From Z-Machines to ALMA: (Sub)millimeter Spectroscopy of Galaxies
ASP Conference Series, Vol. TBD, 2006
A. J. Baker, J. Glenn, A. I. Harris, J. G. Mangum, and M. S. Yun, eds.

Redshift Determination Algorithms for Broadband Spectroscopic Data

Min S. Yun[1], Mark Heyer[1], Itziar Aretxaga[2]

(1) University of Massachusetts and Five College Radio Astronomy Observatory (FCRAO); (2)Instituto Nacional de Astrofísica, Óptica, y Electrónica (INAOE)

Abstract. The LMT Redshift Search Receiver is a powerful new instrument to investigate the population of submillimeter galaxies and the star formation history of the universe. With large instantaneous bandwidth (40%) and high spectral resolution ($R \sim 3000$), it can measure a multitude of redshifted molecular and atomic submillimeter lines simultaneously to provide a millimeter counterpart to medium resolution UV and optical spectra. The use of the redshift receiver system on the 50 meter aperture of the Large Millimeter Telescope (LMT) enables an unbiased survey of a large sample of SMGs to determine redshifts and an accounting of the physical state of the cold ISM in these distant galaxies. In this contribution, we describe a cross-correlation redshift analysis algorithm designed to exploit the spectral information generated by the Redshift Search Receiver. The effects of signal to noise, source confusion and the use of priors such as photometric redshifts are also explored.

1. Introduction

The resolution of the cosmic IR background into discrete sources by the SCUBA camera on the JCMT has firmly established the existence of a substantial population of dusty galaxies with inferred star formation rates in excess of 1000 $M_{\odot}$/yr (see review by A. Blain in this volume). Determining the redshift and evolution of submillimeter galaxies (SMGs) remains a challenging task for astronomers. These systems are faint in the rest frame UV and optical wavelengths. Therefore, measuring their redshifts using traditional optical spectroscopy has been difficult (but see Chapman et al. 2005) and has motivated efforts to define redshifts from molecular line emission from these systems. To date, CO transitions in the millimeter and submillimeter bands have been detected in ∼25 sources (Solomon & Vanden Bout 2005). In addition, the fine structure lines of C I (Weiss et al. 2005) and C II (Maiolino et al. 2005; Iono et al. 2006) have also been recently measured. Detection of these atomic and molecular transitions provide an unambiguous confirmation of the source redshift. In addition, these lines can be used to determine the molecular mass and diagnose the physical state of the cold gas component that is responsible for the high level of star formation in these systems.

To investigate the star formation history of the universe, it is imperative to secure redshift measurements for a large sample of SMGs. This requirement has led to the development of several special purpose instruments specifically designed to detect redshifted CO and C I transitions from distant SMGs (see

Table 1. A comparison of broadband spectrometer systems

	frequency (GHz)	bandwidth	resolution
Zpectrometer (GBT)	$26-40$	42%	2000
Redshift Receiver (LMT)	$74-111$	40%	3000
COBRA (CARMA)	$80-115$	4%	6400
Z-Spec (CSO)	$195-310$	46%	300
ZEUS (CSO/JCMT)	$789-900$	3.2%	1000
ALMA Band 3	$84-116$	8.0%	25600
Band 4	$125-163$	5.6%	29000
Band 6	$211-275$	3.3%	63000
Band 7	$275-373$	2.5%	83000
Band 9	$602-720$	2.4%	170000

Table 1). Such instrumentation provides a valuable resource for the critical, high angular resolution follow-up studies by ALMA. These new instruments are currently in development or in the early phase of operation.

The UMass/FCRAO instrumentation group has developed an ultra-wide bandwidth redshift search receiver system for the Large Millimeter Telescope to operate within the 3mm window (74-111 GHz). This system has sufficient spectral resolution to resolve the molecular emission lines from massive galaxies while simultaneously covering a large range in redshift space. In addition to CO, a variety of molecular transitions from species such as HCN, HCO^+, HNC, CS, CN, HC_3N, and NH_3 (e.g. Henkel & Bally 1985; Nguyen-Q-Rieu et al. 1989; Aalto et al. 2002; Gao & Solomon 2004; Garciá-Carpio et al. 2006) have been detected in nearby starburst and normal galaxies. The Redshift Search Receiver on the LMT should have sufficient sensitivity to detect these secondary lines to greatly improve its ability to determine the redshift even when only a single CO line falls within the passband. This is likely a common feature for other broadband instruments with sufficient spectral resolution operating on large telescopes such as ALMA. An important realization is that observational and analysis schemes relying on the identification of a single bright CO or C I transition (i.e., "redshift by eye") are too limited and do not exploit the available information provided by these new instruments.

Here, we report the analysis of our algorithm designed for robust redshift determination from the broadband spectra obtained using the Redshift Search Receiver. A method for constructing a simulated mm/submm spectrum based on the observed line strengths in Galactic massive star forming regions is described. Based on the review of the traditional methods for analyzing broadband UV and optical spectra, a cross-correlation algorithm is designed. The effects of signal to noise, source confusion, and the use of priors such as photometric redshifts are also explored. Lastly, we discuss the lessons and implications for the studies of high redshift galaxies using ALMA.

2. LMT Redshift Search Receiver System

The Redshift Search Receiver system for the LMT is a combination of a MMIC-based front-end receiver designed to maximize the instantaneous bandwidth with

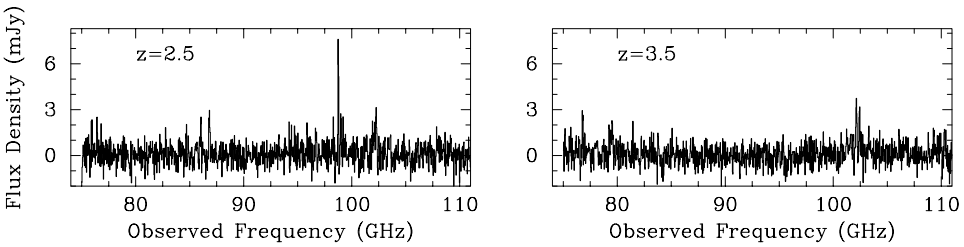

Figure 1. *Left:* A simulated spectrum for a $z = 2.5$ source with a peak CO (3–2) line flux density of 6 mJy (10σ) observed using the LMT with $T_{sys} = 100$ K. *Right:* A simulated spectrum for a $z = 3.5$ source with a peak CO (3–2) line flux density of 3 mJy (5σ).

a single tuning and an innovative wideband analog autocorrelator system that covers the entire 36 GHz bandwidth with 31 MHz (about 100 km s^{-1} resolution at 90 GHz) frequency resolution (see article by N. Erickson on this volume). Covering ∼40% of the total bandwidth at 90 GHz instantaneously, it is well suited for blind searches for redshifted CO and C I lines from gas-rich galaxies in the distant universe. At least one CO/C I transition is expected to fall within the instrument bandpass at redshifts between $z = 0 - 0.4$ and $z = 1.0 - 3.2$, For these z intervals, the detection of other molecular and atomic transitions and the application of prior information such as spectral energy distributions, enable a reliable determination of the galaxy redshift (see § 4.2.). Further confirmation of the deduced redshift can be accomplished by verifying the presence of other molecular or far-IR atomic transitions at other wavelengths. For $z > 3.2$, two or more of CO transitions fall within Redshift Receiver bandpass providing a unique determination of the galaxy redshift.

3. Simulated Spectrum

Apart from the low excitation CO lines, observational data on molecular transitions in extragalactic sources at the millimeter and submillimeter wavelengths are extremely limited. In a recent molecular line survey in the nearby nuclear starburst galaxy NGC 253, Martín et al. (2006) found that the chemistry of the ISM, as traced by the molecular transitions in the 2mm window (129-175 GHz), is strikingly similar to the Galactic center molecular clouds. To construct simulated spectra of galaxies at high redshift, we have selected over 500 transitions at wavelengths between 70 GHz and 1000 GHz, with observed line strengths $\geq$ 10% of the adjacent CO transitions in Sgr B2 and/or OMC-1 line surveys (Johansson et al. 1984; Sutton et al. 1985; Blake et al. 1986; Jewell et al. 1989; Serabyn & Weisstein 1995; Schilke et al. 1997; Lee et al. 2001; White et al. 2003; Comito et al. 2005).

Simulated redshift receiver spectra of a source Doppler shifted to $z = 2.5$ and $z = 3.5$ are shown in Figure 1. We have added Gaussian-distributed noise corresponding to sensitivity expected in each frequency channel after one hour of integration on the LMT (0.6 mJy assuming $T_{sys} = 100$ K). The detection of the secondary molecular lines in these model spectra within reasonable integration times demonstrate the power of the Redshift Search Receiver on a large aperture, single dish telescope.

4. Redshift determination

The complex distribution of lines within the Redshift Search Receiver bandpass demand sophisticated tools to reliably determine the galaxy redshift. Such spectral complexity is well documented with UV and optical bands so it is prudent to consider analysis methods developed within the optical community to obtain the appropriate redshift. The four most common approaches are:

- *Visual inspection of spectra.*
- *Emission line redshifts.* Quantitative analysis techniques such as wavelet transformation of continuum subtracted spectra are used to identify lines that are matched to a list of 30-40 possible lines.
- *Cross-correlation redshifts.* Template spectra (stars, galaxies, AGNs) are cross-correlated with the observed spectra to derive the redshift and the spectral types.
- *Principal component analysis (PCA).* A statistical approach of classifying spectra using a set of template spectra.

Given the large number of lines (> 500) and poor understanding of their expected strengths, a cross-correlation technique is likely the most fruitful for the analysis of the redshift receiver spectra. The cross-correlation analysis requires a rest frame template spectrum. In the absence of empirical data from high redshift systems, we construct a sequence of emission line amplitudes, using the same spectral line database described in § 3., based on observed line ratios with respect to CO. The cross-correlation product,

$$\zeta(z) \equiv \int S(\nu) M(\nu, z) W(\nu) d\nu$$

is computed for each step in z, where $S(\nu)$ is the observed spectrum, and $W(\nu)$ is a weight. The Doppler shifted template spectrum at redshift z, $M(\nu, z)$, is derived as

$$M(\nu, z) = \int_{(\nu - \Delta\nu/2)(1+z)}^{(\nu + \Delta\nu/2)(1+z)} T(\nu')/T_{CO}\ d\nu'$$

where $T(\nu')$ is the rest frame template spectrum (see § 3.), and $\Delta\nu$ is the channel width. Three different weighting schemes are considered (see Table 2).

Table 2. Different weights for cross-correlation analysis

weighting scheme	functional form	
unweighted	$W_i = 1/M_i$ if $M_i \neq 0$ $W_i = 0$ otherwise	least biased by the template
bright lines only	$W_i = 1$ if $T_i \geq T_{cr}$ (e.g., $T_{cr} = 0.2T_{CO}$) $W_i = 0$ otherwise	
flux weighted	$W_i = M_i$	most biased by the template

The redshift at which $\zeta(z)$ is strongly peaked with respect to the statistical error, $\sigma_\zeta(z)$, corresponds to the value of z at which the redshifted template best matches the observed spectrum. The cross-correlation products normalized by the statistical error for the $z = 3.5$ spectrum (shown in Fig. 1) are shown

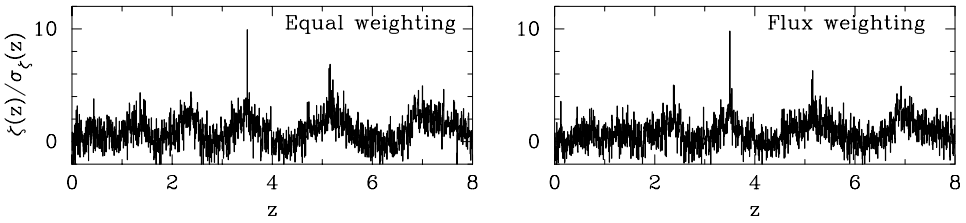

Figure 2. The cross-correlation products normalized by the statistical error, $\zeta(z)/\sigma_\zeta(z)$, for a $z = 3.5$ spectrum shown in Fig. 1 using *Left:* equal weighting and *Right:* weighting by expected line strengths. The input spectrum has two CO lines detected with only S/N=3.3, but the redshift of the source is well determined in both cases.

in Figure 2 using uniform (left) and flux (right) weighting. While the input spectrum has two CO lines that are only marginally detected with S/N$\sim$ 3, the source redshift is robustly determined by both weighting schemes, largely because of the power in the large number of fainter transitions along with the two CO lines that yield a unique redshift estimate. The semi-periodic low level broad peaks seen in these plots reflect the non-uniform distribution of spectral line density in the submillimeter bands. The value of the cross-correlation product depends on the weight W_i used, but the impact of the particular weight adopted is reduced when the cross-correlation product is normalized by the statistical error (see Fig. 2).

4.1. S/N dependence

Observations in general contain a random noise component due to instrumental and atmospheric fluctuations. Success of the cross-correlation algorithm is examined against the signal-to-noise (S/N) ratio of the input spectra in Figure 3. This algorithm works remarkably well even when CO lines are detected with a S/N of only 3.3 owing to the redundancy afforded by the detection of the secondary molecular lines. Increasing the signal strengthens the confidence level as more lines are detected and the amplitudes are more readily discriminated from noise fluctuations. Since the S/N ratio increases only as a square root of integration time, integrating this analysis algorithm with the data collection program could significantly increase the efficiency of redshift surveys.

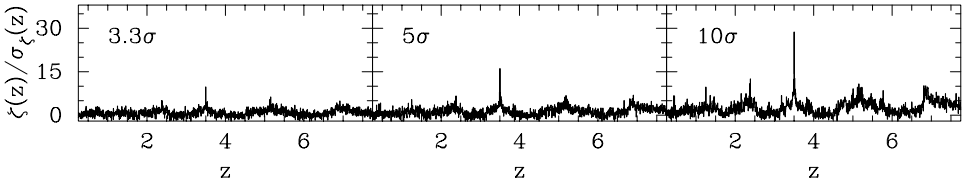

Figure 3. $\zeta(z)/\sigma_\zeta(z)$ for a $z = 3.5$ spectrum where the CO lines are detected with a S/N of 3.3 (*left*), 5.0 (*middle*), and 10 (*right*), respectively. The cross-correlation method performs remarkably well even with low S/N spectra.

4.2. Redshift ambiguity for a single CO line

At $z < 3.2$, only one CO line falls within the bandpass of the redshift receiver. In these cases, incorporating other known priors should significantly improve the

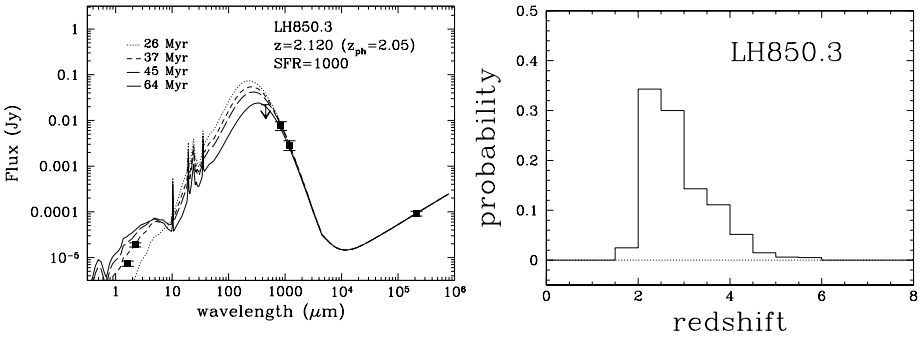

Figure 4. *Left:* The spectral energy distribution of the z=2.15 Lockman Hole source. *Right:* the photometric redshift likelihood distribution.

redshift estimation The photometric redshift likelihood distribution for a given source, discussed by I. Aretxaga in this volume, offers a critical complement to spectroscopic redshifts. As an example, the observed spectral energy distribution for a $z = 2.15$ submillimeter source in the Lockman Hole region is shown in Figure 4 (left panel) along with the photometric redshift likelihood distribution derived for the same source (right panel).

A simulated spectrum with a weak CO line (5σ integrated over the entire line) is shown in Figure 5. Only the CO J=3–2 line is readily identified in this spectrum near ν=110 GHz. Without knowing the redshift, this feature could also be interpreted as the CO J=1–0 line at $z = 0.05$ or the J=2–1 line at $z = 1.10$. The normalized cross correlation function $\zeta(z)/\sigma_{\zeta(z)}$ plotted on the right panel shows a clear peak at the correct redshift with a confidence level of $4.5\sigma_\zeta$. Using both independent analyses can provide a secure measure of the source redshift.

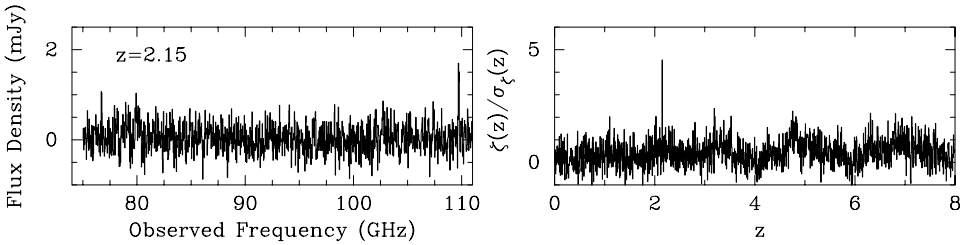

Figure 5. *Left:* A simulated spectrum for z=2.15 source with a peak CO (3–2) line flux density of 1.5 mJy (5σ). *Right:* Profile of $\zeta(z)/\sigma_\zeta(z)$ for the simulated spectrum. Spectral energy distributions and Resdshift Receiver spectroscopy provides a powerful combination of analyses to derive galaxy redshifts.

4.3. Source Confusion

The potential for source confusion rapidly increases with the decreasing angular resolution of the measurements. Based on the derived submillimeter source count, we estimate that multiple continuum sources brighter than 1 mJy at 850 μm are present within the 3mm beam of the redshift receiver on the LMT in less than 5% of time. This fraction increases to ∼10% if the receiver is installed on a 30-m diameter telescope. Our ability to determine redshifts and to distinguish

the presence of a second source is examined by superposing simulated spectra from two galaxies at redshift 2.5 and 3.9 (Figure 6). In these cases, a "redshift by eye" approach would be particularly problematic owing to the confusing pattern of lines. The cross-correlation algorithm readily picks out the $z = 3.9$ system. The presence of the $z = 2.5$ system is also seen, but isolating this feature from the remaining peaks would require additional information.

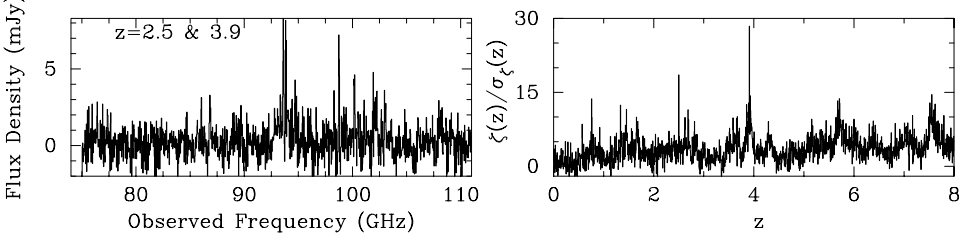

Figure 6. *Left:* A simulated spectrum for one source at $z = 2.5$ and a second source at $z = 3.9$. Each CO line is detected with S/N=5. *Right:* A cross-correlation product for the spectrum shown. Significant peaks corresponding to both z=2.5 and 3.9 are present.

5. Summary and implications for ALMA

A unique combination of a large bandwidth and high spectral resolution for the redshift search receiver system should produce a rich spectrum resembling a medium resolution spectrum at UV/optical wavelengths. The use of the redshift receiver system enables an unbiased redshift survey of a large sample of SMGs and an astrophysical analysis of cold ISM in these distant galaxies at the same time. We have demonstrated that the cross-correlation redshift analysis algorithm can effectively exploit the numerous faint lines appearing within the spectra in addition to the brighter CO and C I lines to produce reliable redshifts. An important lesson learned from our analysis is that utilizing the fainter lines is particularly important for the redshift range where only a single CO line falls within the instrument passband.

One of the key science goals of ALMA is imaging the redshifted dust continuum and line emission from evolving galaxies at epochs of formation and using the CO lines to measure their redshift (see the discussion by A. Wootten in this volume). The utility of the fainter lines and the importance of broad frequency coverage, as demonstrated here, should serve as an important lesson for future ALMA studies. As shown in Table 1, ALMA has a much higher spectral and spatial resolution but with a more limited instantaneous bandwidth, and new analysis algorithms should be developed to take full advantage of its unique capabilities and to address its limited bandwidth at the same time. The larger relative bandwidth makes the ALMA 2mm and 3mm bands attractive choices for this type of study. However, the spectral line density is larger at higher frequencies, and the utility of the fainter secondary lines demonstrated by this study suggests that a careful analysis weighing both of these effects is needed to derived an optimum redshift search strategy for ALMA.

Acknowledgments. Members of the Large Millimeter Telescope Observatory at both UMass and INAOE have contributed to this presentation. The

Five Colleges Radio Astronomy Observatory is supported by the NSF grant AST 02-28993.

References

Aalto, S., Polatidis, A. G., Huttermeister, S., Curran, S. J. 2002, A&A, 381, 784
Blake, G. A., Sutton, E. C., Masson, C. R., Phillips, T. G. 1986, ApJS, 60, 357
Chapman, S. C., Blain, A. W., Smail, I., Ivison, R. J. 2005, ApJ, 622, 772
Comito, C., Schilke, P., Phillips, T. G., Lis, D. C., Motte, F., Mehringer, D. 2005, ApJS, 156, 127
Gao, Y., Solomon, P. M. 2004, ApJ, 606, 271
Garciá-Carpio, J., García-Burillo, S., Planesas, P., Colina, L. 2006, ApJL, 640, L135
Henkel, C., Bally, J. 1985, A&A, 150, L25
Iono, D. et al. 2006, ApJL, submitted
Jewell, P. R., Hollis, J. M., Lovas, F. J., Snyder, L. E. 1989, ApJS, 70, 833
Johansson, L.E.B., Andersson, C., Ellder, J., Fridberg, P., Hjalmarson, A., Hoglund, B., Irvine, W.M., Olofsson, H., Rydbeck, G. 1984, A&A, 130, 227
Lee, C. W., Cho, S.-H., Lee, S.-M. 2001, ApJ, 551, 333
Maiolino, R. et al. 2005, A&A, 440, L51
Martín, S., Mauersberger, R., Martín-Pintado, J., Henkel, C., García-Burillo, S. 2006, ApJS, in press (astro-ph/0602360)
Nguyen-Q-Rieu, Nakai, N., Jackson, J. M. 1989, A&A, 220, 57
Schilke, P., Groesbeck, T. D., Blake, G. A., Phillips, T. G. 1997, ApJS, 108, 301
Serabyn, E., Weisstein, E. W. 1995, ApJ, 451, 238
Solomon, P. M., vanden Bout, P. A. 2005, ARAA, 43, 677
Sutton, E. C., Blake, G. A., Masson, C. R., Phillips, T. G. 1985, ApJS, 58, 341
Weiss, A., Downes, D., Henkel, C., Walter, F. 2005, A&A, 429, L25
White, G. J., Araki, M., Greaves, J. S., Ohishi, M., Higginbottom, N. S. 2003, A&A, 407, 589