded with large resource FPGAs).
   a. This is especially true for the FX ‘corner-turn’ (where portions of the F spectrum are interconnected with the X-engine processors for all antennas) whose optimal design depends on all details adopted for the F- and X-platforms which may evolve with technology advances (FPGAs or GPUs).
6. Preventive actions or failure maintenance are probably better managed when all of the correlator sub-systems are located in the same room. The full correlator performance and power dissipation might be better understood and checked in the same room.

There are also some **open questions** which would need to be addressed and are design-dependent. In particular,

1. It would be useful to: (i) compare the total cost of the resources needed to perform the full F-portion of the FX design in a single room with a design where resources are distributed in the correlator room and, for a first F, at antennas; (ii) estimate the extra power needed to dissipate the extra heat resulting from a first F at antennas.

2. Concerning beamforming to operate the array in a phased mode (VLBI, pulsars) it would be useful to confirm that it is indifferent to a first F at antennas (as suggested for transposed F data (i) summed before VLBI recording and/or (ii) sent to the X-engines for correlation).
    a. We further note that for pulsar observations in the ALMA lower bands an upper limit on channelization may be needed due to de-dispersion measure constraints.

A.3 Considerations related to Delay Correction

As described in S6.1.4.2, the prodigious number of samples per FFT segment that could be generated by the 2nd Generation Correlator will present challenges to the accurate correction of the delay if attempted in a single stage. As mentioned in S5.2, an overall delay correction strategy for ALMA 2030 is currently unknown. In this section we present some of the options/strategies that could be pursued.

A.3.1 Comparison of FFT segment time as a function of assumed sample rate and finest spectral resolution

Section 6.1.4.2 describes the required delay correction rate and FFT segment time using the most challenging of the 2nd Generation Correlator requirements in combination. Here we explore some additional options, if for example the correlator sampling rate were reduced via digital downconversion of the native signal from the ADCs, compared to the digitizer goal rate of 40 GSa/s (S5.1.1).

Max Baseline (km)	Sampling Rate [GSa/s]	Required Delay Update Rate, 1% sensitivity loss (ms)
35	16	0.574
35	32	0.287
35	40	0.23
60	16	0.335
60	32	0.167
60	40	0.134

Next we assess the segment length (in time) at a variety of correlator sampling rates and spectral resolution at maximum correlated bandwidth.

Finest channel spacing goal	Digital IF sky signal range[a] (GHz)	Digital IF signal bandwidth (GHz)	Digitizer sampling rate (GS/s)	Sample rate of signal after digital processing (GS/s)	Actual power of 2 channel spacing (kHz)	Segment length (power of 2)	Segment length (samples)	Segment length time (ms)
Finest resolution at max bandwidth (23.35 kHz)	0-8	8	24	16	15.25879	20	1,048,576	0.065536
	4-12	12	24	24	22.88818	20	1,048,576	0.043691[b]
	0-16	16	40	32	15.258789	21	2,097,152	0.065536
	4-20	20	40	40	19.073486	21	2,097,152	0.052429
Finest possible	0-8	8	24	16	1.90735	23	8,388,608	0.524288
	4-12	12	24	24	1.43051	24	16,777,216	0.699051

resolution min goal (2.33 kHz)	0-16	16	40	32	1.90735	24	16,777,216	0.524288
	4-20	20	40	40	1.19209	25[b]	33,554,432[b]	0.838861[b]
Finest possible resolution stretch goal (1 kHz)	0-8	8	24	16	0.95367	24	16,777,216	1.048576
	4-12	12	24	24	0.715256	25	33,554,432	1.398101
	0-16	16	40	32	0.95367	25	33,554,432	1.048576
	4-20	20	40	40	0.95367	25	33,554,432	1.048576

[a] *Ranges that start at 0 GHz rather than 4 GHz will require a digital downconversion stage to baseband **after** the digitizer.*
[b] *For this row, a factor of 2 smaller segment length would yield a channel spacing which would also **nearly** meet the minimum requirement.*

So from this table it is clear that even reducing the sample rate (via digital signal processing) from the ADC sample rate to 16 GSa/s, there is only one option (row 5) that can meet the minimum goal for the finest possible resolution (2.33 kHz) at 35 km. It has an FFT length of 2^23 samples. Additionally, the required digital processing would inevitably come with additional sensitivity loss that is not presently accounted for in S6.1.4. In contrast, if we only consider the finest resolution that is needed at the max correlated bandwidth, (23.35 kHz or 0.2 km/s at 35 GHz is the minimum requirement), even the full 40 GSa/s from the ADCs (row 4) has a segment length in time that is less than the required delay update rate, which is even true for a 60 km max baseline, and the FFT length is 2^21 samples. This choice would also significantly decrease the losses incurred from imperfect delay correction.

It is worth noting that we know of no science driver (as yet) that would motivate anything remotely close to requiring 1.17 kHz (or even 2.33 kHz) channels at the full maximum correlated bandwidth, and as noted in S6.1.4.2, the data rate would be astronomical. The main driver for such a capability is simplicity -- if all observing modes can be directly derived from a single constant frequency division FFT and then spectrally averaged as needed, then the concept of correlator "mode" is almost non-existent (except on the user-side). The alternative is to use something like a spectral resolution of 23.35 kHz for "normal" observing modes (S6.2.2), and have a specialized "ultra-high" spectral resolution correlator mode(s) that gives up bandwidth for resolution (see e.g. S6.2.4).

A.3.2 Delay correction strategies

The preceding section and S6.1.4.2 considered the requirements if the full delay correction were restricted to the fine channelization portion of the correlator processing. An alternative is to apply the delay in two or more stages; recall that ALMA presently applies portions of the delay in several different stages (S4.1.1, Q3): Ultra-fine (Digitizer clock-phase), Fine (TFB phase shift used for VLBI implementation of baseband delays normally done in the CDP), Bulk (Correlator Station Card), and finally Residual delays (CDP). As was noted at the 2020 Correlator Workshop there is a strong desire to simplify the delay tracking strategy, but going from 4 to 2 stages would still be an improvement.

Another possible way to achieve the fine delay correction within the correlator is to rotate the phase of time-series data just before the Fourier transform, which is multiplication with cosine and sinusoidal functions in the complex domain. This operation would need to be performed with a sufficient number of bits (≥ 6) to maintain the required accuracy[48]. However, since the operation is followed by Fourier transform in the FX method, the increased number of bits can be accommodated. Care must be taken in this operation to ensure return to phase.

A.4 Additional Findings Related to Parallel Deployment: Front-End Assembly, Space availability for New Digitizers and IF Switches, and Fiber Availability

This section elaborates on the findings of S8.1. The details presented here result from discussions and consultation with several ALMA engineers, especially with K. Saini (FE Electronics Chassis), G. H. Tan (BE racks) and C. Jacques and B. Shillue (Fibers).

Information regarding the possible location of new digitizers and IF switches is given below:

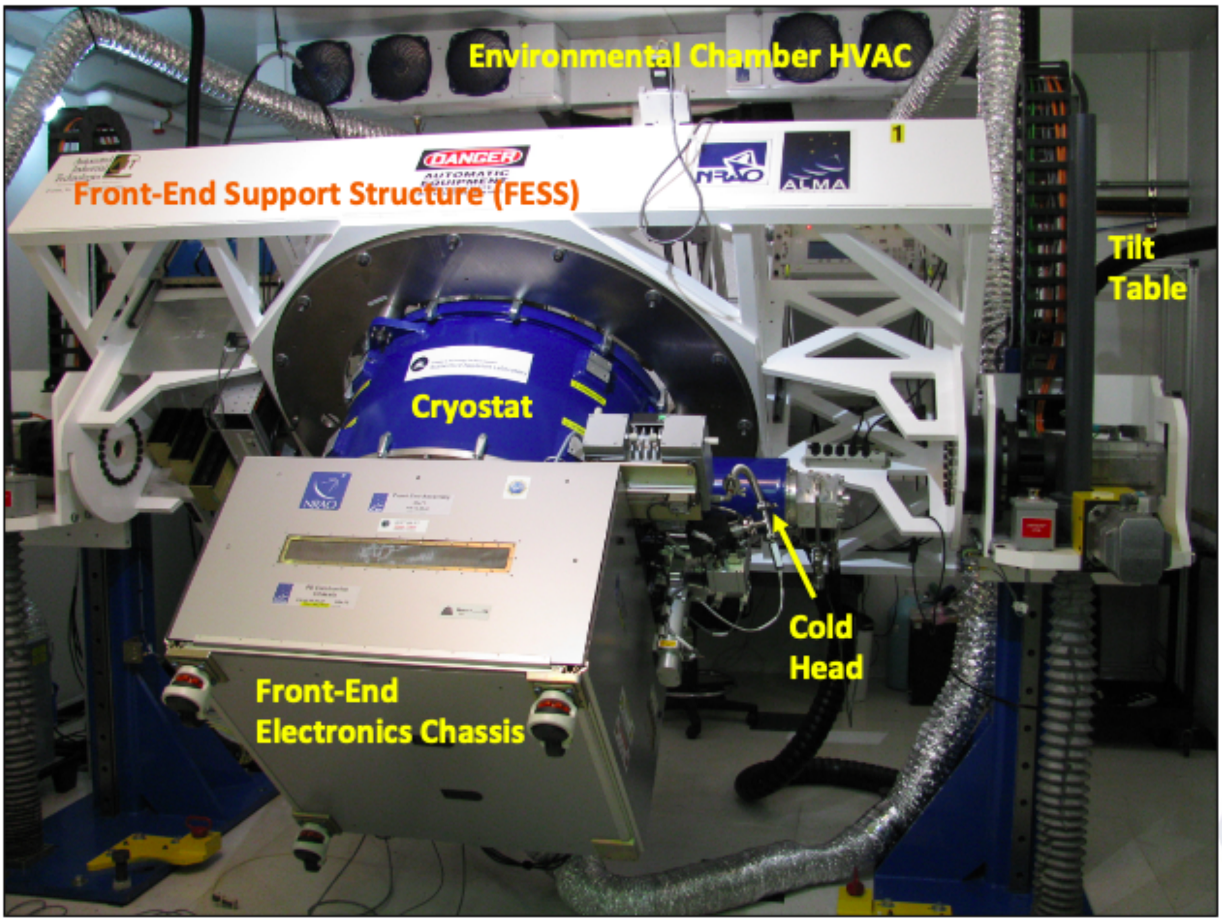

Figure 6: ALMA Front-End assembly from the NRAO Central Development Lab with major components listed. Assembly is tilted at an angle of 30 degrees.

---

[48] The current Tunable Filter Bank cards employ 9 bits to apply the fine-delay correction.

Figure 6 shows a picture of the ALMA Front-End Assembly located at the NRAO Central Development Laboratory. It is notable that while the picture shows the assembly tilted at an angle of 30 degrees, when the assembly is removed from an antenna it is at a tilt of 0 degrees, i.e. the “Front End Chassis” label would be pointed toward the floor. The maximum width of the assembly is constrained by the Cold Head Assembly sticking out from the right-hand side, which is a critical component of the receiver cooling system that must be serviced on each antenna annually. During removal, the whole assembly is lifted vertically so that the Cold Head can clear the Back-End rack (not shown).

Figure 7 shows a close up view of the Front-End Electronics Chassis with the bottom panel removed, as it would be during servicing (the FE assembly is tilted at 90 degrees for this photo). The location of the current IF Switches is labeled for reference, and indeed, space inside the chassis for new digital components is severely limited. K. Saini estimates that the available space just to the left of the current IF Switches is at most approximately 9 x 12 x 3 u (where u = 1.75 inches). Additionally, there is concern about heat dissipation, RFI shielding, and the proximity to other critical components inside the Chassis. However, there appears to be room for a decent sized “box” to be bolted on to the side of the Chassis, taking advantage of the clearance already required by the Cold Head Assembly. This box could harbor the new digitizers and IF Switches (which are likely to be integrated), with ample room for shielding. Note that the IF switch (and likely some IF cabling) will need to be replaced anyway to support higher IF frequencies, so the existing IF Switch space is potentially available.

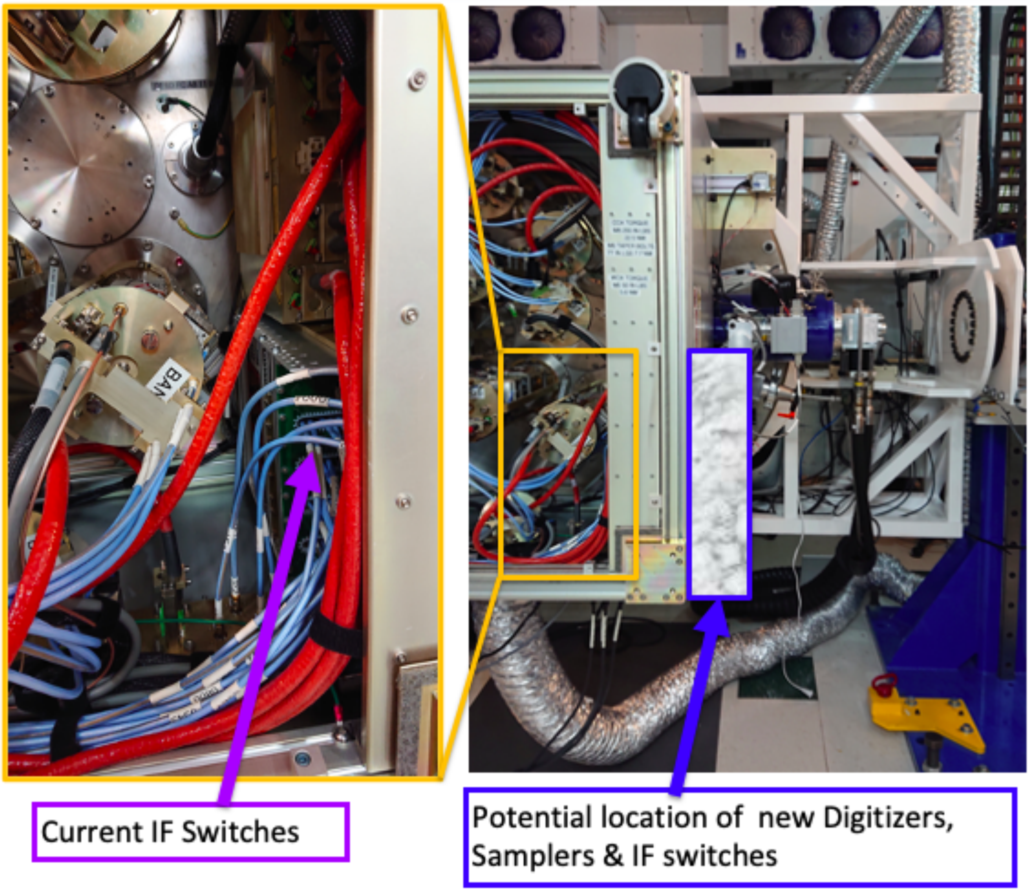

Figure 7: View Inside Electronics Chassis. The inset (left) shows the area indicated by the gold box on the right where the current IF switches are located and demonstrates that there is little room left inside the Electronics Chassis. The blue

box (shaded grey) in the right figure shows a potential option for locating a new “Digital Assembly Box” that could harbor the new IF switches and digitizers.

The concerns noted regarding the availability of fibers from the antennas to the AOS and thence to the OSF are listed below:

1. For the antenna pad to AOS fibers, the four best fibers were already “cherry-picked” for each pad. For some pads it is possible that there isn’t even one more fiber that is usable. It was noted that for some pads the variation in power between the best and worst fibers is as much as 7dB. They have not been re-evaluated for many years.
2. The number of “extra” fibers from the AOS that will be required to “feed” a correlator located at the OSF is not presently unknown. Though unused fibers do exist (see S8.1.2), they have not been assessed for many years, it is likely that some of them are not viable.
3. Regarding the AOS to OSF fibers, with the purchase of new telecommunications equipment it might be possible to free up some of the fibers currently in use. This equipment is getting older and could face obsolescence issues in the future anyway.