CHILES

CHILES, the COSMOS HI Large Extragalactic Survey

1002 hours JVLA B array

SKA Science   and a path finder for the path finders

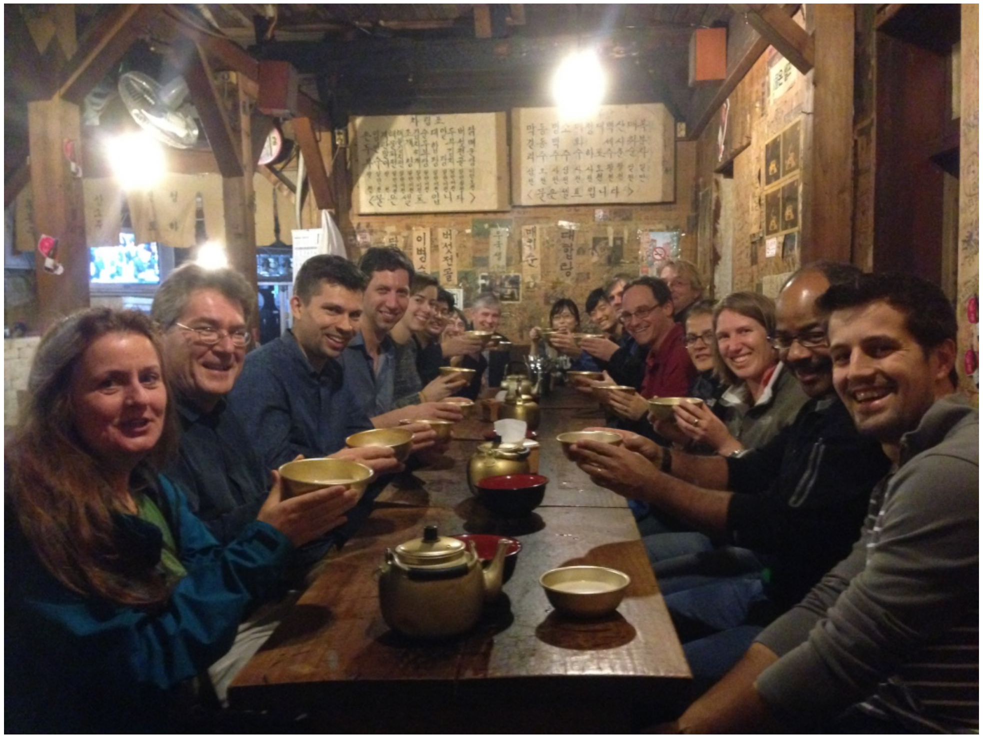
USA

Columbia
Jacqueline van Gorkom
Ximena Fernandez
Julia Gross
Nick Luber

West Virginia University
D.J. Pisano
Lucas Hunt
Evan Smith

University of Wisconsin
Eric Wilcots
Julie Davis
Catherine Witherspoon
Matt Bershady

University of Massachusetts
Min Yun
Hansung Gim



University of New Mexico
Trish Henning
Monica Sanchez Barrantes

NRAO
Emmanuel Momjian,
John Hibbard
Jennifer Donovan-Meyer

Virginia Tech
Danielle Lucero

Germany

Kathryn Kreckel

Korea

Aeree Chung

The Netherlands

Astron, Kapteyn Institue
Kelley Hess
Tom Oosterloo
Thijs van der Hulst
Marc Verheijen
Natascha Maddox

Australia

Richard Dodson
Kevin Vinsen
Chen Wu
Attila Popping
Martin Meyer
Luke Davies

Chile

Yara Jaffe

Unique aspects of VLA among SKA path finders

- Strengths (what we thought when we started)
- It is up and running
- Correlator is more powerful than planned for any of the path finders
- Sensitivity and angular resolution comparable to MeerKAT
- Baseline distribution, angular resolution of 5”

  and most collecting area at spacings > 2 km

Weaknesses

It is a multi user instrument and it

is hard to schedule large amounts of time

Relatively small FOV

Uniquely suited to do deep imaging at high redshifts

HI Deep Fields

The big question is, how does galaxy growth and evolution depend on the location in the large scale structure?

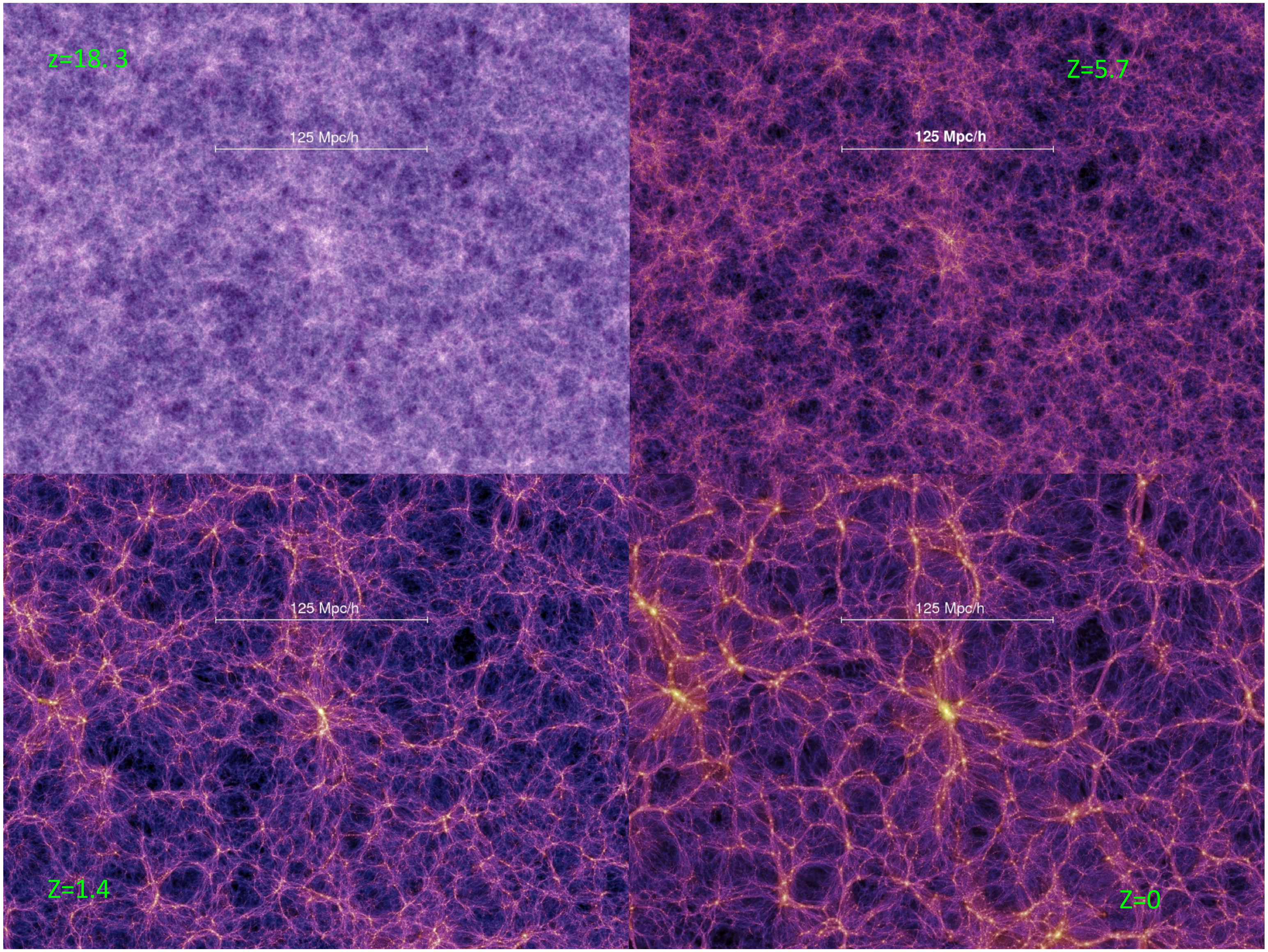
We know the large scale distribution of galaxies quite well.

Theoretically:

from cosmological simulations of the growth of dark matter structures such as millennium simulation

Observationally:

from wide area surveys (e.g. SDSS) and deep surveys (e.g.2DF) simulations



Predictions from theory

Hierarchical galaxy formation in "standard" ΛCDM used to make galaxies grow by merging. The importance of gas accretion was underestimated, and the physics misunderstood.

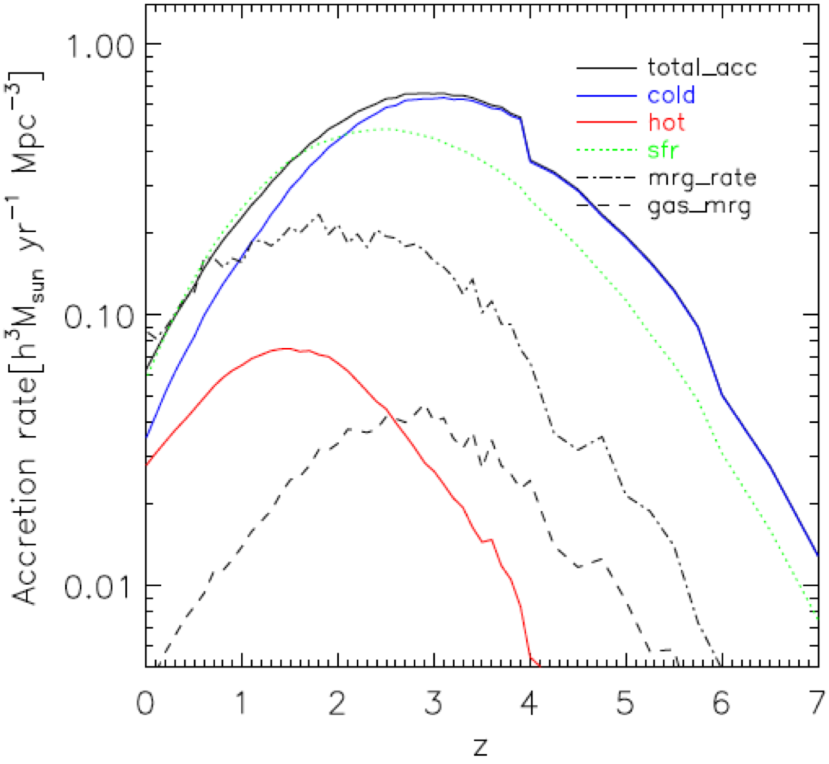
There are two ways for galaxies to grow

1) Merging with smaller galaxies can add gas and stars

2) Smooth accretion of cool gas dominates galaxy growth at z>1.

Keres et al 2005, Dekel and Birnboim 2006, Binney 1977

Many recent papers, much debate, new code Arepo (Springel, Hernquist and collaborators)



At high z, gas accretion dominates, even at low z it is important.

More specific predictions:

1. Mode of accretion depends on redshift. Cold mode accretion dominates in all galaxies at $z>1$

2. Mode of accretion depends on galaxy mass. As galaxies grow bigger a transition from cold to hot mode accretion occurs. At $z=0$ Milky Way mass is transition point.

3. Mode of accretion depends on local galaxy density. At $z=0$ cold mode accretion still dominates in lowest density regions, the voids.

Alignment between cosmic web filaments and galaxy spins?

Dark matter halos acquire their angular momentum through tidal torquing by neighbouring large scale structure. Filaments form by collapse in two directions. Halos should acquire spin parallel to the filament, as matter collapses and rotates in plane perpendicular to the filament.

Simulations find this to be true for low mass halos, while mergers align high mass halo spins perpendicular to filaments by converting motion along the filament into spin.

Observations however probe the spin of the baryonic matter. Initially spin of baryons and dark matter share same angular momentum, but further evolution depend on the details of the baryonic physics.

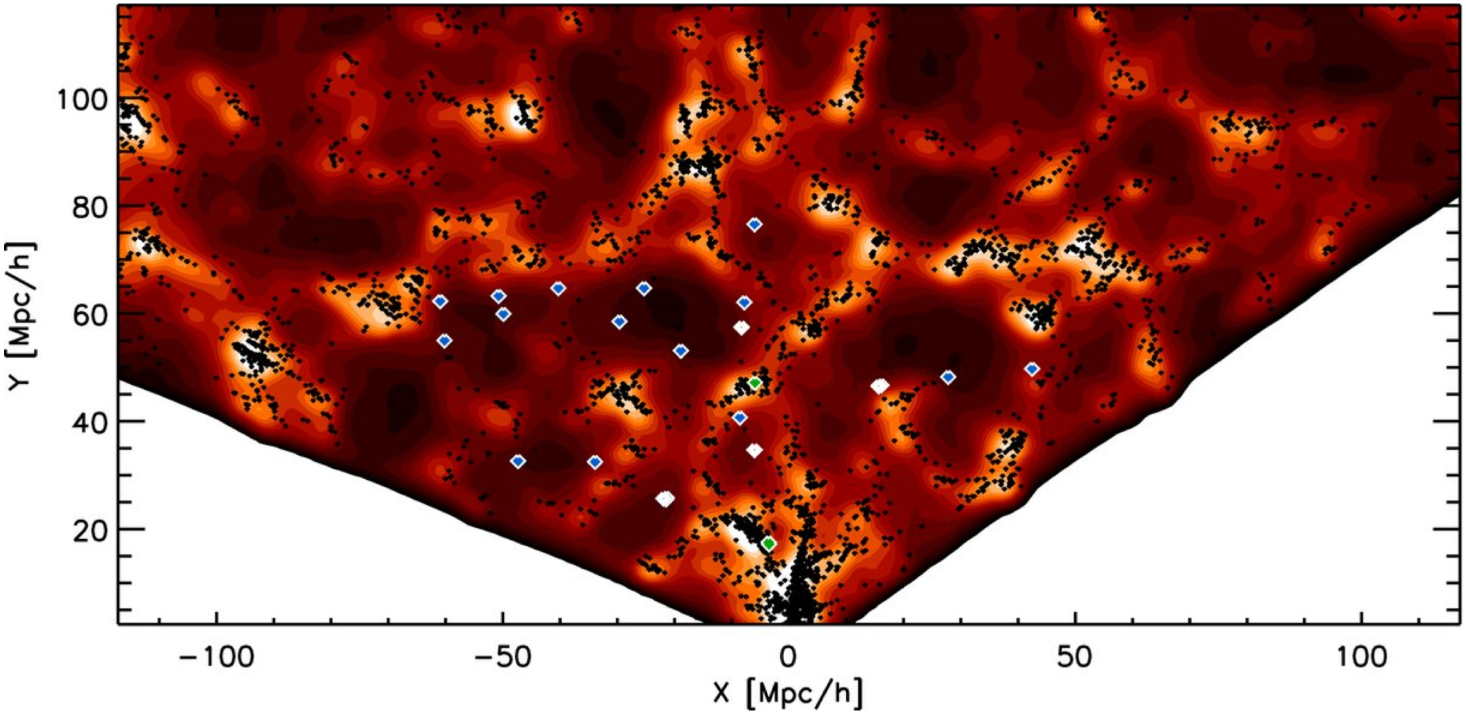
So far observations give mixed results.

See for example Krolewski et al 2019, for observations and Kraljic et al 2019 for simulations

The Local Universe: Galaxies in voids

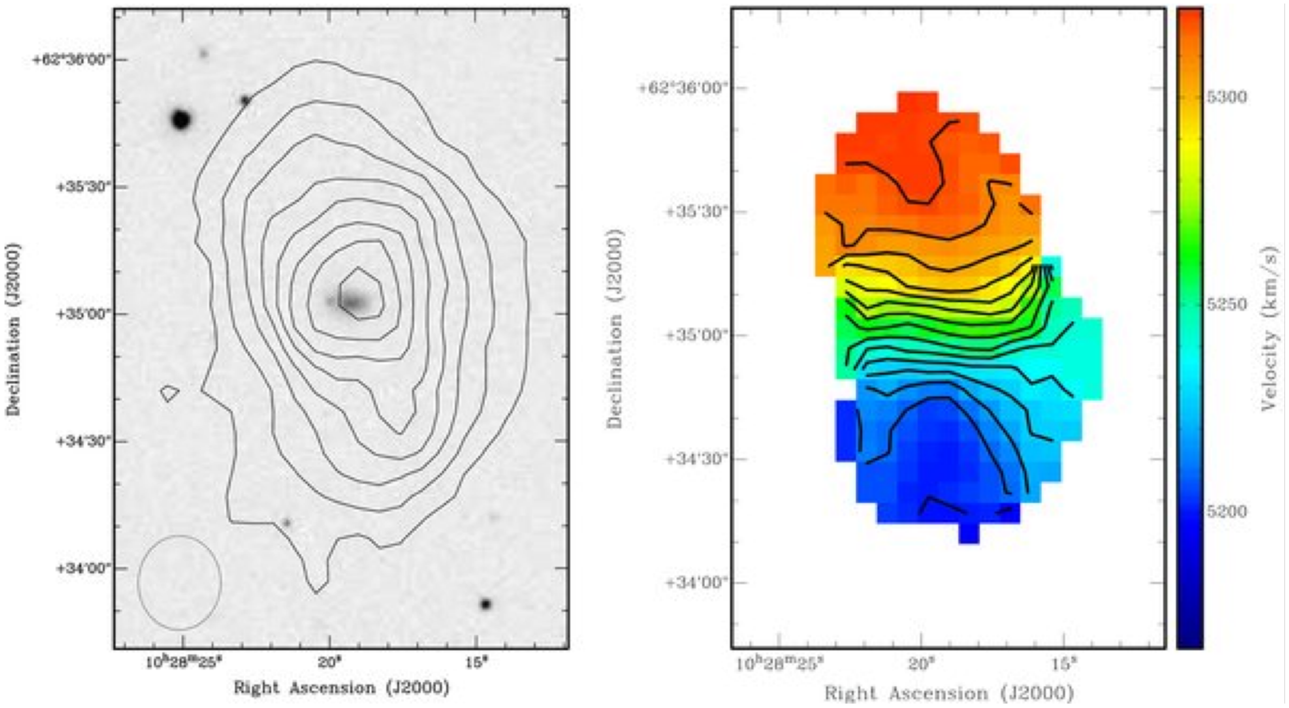
Kreckel et al 2012

Using a Voronoi tessellation to define density field and the watershed void finder to find the deepest under densities, a sample of 60 galaxies in the voids is defined



Reconstructed density field, black SDSS, diamond void galaxies, green control sample

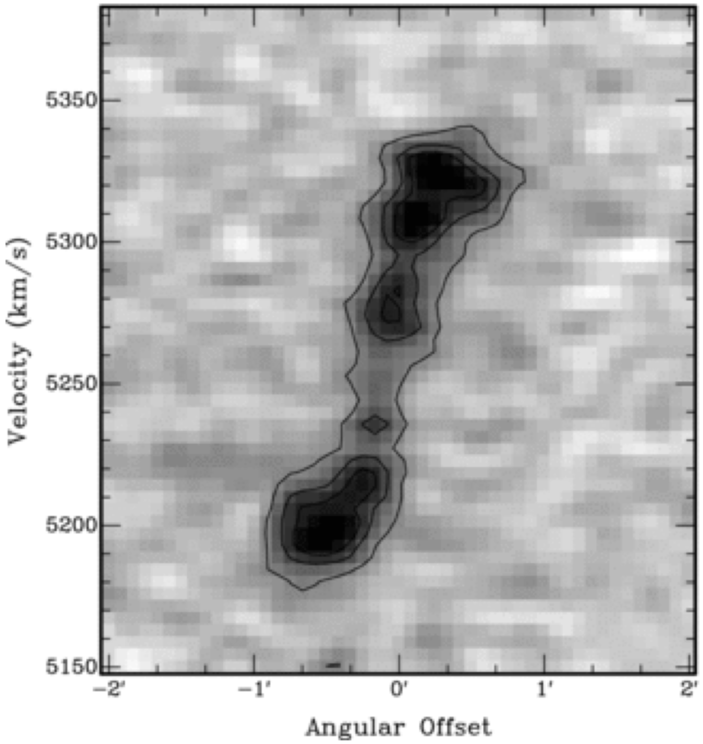
The void galaxy survey.. Some tantalizing results

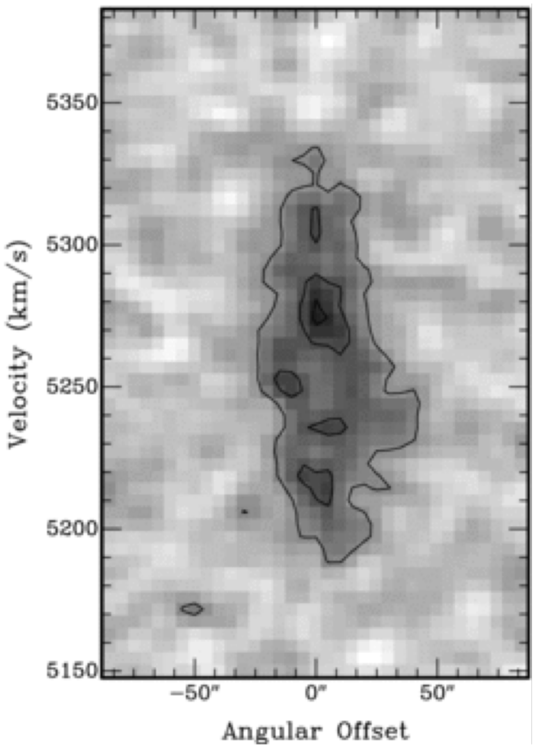


A polar disk

Stanonik et al 2009

Kreckel et al 2012

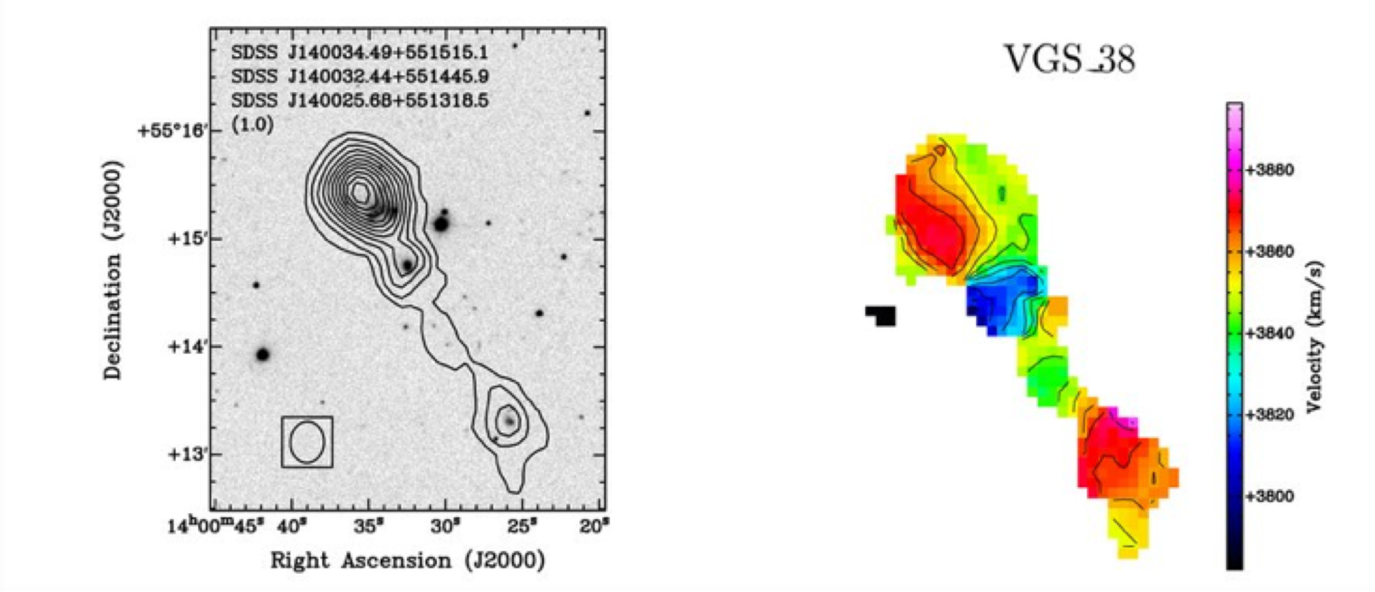




Polar ring. Mass in HI (3×10^9 Msun) > Mass in stars (1×10^9 Msun)

HI much more extended than stellar disk. No optical or UV counterpart to polar ring. Tidal interaction would destroy rotation in disk.

Possible example of cold mode accretion. In this case, gas flows out of the void



Several examples of galaxies embedded in filaments. Note that the galaxy spins are aligned with filament

CONCLUSIONS Void Galaxy Survey

Kreckel et al 2011, 2012,2014, Beygu et al 2013, 2014, 2016

By looking in voids you select an interesting sample of small galaxies (no stellar masses > $3x10^{10}M_{sun}$).

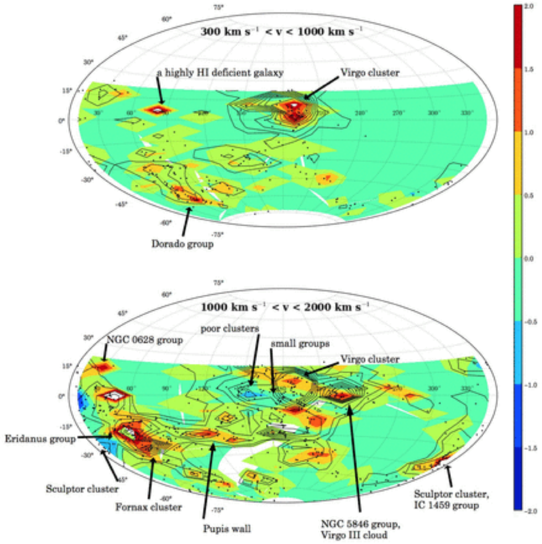
Most of these galaxies are gas rich. Many show kinematic signs of disturbances and possibly evidence forongoing accretion.

Some evidence that these galaxies are metal poor.

Several cases are found where galaxies are embedded in larger HI filament possibly with spin alignments

There are other amazing hints that galaxies in voids maybe growing through smooth accretion.

Are we close to making a neutral hydrogen image of the (local) universe?



Red is very gas poor, blue is gas rich

Contours galaxy density

May explain galaxy conformity?

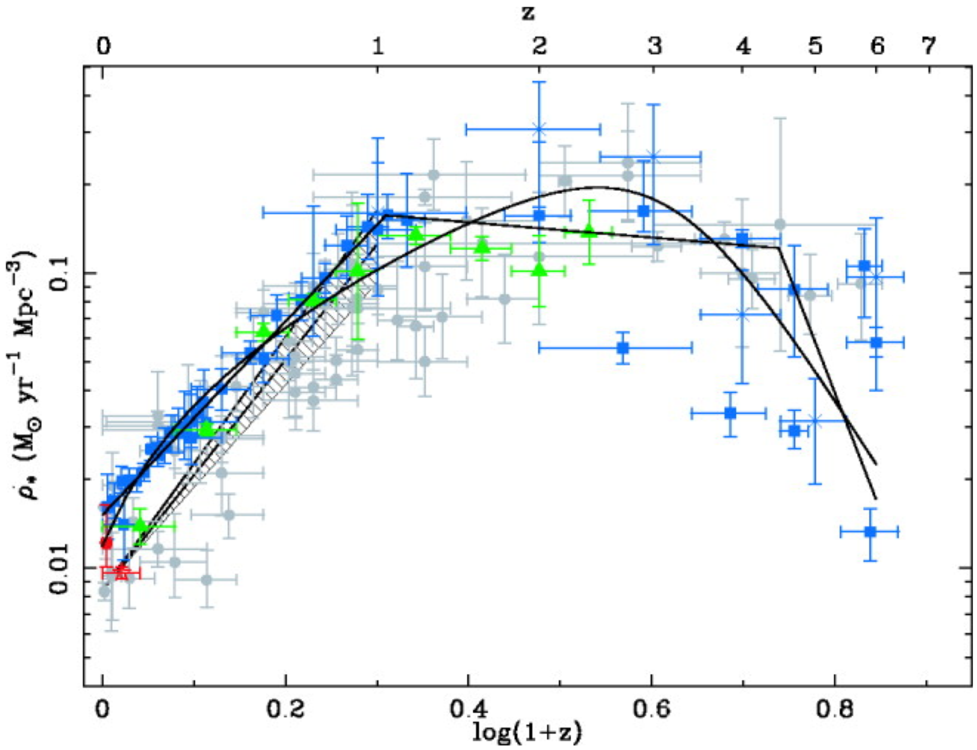
Figure 7.
Sky distribution of the H I deficiency parameter in two-dimensional bins overlaid with HyperLEDA density contours. The colours represent average H I deficiencies of different areas. Red and orange regions have on average more H I-deficient galaxies and dark blue regions have on average more H I-rich galaxies than the green and light blue regions. Density contours are 10, 30, 50, 70, 90, 110 galaxies. Black dots represent the individual galaxies of our HOPCAT and NOIRCAT samples.

HI deficiency maps.. Denes, Kilborne and Koribalski, 2014

Evolution with redshift

What we know

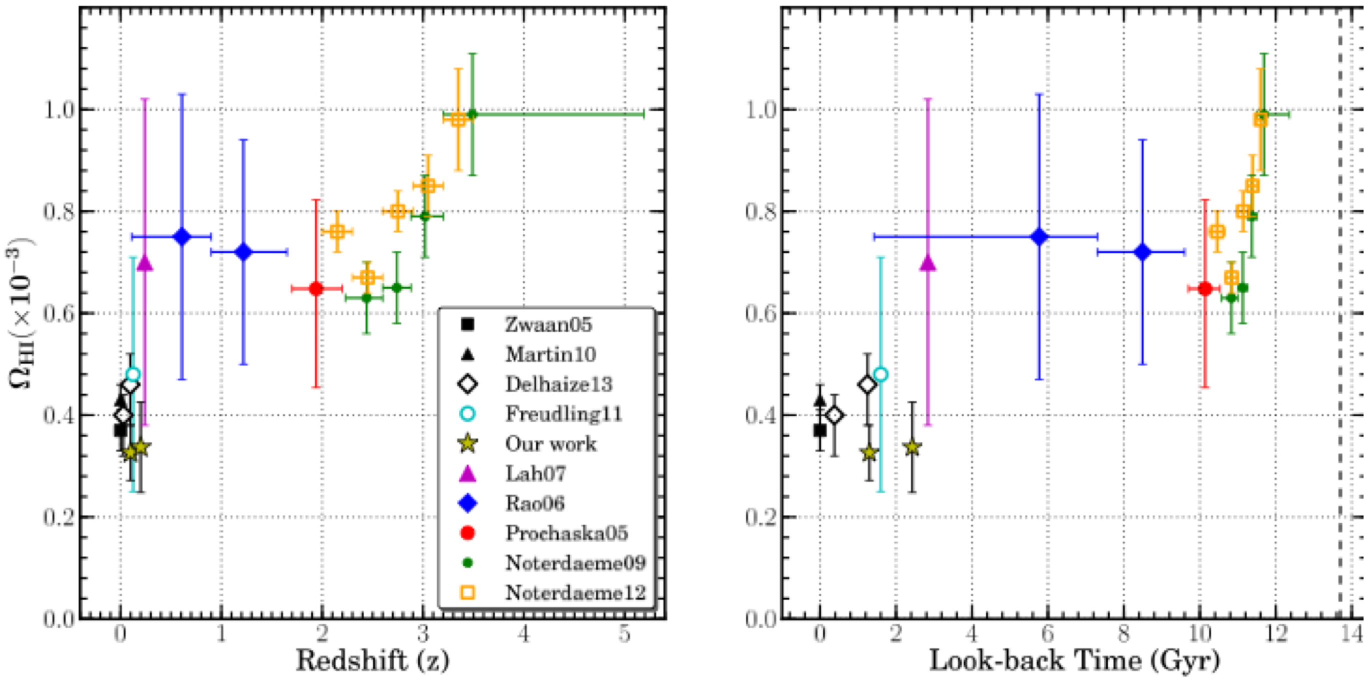
The evolution of Star Formation Rate Density (Hopkins and Beacom 2006)



Images can tell how galaxies get and lose their gas

Evolution with redshift

What we don't know

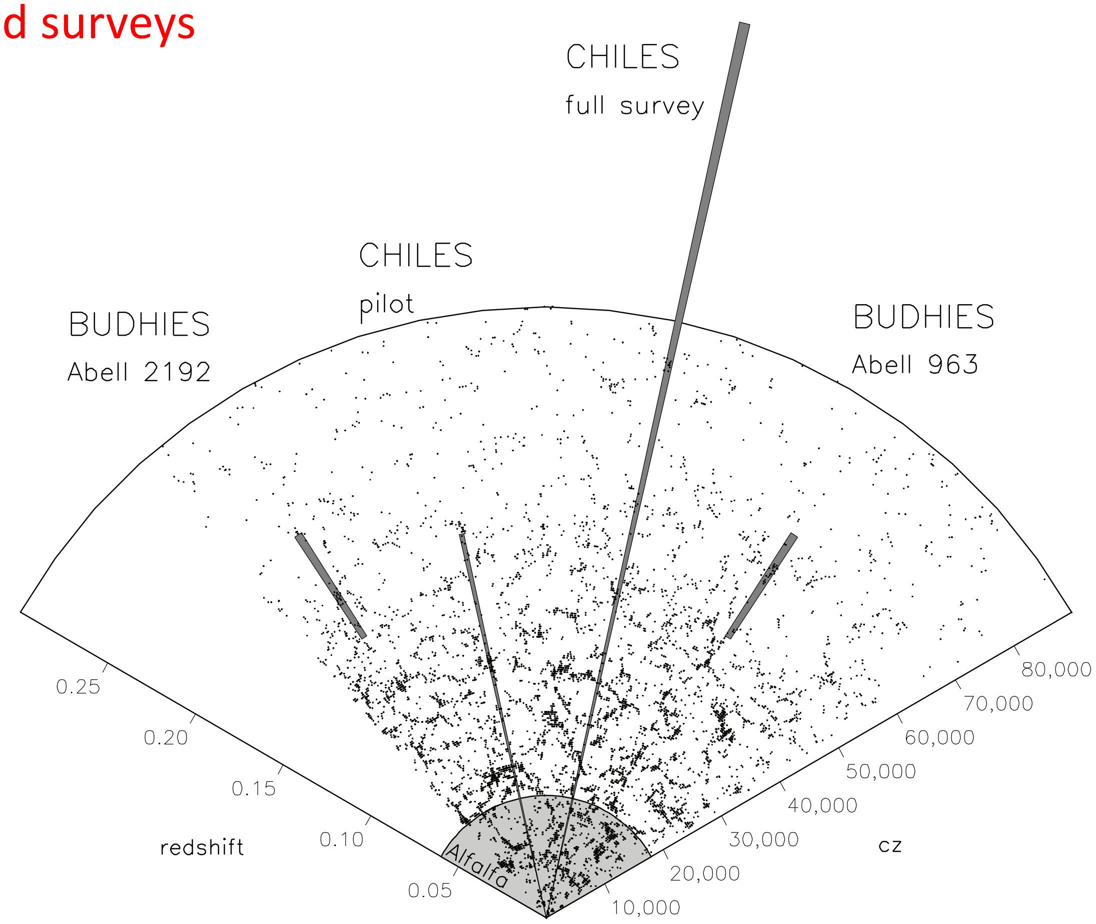


Zwaan HIPASS; Martin ALFALFA all sky surveys out to z=0.06
0.1<z<0.4 HI stacking with optically known galaxies
Z>0.4 Damped Lya alpha systems

Omega HI well constrained at z=0 and at z>2

Observations suggest no evolution between z=0 and z=0.2. Yet integral SFR drops steeply between z=0.2 and 0

HI blind surveys



What do we know about evolution of HI with redshift

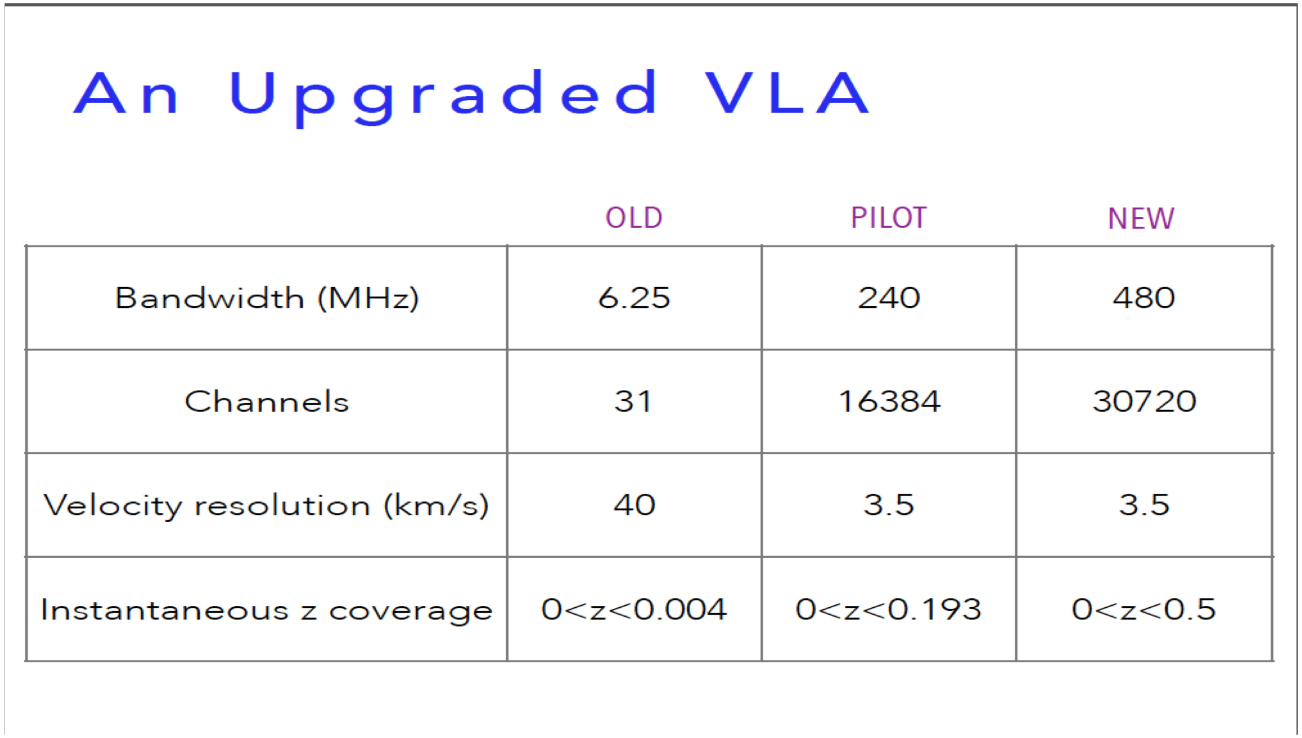
Main scientific motivation for CHILES

HI morphology as function of location in underlying large scale structure in a blind survey. Do the accretion modes change with redshift? At what point does the gas cool enough to become neutral hydrogen?

note that even at z=0.45 we will probably be able to say whether HI is inside or outside a galaxy

HI content, morphology and kinematics of individual galaxies
HI mass function as function of z and environment
Cosmic neutral gas density as function of z
Evolution of Tully Fisher relation

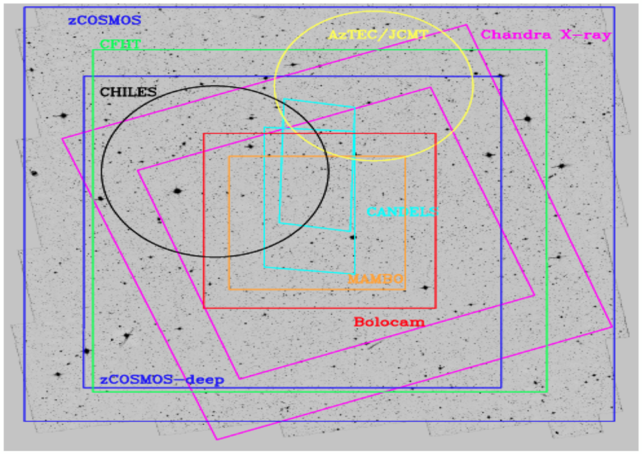
Chiles could only be done because of VLA upgrade



An Upgraded VLA

	OLD	PILOT	NEW
Bandwidth (MHz)	6.25	240	480
Channels	31	16384	30720
Velocity resolution (km/s)	40	3.5	3.5
Instantaneous z coverage	0<z<0.004	0<z<0.193	0<z<0.5

Target: COSMOS field



Deep multiwavelength data available

A pilot for an EVLA HI Deep Field

One pointing in COSMOS field

Fernandez, Hess, Momjian, Pisano, Oosterloo, JvG (the human calibration pipeline)
Popping, Chung, Henning, Verheijen, Schiminovich, Scoville

60 hours in B array (5 arcsec at z=0) , data taken in 2011.. 2.5 Tbyte

32 sub bands 16384 channels (1420-1190 MHz; z=0 to 0.2) vel resolution 3.3 km/s

Detection limits	z=0.07	$7\text{x } 10^8\ M_{sun}$
	z=0.13	$4\text{x } 10^9\ M_{sun}$
	z=0.2	$1.3 \text{ x } 10^{10}\ M_{sun}$

Column density sensitivity $3\text{x}10^{19}\ \text{cm}^{-2}$

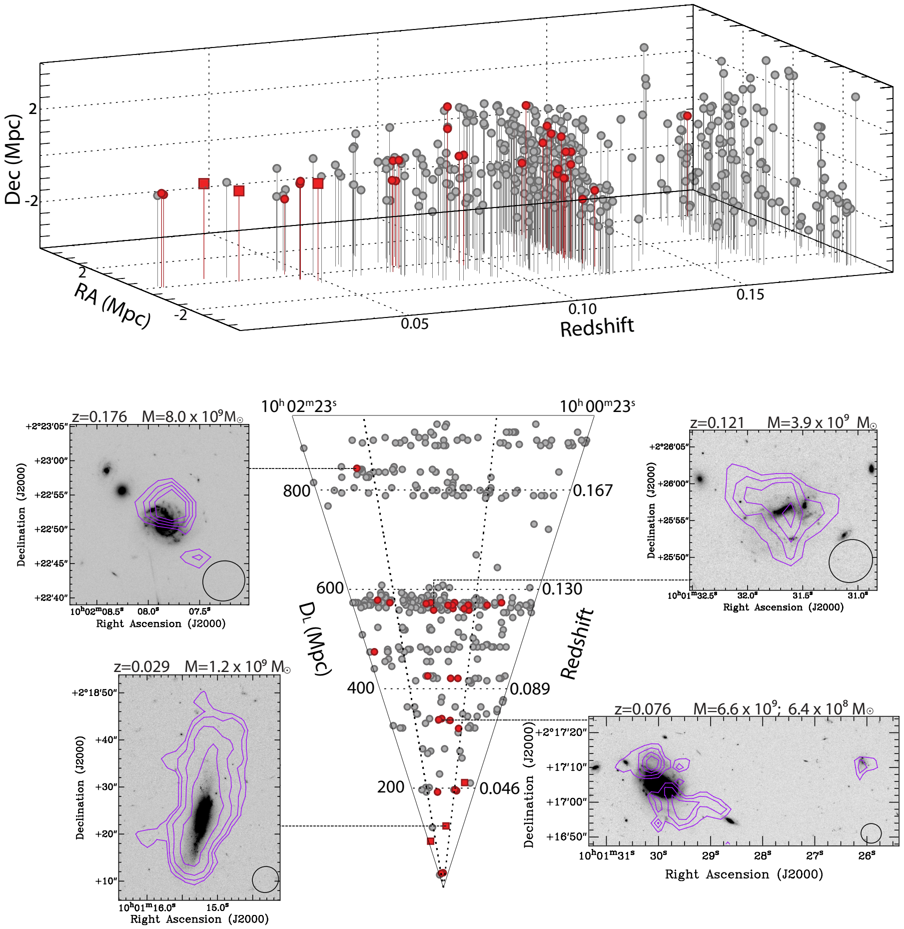
Resolution	350 pc at 16 Mpc	17 kpc at z=0.2
FOV	150 kpc	7.5 Mpc

Pilot was done during commissioning, Using half the correlator, 16000 channels, covering z=0 to 0.2 and Integrating for 60 hours

We detected galaxies across the entire redshift range, achieved the planned sensitivity and reduced data in a year.

We find interesting morphology in different environments

This was good enough to convince the TAC to give us 1000 hours for CHILES



Results of the pilot FERNANDEZ et al 2013

CHILES 1002 hours in B array spread over five B array configurations

Same detection limit at z=0.45 about 3×10^{10} M_{sun} as for pilot (60 hours) at z=0.2

Cover z=0 to 0.5 with 31 000 velocity channels

15 subbands of 32 MHZ 2048 channels each 3.3 km/s
use frequency dithering

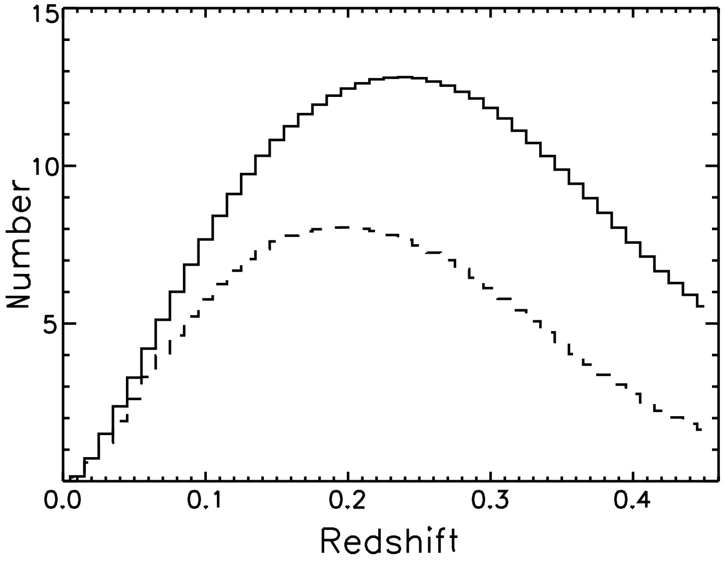
Calibrate data in Socorro
6 hours of data 1.5 Tbyte.. Pipeline 60 hours
inspection takes roughly 1 astronomer week (few hours a day)

ship data to Perth.. (both calibrated data and raw data with tables)
imaging in Perth
they had a computer, but telescope was still under construction

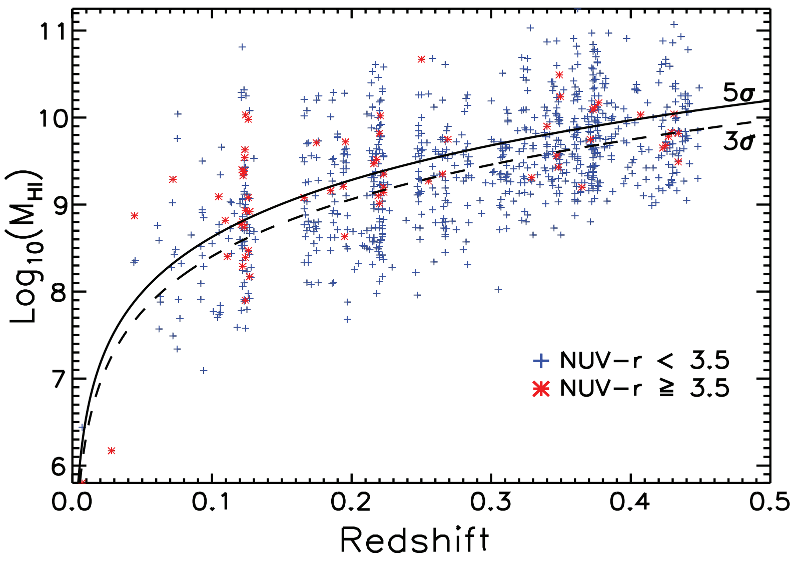
Data taking will be completed in summer 2019

Expected detection rates for 1000 hour project

We expect at least 300 direct detections.. i.e. HI IMAGES



Estimate based on HI mass function



Estimate based on photometric gas fraction

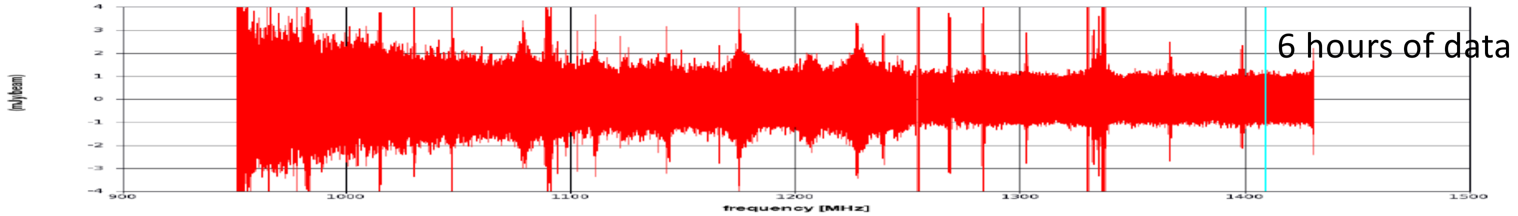
Technical Issues

Does frequency dithering work

Can we integrate down to the noise

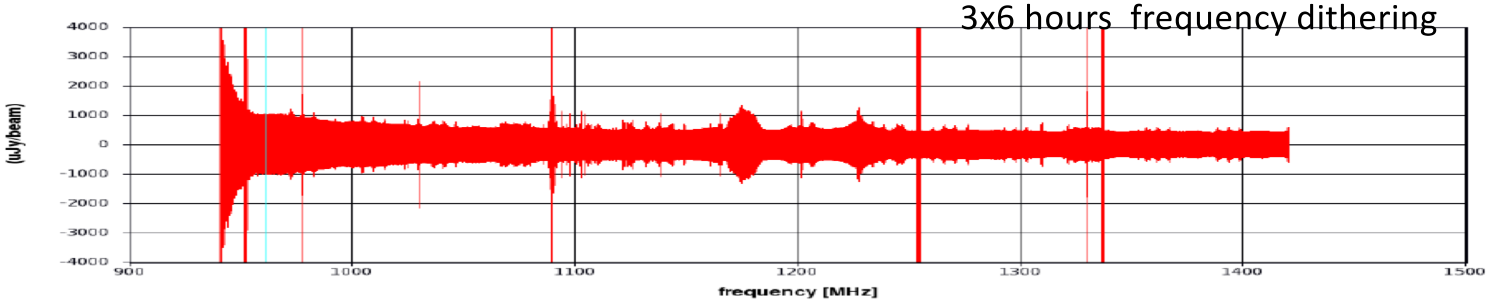
How to deal with radio frequency interference flagging and calibration

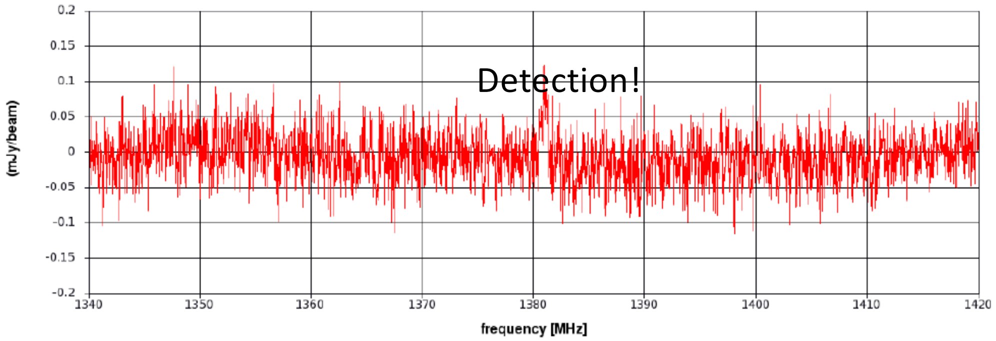
Problems associated with data volume and imaging



6 hours of data

3x6 hours frequency dithering





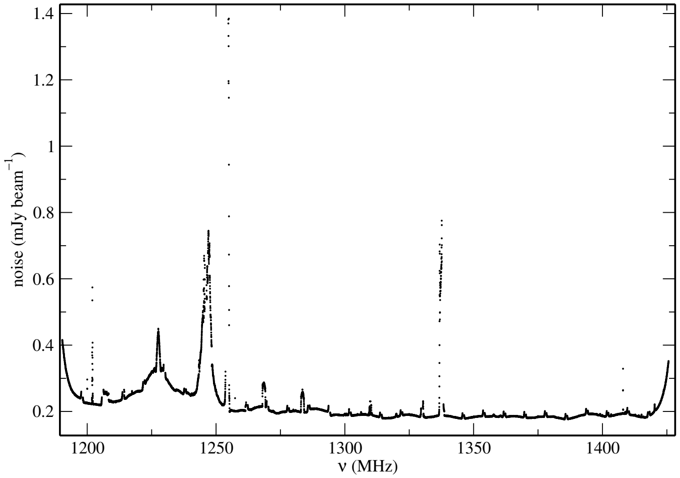
Detection!

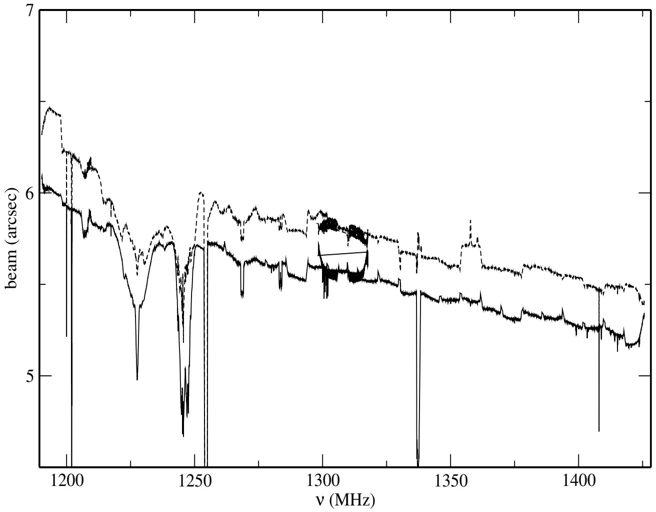
The good news is.. RFI less below 1170 MHz,   frequency dithering works beautifully

But.. Things aren’t always easy

Result of the pilot

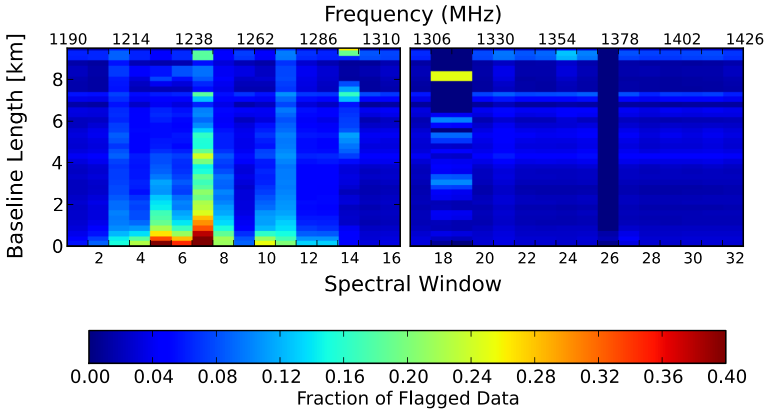
Rms noise as function of frequency looks really good



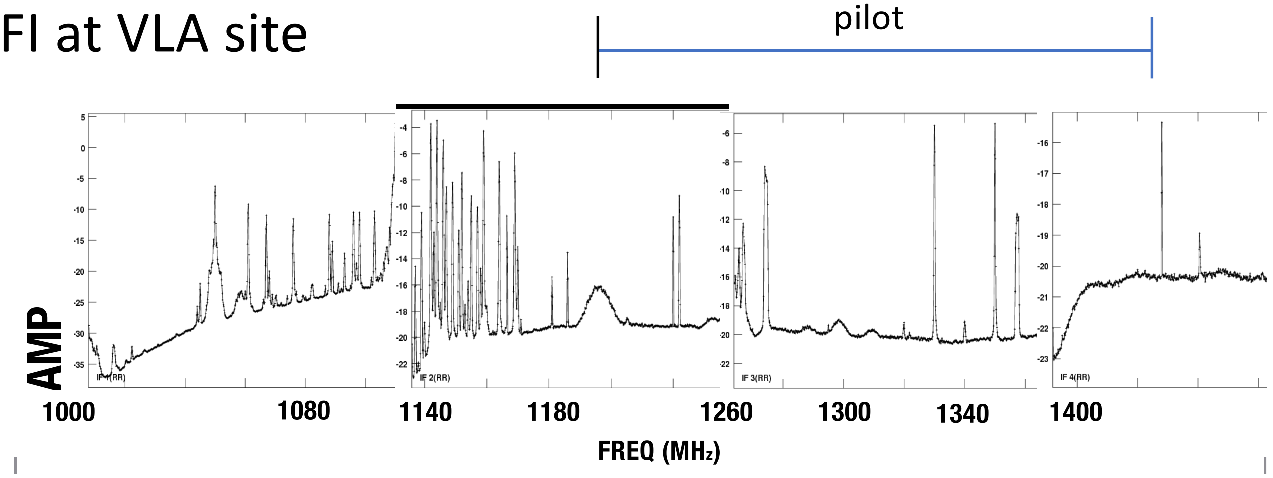


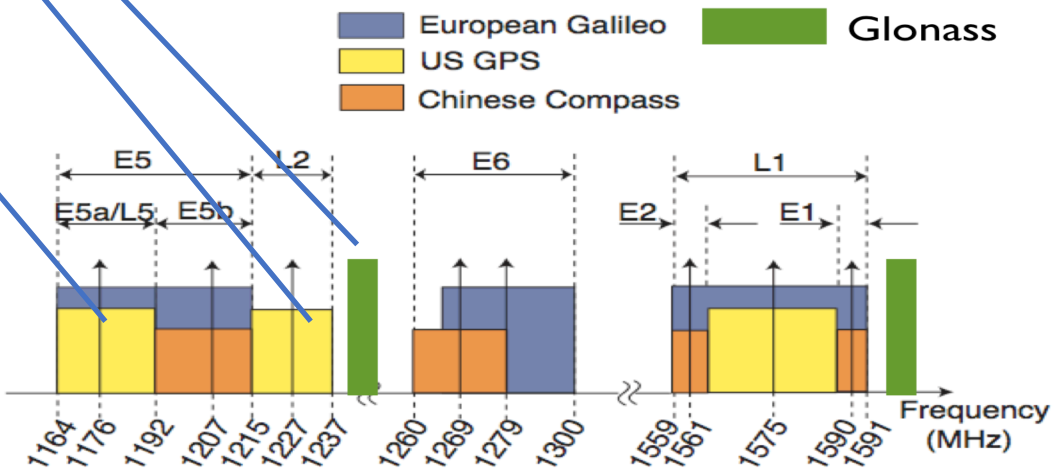
Synthesized beam

But, baseline distribution matters

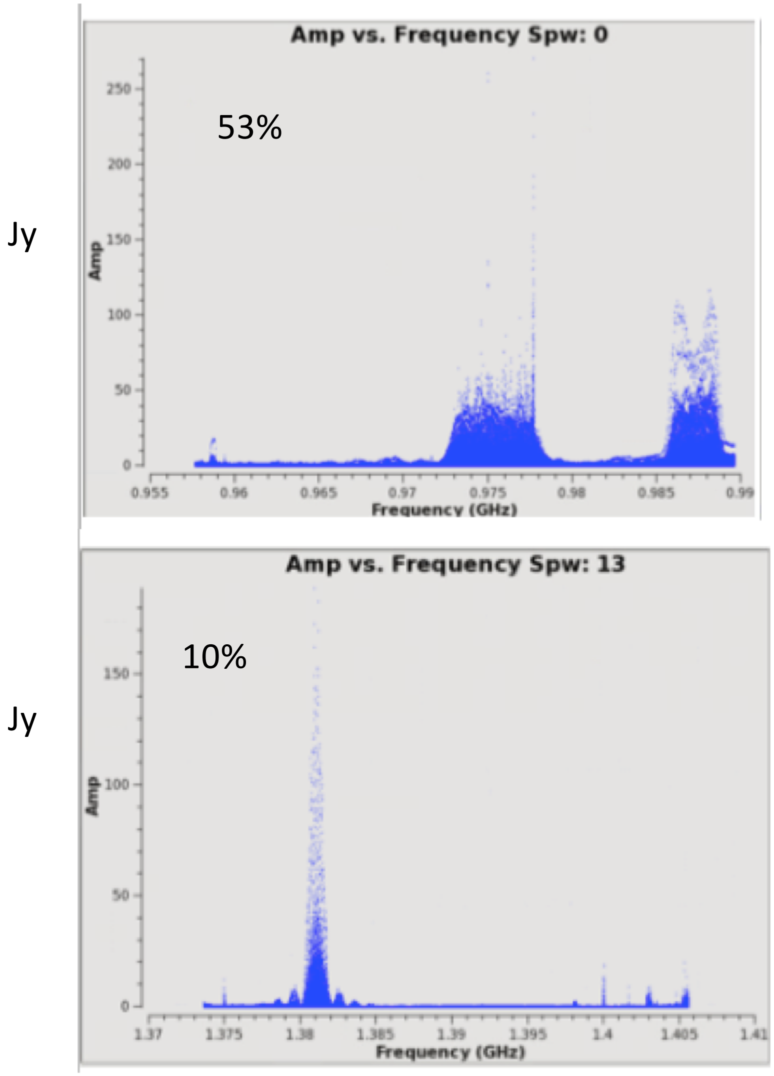


RFI at VLA site

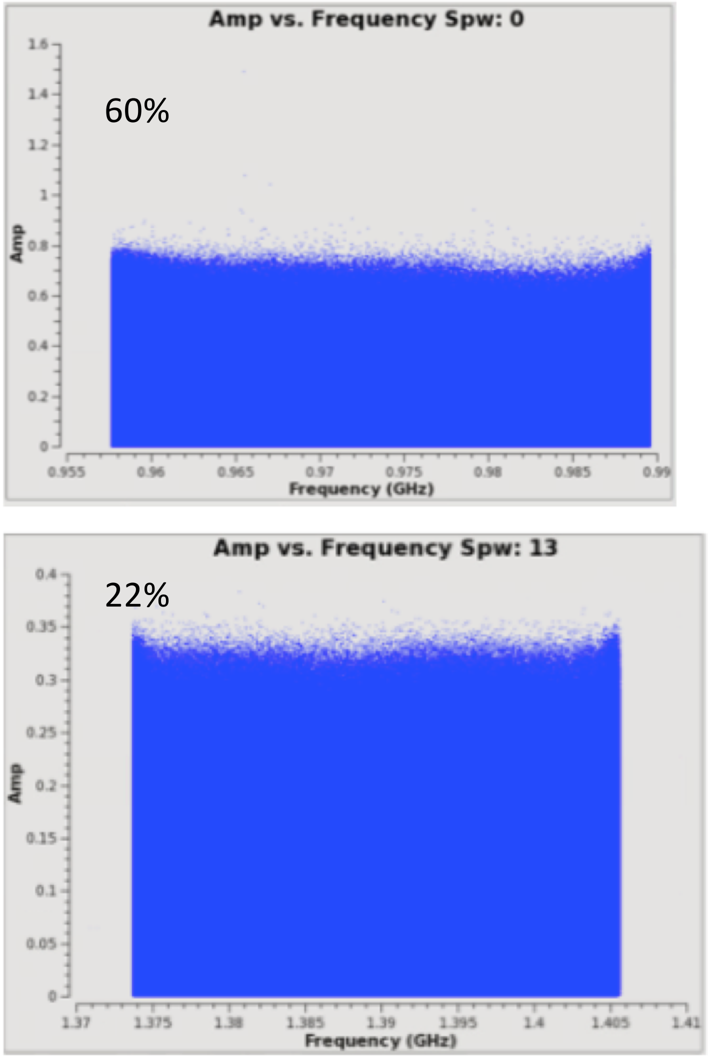




It is the satellite RFI that matters
1170 MHZ to 1307 MHz pretty bad



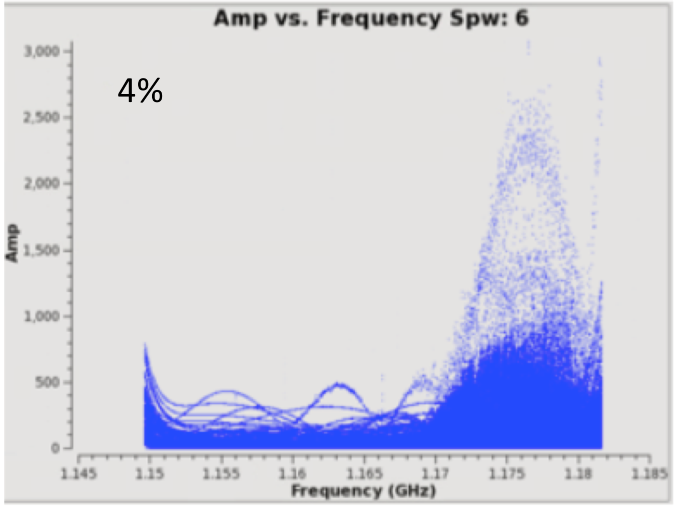
Data after online flagging

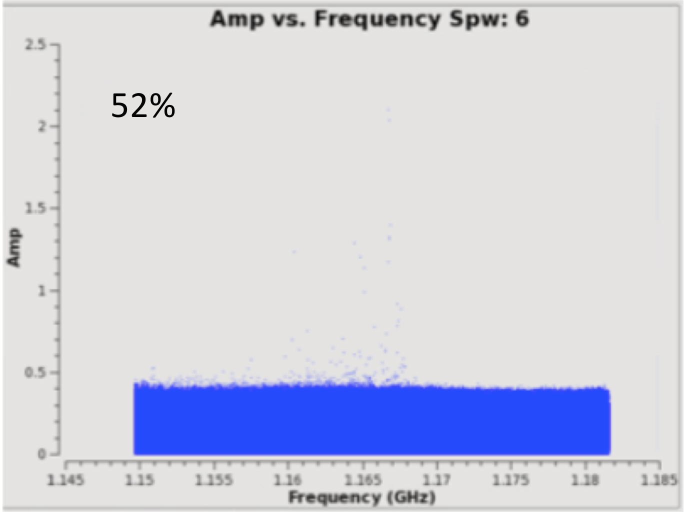


Data after pipeline flagging

GPS

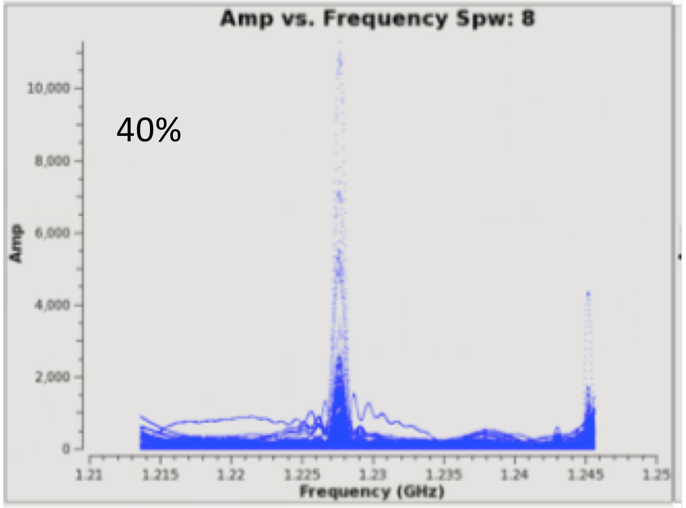
Jy

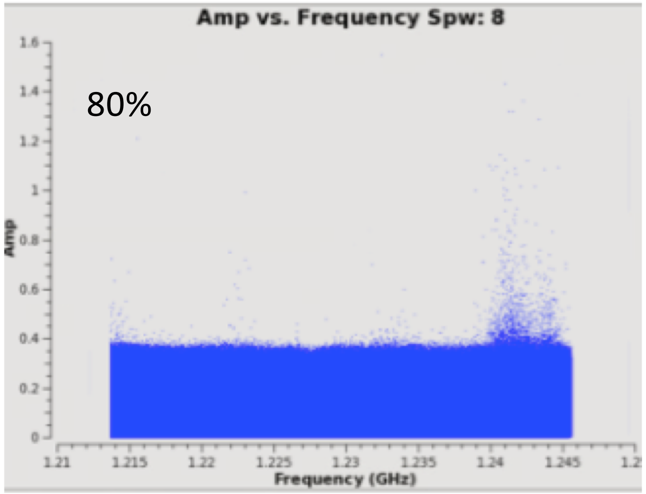




GPS

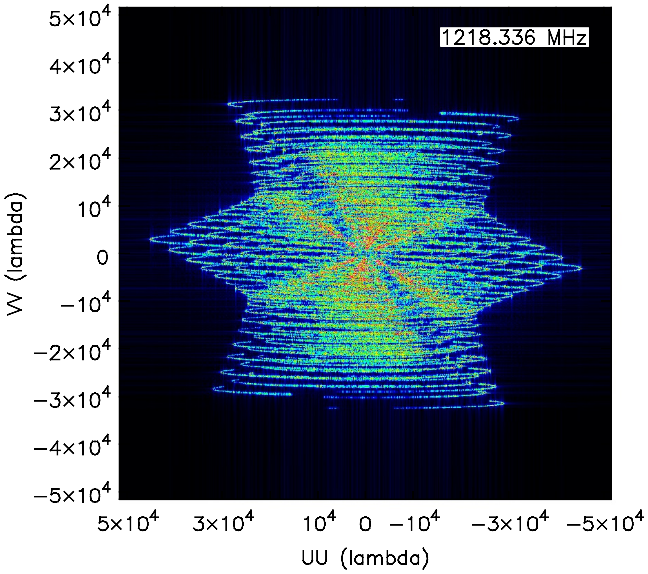
Jy





Data after online flagging

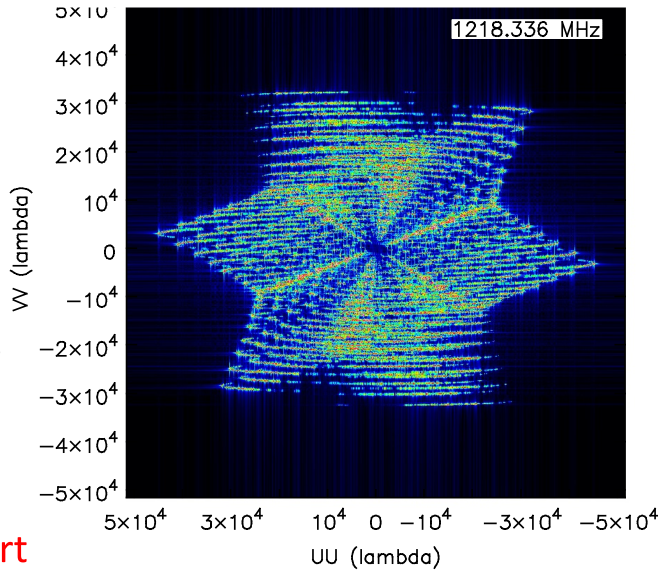
Data after pipeline flagging



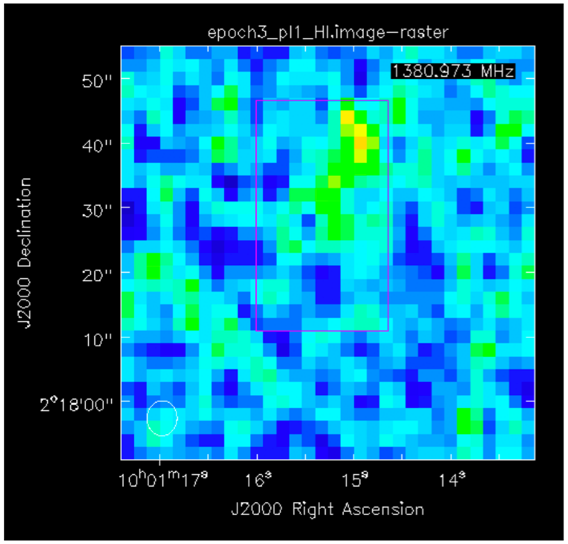
Epoch 1 data 2014

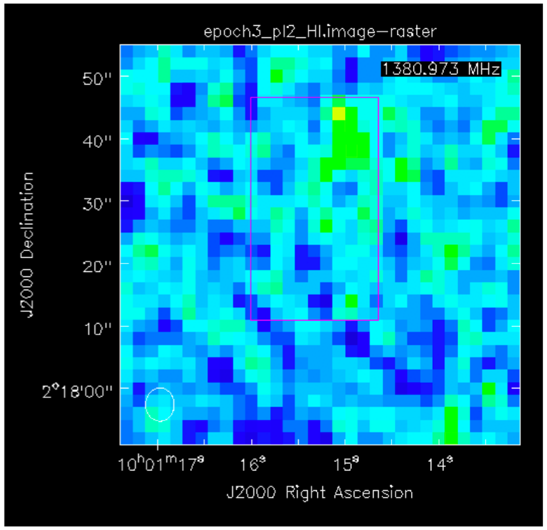
Then we tried to flag better

Epoch 2 data 2015



Astronomers are not that smart





Flagging and calibration

Pilot was done in AIPS… Each frequency sub band was calibrated separately.
Was still manageable. New algorithm development was remarkably easy thanks to Eric Greisen.

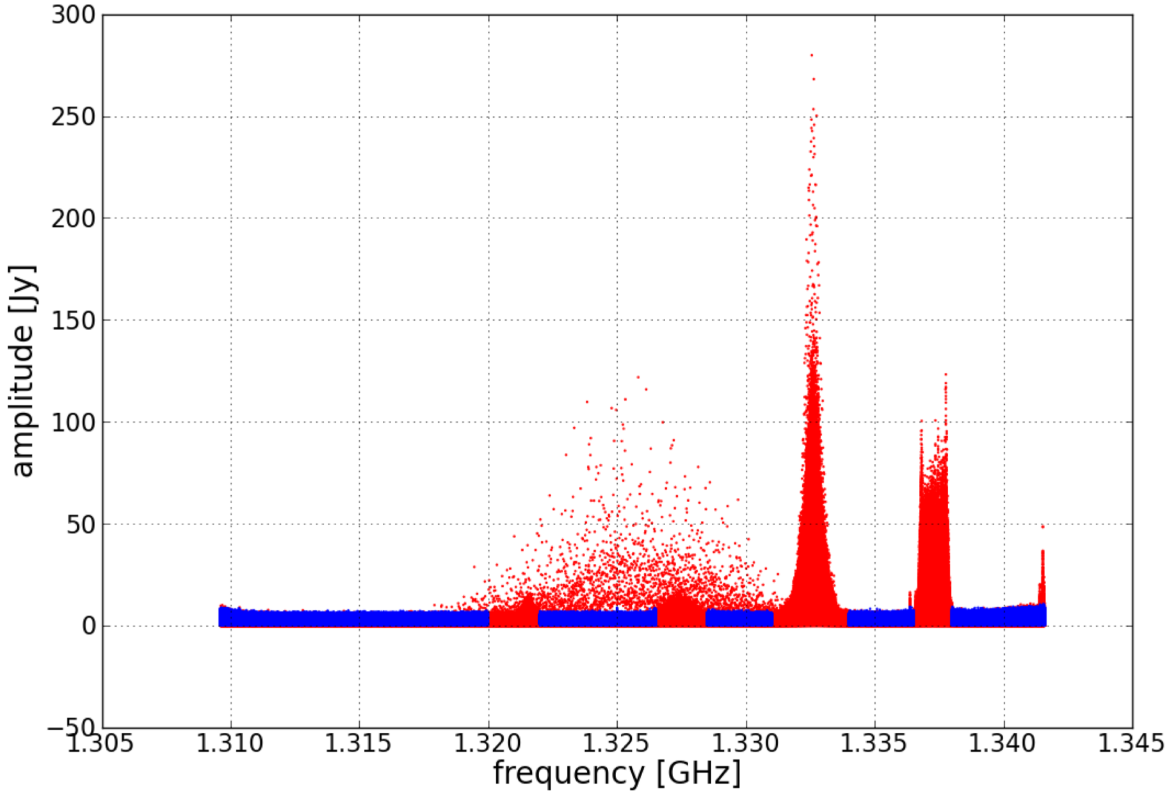
CHILES is done in CASA… Frequency dithering, large data volume, software of the future?
But, CASA is still being developed. Multi national project. Very slow response time.

First pipeline: for epoch 1 . To calibrate and flag 6 hours of data takes 60 hours of computing time and one week of astronomer time. Very hard (impossible) to go back and forth between UV and image plane. This is an issue of large data volume.
Good easy way to look at data is really important

Second pipeline modular, more easy to interact with the data and diagnose problems. We used new flagging options in CASA and tried to reduce data volume by smoothing in velocity. BUGS..
We overflagged especially on short spacings and got hit by serious software bugs

Third pipeline.. Use very conservative flagging (no extend, no smoothing)
Introduce the use of masks..Bad RFI stretches are masked iin calibration and gaintables interpolated and extrapolated over masks.
Masks are chosen per epoch. Flagging done after calilibration

Baselines between 1500m and 1600m

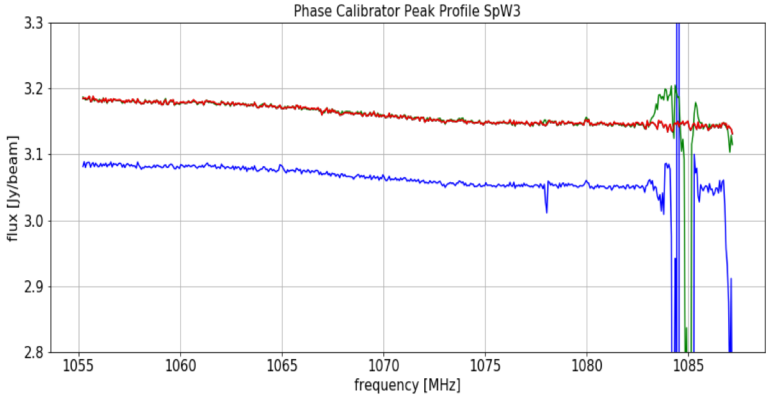


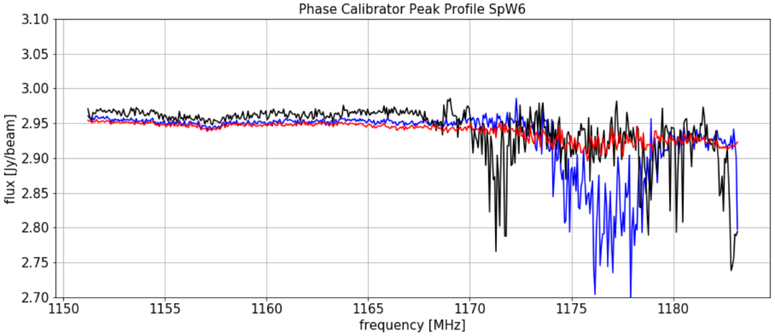
Using masks to block out bad frequency range.

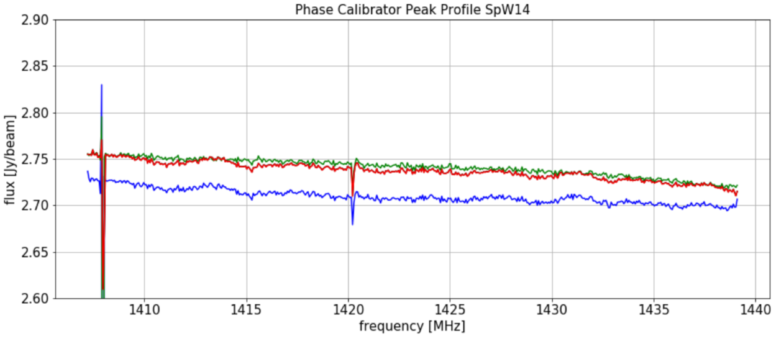
Interpolate and extrapolate calibration on good data

Then flag

We are on our fourth version of the pipeline


Phase Calibrator Peak Profile SpW3


Phase Calibrator Peak Profile SpW6


Phase Calibrator Peak Profile SpW14

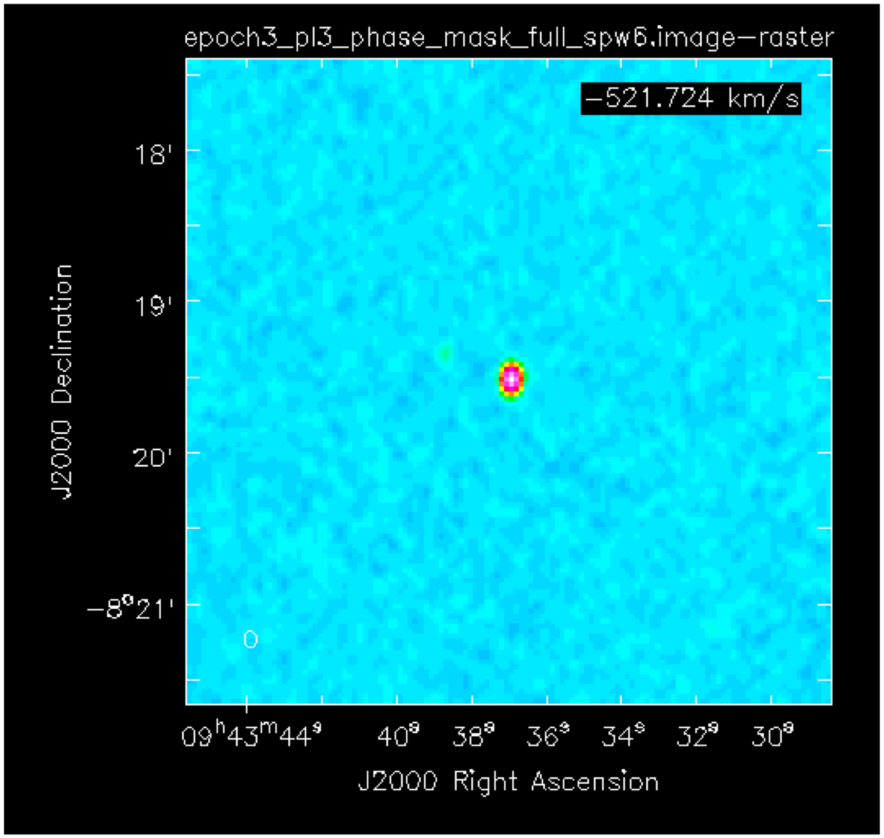
PL1

More flagging **PL2**

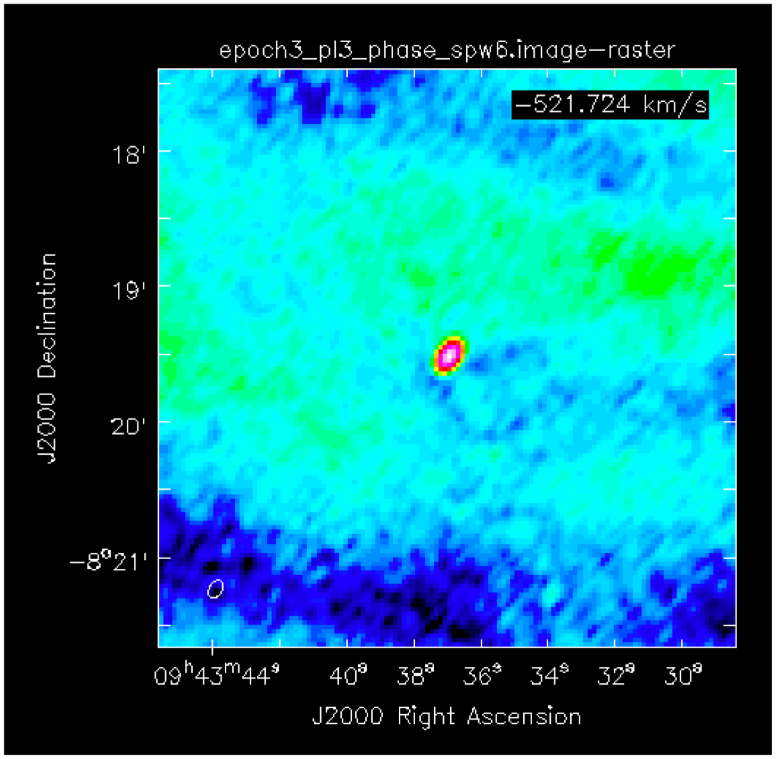
PL3 Less flagging

PL3 Mask

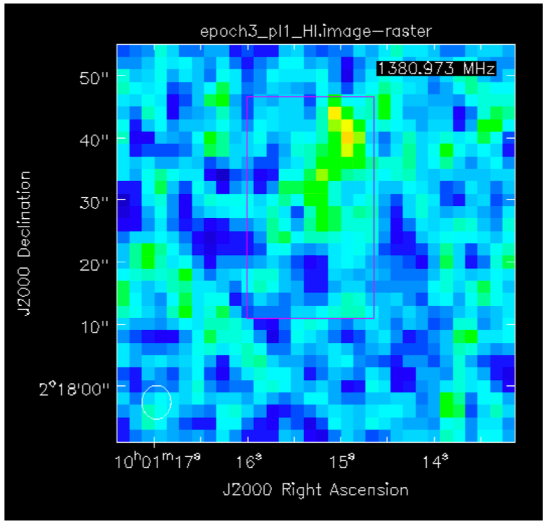
Pl3 works great



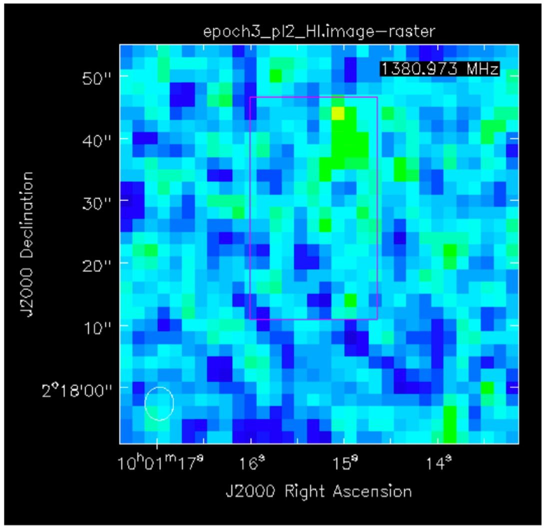
PL3 with mask



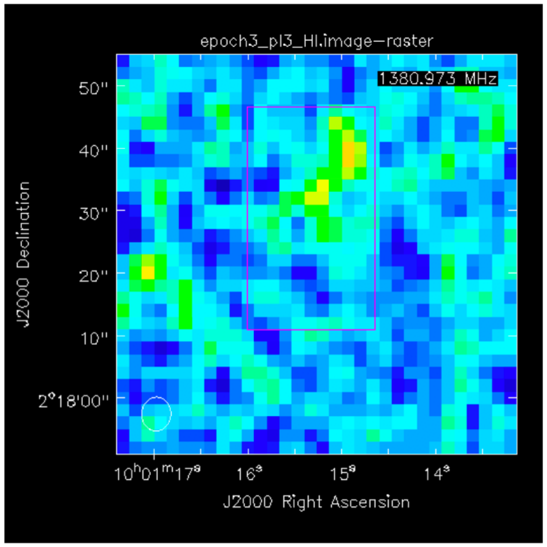
PL3 without masks



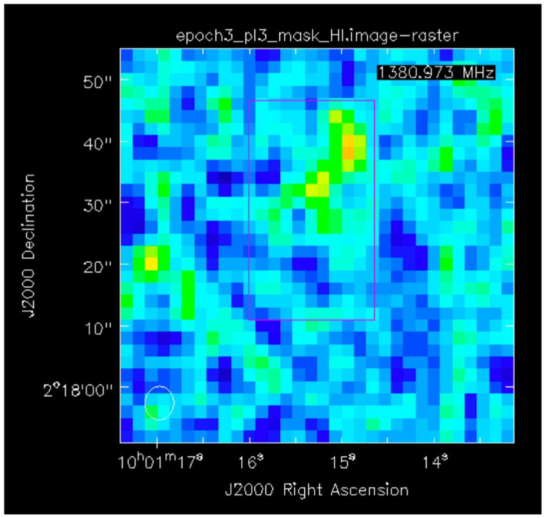
PL 1



PL 2



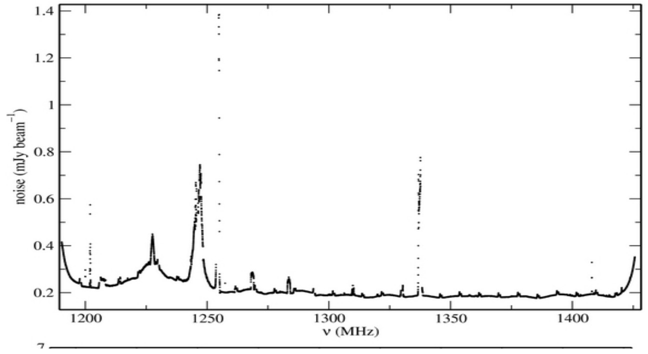
PL 3



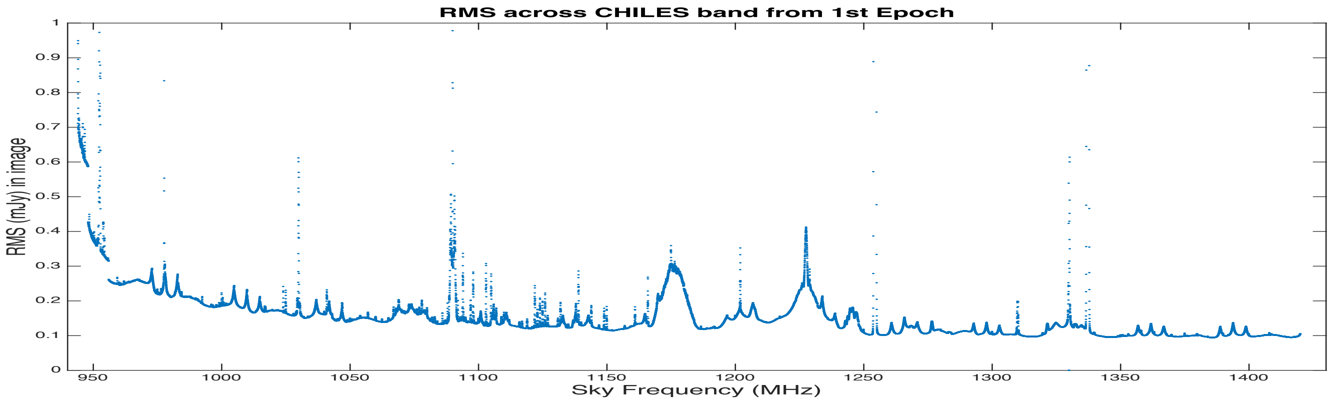
PL 3 plus mask

The noise in final data cubes

pilot

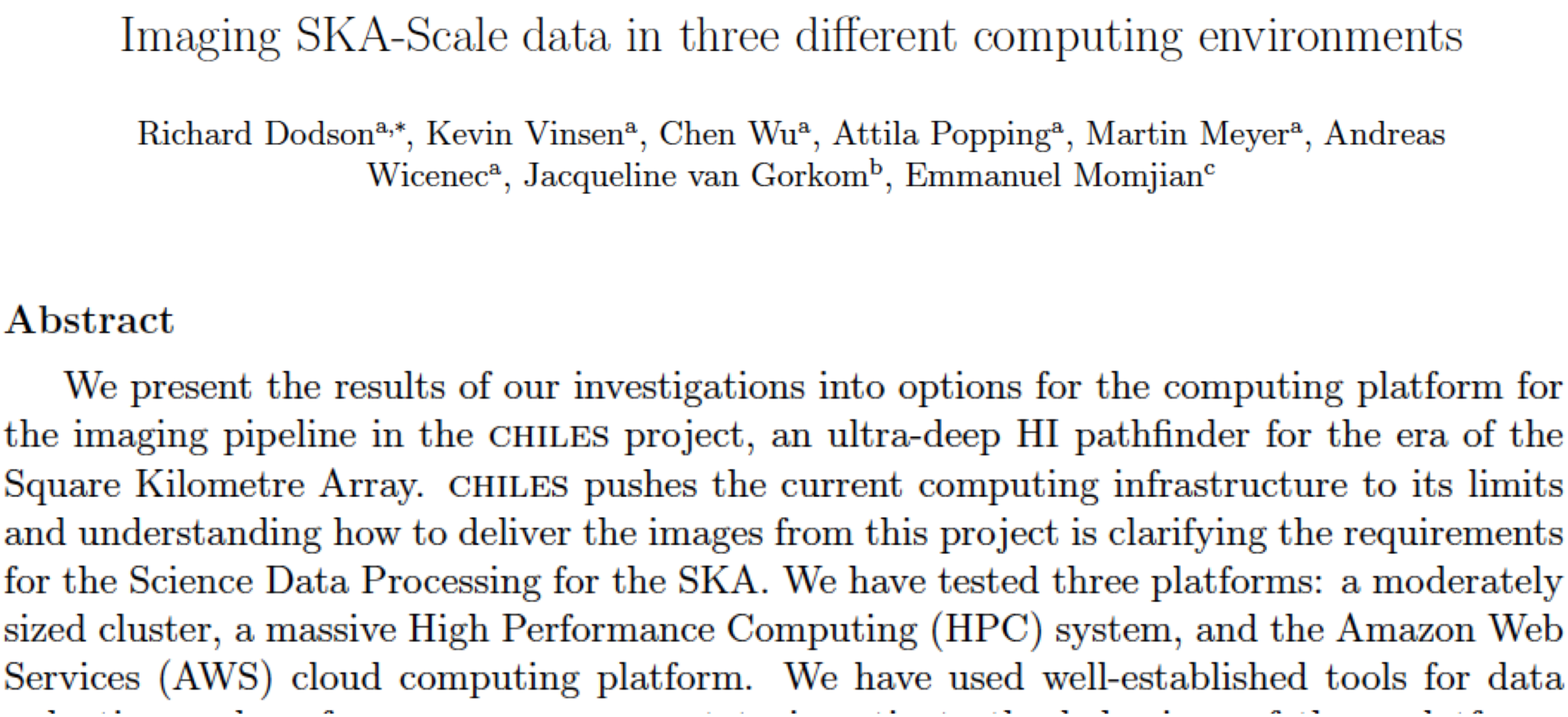


CHILES 178 hours



We do integrate down to the noise.. RFI stays same at most places or get worse, new RFI shows up. Some disappear.

Imaging in Australia

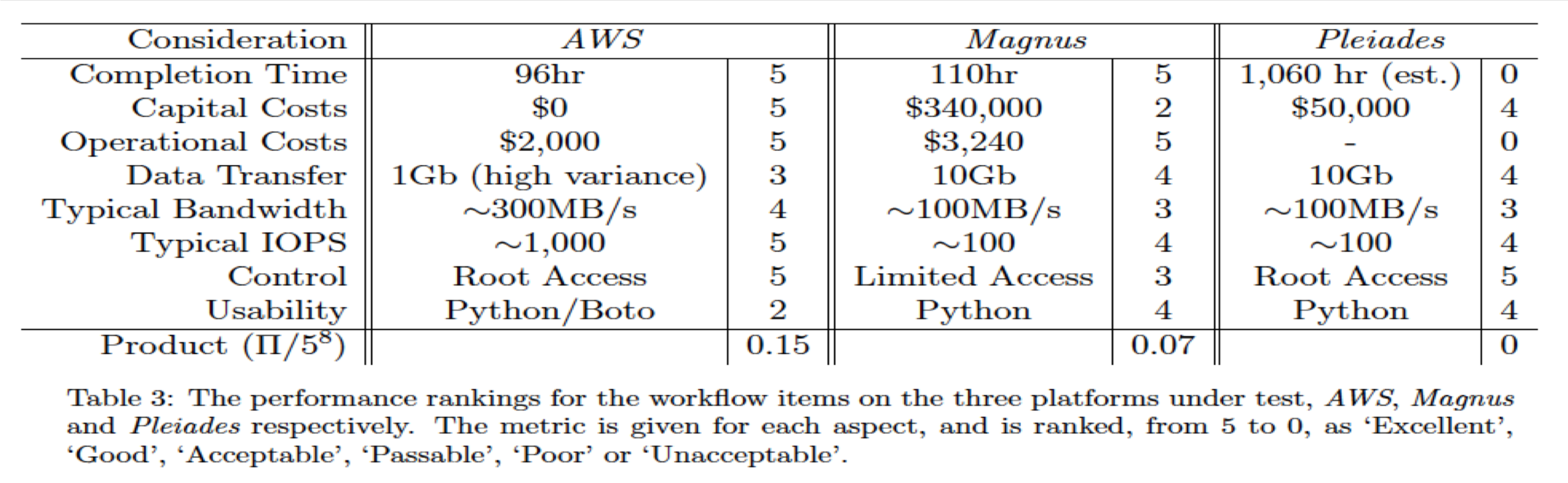


Imaging SKA-Scale data in three different computing environments

Richard Dodson[a,*], Kevin Vinsen[a], Chen Wu[a], Attila Popping[a], Martin Meyer[a], Andreas Wicenec[a], Jacqueline van Gorkom[b], Emmanuel Momjian[c]

Abstract

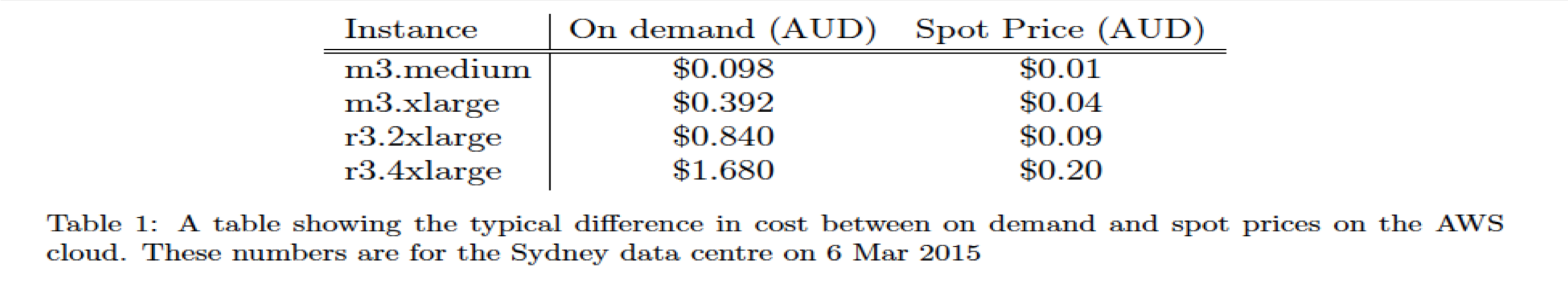
We present the results of our investigations into options for the computing platform for the imaging pipeline in the CHILES project, an ultra-deep HI pathfinder for the era of the Square Kilometre Array. CHILES pushes the current computing infrastructure to its limits and understanding how to deliver the images from this project is clarifying the requirements for the Science Data Processing for the SKA. We have tested three platforms: a moderately sized cluster, a massive High Performance Computing (HPC) system, and the Amazon Web Services (AWS) cloud computing platform. We have used well-established tools for data



Consideration	*AWS*		*Magnus*		*Pleiades*	
Completion Time	96hr	5	110hr	5	1,060 hr (est.)	0
Capital Costs	\$0	5	\$340,000	2	\$50,000	4
Operational Costs	\$2,000	5	\$3,240	5	-	0
Data Transfer	1Gb (high variance)	3	10Gb	4	10Gb	4
Typical Bandwidth	~300MB/s	4	~100MB/s	3	~100MB/s	3
Typical IOPS	~1,000	5	~100	4	~100	4
Control	Root Access	5	Limited Access	3	Root Access	5
Usability	Python/Boto	2	Python	4	Python	4
Product ($\Pi/5^8$)		0.15		0.07		0

Table 3: The performance rankings for the workflow items on the three platforms under test, *AWS*, *Magnus* and *Pleiades* respectively. The metric is given for each aspect, and is ranked, from 5 to 0, as 'Excellent', 'Good', 'Acceptable', 'Passable', 'Poor' or 'Unacceptable'.

Computing and capitalism



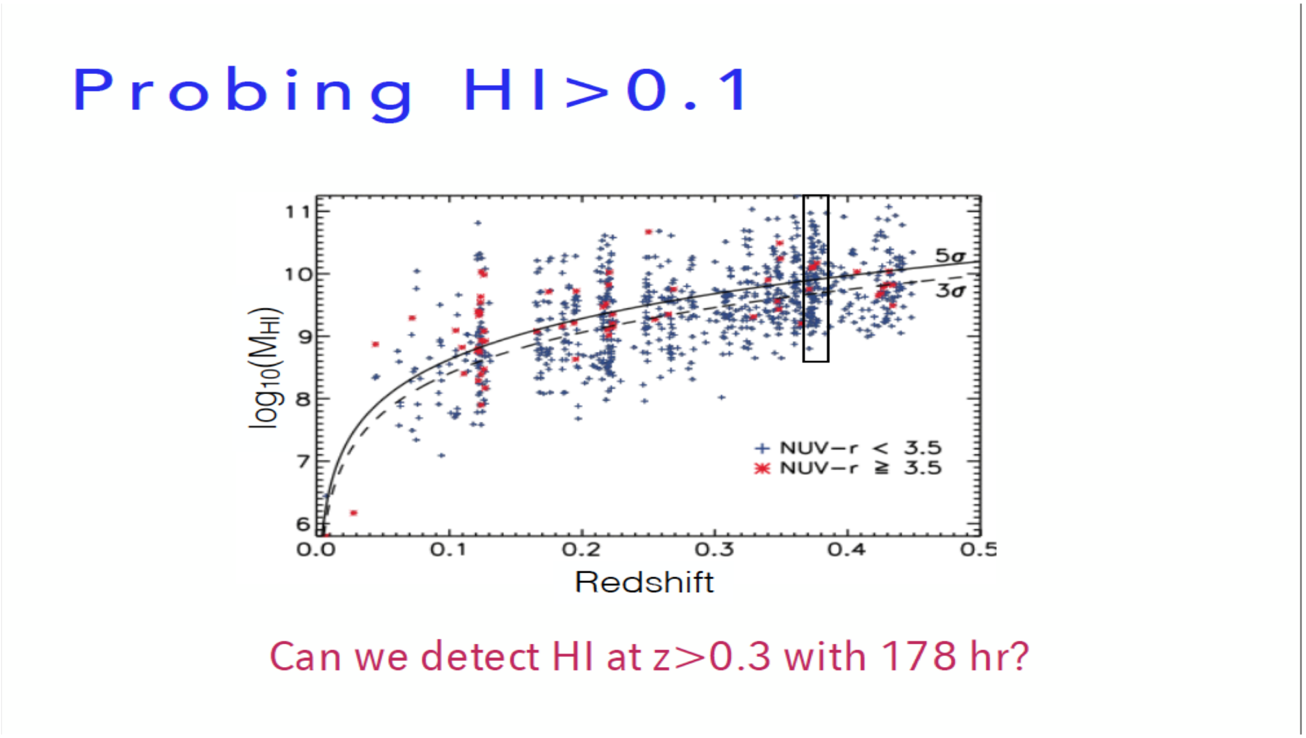
Instance	On demand (AUD)	Spot Price (AUD)
m3.medium	$0.098	$0.01
m3.xlarge	$0.392	$0.04
r3.2xlarge	$0.840	$0.09
r3.4xlarge	$1.680	$0.20

Table 1: A table showing the typical difference in cost between on demand and spot prices on the AWS cloud. These numbers are for the Sydney data centre on 6 Mar 2015

If we would do this for real, you need person to analyze efficiency instead of hardware person

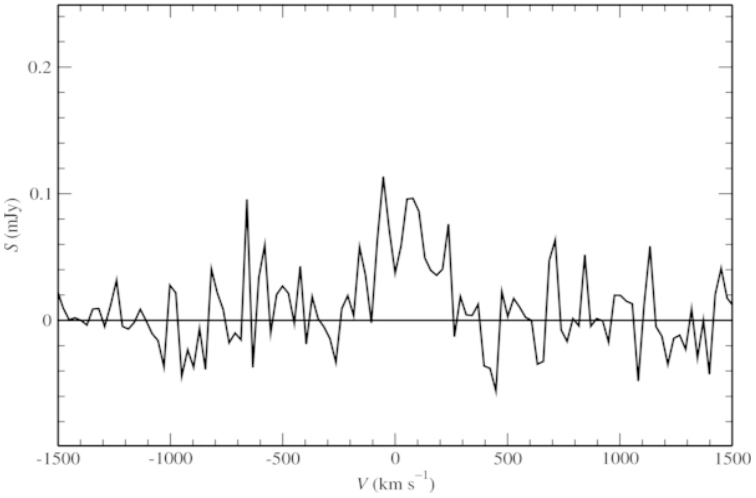
Greatest achievement so far is that we have not paid amazon a penny

Some science results of epoch 1 observations .. 178 hours



Probing HI>0.1

Can we detect HI at z>0.3 with 178 hr?

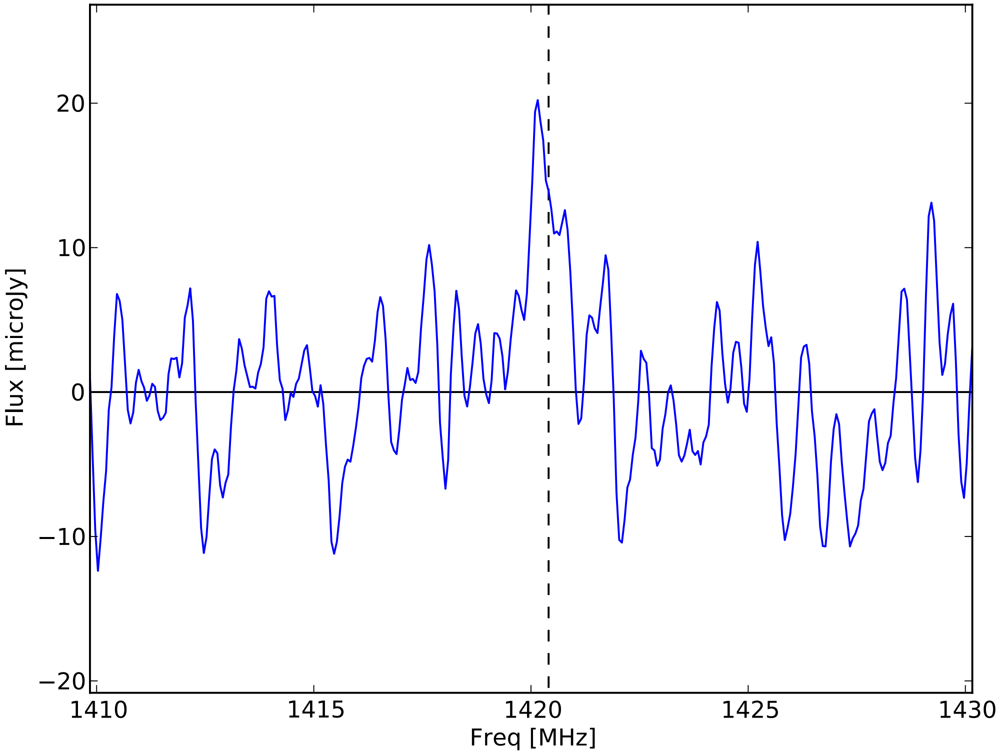


Stacking at z=0.12

M_{HI} = 1.8x10^9 M_{sun}

Stacking at z=0.37

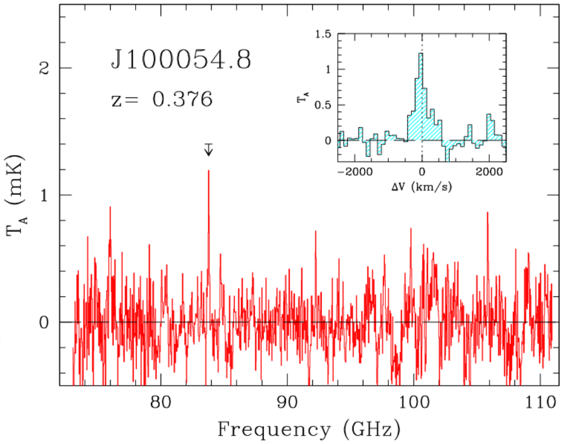
M_{HI}= 3x10^9 M_{sun}

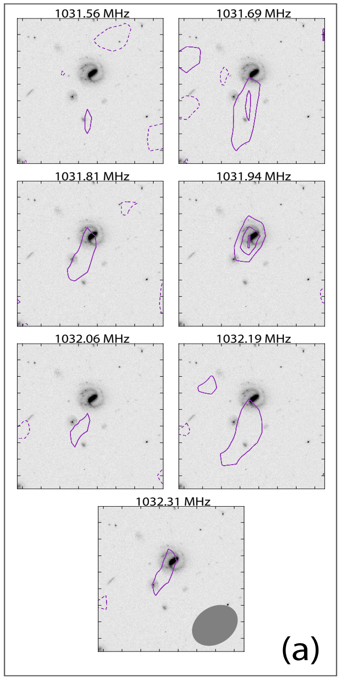


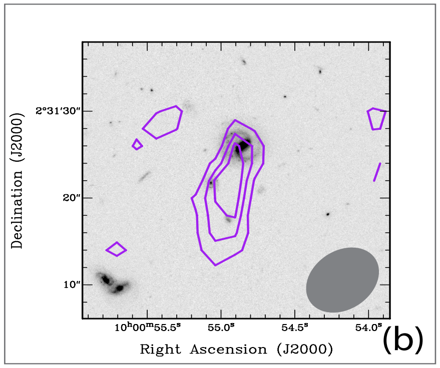
Ximena Fernández

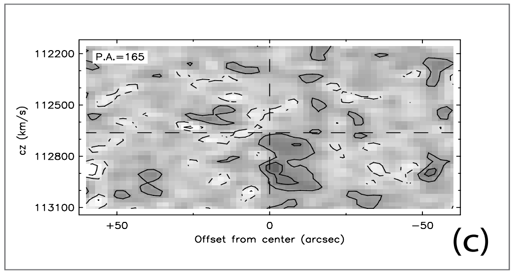
Hansung Gim

CO with LMT









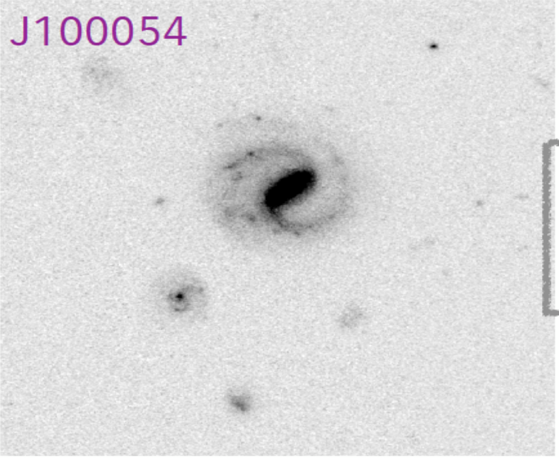
HI at z=0.376

Fernandez, Gim et al 2016, ApJ 824, L1

HI mass 3×10^{10} M_{sun}

H2 mass 5×10^{10} M_{sun}

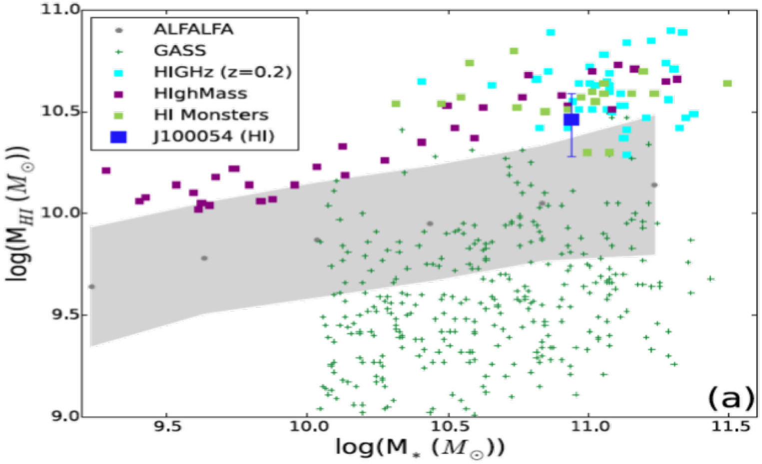
Starbursting spiral



LIRG
$M_{*} = 8.7 \times 10^{10}\,M_{\odot}$
$SFR_{IR} = 85\,M_{\odot}/yr$

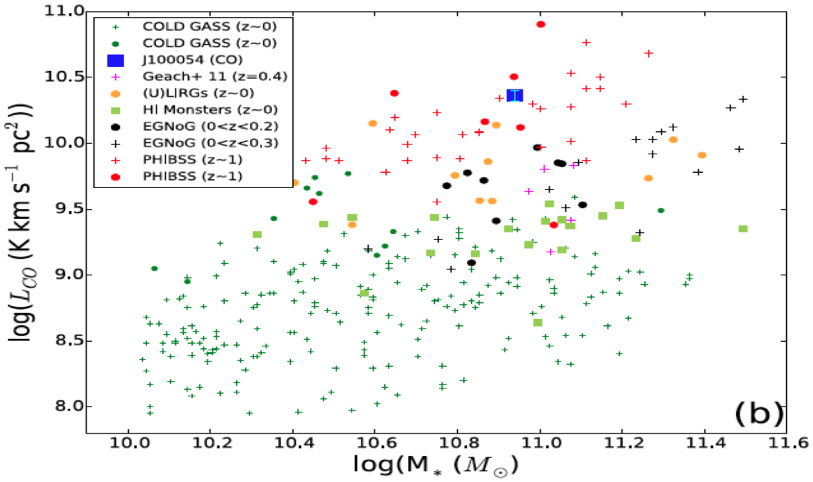
Star formation is asymmetric

Gas-rich: HI



Comparable to other gas-rich galaxies

Gas-rich: CO



Only comparable in CO to PHIBBS (z~1)

YF+

First detection is a very interesting system

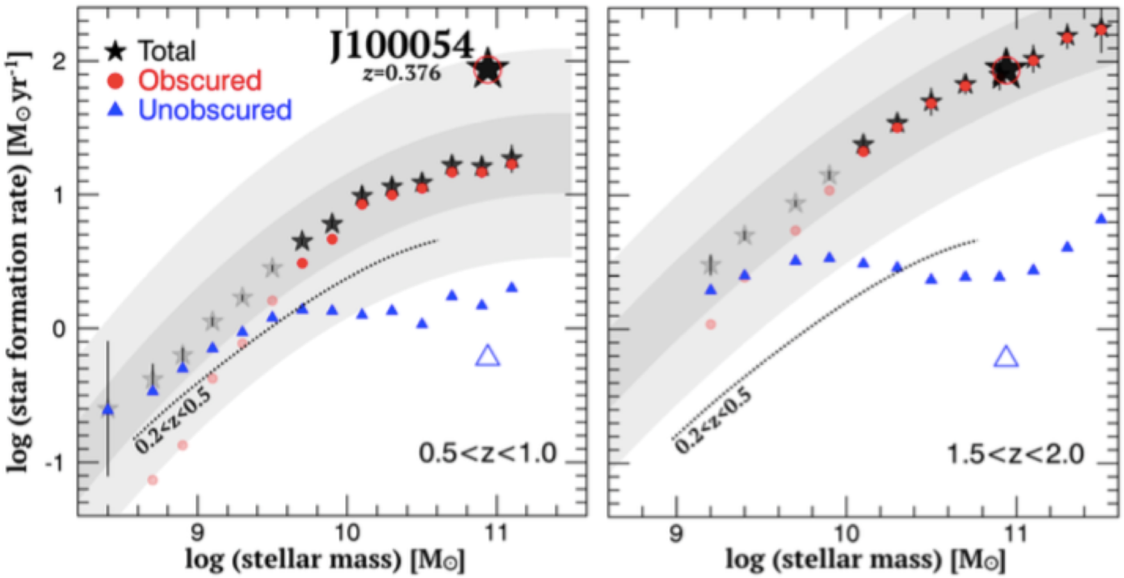
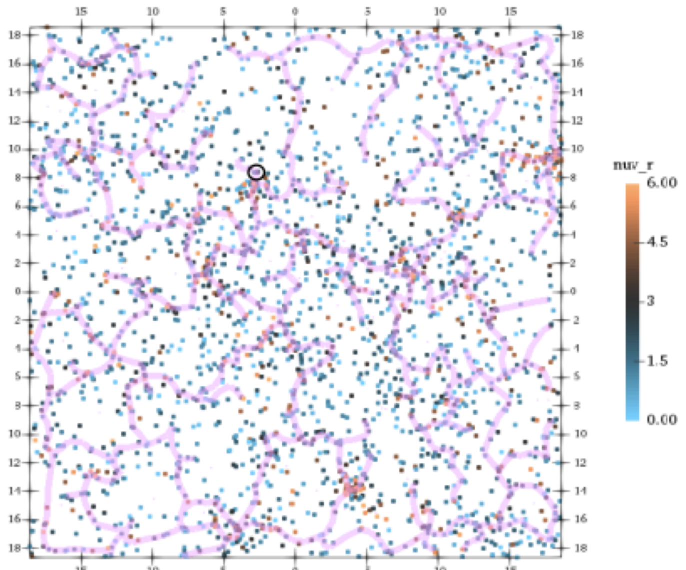


Figure 1: The high SFR of J100054 is consistent with the average log(SFR)-log($M_\star$) relation at $1.5<z<2.0$ in the right panel from Whitaker et al. (2014). The total SFR is divided into the

Reminiscent of clumpy disks at z>1 asymmetric SF and very gas rich
Possible evidence for gas infall and off nucleus enhanced starformation?

Identifying the Cosmic Web in the CHILES volume

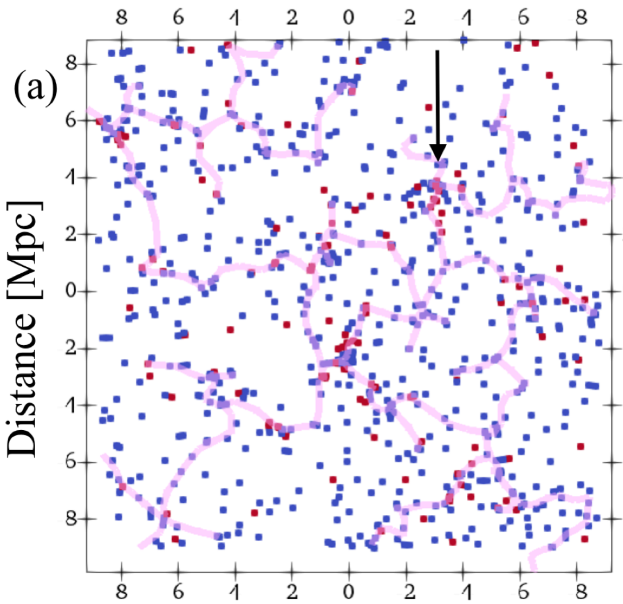


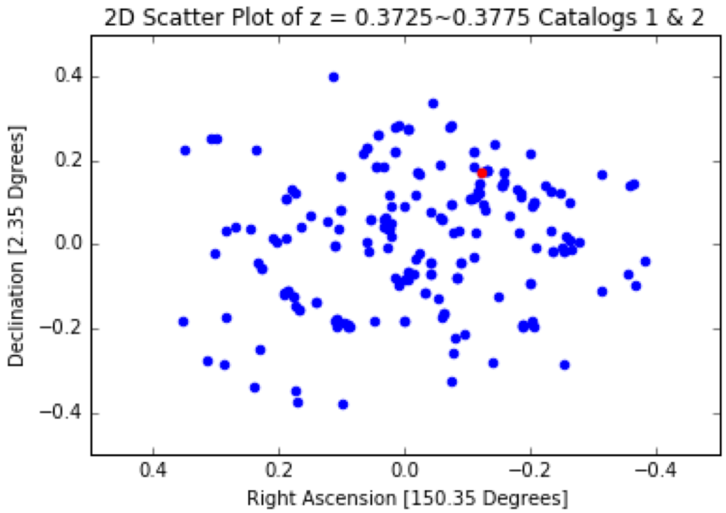
0.37<z<0.38

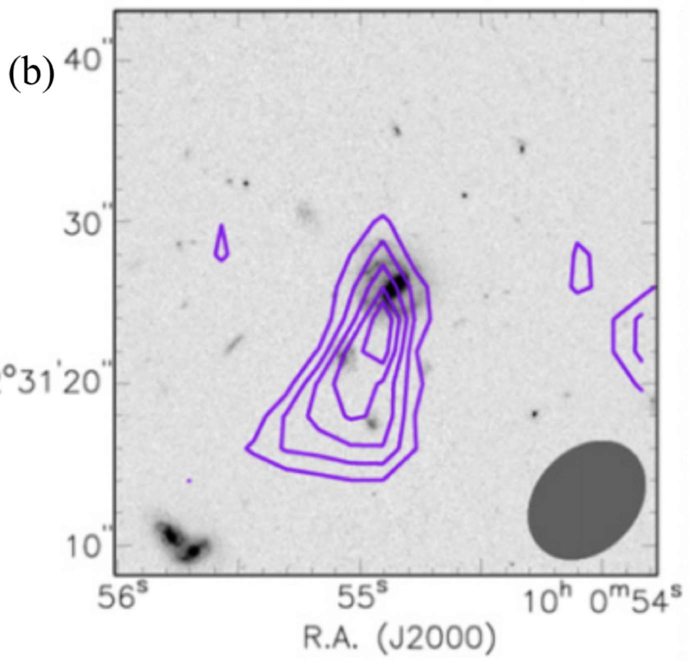
Cosmic web identified by DisPerSE

DisPerSE - a scale free, topological structure finding algorithm

Luber et al 2018, submitted

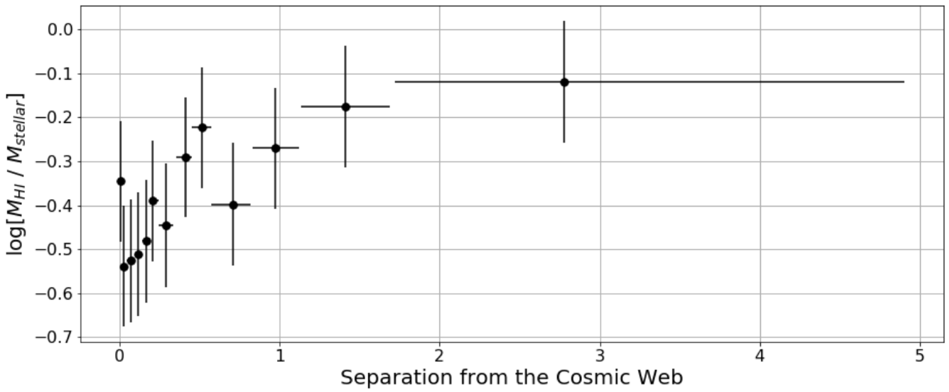




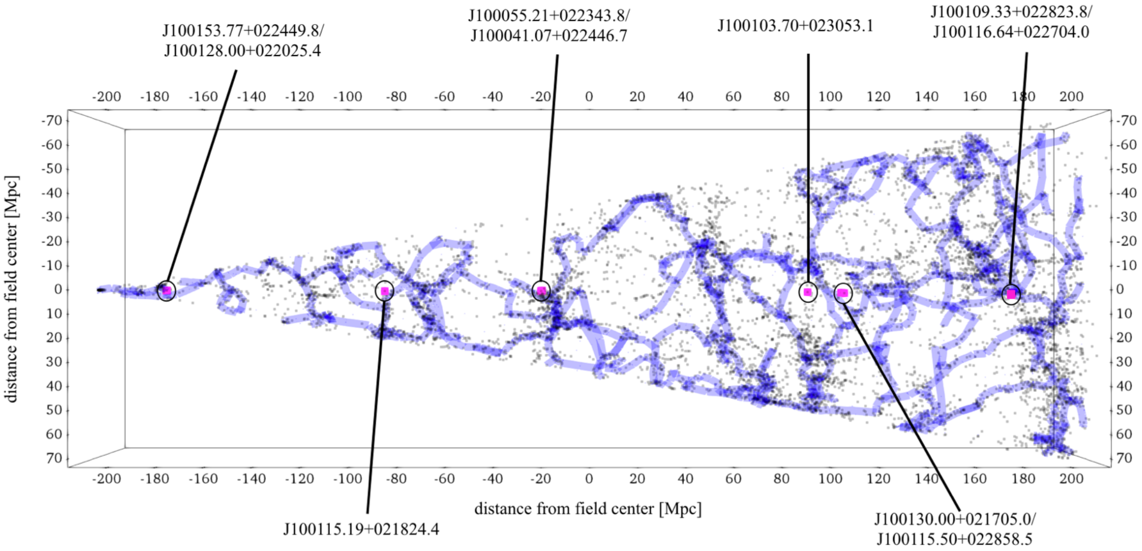


Z=0.376 galaxy extended in direction of filament

Using DisPerSE to define LSS in CHILES volume Luber et al 2018, submitted

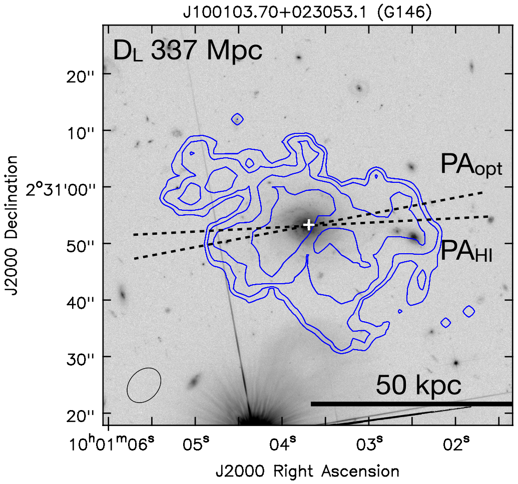


Predicted gas fraction as function of distance from filaments in Mpc



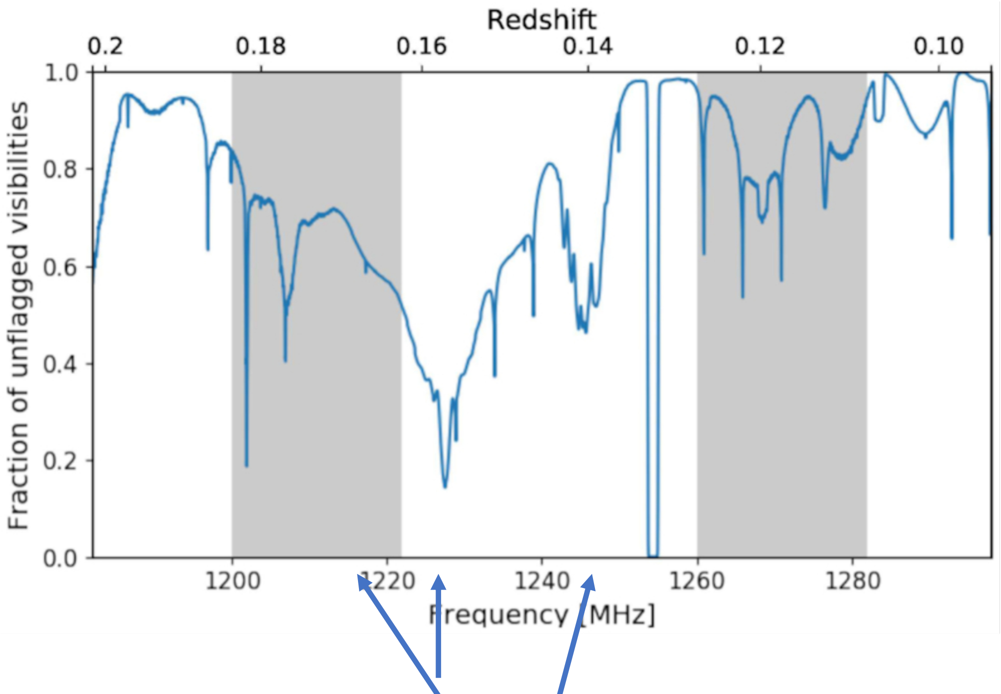
Gross, Davis et al, in prep

178 hours on lowest velocity range. Interesting results on nearby galaxies



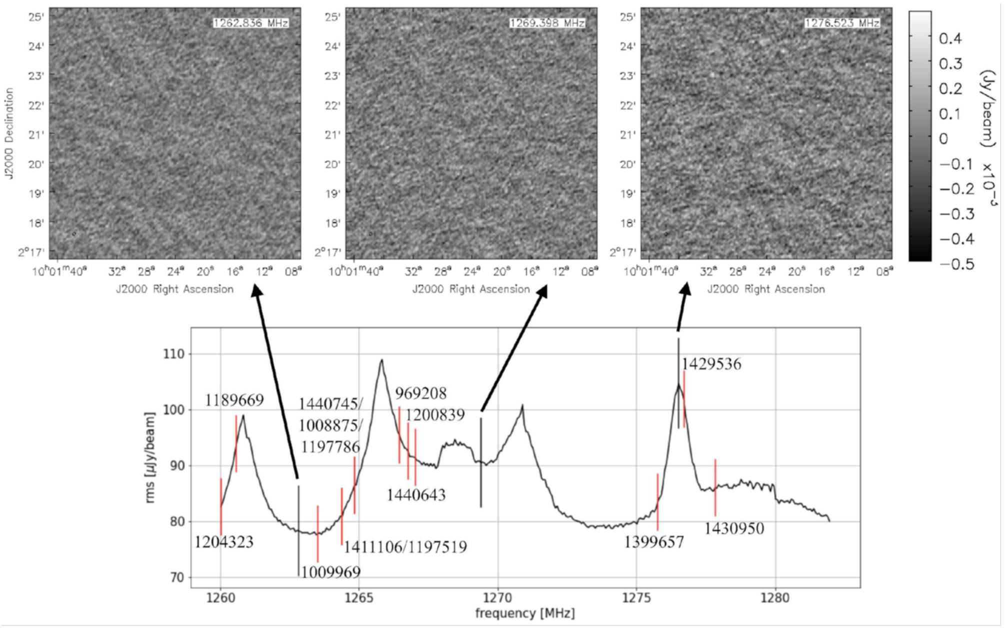
N_{HI} (1σ) = 4.3E+19 cm^{-2}, contours start at 2σ

Doing the impossible

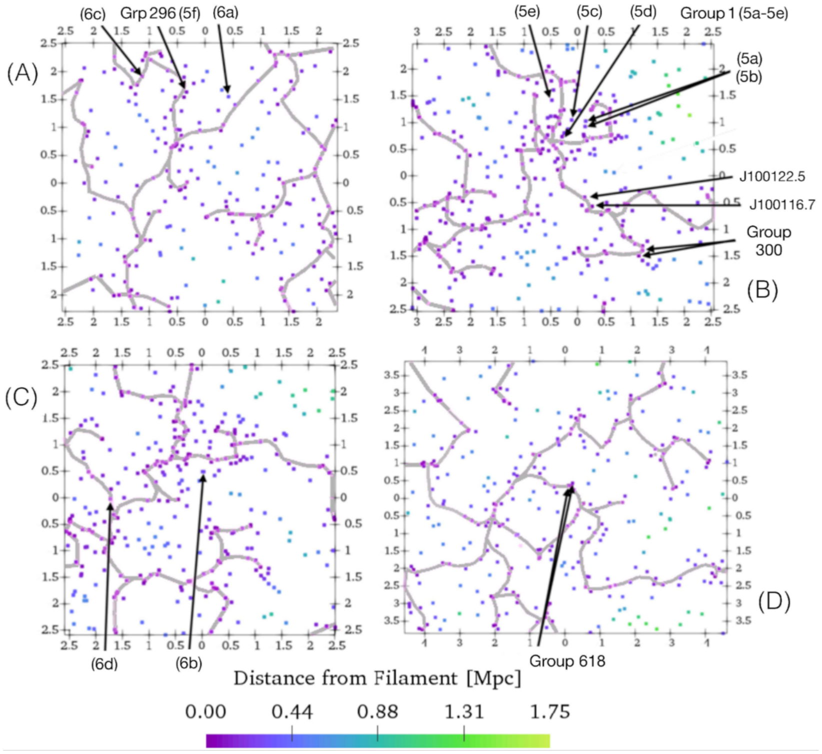


GPS Glonass

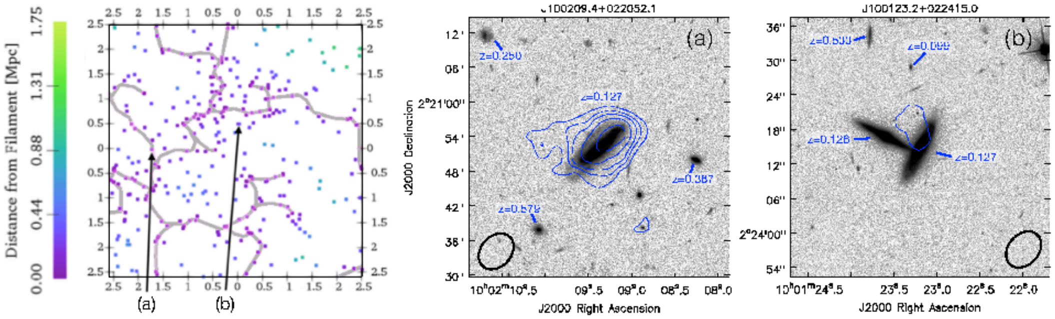
Hess et al 2019, MNRAS look at worst possible place in frequency



Rms varies due to frequency dependent flagging, but final images appear to be noise dominated


(A)
(6c)
Grp 296 (5f)
(6a)
(B)
(5e)
(5c)
(5d)
Group 1 (5a-5e)
(5a)
(5b)
J100122.5
J100116.7
Group
300
(C)
(6d)
(6b)
(D)
Group 618
2.5 2 1.5 1 0.5 0 0.5 1 1.5 2 2.5
3 2.5 2 1.5 1 0.5 0 0.5 1 1.5 2 2.5
4 3 2 1 0 1 2 3 4
3.5 3 2.5 2 1.5 1 0.5 0 0.5 1 1.5 2 2.5 3 3.5
Distance from Filament [Mpc]
0.00
0.44
0.88
1.31
1.75

HI morphology at z=0.12 and z=0.17, Hess, et al 2019, MNRAS



Current Status

Data taking will be complete by June.

We think we can do reasonable job in calibrating, use of masks help quality and people

Still many imaging issues to be dealt with.. Subtraction of sources very far from field center

Strongly varying PSF due to frequency dependent flagging remains a challenge

Release of cube of epoch 1 will be done this summer

Some thoughts..

Short spacing are important for HI..
Debugging instrument and software at same time is not a good idea

We probably should not do what “can be done”, but start with what we need for the science
For example use much lower velocity resolution, to improve science and reduce data volume

Reduction of data volume at all steps is important

Future looks great indeed

SKA path finders have started taking data

These are wide area survey telescopes

ASKAP and Apertif have about 15 arcsec resolution.. Will image about 300 galaxies a day over entire sky out to z=0.2

MeerKAT will go deeper .. Direct imaging to z=0.5, stacking to z=1.2
resolution eventually comparable to JVLA

GMRT and MeerKAT both have probably bettert baseline distribution for HI

Square Kilometre Array will image HI at redshifts beyond z=1