Analysis of Molecular Spectra with SOFIA/EXES

Sarah Nickerson

Bay Area Environmental Research Institute/NASA Ames Research Center

Bay Area Environmental Research Institute

Stratospheric Observatory for Infrared Astronomy

SOFIA Spring School
April 20th, 2023

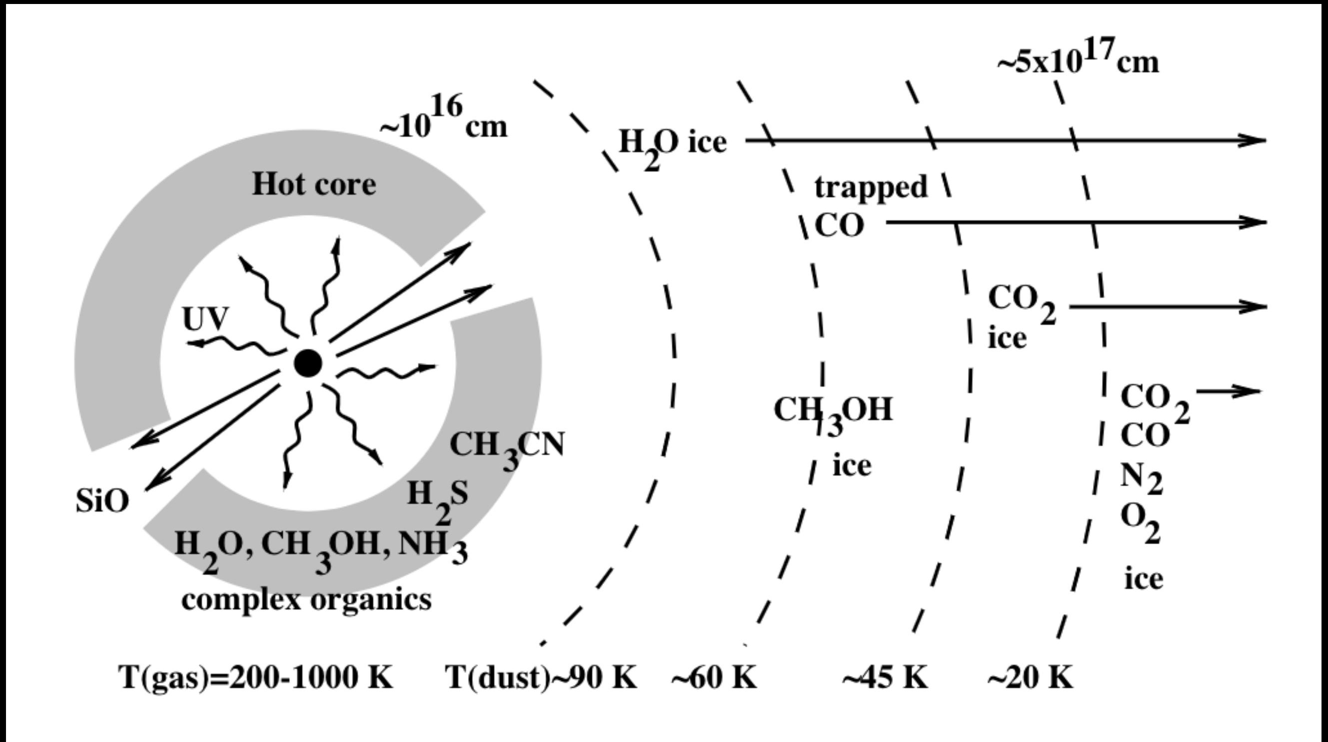
Overview

a.k.a. everything I wish I knew in the beginning

1. Introduction to hot molecule cores and their importance
2. Brief summary of our EXES survey towards Orion IRc2. This will used to as an example to demonstrate the analysis methods.
3. Analysis methods:
    I. Getting started with plotting EXES data
    II. Normalization
    III. Identifying molecules
    IV. Gaussian fits and rotation diagrams
    V. Crowded lines and simulated spectra
4. Applications and complex situations

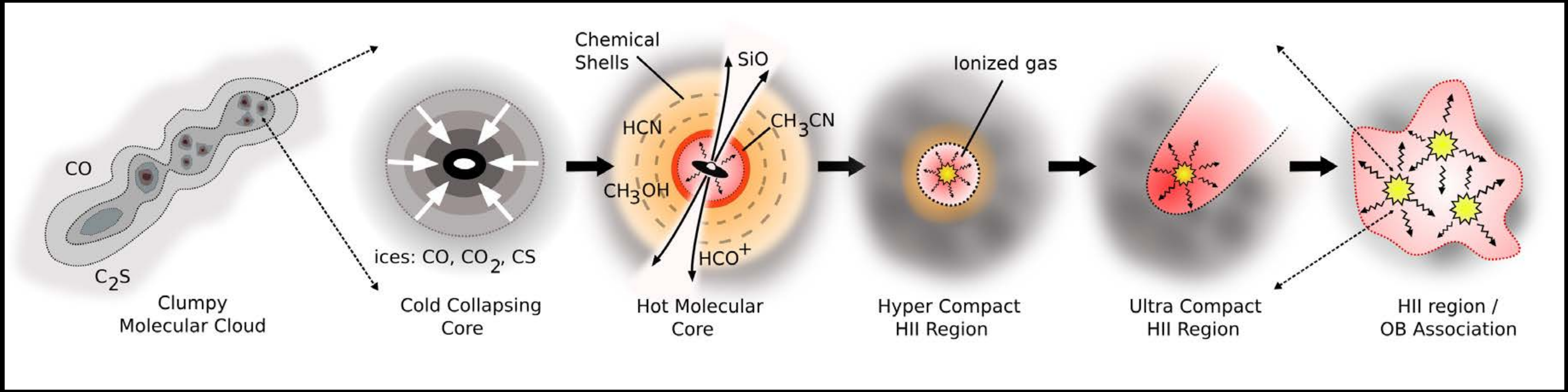
Hot Molecular Cores

- Warm (≥ 100 K), small (≤0.1 pc) and dense (10^5 to 10^8 cm^{-3}) gas near young, associated with high mass protostars (Ohisi 1997)
- Intermediate stage in star formation: stellar radiation evaporates ice on dust grains in molecular clouds
- Unlocks chemically rich reservoirs of complex and prebiotic molecules



Schematic illustration of the chemical environment of massive YSOs. The variation in the chemical structure of the ice mantle in the envelope due to thermal desorption is shown (based on Tielens et al 1991, Williams 1993).

van Dishoeck & Blake 1998



Cormac R. Purcell, https://crpurcell.github.io/projects.html

What Do We Learn From Hot Cores?

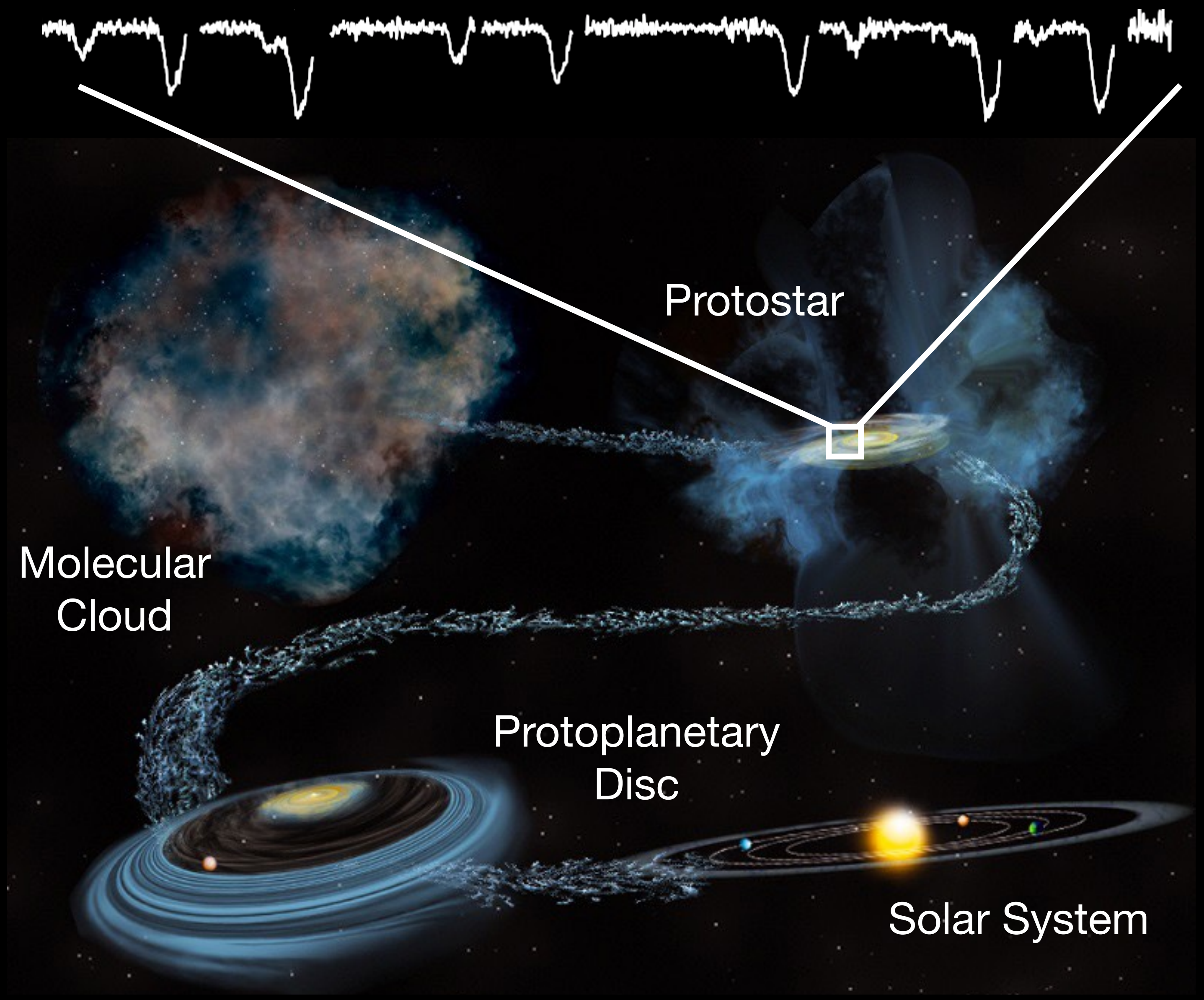
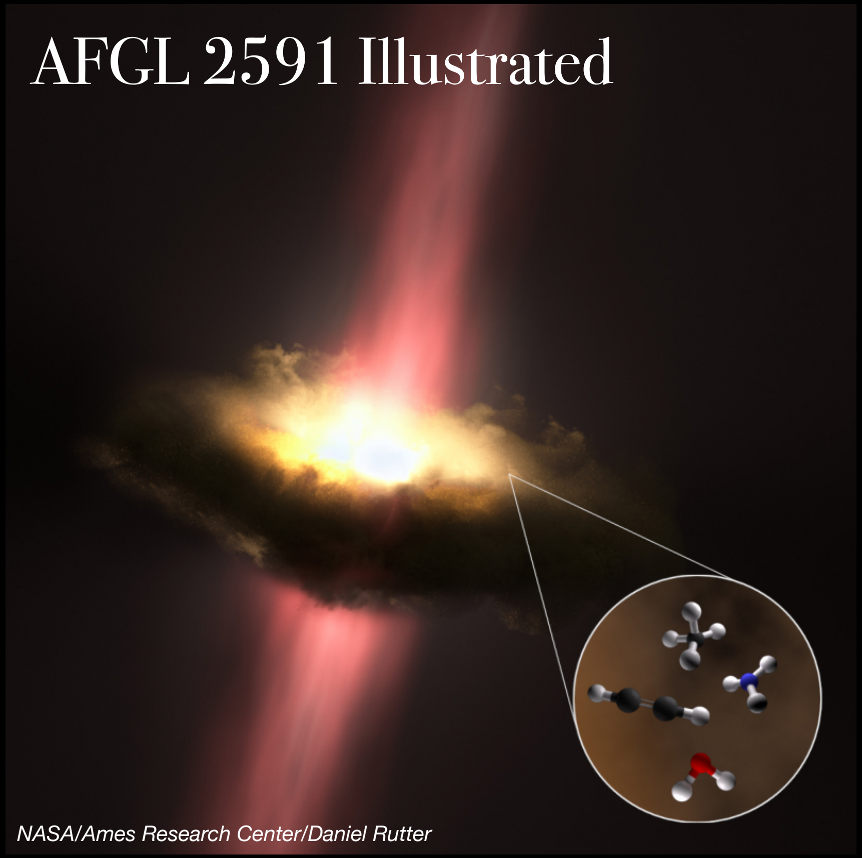


Illustration: Bill Saxton, NSF/AUI/NRAO;

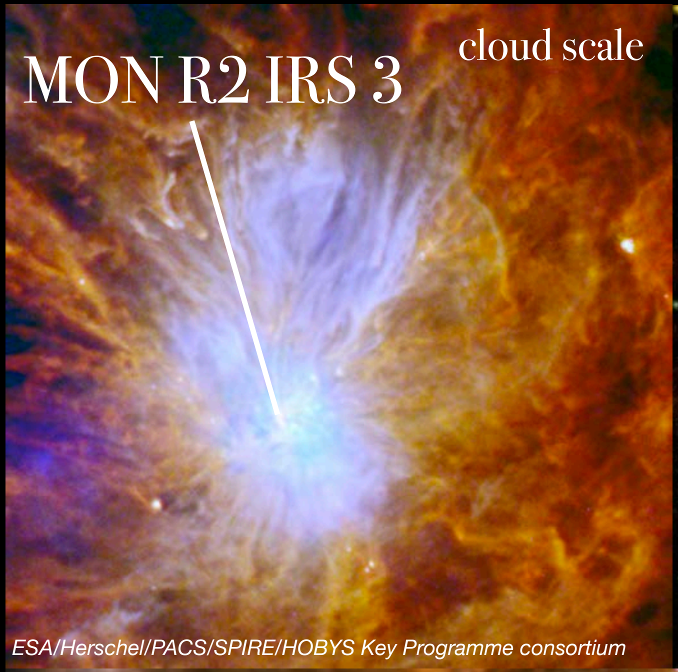
Spectra: H_2O towards AFGL 2136 (Indriolo+ 2020)

- Massive protostars probe the state of the interstellar medium at the earliest stages of star formation
- Our own sun may have formed in a massive star-forming region and inherited its molecular inventory from a natal hot core region (Adams, 2010; Drozdovskaya et al., 2018; Beltran & Rivilla, 2018)
- This gas contains the precursors to probiotics that will form planetary systems such as our own
- Studying hot cores will elucidate the origins of prebiotic molecules and inform chemical modelling

- Dozens of hot cores discovered in the Milky Way, handful in the LMC and the SMC
- Hot Core Resources:
  - Overview Papers: Ohishi 1997, van Dishoeck & Blake 1998, Kurtz et al. 2000, van der Tak 2004, Cesaroni 2005, Beltrán & Rivilla 2018
  - Textbook: A. G. G. M. Tielens 2021, Molecular Astrophysics



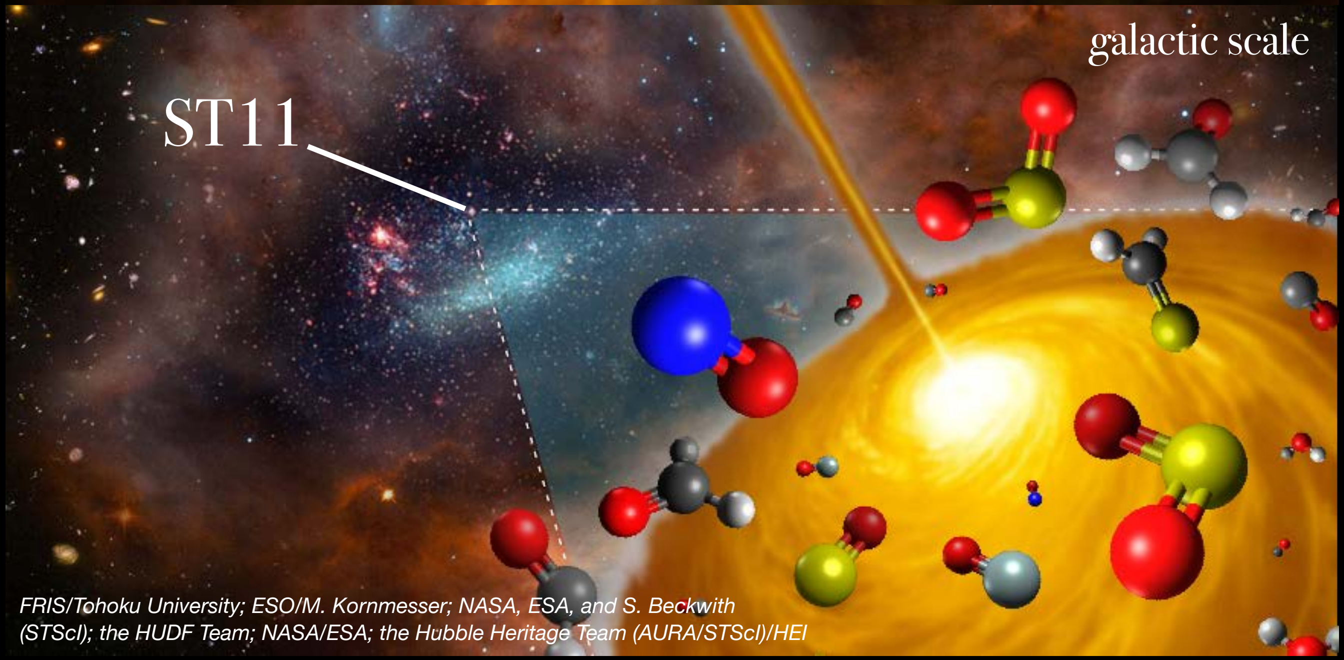
NASA/Ames Research Center/Daniel Rutter



ESA/Herschel/PACS/SPIRE/HOBYS Key Programme consortium

C. Aspin+, NIRI, Gemini Obs., NSF

ESA/NASA/JPL-Caltech/Whitman College

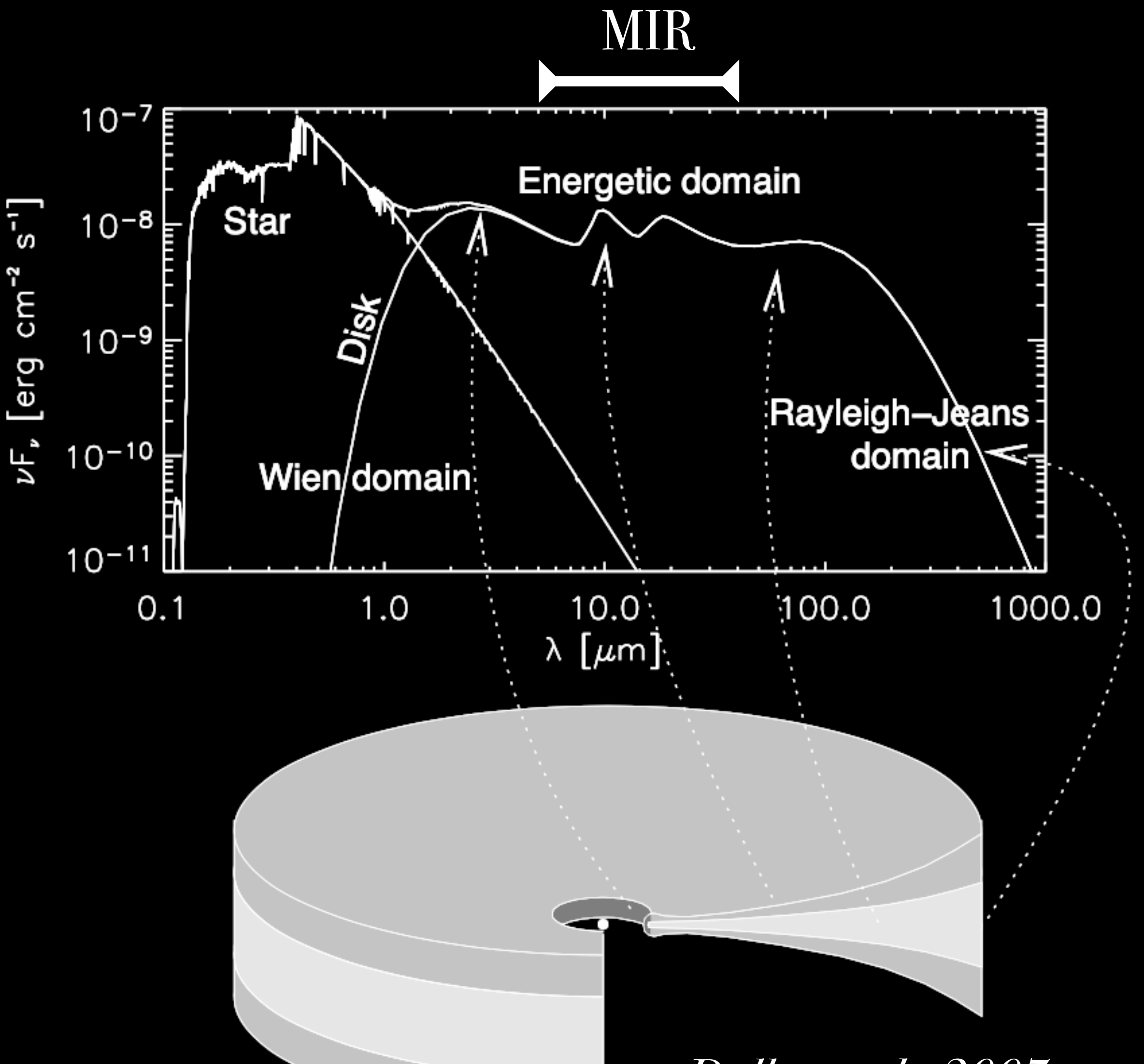


FRIS/Tohoku University; ESO/M. Kornmesser; NASA, ESA, and S. Beckwith (STScI); the HUDF Team; NASA/ESA; the Hubble Heritage Team (AURA/STScI)/HEI

Hot Cores in the MIR

- Previous high spectral resolution surveys limited to radio, sub-mm, mm, and far-infrared wavelengths
- These longer wavelengths capture rotational transitions of molecules with permanent dipole moments
- Easily accessible from the ground with facilities such as ALMA and SMA

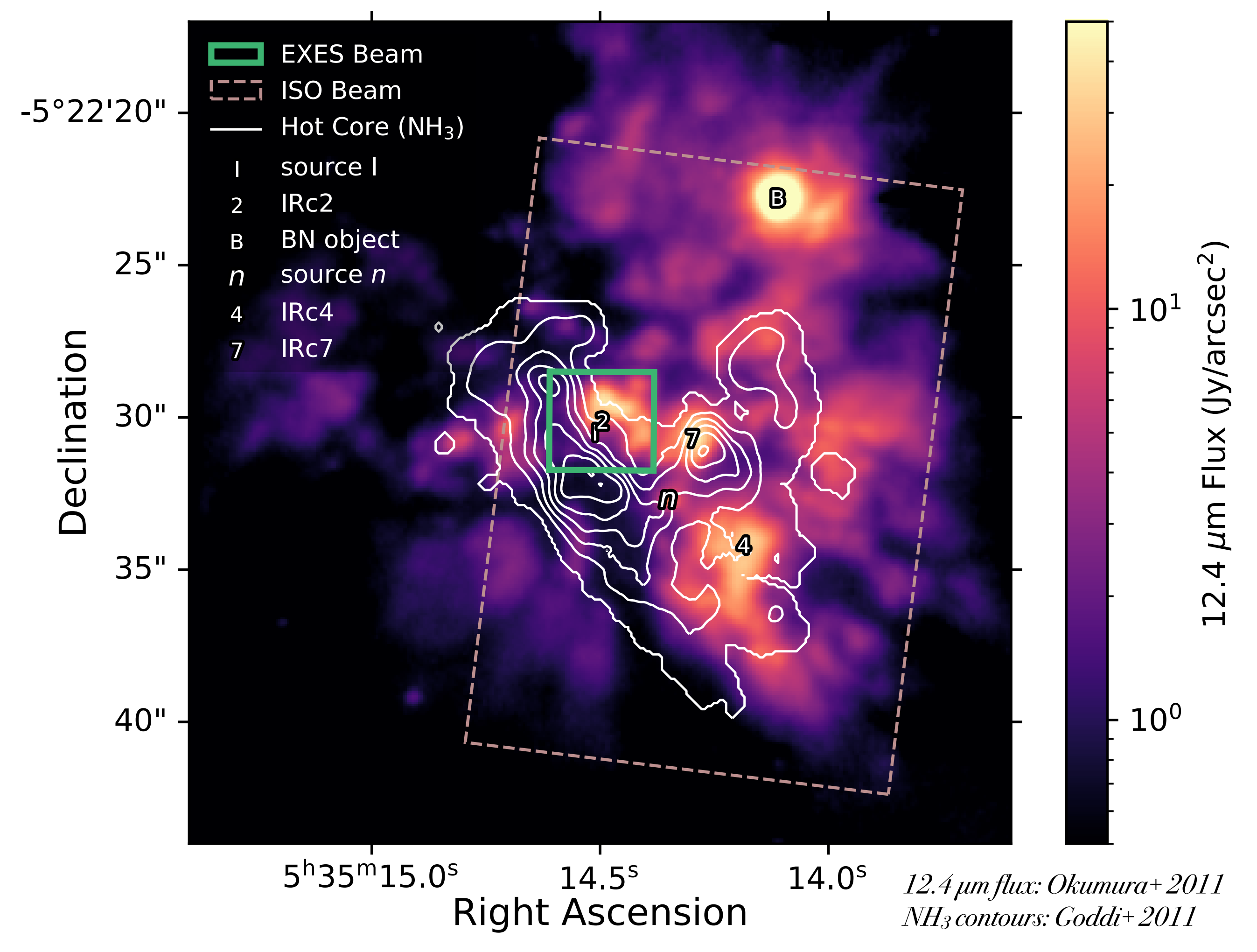
- Only the mid-infrared (MIR) can observe rovibrational transitions and molecules with no permanent dipole moment (e.g. C_2H_2 and CH_4)
- Radio to FIR captures molecules in cooler, outer regions of discs while the MIR to NIR covers the inner regions (Dullemond+ 2007, Barr+ 2020)
- MIR difficult to access because of atmospheric interference
- Past space telescopes *ISO* and *Spitzer*, and present *JWST* cannot resolve individual lines of hots cores in the MIR



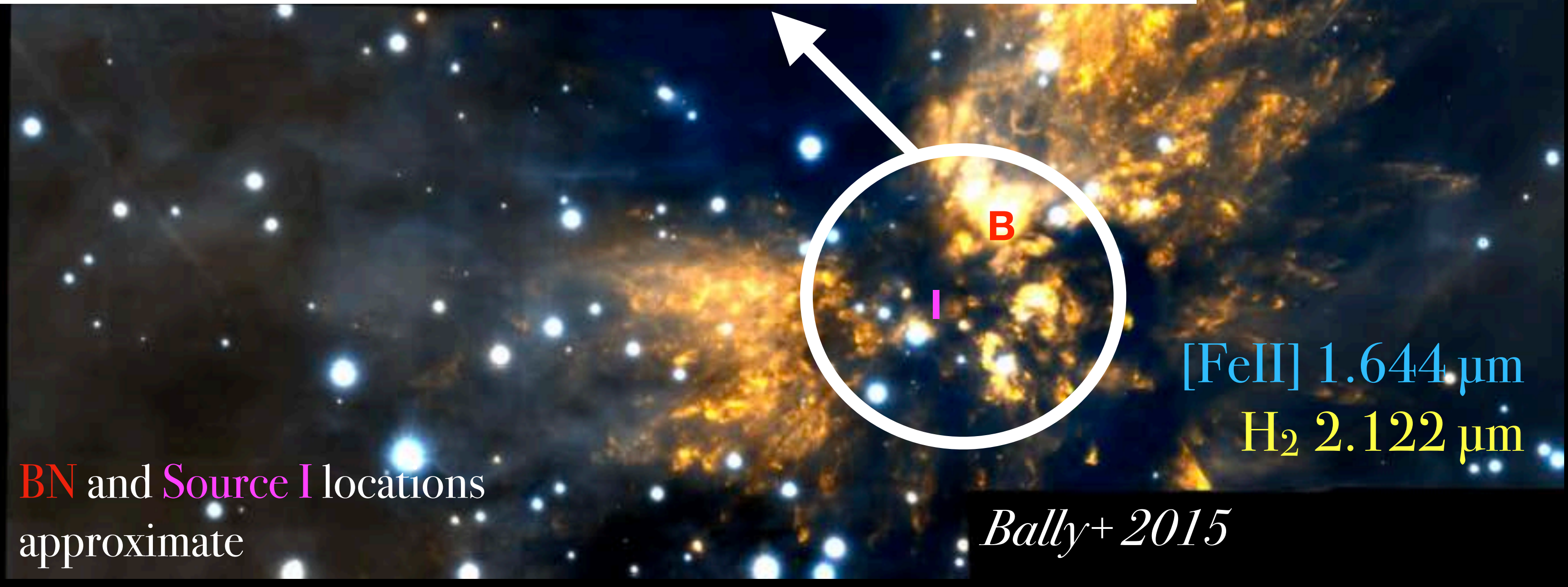
Dullemond+ 2007

Schematic SED with origin of wavelength regimes emitted spatially along a protoplanetary disc

Atypical: The Orion Hot Core



12.4 μm flux: Okumura+ 2011
NH_3 contours: Goddi+ 2011



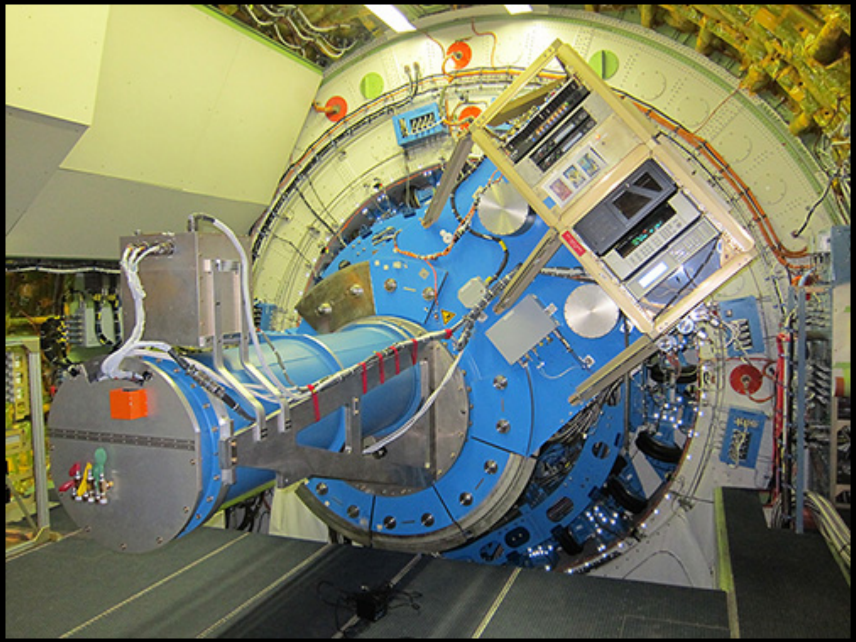
BN and Source I locations approximate

Bally+ 2015

- Orion BN/KL within closest and most studied massive star formation region
- Site of explosion ~500 years ago from multi-body encounter; pushed BN and source I apart (Bally+ 2015)
- Orion hot core was first hot molecular core discovered, via NH_3 emission (Ho+ 1979)
- Orion hot core: atypical, externally heated and has no embedded protostar
- The edge of the Orion hot core is illuminated in MIR by IRc2
- IRc2 is possibly scattered radiation from radio source I (Okumura+ 2011) or source *n* (Simpson+ 2006)
- IRc2 may be cavity in the Orion BN/KL nebula (Wynn-Williams+ 1984)

Orion IRc2 with SOFIA/EXES

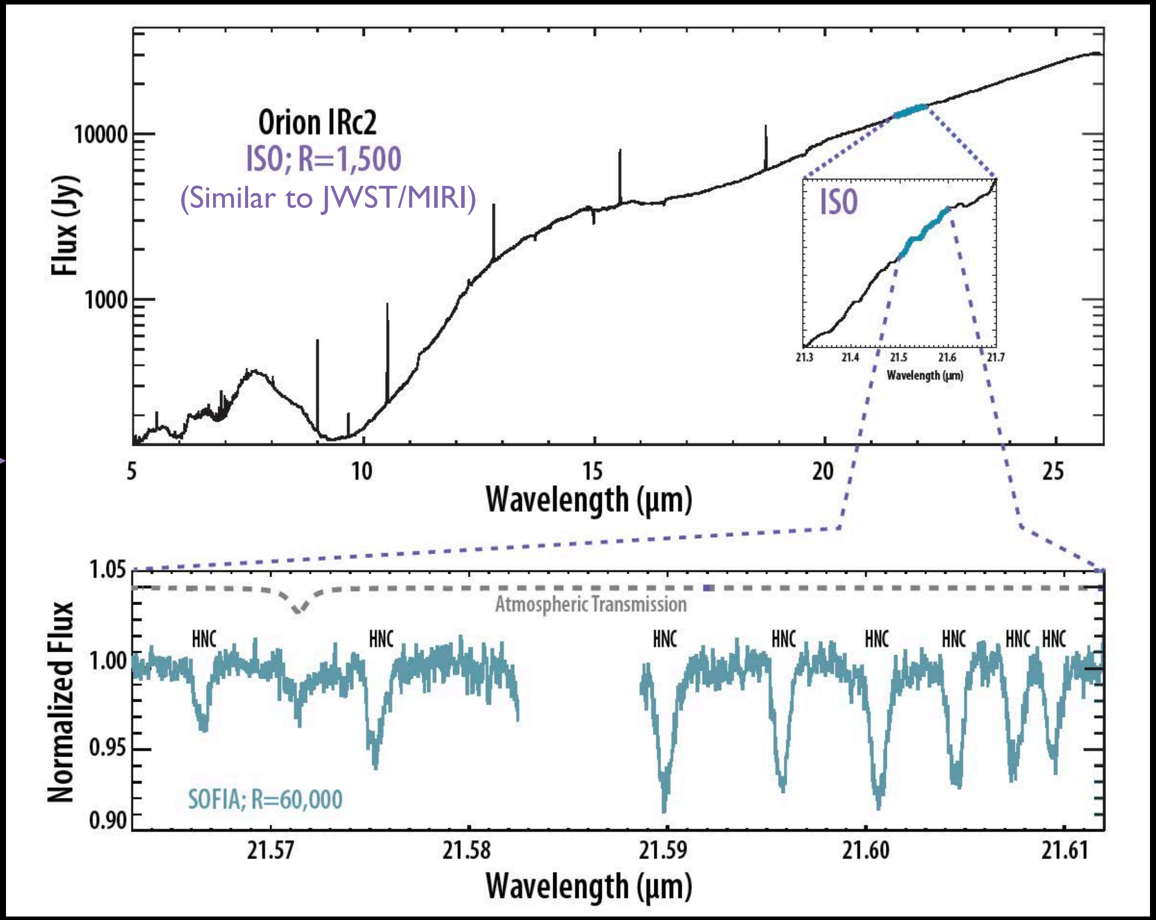
- We conducted an unbiased, MIR line survey at high resolution (R ~ 60,000) from 7.2 to 28.3 μm towards Orion IRc2
- For this tutorial, I will be using data from this sample as an example to demonstrate analysis techniques



Nickerson+ 2021, ApJ, 907, 51
Nickerson+ 2023, ApJ, 945, 26

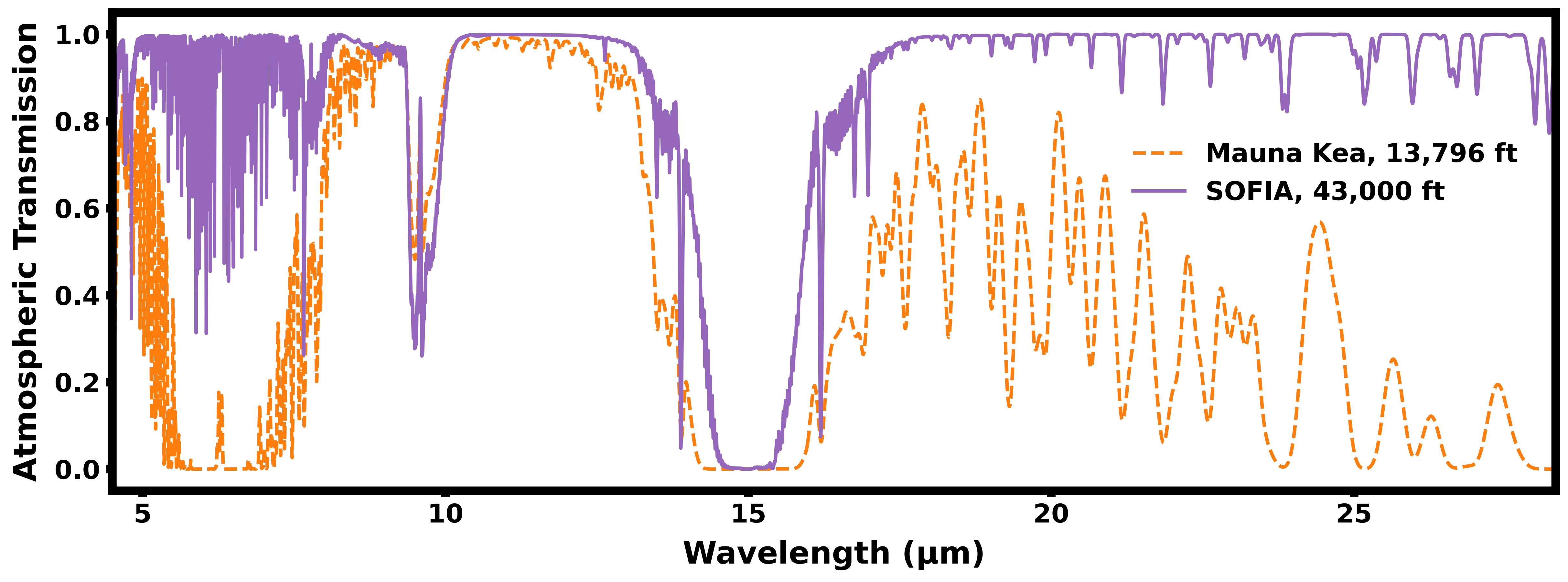
Compare resolution between MIR surveys towards Orion IRc2:

- Top: ISO/SWS, resolution ~1,500 (van Dishoeck et al. 1998), similar to JWST/MIRI
- Bottom: SOFIA/EXES, this survey, HNC absorption lines, resolution ~60,000
- With JWST/MIRI, these lines would be indiscernible from the continuum



NASA/SOFIA/M. Rose/N. Rangwala

Atmospheric Transmission



Comparison of atmospheric transmission across the EXES Range, between SOFIA and Mauna Kea

Importing EXES data

EXES spectra are stored in two different file types:

the orders are merged (MRD)

```
from astropy.io import fits
infits= fits.open(infile)
wavenumber  = infits[0].data[0,:]
flux  = infits[0].data[1,:]
error = infits[0].data[2,:]
atran = infits[0].data[3,:]
infits.close()
```

← *EXES data is stored in wavenumbers, cm^{-1}*

← *unormalized flux*

← *variance (not usually needed)*

← *atran model, untuned to specific observation*

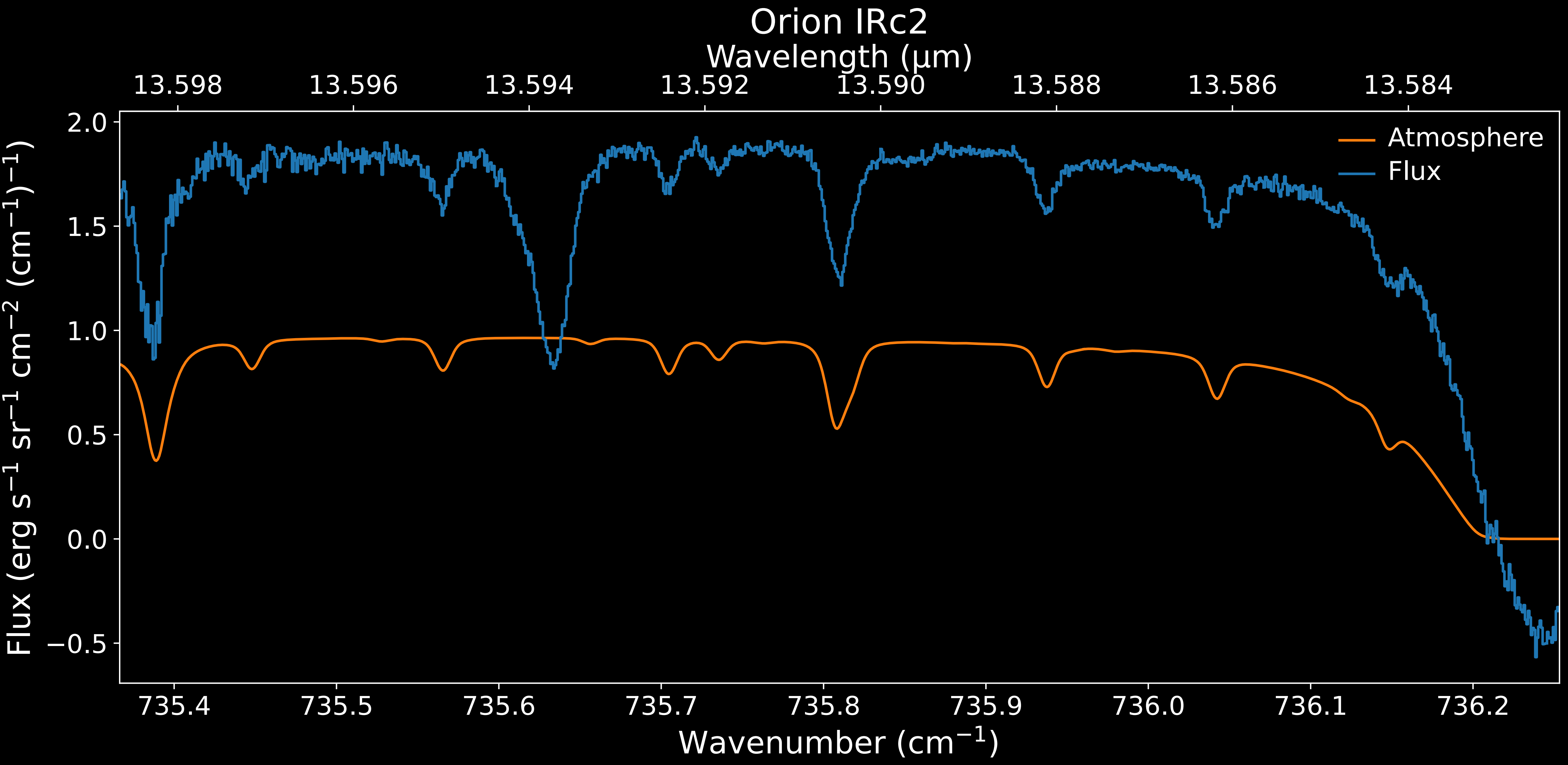
the orders are stored in separate arrays (CMB or SPC)

```
from astropy.io import fits
infits= fits.open(infile)
wavenumber  = infits[0].data[:,0,:]
flux  = infits[0].data[:,1,:]
error = infits[0].data[:,2,:]
atran = infits[0].data[:,3,:]
infits.close()
```

Order ↑ ↑ *Data*

(Examples given in python, but procedure similar in IDL or language of your choice)

Plotting EXES data



- If you plot the Flux vs. Wavenumber straight from the file, you will noticed that is not normalized
- The Atran (atmosphere) model found in the fits file is normalized, but is not tuned to this data

Normalization Methods

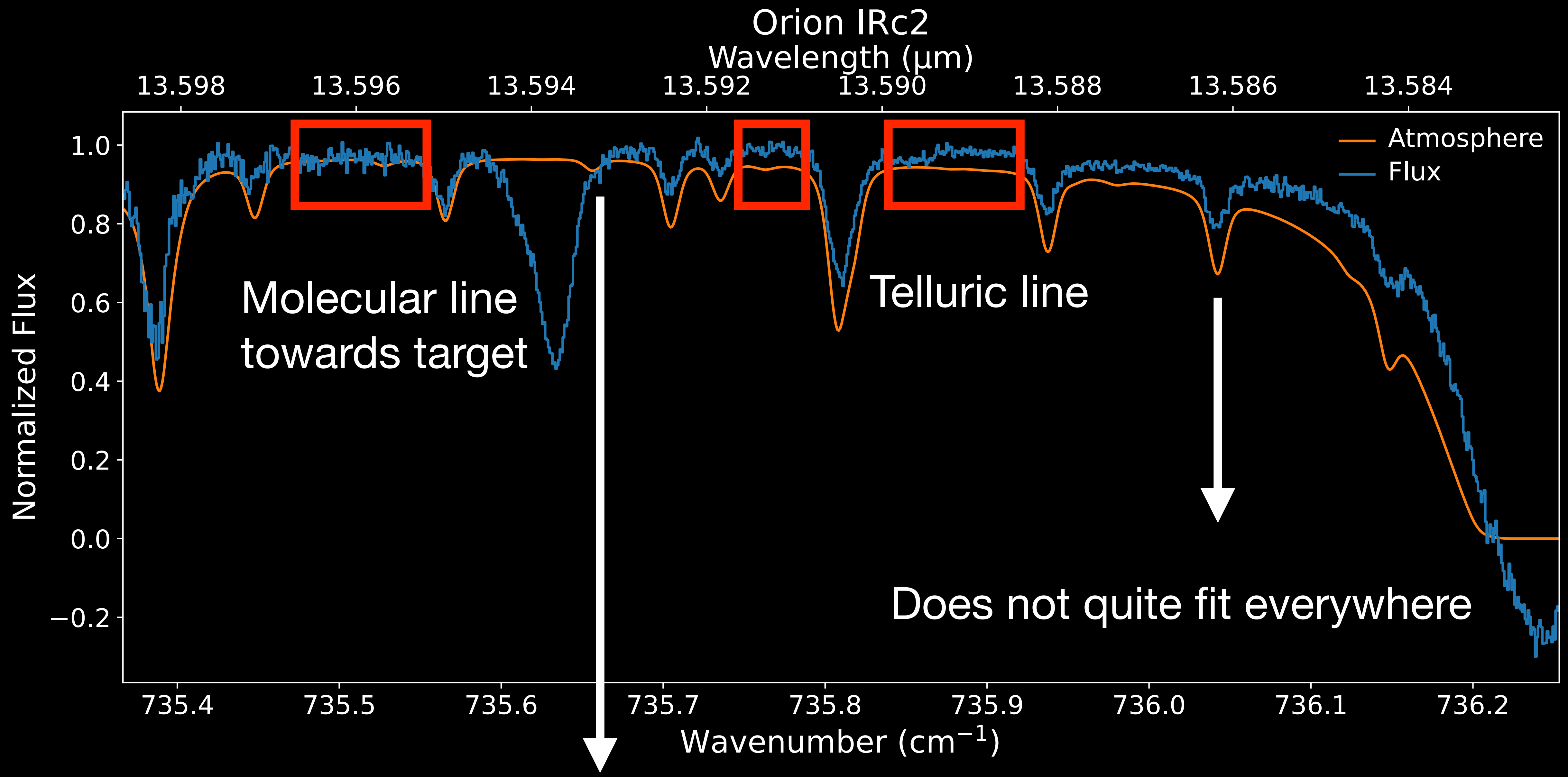
1. Quick
    - Find a portion where both the flux and atran model are flat, and divide that flux
    - Only good for quick visualization, or for regions without much atmosphere, where flux does not overlap with telluric features
2. Divide by Calibrator (an observation towards an object without spectral features, rarely taken with EXES
3. Divide by Atmospheric Model
    - ATRAN model (https://atran.arc.nasa.gov/cgi-bin/atran/atran.cgi) by entering in observation parameters in fits header
    - NASA Planetary Spectrum Generator (https://psg.gsfc.nasa.gov/)
4. If baseline is uneven, need to divide by polynomial

Covered in detail by Curtis DeWitt in the 2022 SOFIA School

He also has a great introduction into what EXES is and the available data products

Quick Normalization

- Identify regions where the flux is flat and the atran model is close to 1
- Divide the entire flux by the mean flux of these regions
- Can easily see which lines are telluric features, and which are towards the target



Molecular line towards target

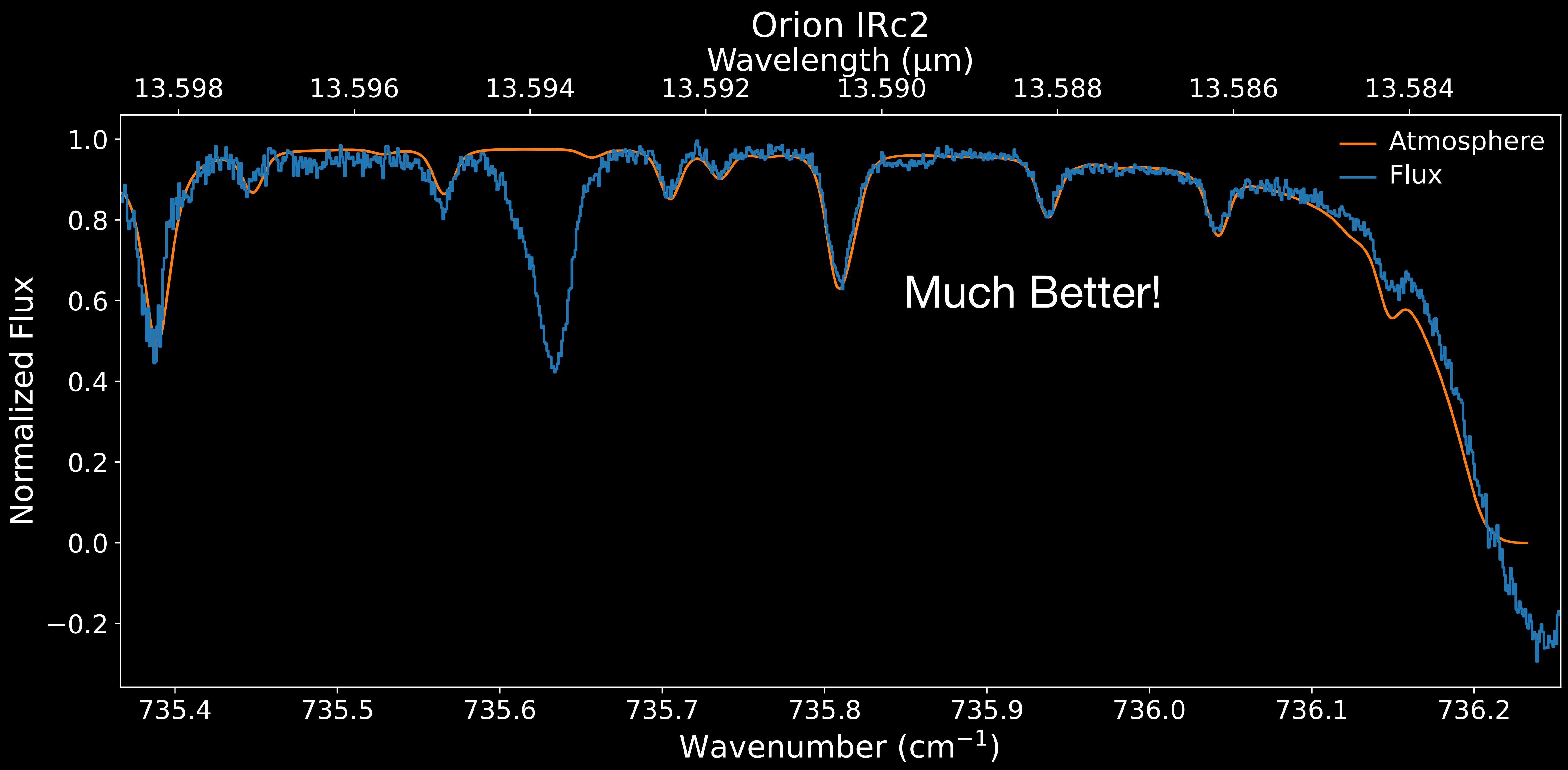
Telluric line

Does not quite fit everywhere

Atmosphere overlap with line, will need better normalization method to divide out when analyzing

Sarah Nickerson — EXES Analysis — nickerson@baeri.org

Normalization by Tuned ATRAN Model

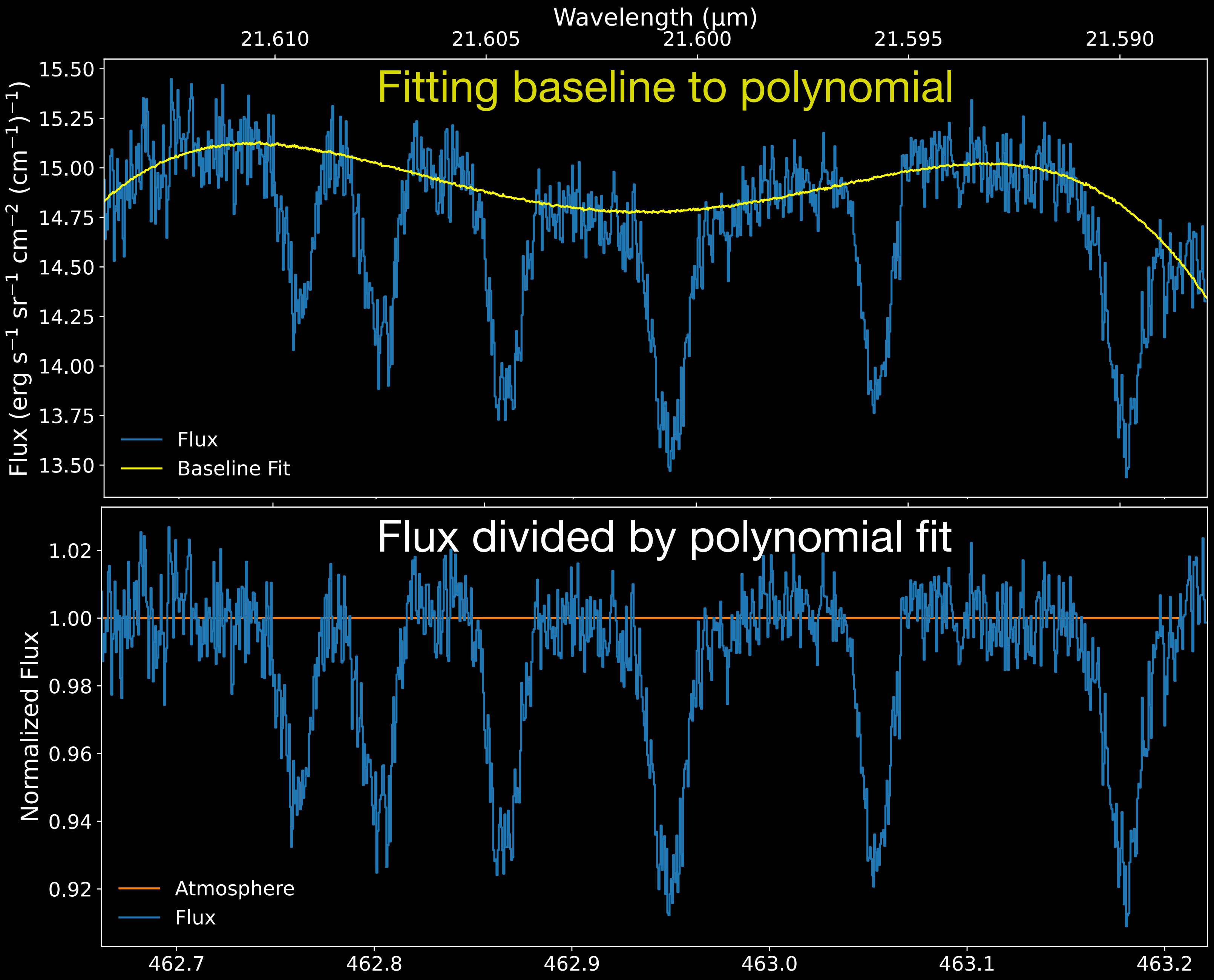


- Download unsmoothed ATRAN model (https://atran.arc.nasa.gov/cgi-bin/atran/atran.cgi) by entering in observation parameters (altitude, latitude, zenith) from fits header; R=0 for no smoothing
- Interpolate to move ATRAN model onto same wavenumber grid as normalized flux
- Find normalization (n) and smoothing parameter (σ) to best fit ATRAN model to observed telluric lines
- scipy minimize function to find n and σ that minimize the difference between normalized flux and smoothed ATRAN model (with scipy gaussian_filter1d):

  minimize(gaussian_filter1d(atran,σ) - flux/n)

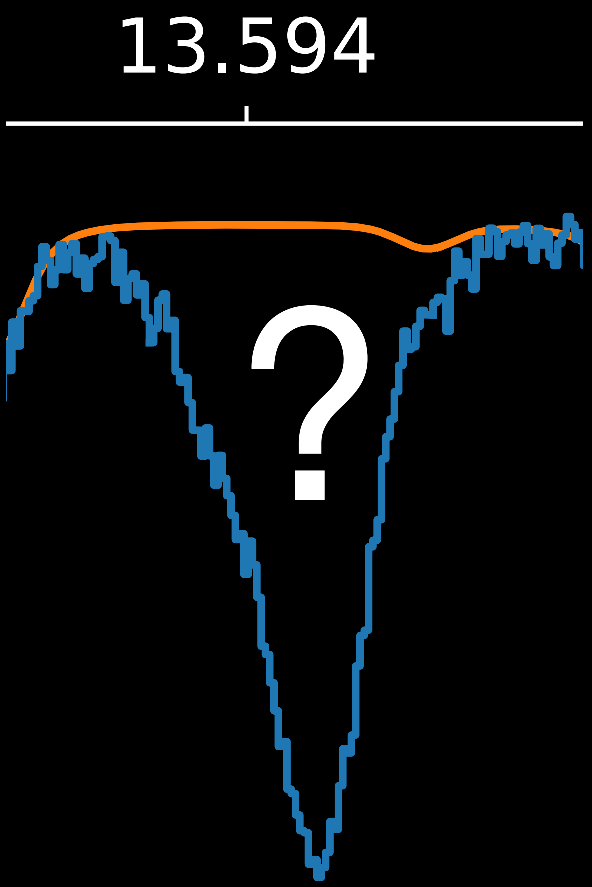
Sarah Nickerson — EXES Analysis — nickerson@baeri.org

Uneven Baseline



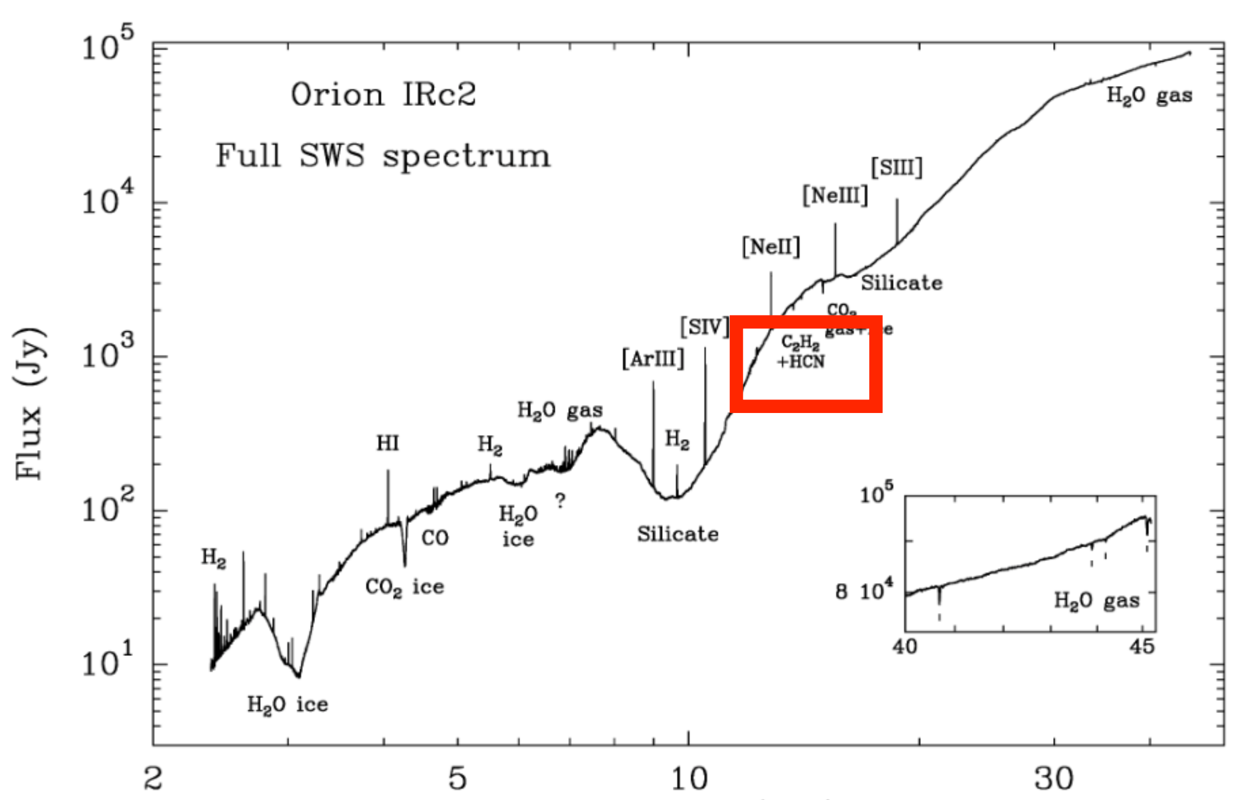
- Sometimes an order, or setting, may have an uneven baseline
- This particularly happens for the settings around 300 to 400 cm^{-1}
- In this situation, you need to fit a polynomial to the baseline and then divide to straighten it out

What's that Line?



- Search the literature for common MIR species in hot cores/object of interest
- Older ISO papers
- High resolution MIR hot core papers with many species:
  - NGC 7538 IRS 1, Knez+ 2009 (TEXES, simialr to EXES but ground-based)
  - AFGL 2591, 2136 Barr+ 2020 (EXES)
  - Orion IRc2, Nickerson+ 2023 (EXES)

Previously, Orion IRc2 was observed in the MIR with ISO…



van Dishoeck+ 1998

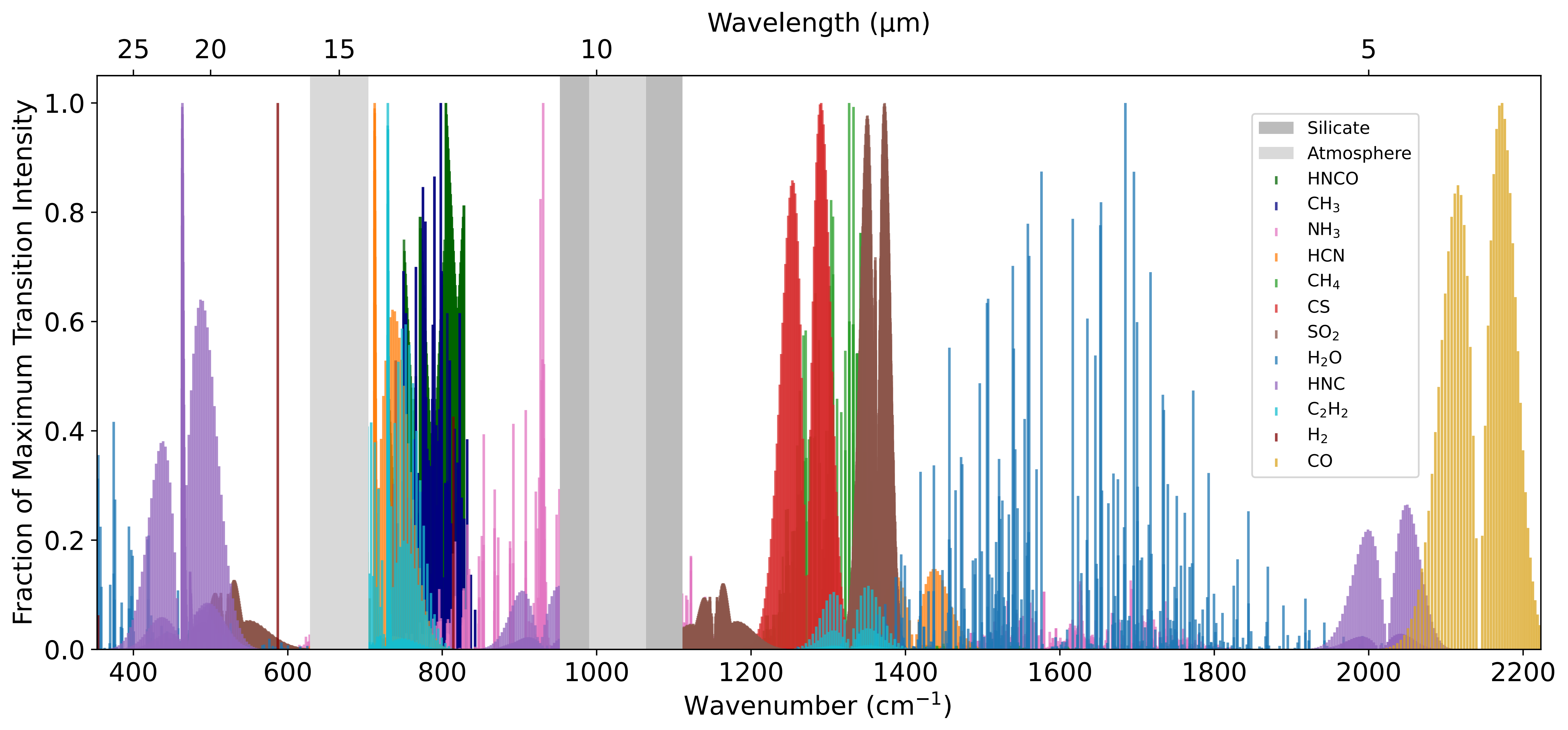
…C_2H_2 and HCN detected around 13 μm!

Databases with MIR Lines

To identify molecules, we need lines lists of transitions for each molecule. These databases also contain quantum parameters we will need for later analysis.

- HITRAN https://hitran.org/
  - Molecules relevant to the Earth's atmosphere
  - Most user friendly database and contains almost all molecules observed towards hot cores with EXES
  - Can be accessed through python package, hapi, to easily integrate into one's own code
- GEISA https://geisa.aeris-data.fr/
  - Like HITRAN, molecules relevant to the Earth's atmosphere
  - Molecules mostly overlap with HITRAN, but contains a few HITRAN does not have (e.g. HNC)
- ExoMol
  - Molecules relevant to exoplanet atmospheres
  - Contains molecules not found in Earth's atmosphere (e.g. SiO)

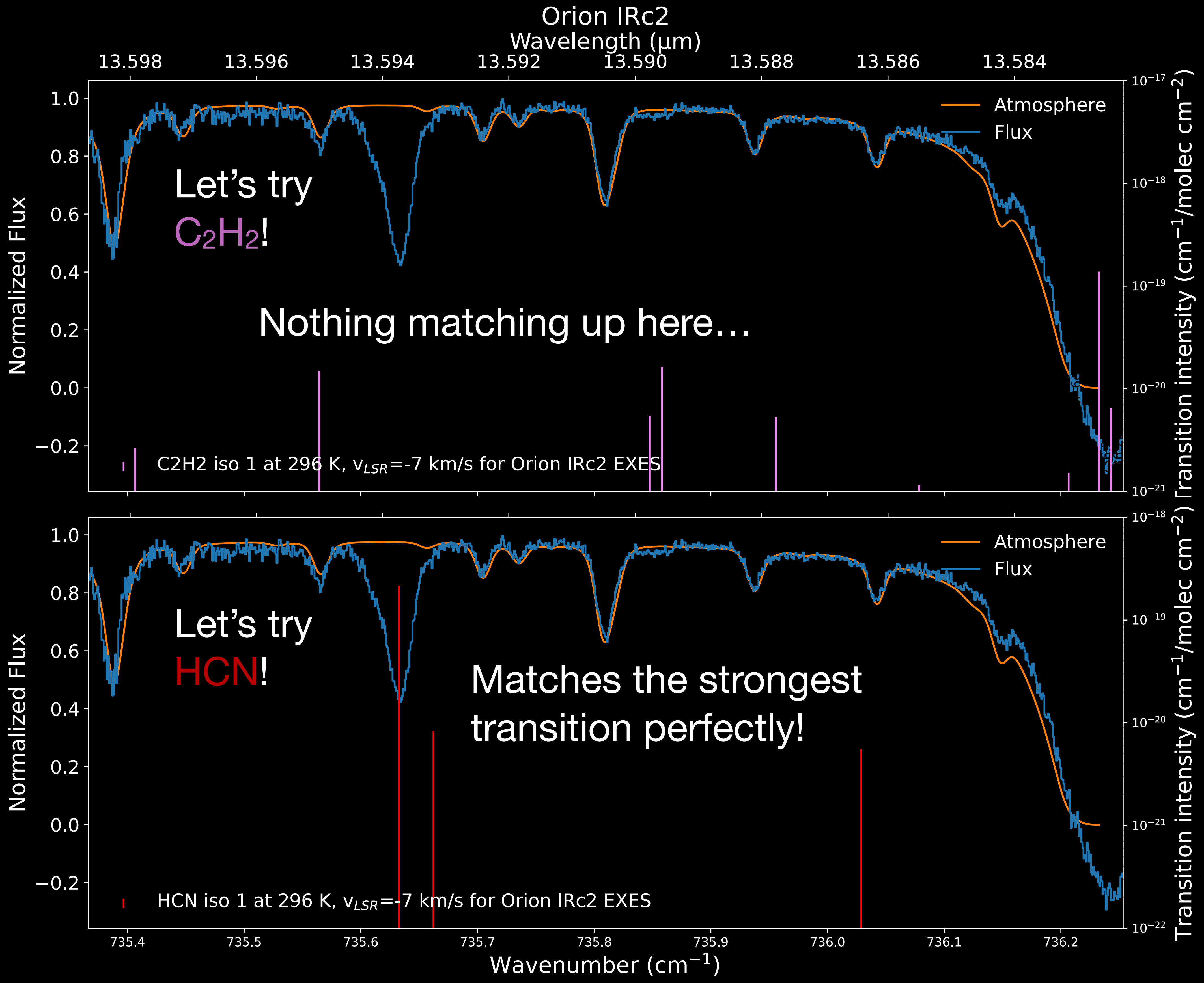
Common Hot Core Molecules



- Relative transition strengths of molecules within the EXES and TEXES range, commonly found towards hot cores in the MIR
- Greyed out regions unobservable due to atmosphere or deep silicate features in hot cores (approximate)
- Line lists used: GEISA (HNC), Knez+ 2009 (HNCO, CH_3), HITRAN (rest)

Line ID

- Download line list from database (in this case HITRAN)
- Shift database's rest wavelength according to a best guess v_{LSR}
- Plot and see if strongest transitions match up with unidentified lines
- Look through entire setting to make sure all strong transitions match



Let's try C_2H_2!

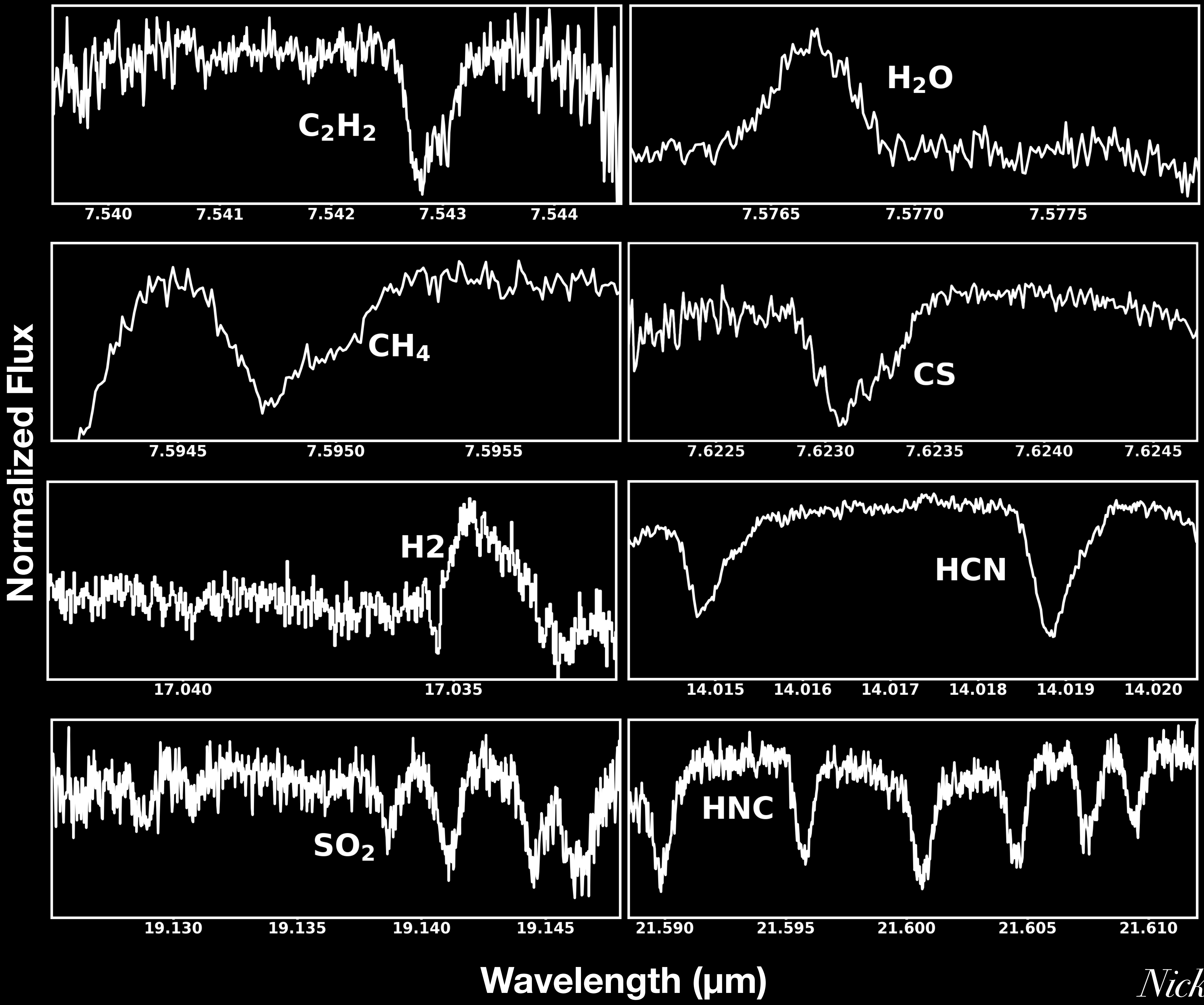
Nothing matching up here…

Let's try HCN!

Matches the strongest transition perfectly!

You can guess the v_{LSR} by using a previous high resolution NIR/MIR publication for the same object (here Rangwala+ 2018); second choice is publications from longer wavelength regimes, but those do not necessarily match MIR velocities

What we found towards Orion IRc2



- Survey 7.2 to 28.3 µm
- Over 350 unique features
- Molecular species identified:
  - absorption: HCN, HNC, C_2H_2, H_2O, NH_3, CH_4, SO_2, CS, $H^{13}CN$ and $^{13}CCH_2$
  - emission: H_2, H_2O and SiO
- Detect two velocity components in some absorption species (e.g. C_2H_2, CH_4, and HCN) and H_2 emission

Nickerson+ 2021, ApJ, 907, 51
Nickerson+ 2023, ApJ, 945, 26

Gaussian Fits

- Once we have normalized our flux and identified the lines, we fit our lines to Gaussians to measure important quantities
- If the line overlaps with a little atmosphere, divide out the tuned ATRAN or PSG atmospheric model
- Select the region of flux for fitting to the Gaussian; the amount of baseline is important to ensuring a good fit
- Following Indriolo+ 2015, fit the flux to: $I = I_0 e^{-\tau_0 G}$ where,

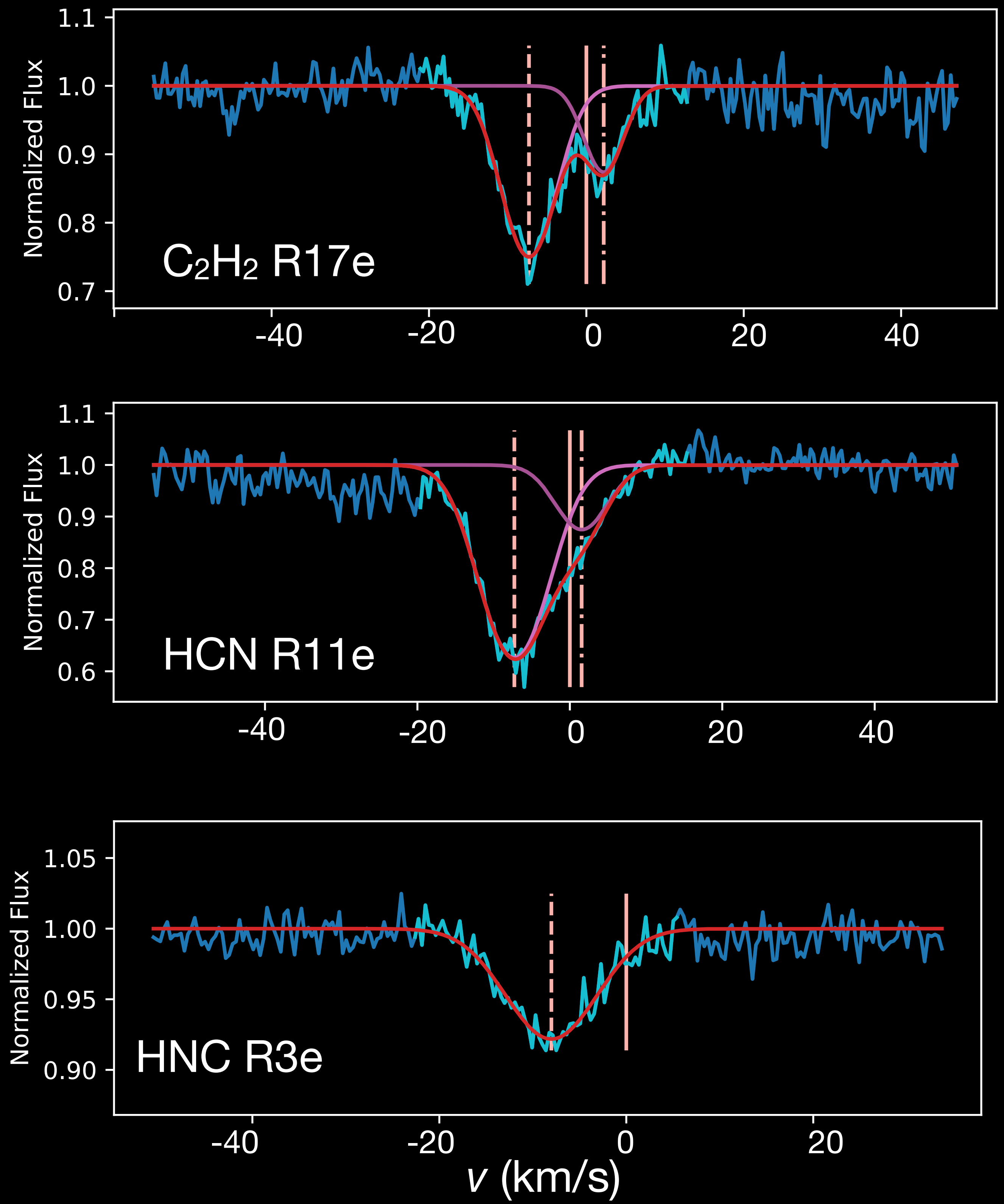
$$G = \exp\left[-\frac{(v - v_{\mathrm{LSR}})^2}{2\sigma_v^2}\right]$$

and calculate column density as integral under $\tau_0 G$:

$$dN_l/dv = \frac{g_l}{g_u}\frac{8\pi}{A\lambda^3}\tau_0 G$$

- If seeing more than one velocity component, then fit to the same number of Gaussians!

I_0: baseline
τ_0: optical depth
v: LSR velocity
v_{LSR}: LSR velocity at line centre
σ_v: velocity dispersion
N_l: lower state column density
$g_{l/u}$: lower/upper statistical weight
λ: rest wavelength
A: Einstein coefficient



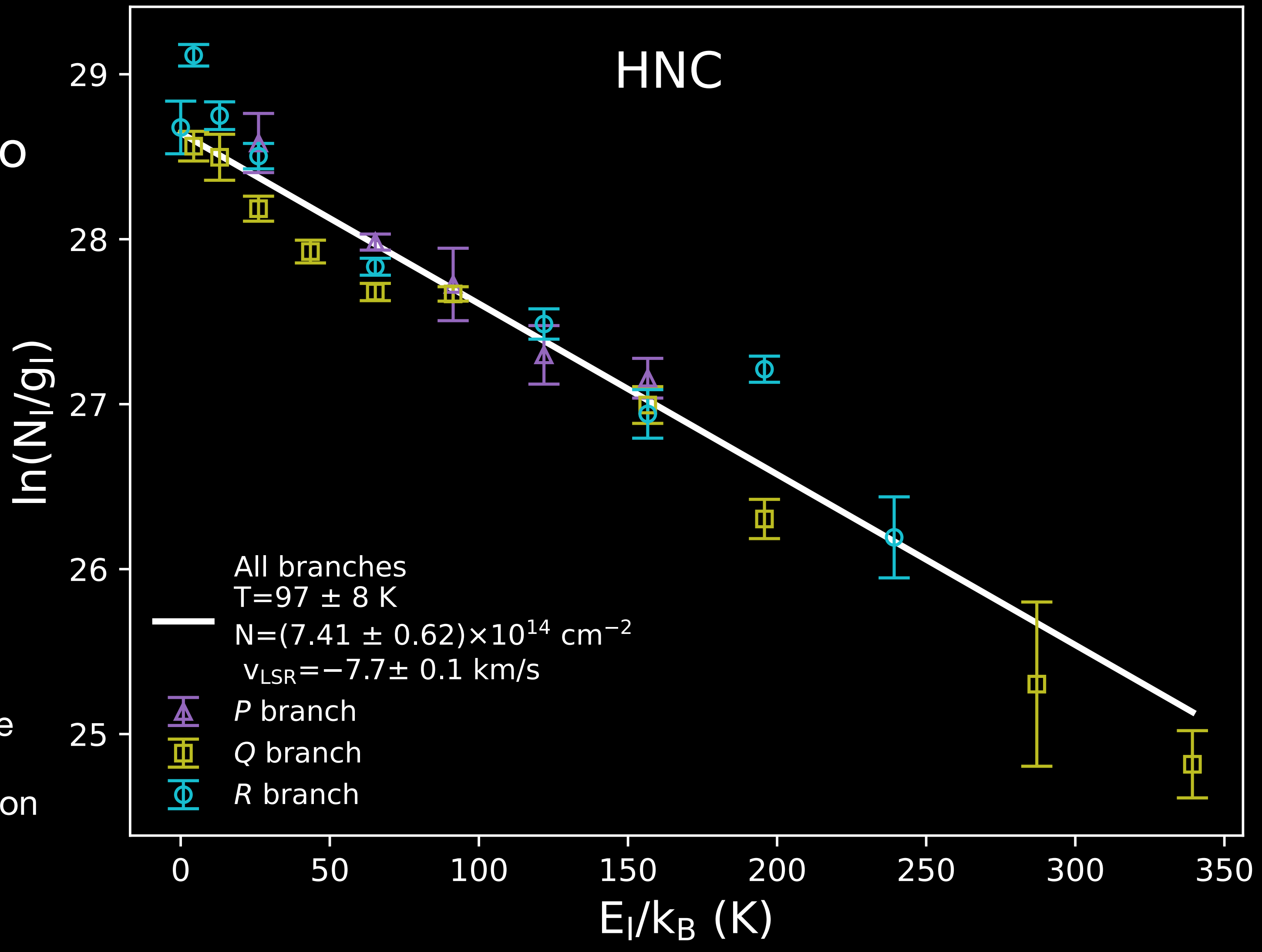
Rotation Diagram

- Assuming local thermodynamic equilibrium (LTE), can fit the column densities and energies of all transitions to Boltzmann's equation (Goldsmith & Langer 1999) to obtain overall column density and excitation temperature of species:

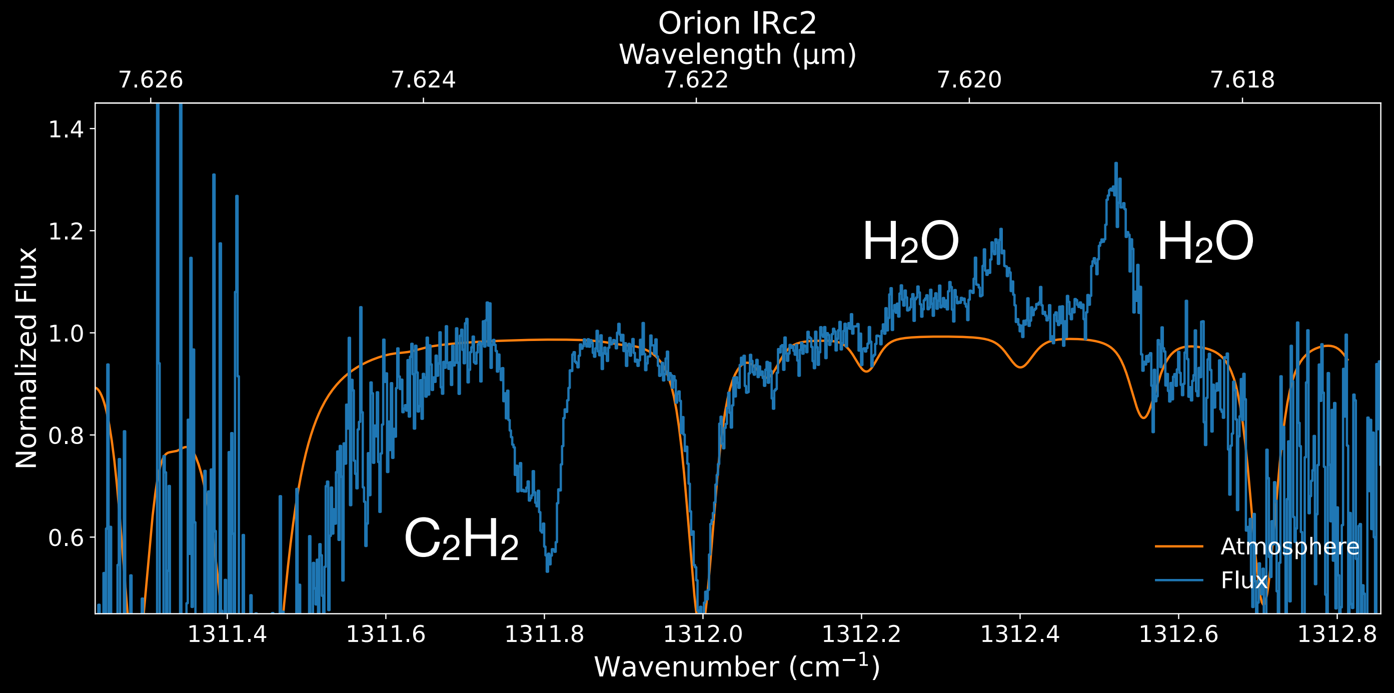
$$\ln\frac{N_j}{g_j} = \ln\frac{N}{Q_R(T_{\mathrm{ex}})} - \frac{E_l}{kT_{\mathrm{ex}}}$$

N_j: transition column density
g_j: transition lower statistical weight
N: total column density

T_{ex}: excitation temperature
$Q_R(T_{ex})$: partition function
E_l: lower energy of transition
k: Boltzmann constant



Emission Lines



- Gaussian fitting is similar to absorption lines, though the equations are slightly different (see Mangum & Shirley 2015, Nickerson+ 2023)

$$S_\nu(v) = B_\nu + S_{\nu 0}\ \exp(\frac{-(v - v_{\mathrm{LSR}})^2}{2\sigma_v^2})$$

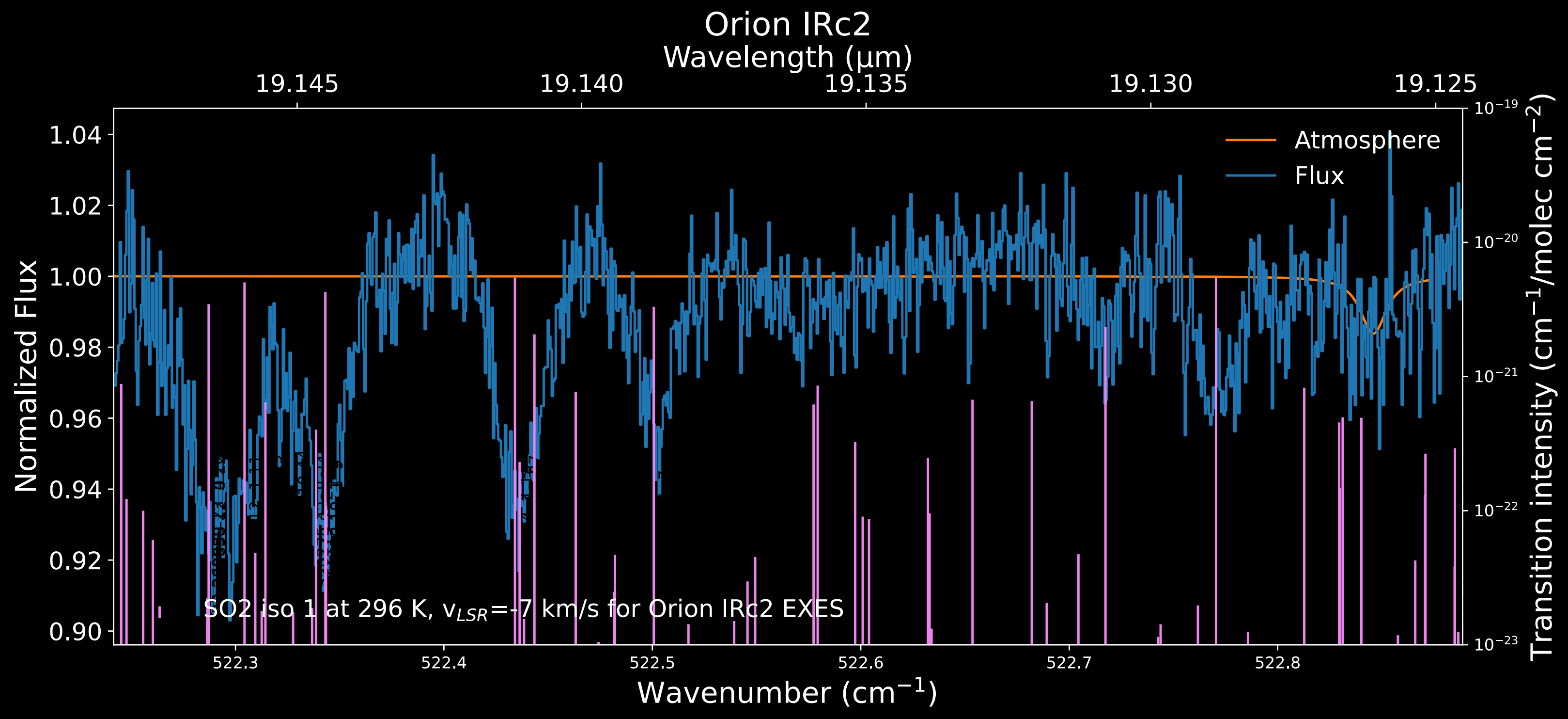
$$N_u = \frac{4\pi\sqrt{2\pi}\, S_{\mathrm{Ja}}\, S_{\nu 0}\, \sigma_\nu}{hcA}$$

- Rotation diagram analysis is the same as previous slide, replacing lower values with upper

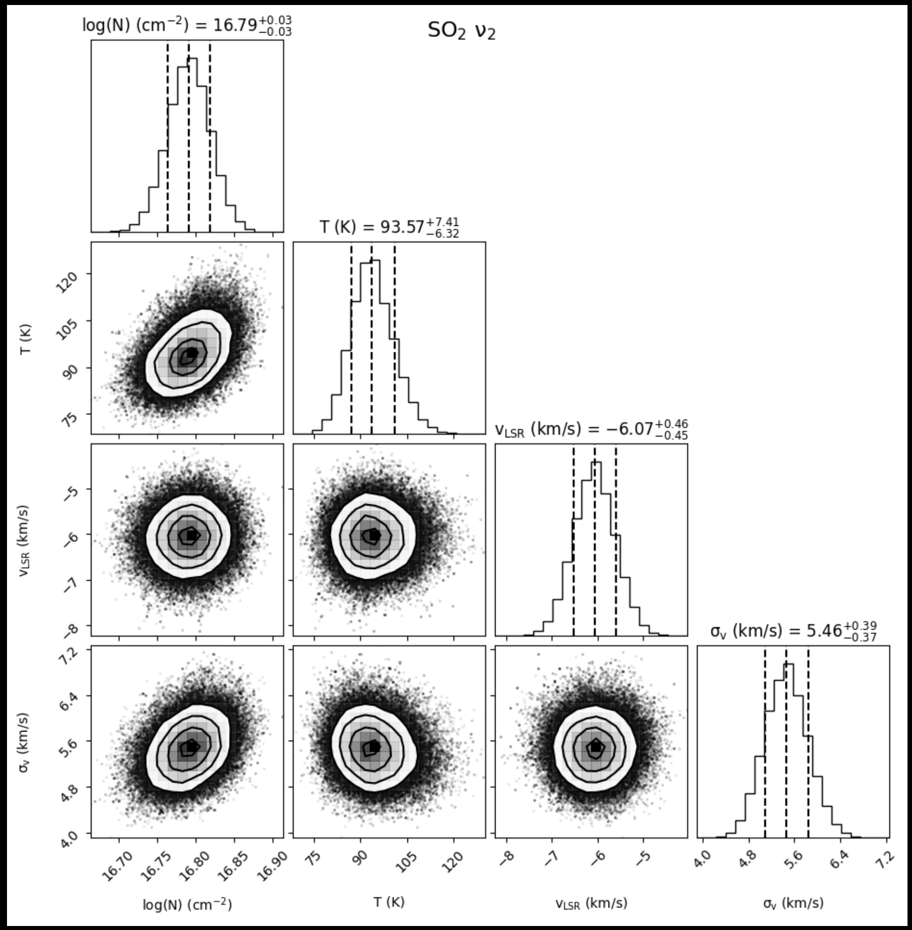
- The challenge is finding the conversion factor (S_{JA}) to convert the EXES flux into an absolute flux in units of Jy
- If there was a calibrator observed (e.g. Vega) on the same night, this may be useful
- May also find ISO data (on IRSA) taken of same target to be useful, though keep in mind ISO has a much larger beam than EXES and will high a higher flux is the target is in a crowded region
- Find other clues in literature about target. MIR observations with other instruments at simialr wavelengths (e.g. SUBARU/COMICS) may have absolute flux measurements

B_ν: the continuum level
$S_{\nu 0}$: amplitude
v: LSR velocity
v_{LSR}: LSR velocity at line centre
σ_ν: velocity dispersion
N_u: upper state column density
S_{JA}: conversion factor
A: Einstein coefficient
h: Planck constant
c: speed of light

What if there are too many lines to fit?

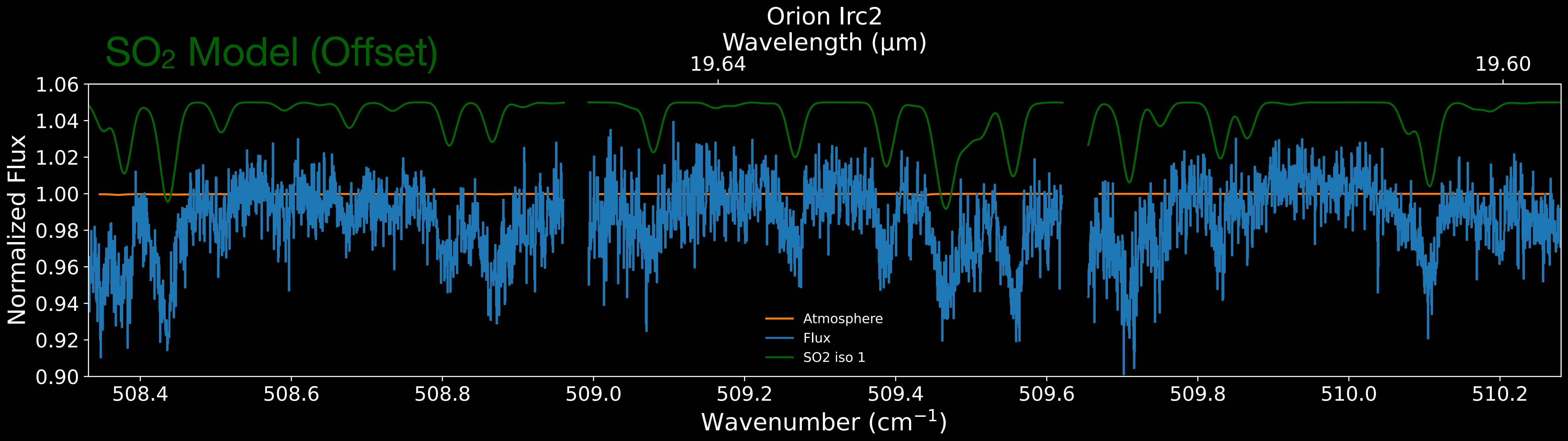


- SO_2 transitions too numerous and close together to fit individual transitions to Gaussians
- Solution: work backwards from the Boltzman equation, treating N, T_{ex}, v_{LSR}, and σ_ν as inputs to generate simulated spectra



- Use an optimization function to find the input parameters that generates the simulated spectra that most closely fits the EXES flux
- emcee (Foreman-Mackey+ 2013) is a python implementation of a Markov chain Monte Carlo ensemble sampler that is very popular in astronomy for these optomization problems

Here is the result: the best fit of the simulated SO_2 spectra the EXES flux, as computed by emcee



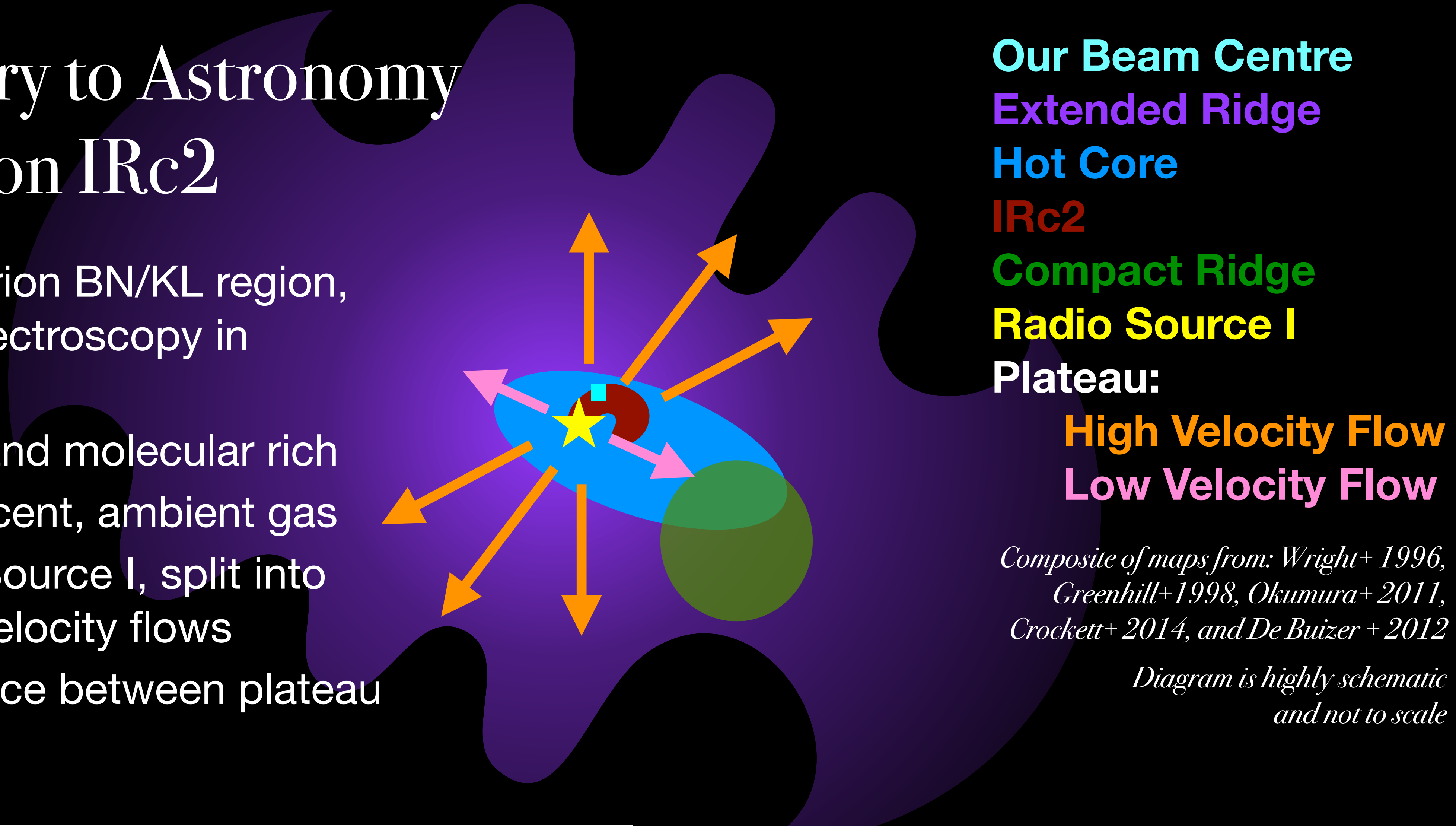
The Big Picture:

- Great! We've gathered all these column densities, temperatures, and velocities for our molecules, but what does it all mean?
- We can use what we learned about the chemistry to shed light on the sturcture of these regions
- The chemistry links back to astronomy…

Linking Chemistry to Astronomy Example #1: Orion IRc2

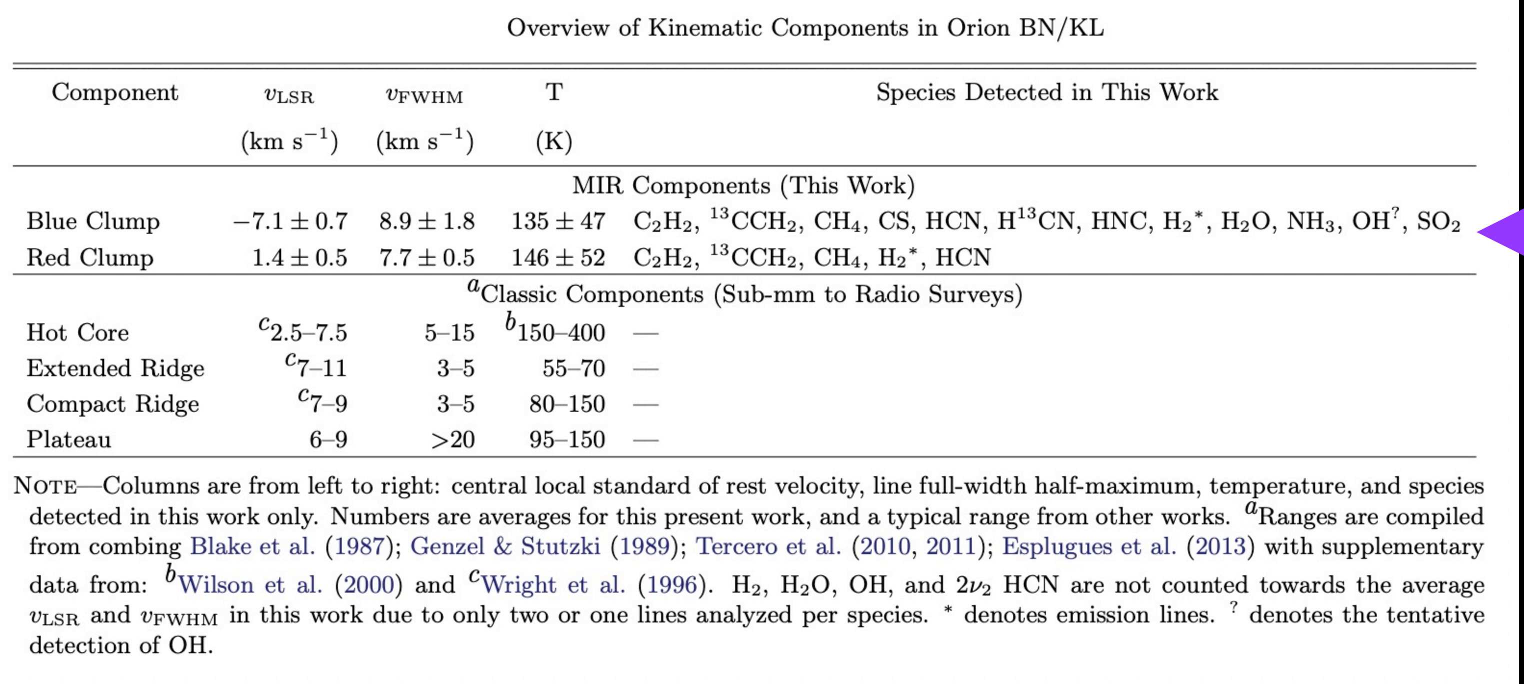
Classic components in Orion BN/KL region, from sub-mm to radio spectroscopy in emission:

- Hot Core: hot, dense, and molecular rich
- Extended Ridge: quiescent, ambient gas
- Plateau: outflow from Source I, split into low velocity and high velocity flows
- Compact Ridge: interface between plateau and extended ridge



Composite of maps from: Wright+ 1996, Greenhill+1998, Okumura+ 2011, Crockett+ 2014, and De Buizer + 2012

Diagram is highly schematic and not to scale



Overview of Kinematic Components in Orion BN/KL

Component	v_{LSR} (km s^{-1})	v_{FWHM} (km s^{-1})	T (K)	Species Detected in This Work
MIR Components (This Work)				
Blue Clump	-7.1 ± 0.7	8.9 ± 1.8	135 ± 47	C_2H_2, $^{13}CCH_2$, CH_4, CS, HCN, $H^{13}CN$, HNC, $H_2{}^*$, H_2O, NH_3, $OH^?$, SO_2
Red Clump	1.4 ± 0.5	7.7 ± 0.5	146 ± 52	C_2H_2, $^{13}CCH_2$, CH_4, $H_2{}^*$, HCN
[a]Classic Components (Sub-mm to Radio Surveys)				
Hot Core	[c]2.5–7.5	5–15	[b]150–400	—
Extended Ridge	[c]7–11	3–5	55–70	—
Compact Ridge	[c]7–9	3–5	80–150	—
Plateau	6–9	>20	95–150	—

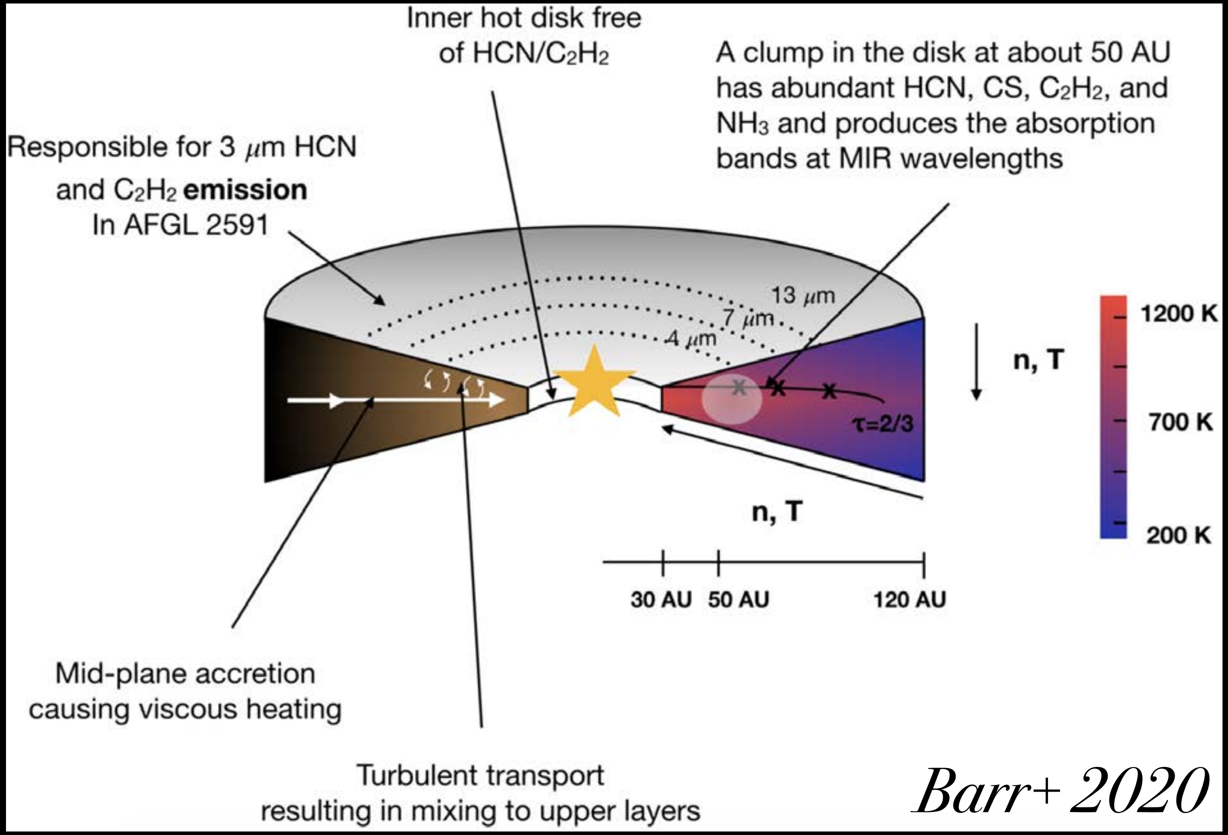
NOTE—Columns are from left to right: central local standard of rest velocity, line full-width half-maximum, temperature, and species detected in this work only. Numbers are averages for this present work, and a typical range from other works. [a]Ranges are compiled from combing Blake et al. (1987); Genzel & Stutzki (1989); Tercero et al. (2010, 2011); Esplugues et al. (2013) with supplementary data from: [b]Wilson et al. (2000) and [c]Wright et al. (1996). H_2, H_2O, OH, and $2\nu_2$ HCN are not counted towards the average v_{LSR} and v_{FWHM} in this work due to only two or one lines analyzed per species. * denotes emission lines. $^?$ denotes the tentative detection of OH.

- The two kinematic component in our survey are detected in absorption lines and H_2 emission
- Have distinct properties from the classic components and are only dtected in the MIR
- Through chemistry, our survey and the ones before it have decyphered the many compoents in the region

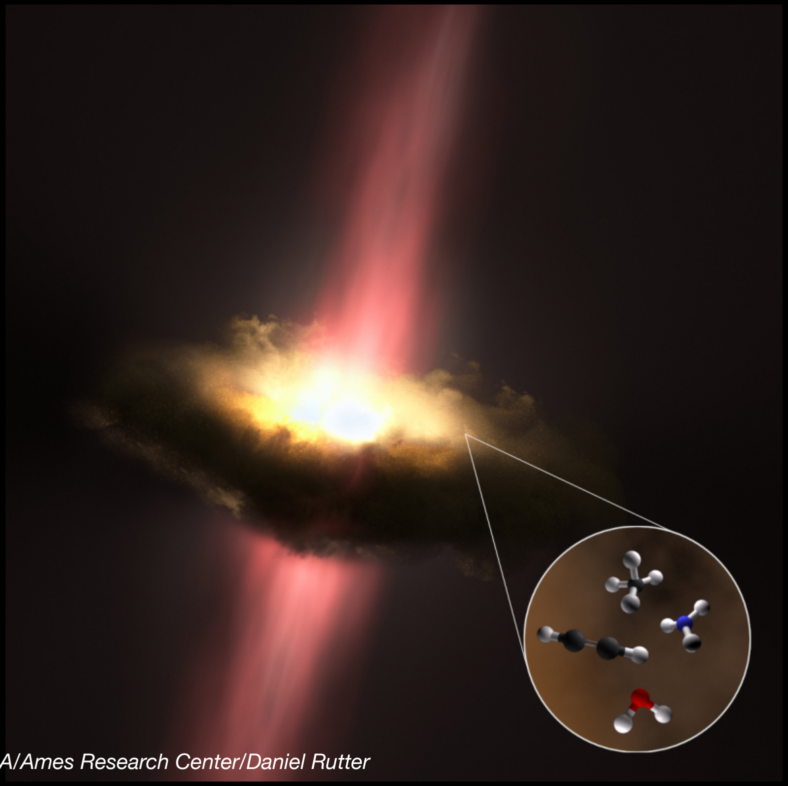
Nickerson+ 2023, ApJ, 945, 26

Linking Chemistry to Astronomy Example #2: AFGLs 2136 and 2591

- Both AFGL 2136 and 2591 are traditional hot cores in which the gas is surrounding a massive protostar
- Barr+ 2020 high resolution survey of these two hot cores, from 4 to 13 μm
- Combined data from EXES, IRTF/TEXES (similar to EXES, but ground-based), and IRTF/iSHELL (near infrared, shorter wavelengths than EXES)



Barr+ 2020



NASA/Ames Research Center/Daniel Rutter

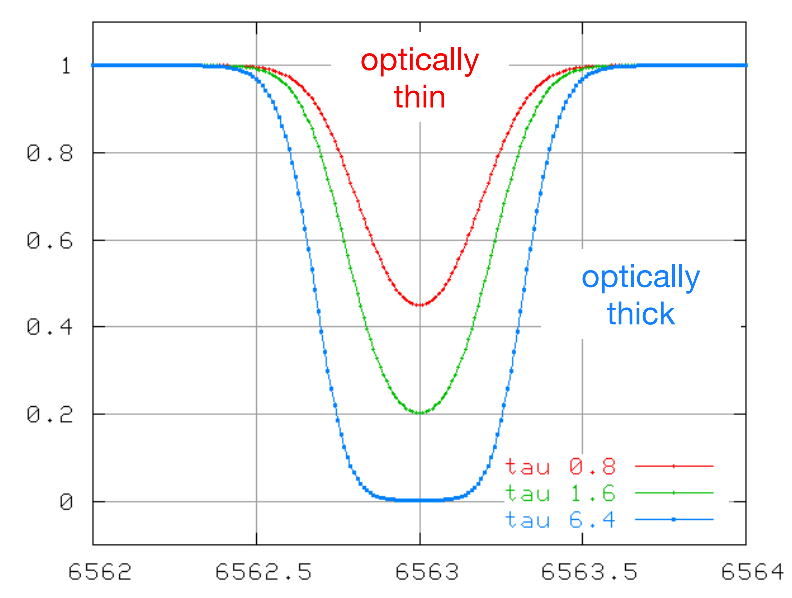
- Absorption lines: CO, HCN, C_2H_2, NH_3, and CS with temperatures on order of 600 K; emission lines: HCN and C_2H_2
- Proposed that the absorption lines trace a disc around the protostar with an outwardly decreasing temperature profile
- Shorter wavelengths trace inner regions of the disc, longer trace outer; outer envelope traced by sub-mm emission lines
- As with Orion IRc2, the MIR reveals a region unseen by longer wavelengths

Barr+ 2020, ApJ, 900, 104

A Note on More Complex Situations

Optically Thick Lines

- Signs of optically thick absorption lines:
  - Flat bottoms
  - Lines saturate at a certain value
  - Large scatter in rotation diagram
- Solutions:
  - Stellar atmosphere model to find abundances (Mihalas 1978, Barr+ 2020)
  - Curve of growth analysis (Draine 2011, Tielens 2021, Barr+ 2022)



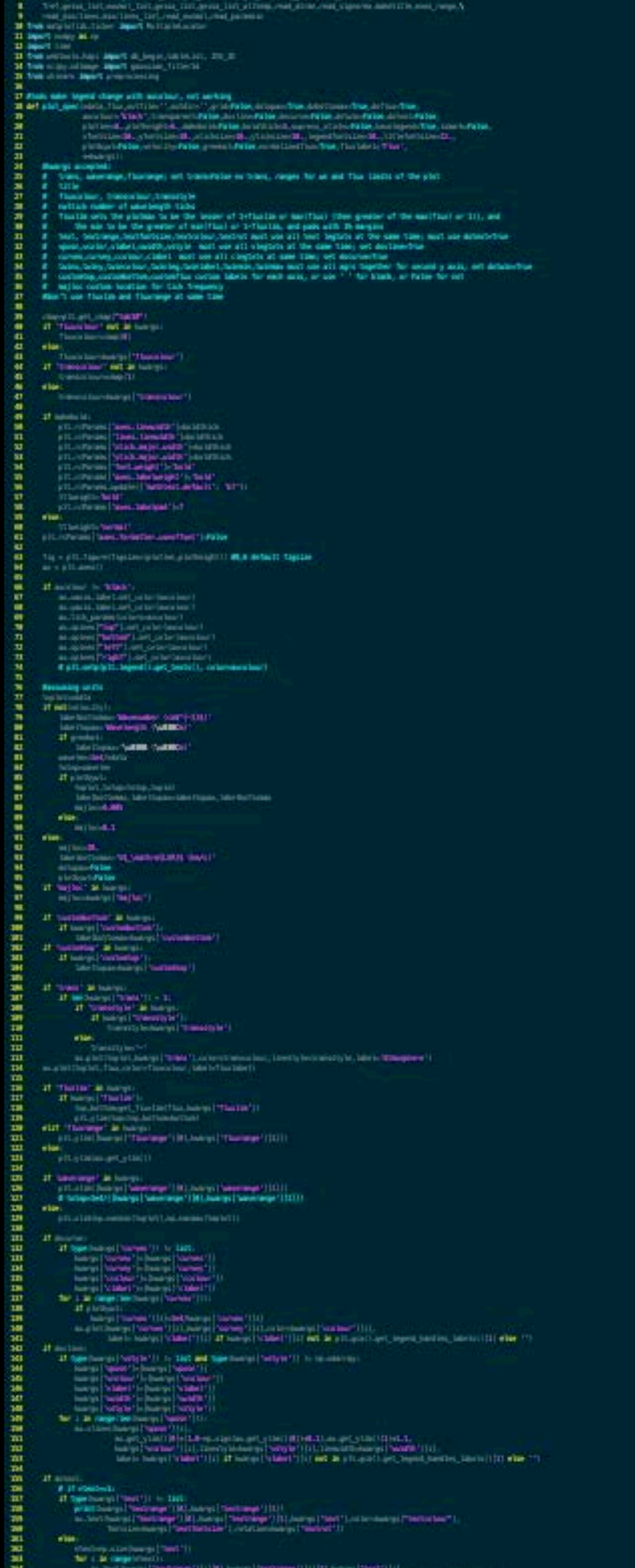
http://spiff.rit.edu/classes/phys440/lectures/curve/curve.html

Non-LTE

- If a rotation diagram is non-linear and is not well fit by a straight line, then the gas is not in local thermodynamic equilibrium
- Can use a non-LTE model to fit the spectra (e.g. RADEX, van der Tak 2007)
- However, these require collision coefficients and presently no database with collision coefficients covers the MIR

MIR Python Analysis Toolkit

- Publicly available analysis tools for spectroscopy are focussed on longer wavelengths, FIR to sub-mm
- Nothing for the MIR
- Developed Python Toolkit to analyze EXES and TEXES high resolution spectra; includes:
  - Visualization
  - Molecule identification using MIR line lists from databases (HITRAN, GEISA, and ExoMol), and custom line lists
  - Peak finder and noise finder
  - Fitting atmospheric models to data for subtraction
  - Rotation diagram analysis of both absorption and emission lines
  - Production and fitting of simulated spectra
- Presently not ready for public availability but happy to collaborate on request
- Intend to publicly release in future with documentation



Thank you!