# ALMA Science



## How can I reduce my data rates for an observation (and reduce the size of my dataset)?

Brenda Matthews - 2020-09-21 - ALMA Observing Tool (OT)

The total data rate in an observation is an estimate of the total instantaneous data rate summed over all basebands based on the current parameter settings. Please see the OT Reference Manual and the OT User Manual, available from the Science Portal under 'Documents & Tools' page in the ALMA Science Portal, regarding baseband configuration. Basically, it is a measure of how much data can be sent to the archive per second, a number which can be considerably less than ALMA is capable of producing. When proposing a FDM (Frequency Division Mode, see OT User Manual) observation in the OT, you may therefore encounter a warning about the data rates your observation will require. The FDM raw data rate (in Mbytes/sec) is approximately

$N_A(N_A - 1)/2 * N_{bb} * 8192$ channels * (4/10)Mbytes/$I_T$

where $N_A$ is the number of antennas, $N_{bb}$ is the number of basebands (normally 4) and $I_T$ is the spectral integration time (see ALMA Technical Handbook).

For the 64-input correlator (used for non-ACA observations) the maximum data rate is 70 MB/s and a validation warning is issued if this is exceeded - a similar check has been introduced for the ACA data rate. Average data rates (reduced from the peak value due to time not spent observing) >40 MB/s (only relevant to non-ACA observing) may be examined for changes that might lower the data rate at Phase 2. In general, if you can lower the date rate (using the methods below) without compromising the science, then you should do so since the final size of your dataset will be reduced and become more manageable.

In Phase 1 (proposal submission), the proposer is able to select spectral averaging of channels within a subband (independently for different lines in the same subband provided they all start with the same spectral resolution). This factor is the number of adjacent channels to average together and must be a power of 2, e.g., 1, 2, 4, 8 or 16. The degree of averaging can be different for different lines. This is useful if you are interested in a relatively small bandwidth but do not require the best spectral resolution that bandwidth can offer or you know you will need to average the data for a significant detection. In this case, opting for the narrower band width combined with spectral averaging will reduce your data rate. TDM (Time Division Mode) spectral windows are also available for observations of continuum or wide spectral lines.