North American
ALMA Science
Center

ALMA data rates and archiving at the NAASC

NAASC Memo 110

Mark Lacy, David Halstead

Date: revised 2013 Jan 15, 2015 May 22, 2015 July 17, 2015 August 18; original 2012 April 26

ABSTRACT

The ALMA baseline correlator can send 1 GB/s of lag data to the data processing cluster, which after processing and conversion to the ALMA data format would result in 512 MB/s of ALMA "raw" data (visibilities and autocorrelations). In practice, the data rate from ALMA is much lower due to a combination of scientific and practical considerations. The current ALMA operations plan is based on a data rate of 200TB/yr, an estimate obtained from consideration of specific science cases in the Design Reference Science Program. This study was, however performed some years ago. In this memo we update this estimate based on experience from ALMA operations in Early Science and developments in the scientific field since then. We show that, based on realistic assumptions for ALMA Full Science operations, 200TB/yr remains a good estimate of the steady-state data rate, and a data link speed between Chile and the ARCs of 100Mb/s will be sufficient for the next decade (though increasing the link speed will allow more timely copying of large datasets to the ARCs). The cost of storage has also decreased by a factor of ~3 between 2011 and 2015. In the light of this we recommend the OT warning on observational data rates (currently 12MB/s) is removed, though observers should be prompted to use channel averaging where appropriate. Future upgrades to the correlator and front end, and improvements to operational efficiency could, however, increase the data rate by an order of magnitude on timescales of a decade or more.

1. Introduction

This document is structured as follows. Section 2 discusses the science cases used to estimate data rates for ALMA, and their possible future evolution. Section 3 describes the physical limitations on data transfer and how they might evolve on an ~10yr timescale. We then discuss the likely seasonal variation in the data rate before concluding with a discussion of data management strategies that could be employed.

2. Data rates for science projects

2.1. Early estimates of the ALMA data rate

ALMA memo 501 (Lucas et al. 2004) makes an estimate of the data rate for ALMA based on a selection of projects from the ALMA Design Reference Science Plan (DRSP). This study concluded that the mean data rate from ALMA would be 6MB/s, with a peak rate of 60MB/s. This estimate has increased slightly since then due to an ~10% addition from the compact array (ACA) correlator to 6.7/67MB/s. (The ACA correlator is capable of data rates up to 2GB/s, however, we believe it will be typically used in the same modes as the main array to complement the *uv*-coverage of 12m observations, thus the 10% estimate is reasonable.)

Assumptions were:

1. Images have a spatial sampling 1/3 the beam, and only final images are stored. (This leads to images taking up about 5% of the total data volume.)

2. 4 bytes/visibility

3. Only the part of the spectrum required by the observer to satisfy the DRSP science goal is kept, and it is sampled at the Nyquist frequency corresponding to the required resolution.

4. Only one (WVR corrected or not) dataset is archived

5. Integration (sampling) time $82/b$ where b is the maximum baseline in km, up to a maximum of 45s.

6. Calibration follows standard procedures.

2.2. Estimates based on ALMA Cycles 0-2 Early Science

We found that Early Science programs differed from the DRSP programs in several important respects:

1. Multiple FDM basebands are the rule rather than the exception, thus a typical project has ~1000 spectral channels (240-3840) instead of a few hundred. In many cases extra basebands are

set to contain “bonus lines” that are not the main science goal of the proposal, but which would add value to the observations. This was exacerbated in Cycles 0 and 1 as no channel averaging was available in the OT, meaning that many observations were taken at higher spectral resolution than necessary. Spectral averaging up to a factor of 16 was available from Cycle 2 onwards.

2. Integration times are not currently tuned to the baseline. Instead all FDM observations were obtained with 6s averaging and all TDM ones with 2s.

3. The observational efficiency is lower than anticipated for a variety of reasons. During good weather, the efficiency can be ~80%, but a more typical average is ~60% once losses to weather are included, and it was as low as 25% early in Cycle 1. Also the available science time in Cycles 0-3 have been limited (Cycle 0:1000hr, Cycle 1: 1140hr [including 340hr carried over to Cycle 2], Cycle 2: 1700hr, Cycle 3: 2100hr).

4. The growth of promised 12m antenna numbers in each cycle has been as follows: Cycle 0: 16 antennas; Cycle 1: 32 antennas; Cycle 2: 34 antennas; Cycle 3: 36 antennas. (The actual numbers of antennas used for observations has varied around these nominal figures.)

5. For Cycle 0 only, calibrated measurement sets were stored in the archive, increasing the archive volume by about a factor of three over the raw data alone. This practice was abandoned after Cycle 0 due to data volume concerns.

6. In Cycle 4 online WVR correction may lead to an approximate doubling of the raw data volume until the offline and online corrections can be compared over a range of conditions and a decision made as to which is preferred.

7. In later cycles (>~4), the telescope will move towards using bandwidth switching for calibration, i.e. phase and flux calibrations will be performed in low resolution (TDM) mode, even if the data are taken at full spectral resolution in FDM mode. This would reduce data rates by 30-50%.

Item (1) above reflects the way mm/submm science has changed since the DRSP was written. Many more interstellar molecules are now known, and the wide bandwidth of ALMA means that there is a high probability of multiple lines being available in a single Science Goal setup. It is likely that item (2) will change once long baseline (>2km) observations become common, however for Cycle 3 such observations are still in the 25% “non-standard” pool along with high frequency (Band > 7) and polarization observations, so will have little impact on data rate. Items (3) and (4), observational efficiency, science hours and numbers of antennas should all increase slowly with time, but their evolution is hard to predict. Item (7) will reduce the data rate for high spectral resolution projects, perhaps by as much as a factor of two at high frequencies.

An analysis of a randomly-selected set of 77 Cycle 3 proposals (for which the first author was technical assessor) was made in order to make an estimate of the data rate based on observer requests. The mean data rate requested (taking the highest data rate science goal from each proposal) was 6.8MB/s (54Mb/s) with a median of 6.1 MB/s (49Mb/s). With 2100hr available for science in Cycle 3, and assuming an observing efficiency (including weather) of 0.6, this corresponds to a total data volume of 31TB, which, when spread evenly throughout the 1-year duration of Cycle 3, corresponds to a mean data rate of R_{bl}=0.012Mb/s per baseline.

We have also made an analysis of the actual data gathered during Cycle 0-2. Figure 1 shows panels with (top) the total data accumulation in the archive up to the end of March 2015, (middle) the data rate in TB/month and (bottom) the data rate in Mb/s per baseline (R_{bl}). This latter number is easy to scale to future configurations. During fairly good weather and system stability during the latter half of Cycle 2 this number has reached R_{bl}~0.015-0.02Mb/s of elapsed time, compared to our estimate of R_{bl}=0.012Mb/s per baseline based on the Cycle 3 proposals. (Note the spike towards the end of Cycle 0 was partly due to more than the nominal number of antennas being used, and partly due the archiving of calibrated measurement sets, not done after Cycle 0.) The difference is likely to be due to a combination of not including the ACA data and the data products (about 10% of the data volume each) resulting in underestimating the Cycle 3 data rate, and not including shutdown times resulting in an overestimate of the Cycle 2 rate.

The average data rate in a given annual Cycle will scale as:

$$\text{DR(Mb/s)} = R_{\text{bl}} * N(N-1)/2 * (T_{\text{sci}}/8766)\,\eta$$

Where T_{sci} is the time available per annual cycle for science observations in hr (compared to the 2100hr in Cycle 3), η is the observational efficiency during science runs in average weather (estimated to be 0.6 in recent Cycle 2 campaigns). R_{bl} remains uncertain as it is difficult to predict how observers will react when the current OT warning is removed (or changed), see below.

2.3. Data rate warnings

It is possible that if data rate warnings are removed users will attempt to use the maximum data rate. Although only 10% of Cycle 3 proposals requested data rates above the 12MB/s value that triggered a warning in the OT, anecdotal evidence suggests many users tuned their proposals to stay below that limit. The hard limit of 60MB/s cannot be reached by proposers with the current correlator setups available in the OT combined with the inability to shorten sampling times. In practice the highest data rate that can be requested in the OT corresponds to 4 full FDM windows, 34.5MB/s with 43 antennas in our "realistic" Full Science scenario. In the unlikely event that all our users did this, the data rate would be 560TB/year, assuming 4500hr of observations. In a more realistic case, where 10% of our users required such high data rates (and were put off from applying for them in Cycle 3 by the OT warning), then the data volume would increase by only ~56TB/year. *Nevertheless, we recommend that the OT provides a hint that the user should consider channel averaging if the channelization requested is much smaller than the bandwidth used for sensitivity for FDM modes, at the level of the current "blue warnings".*

2.4. Full Science Scenarios

For Full Science, we assume as a baseline 43 operating antennas and 4500 science hours per cycle, with a science observing efficiency of 0.6 (including weather). In our FS1 scenario (Table 1) we assume the same mean data rate per baseline as for Cycle 3. In FS2, we assume that the balance of FDM (spectroscopic) and TDM (continuum and low resolution) observations remains the same as in Cycle 3 (27% TDM to 73% FDM), but that the FDM observations use the full available correlation resources (e.g. 4x3840 channels x 2 polarization). In the "worst case" FS3 scenario, we make the same assumptions as FS2, but increase the number of available antennas to 48, the number of hours to 5000 and the observing efficiency to 0.8. Our best guess would be somewhere between FS1 and FS2, so around the 200TB/year in ALMA memo 501, though the range could be a factor of two in either direction. Note also that we have not included the volume of the ACA data or the data products, which will increase the data volume by 10-20%.

Scenario	Data rate per baseline during obs (Mb/s)	Number of 12m antennas	Number of hr for science	Science observing efficiency (including weather)	Mean data rate (Mb/s) during obs	Data volume (TB/yr)	Mean data rate over cycle (Mb/s)
Cycle 2	0.1	34	1700	0.6	56	26	6.5
Cycle 3 (predicted)	0.1	36	2100	0.6	63	36	9.1
FS1	0.1	43	4500	0.6	90	110	28
FS2	0.23	43	4500	0.6	208	253	64
FS3	0.23	48	5000	0.8	259	467	118

Table 1: data rate predictions based on Cycle 0-2 experience. Scenarios: **FS1** - full science with "realistic" assumptions of telescope availability and observing efficiency, same mean data rate per baseline as Cycle 3, **FS2** - assuming the current balance of FDM and TDM observations, but with the FDM observations using four full resolution spectral windows, and **FS3** - as FS2, but with more optimistic operational assumptions. None of these include possible savings from using bandwidth switching calibration (item (7) above), but neither do they assume any shortening of current integration times (item (2) above). They also do not include ACA data or data products (so actual rates will be 10-20% higher).

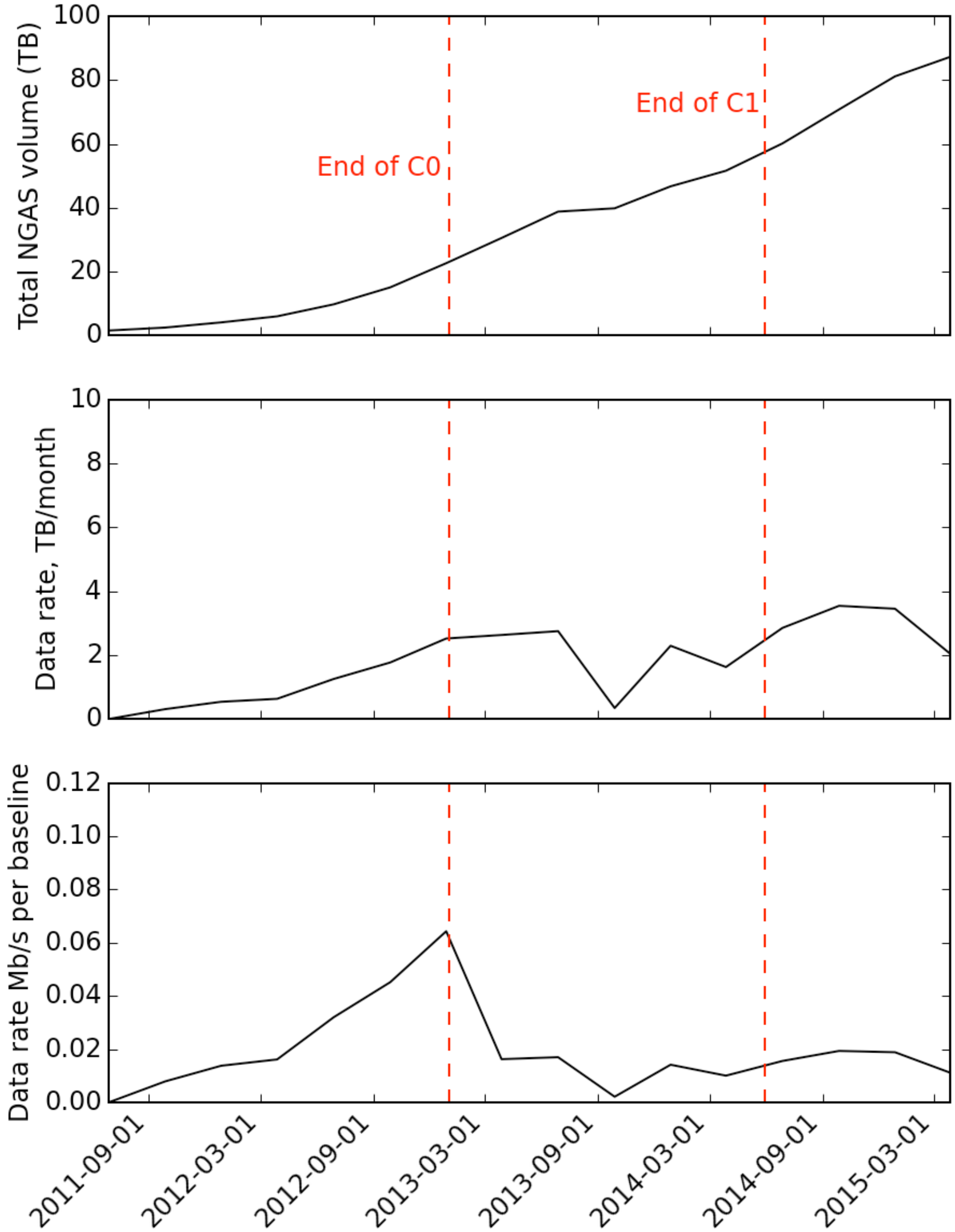

Figure 1: Measured archive growth in Cycles 0-2, top to bottom: cumulative increase in volume, data rate in TB/month, data rate in Mb/s per baseline.

2.5. Effect of future cycle capabilities

The extent to which Cycles 0-2 provide a good template for future cycles is debatable. Future capabilities may both increase and decrease the data rate:

1. *The ability to change the integration time.* Projects in Cycles 0-2 were observed with shorter integration (sampling) times than required by the $82/b$ law. The principal reason for this was that the WVR corrections needed to be applied offline. When online WVR correction becomes available these may be increased. However, for projects requiring high dynamic range, short integration times (~1s) will still be preferred to allow effective self-calibration of the data. Thus it is unclear how the ability to tune integration times will affect the data rates in the future. For the purposes of this document we assume making the integration times variable will have a net zero impact on the overall data rate.

2. *On-the-fly (OTF) mosaics*. As discussed below, this is an example of an observing mode with a strong science case for using high data rates, which will be commissioned in a future Cycle (4+).

3. *Improvements to observing efficiency*. The observational efficiency has increased considerably from Cycle 0 to Cycle 2, principally due to reductions in latencies. Further improvements may be possible by, for example, combining calibrations into sessions. However, we anticipate that due to a combination of many short duration observations and technical issues with long duration scheduling blocks, observational efficiency is unlikely to rise much over the next 2-3 years.

4. *Pipelined image products.* Pipeline image products will ultimately be produced in Santiago and mirrored to the ARCs, for the most part these will be a small fraction of the raw data volume (~5%), however in some cases (such as short integration mosaics) these will be comparable in size to the raw data.

2.6. Science cases for future data rates

On-the-fly (OTF) interferometry provides a specific example of a strong science case that will lead to a severe data rate challenge. OTF modes will require short sampling times, down to the limits of the correlator modes (512ms would allow all FDM correlator modes, 16ms for TDM). Data rates as high as ~300MB/s could be justified scientifically fairly easily, for example in the case of a line survey of a bright starforming region. With the current 60MB/s limit, OTF modes would have to be restricted in either number of channels, scan speed, or both.

Data reuse is also becoming increasingly important in astronomy, even for telescopes such as the Hubble Space Telescope that, like ALMA, are not primarily survey instruments. The archival value of taking large data volumes should therefore not be underestimated. Upgrades to the correlator, receivers and electronics on an ~10yr timescale that allow a significant increase in bandwidth beyond the current 8GHz may be considered. This would considerably expand the science that could be done with the array, and lead to further good science cases for very high data rates. Such an upgrade would put even a 1Gb/s link under pressure during high spikes (Figure 2).

A correlator upgrade study is currently being considered that would increase the available number of spectral channels by a factor of eight (Escoffier et al. 2015). In addition, with commensurate changes to the front-ends the bandwidth could also be increased. For example, the NEOMA array is planning for a 32GHz bandwidth. These would lead to data rate increases on the order of a factor of 4-8 on a ten year timescale. This is not a severe concern provided Moore's Law and related scalings for storage can keep pace. At least for the NAASC, bandwidth to South America is also not a concern given the rapid improvements in the internet infrastructure.

Developments in detector technology may allow for the fitting of focal plane arrays in one or more bands. These could easily introduce a further order of magnitude or more increase in the data rate. Likely timescales for this would be ~ a decade for technology development and a further decade for production, however, by which time we can expect data storage and transfer technologies to have evolved significantly.

3. Data rate limitations

3.1. Correlator/data capture

The ALMA correlator can output 1GB/s (Pisano et al. 2005) (though this assumes 8-byte visibilites, which are the size output by the correlator, in practice, however, 4-byte visibilities are typically archived, making the effective maximum data rate 512MB/s). This is currently limited to 64MB/s due to the speeds of the network interface cards and the 1Gb/s connections used. An upgrade would be possible at relatively modest cost to upgrade the connections and network to 10Gb/s allowing the full potential of the correlator to be realized, and is likely to be proposed as part of an ALMA development plan.

3.2. Transfer

AOS to SCO:

Data is currently transferred from the AOS to the mining town of Calama, between San Pedro and the coastal city of Antofagasta via fibre link at 2.5Gb/s. At Calama it joins the Chilean fiber backbone for transfer to Santiago (fibre is leased commercially from Calama to Antofagasta, then joins the REUNA backbone at Antofagasta). This allows both a high data rate from the telescope, and for a wide bandwidth video connection for remote observing operations.

SCO to the ARCs:

NA has secured a 100Mb/s link from SCO to Florida International University, Miami and hence to the US research backbone (I2/NLR). Cost is $50k/year, negotiated as a share of a 622Mb/s link used by AURA/NOAO-CTIO. The upgrade path would see this AURA link increased to a 1Gb/s link in the near future, and a 10Gb/s link to support NOAO initiatives such as LSST, with NRAO retaining a minority share. All these links are planned to be burstable to capacity, to allow the full bandwidth to be used, for example, to take advantage of the fact that most optical telescope data transfer will take place at night, leaving the daytime free for ALMA.

EU has an agreement through ESO with REUNA for a 40Mb/s link, this may be upgradable in the near future, however. In addition they have available any unused portion of the ESO 35Mb/s link (this link is heavily used at night, but less so during the day). The link to EA is currently 10Mb/s, with a further 10Mb/s on a "best efforts" basis, this will be upgraded to 25Mb/s with a further 25Mb/s "best efforts". EA has a guaranteed 120Mb/s link with 1Gb/s contracted until June 2016. Table 1 suggests that a steady-state rate of 50Mb/s is the minimum to ensure transfer of data in Full Science, and 100Mb/s is preferable to allow for recovery from outages and spikes of high data rate.

3.3. Data storage

The NAASC has 500TB/year of storage in its budget to 2015, corresponding to a steady-state data rate of 100Mb/s. Upgrades beyond 1PB/year with current storage technology would mean outsourcing the archive to a computing center as the computer room cooling capacity will be exceeded. However, the current budget is likely to be ample for our requirements. EU was initially budgeting for 200TB/year, but this was revised downwards based on the achieved Cycle 0-2 data rates.

4. Seasonal/Cyclical variations

ALMA will work on a 1 year cycle, with configurations varying from compact through extended. Extended array observations will use higher data rates as the data need to be sampled more often. Extended array observations also typically require better conditions, as phase coherence needs to be maintained over longer baselines. High frequency (>500GHz) observations are also likely to require high sampling rates, and will also be concentrated towards times of year when the conditions are good. Strong seasonal variations in data rate are thus to be expected. http://almascience.eso.org/about-alma/weather shows the monthly variations in water vapor at the ALMA site. On this basis the best months (for extended arrays and/or high frequencies, and thus high data rates) will be July through October, the worst months (compact arrays, low frequencies, or no data at all during the regular February shutdowns) will be January through March.

The data rate during these periods is likely to be about three times higher than average. Figure 2 shows a notional estimate of the growth of the ALMA data rate with time out to ~2020. The seasonal variations assume that the winter quarter contains most of the high frequency.extended configuration observations, leading to spikes in the data rate at those times of year.

5. Growth of storage requirements in Charlottesville

The NGAS system consists of sets of 24-disk nodes, generally installed in sets of four. The first set of four nodes had 2TB disks, allowing for redundancy these result in 30GB of storage per node. Future nodes will use 6TB disks, or larger as they become available. Each NGAS node takes up 5Us of rack space (in 40U racks) and consumes approximately 0.42kW of power. Power requirements in the ER computing room are dominated by the NAASC compute cluster (and the accompanying Lustre filesystem), each compute node occupies 1U of rack space and consume 0.27kW each.

In addition to the ALMA mirror, Charlottesville also needs to store the EVLA mirror and (selected) GBT data. Estimates are that each archive will reach ~3PB by the end of 2015. UVa has recently opened a computing center where rack space is currently available for nominal cost to collaborations. We will therefore use our collaborations with UVa to try to secure some of this space for data which is public, i.e. past its proprietary period. Alternatively, should negotiations with UVa prove unsuccessful, we will ask NCSA to host data for us. As all our archives are mirrors, and we will only outsource storage of public data, there is no data security risk in this.

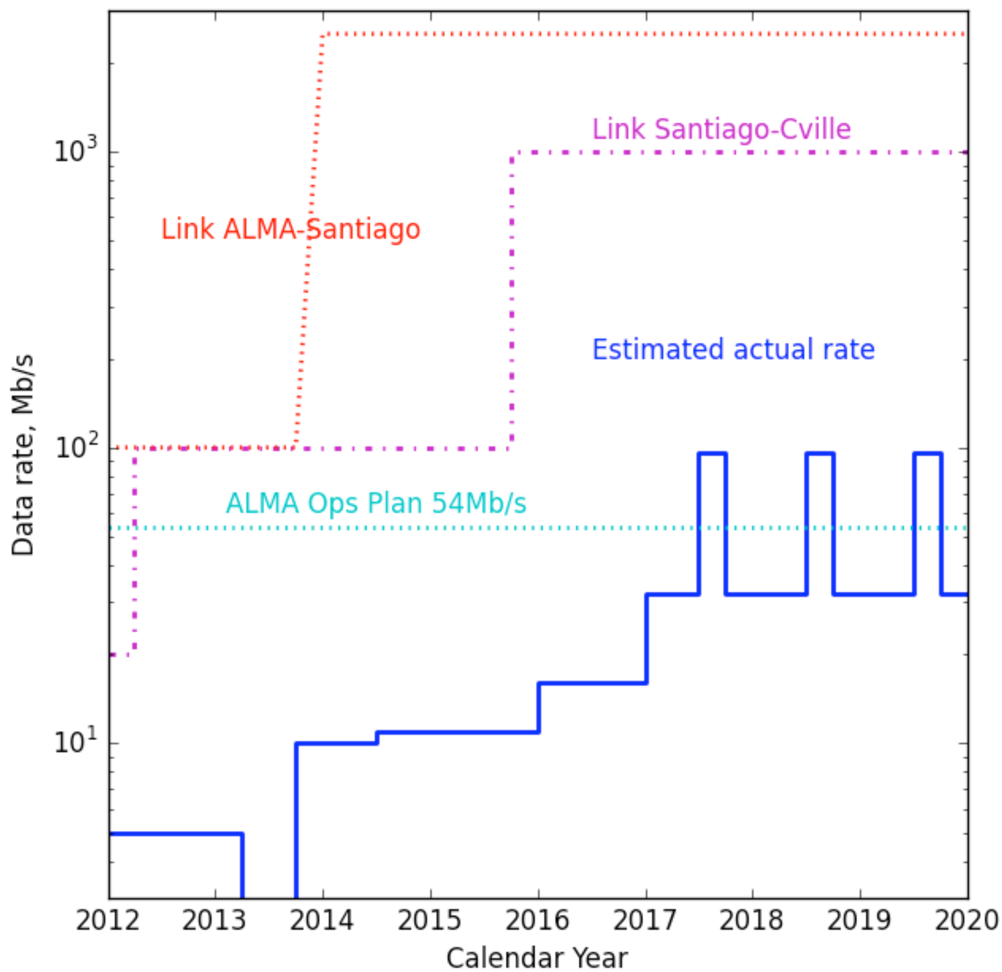

Figure 2: likely growth of the ALMA data rate through 2020 (including CSV data until 2017). The blue solid line is our best estimate of the data rate based on Cycle 0 proposals and likely seasonal variations, assuming an AOS network upgrade in 2016 and a possible observational bandwidth increase on the 2020 timescale ("Poss. telescope upgrades"). The dotted cyan line corresponds to the rate in the ALMA Operations Plan vD (200TB/yr). The red dotted line is the likely JAO target for the ALMA to Santiago link likely to be in place by the end of 2012. The dot-dashed magenta line is the NAASC link to Chile (100Mb/s from ~April 2012, with an upgrade to 1Gb/s mid-decade).

6. Implications for data management strategy

6.1. Managing the data rate growth through to Full Science

Current data rates suggest that no drastic action is required to manage growth of the data rate. Increases in the amount of data taken have been offset by the more realistic estimates of observational efficiency, and hours and numbers of antennas available for science based on our experience in the Early Science cycles. Possibly during Cycle 4, the data rate may be higher as online WVR may be commissioned, and for a short time both corrected and uncorrected ASDMs will be stored, doubling the data volume during that period, however, we expect that to be of relatively short duration.

6.2. Beyond 2020

A new correlator could also be on the horizon on the 10-20 year timescale as software improves. This would lead to an increase in the data rate by a factor of 4-8.

6.3. Evolution of storage costs

The cost per terabyte of disk storage was dropping steadily at a rate of about 40% per year in the first decade of the 21st century ("Kryder's Law"), however, since 2010 the rate of price drop has slowed. Nevertheless, since 2011 we have reduced the cost of storage in NA by one third (equivalent to a price drop of about 27% per year), achieved as the drive sizes for a typical $500 disk have increased from 2TB to 6TB (so 1PB of storage which would have cost ~$315k in 2011 now costs ~$105k, including the costs of the NGAS nodes). Although disk drive development may be close to the floor in terms of $/TB, new (solid state) technologies are making strides, and are likely to provide the next wave of reduction in cost of storage.

6.4. Data processing

Data processing is currently a bottleneck in the data flow from ALMA, this is primarily due to the lack of an automated pipeline that can process data through calibration to imaging. Pipeline calibration is currently attempted for about 75% of the data, this fraction will increase slowly with time. Pipeline imaging is still several months away at the time of writing, and all imaging (and calibration of non-pipelineable data) is performed by hand. Manual imaging and calibration is slow as it requires careful setup and review of the parameters, and often multiple iterations of the reduction scripts (including runs to test scripts and produce "clean" logfiles). Weblog reviews of pipeline calibrated data are also currently taking a long time. Thus, we believe, the incremental increase to run times for the pipeline and scripts from a higher data rate will have some negative effect on reduction times, but the increase will not be linear with the data rate due to the large amount of manual setup and review needed during the reduction process. The manual process also makes very poor use of the computing clusters, with large idle times e.g. overnight and at weekends when the manual reducers are not at work. As the pipeline becomes more capable, the need for manual processing, with its accompanying inefficiencies will reduce. The cluster hardware is optimized for parallel processing, and we are also expecting the pipeline to become increasingly parallelized over time, improving our usage of the cluster in that respect too.

References:

Escoffier, R. et al. 2015, NAASC Memo 115; http://library.nrao.edu/naasc.shtml

Pisano, J., Amestica, R. & Perez, J. 2005 10th ICALEPCS Int. Conf. on Accelerator & Large Expt. Physics Control Systems. Geneva, 10 - 14 Oct 2005, PO2.067-4