Data reduction strategy
for the Effelsberg-Bonn HI Survey

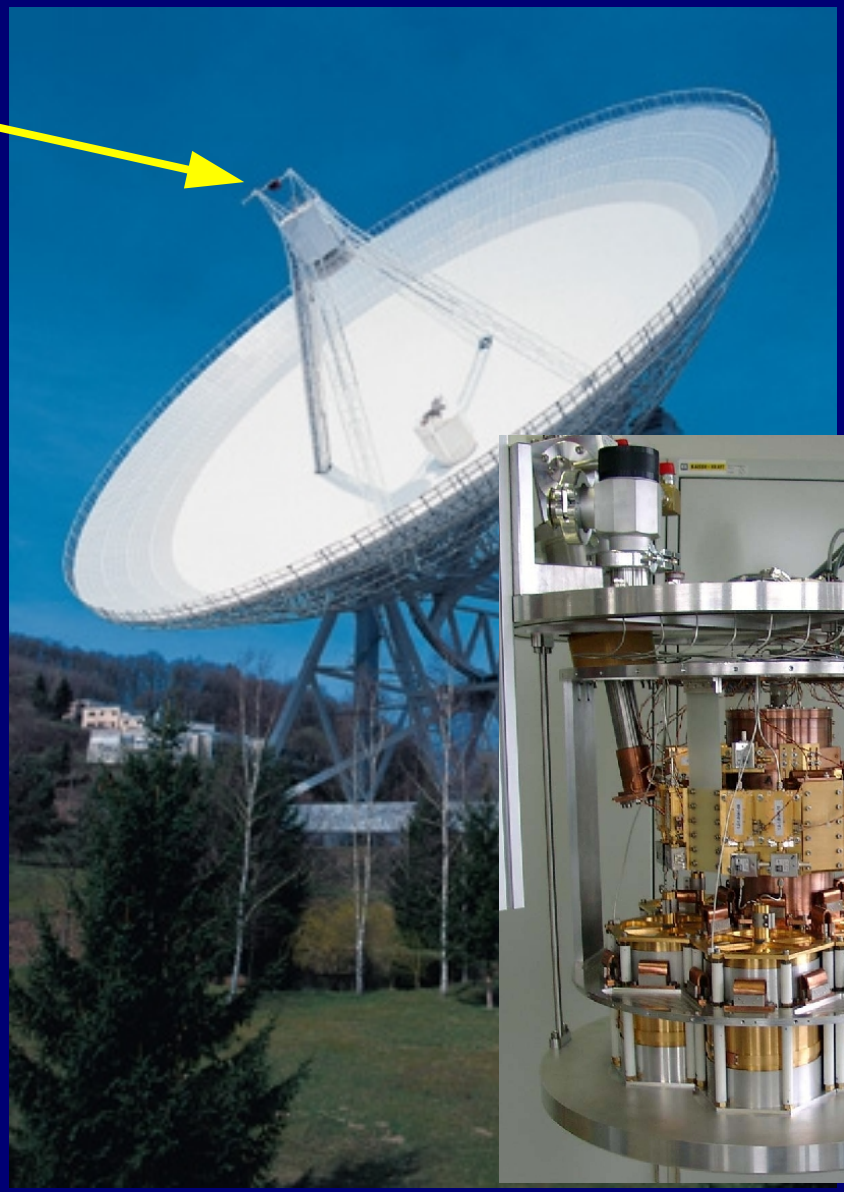
Benjamin Winkel

Max-Planck-Institut für Radioastronomie

Argelander-Institut für Astronomie

The Effelsberg-Bonn HI Survey (EBHIS)

7-Beam L Band Receiver (21-cm)



Team:

J. Kerp, B. Winkel,
P. Kalberla, R. Keller,
A. Zensus

Collaborators:

R. Braun, U. Hopp,
B. Koribalski, M. Kramer,
T. Oosterloo, M. Putman,
P. Richter, F. Walter

Overview

- Data reduction
  - Instrumental parameters
  - Reduction scheme
- Simulation of survey data
  - Influence of RFI
  - Systematic effects
- First observations

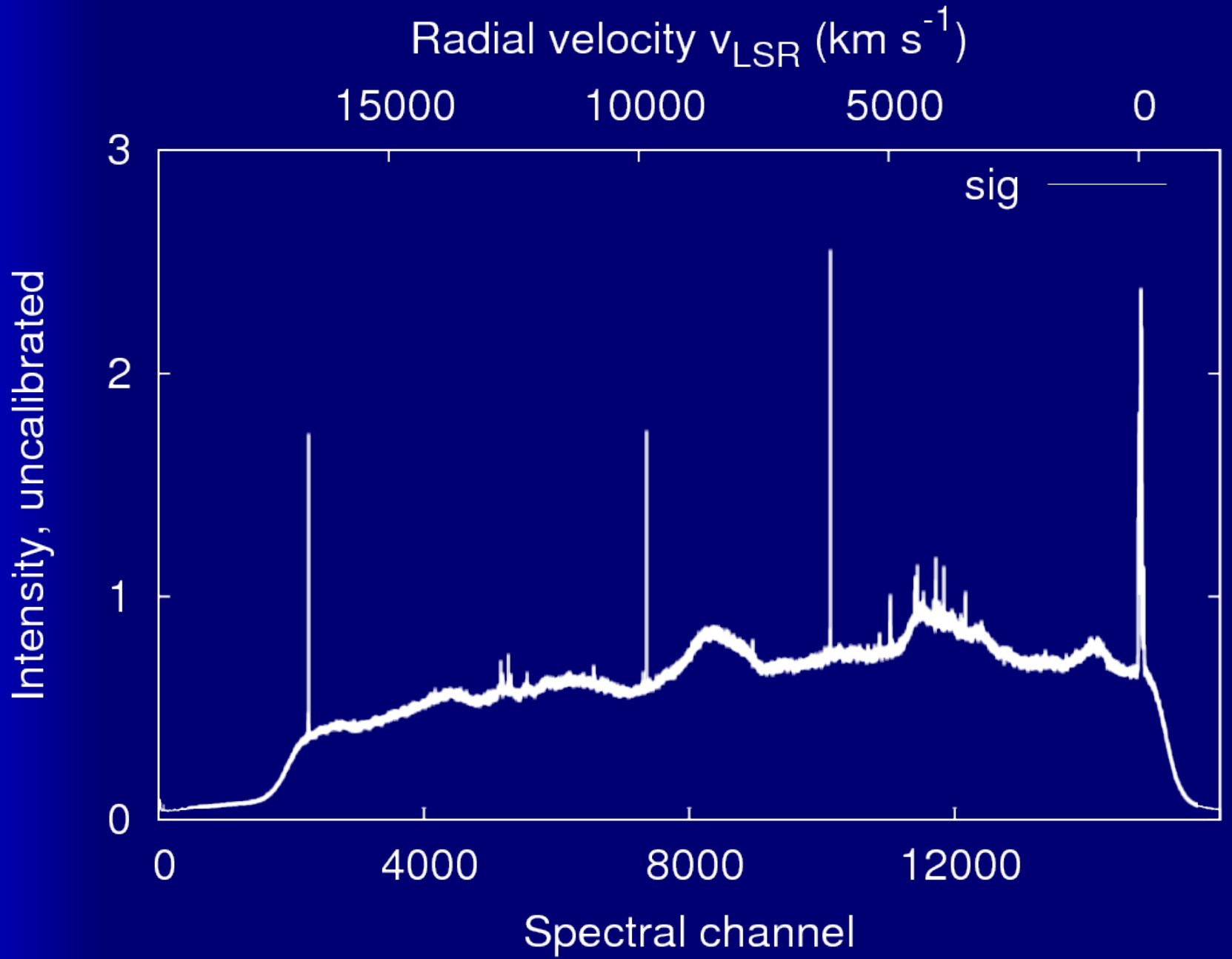
Data Reduction

Survey parameters

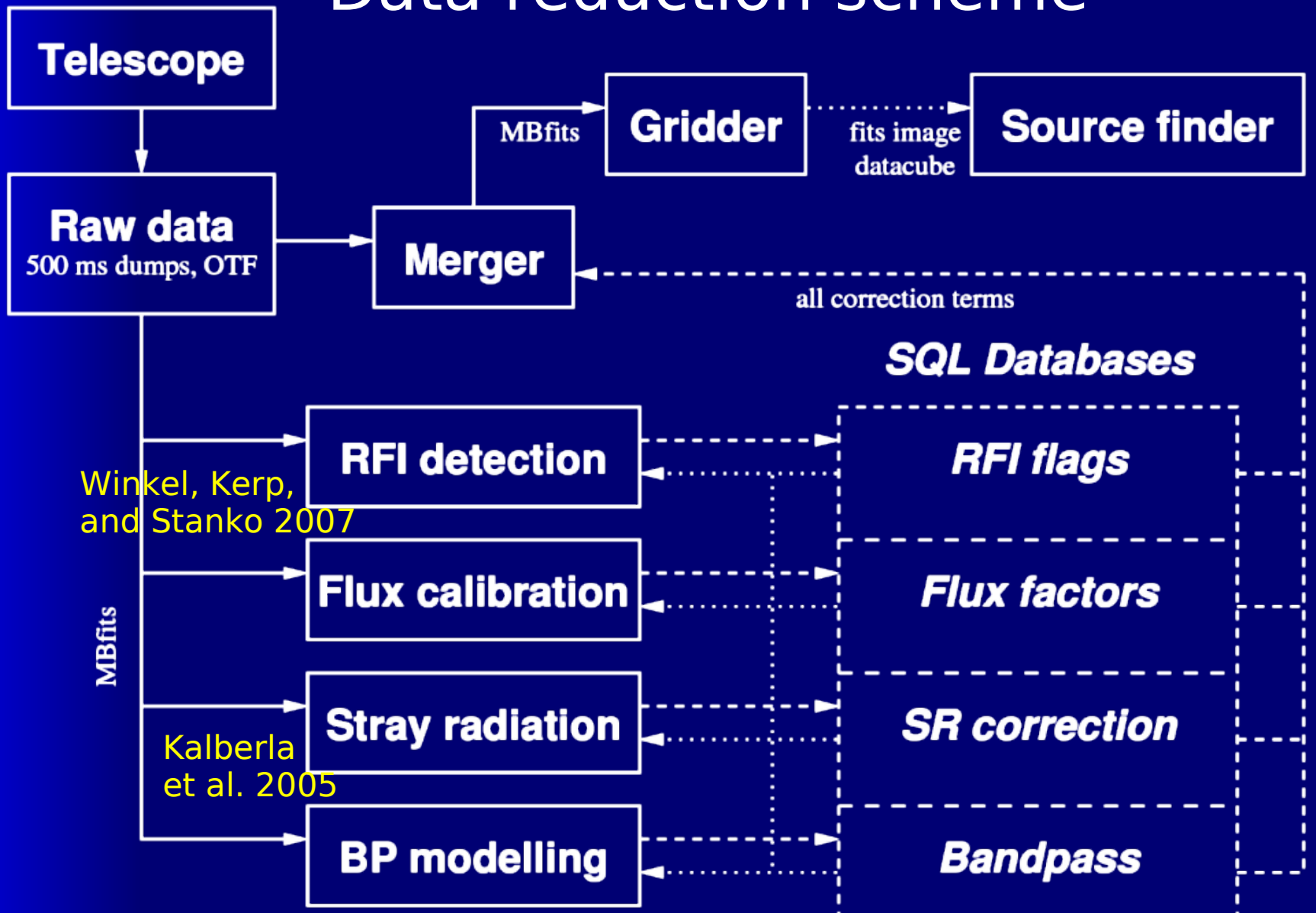
- 14 FPGA based spectrometers (16k channels)
- Bandwidth 100 MHz (1.2 km/s resolution)
- 10' FWHM, fully sampled
- 500 ms dumps (→ RFI detection)
- 10^7 $M_\odot$ detection limit at 16 Mpc

Receiver/data quality

- System temperature: 20-25 K
- Allan times: > 1000 s
- Gain curve stability: good
- Standing wave pattern


Radial velocity v_{LSR} (km s^{-1})
15000
10000
5000
0
sig
Intensity, uncalibrated
3
2
1
0
0
4000
8000
12000
Spectral channel

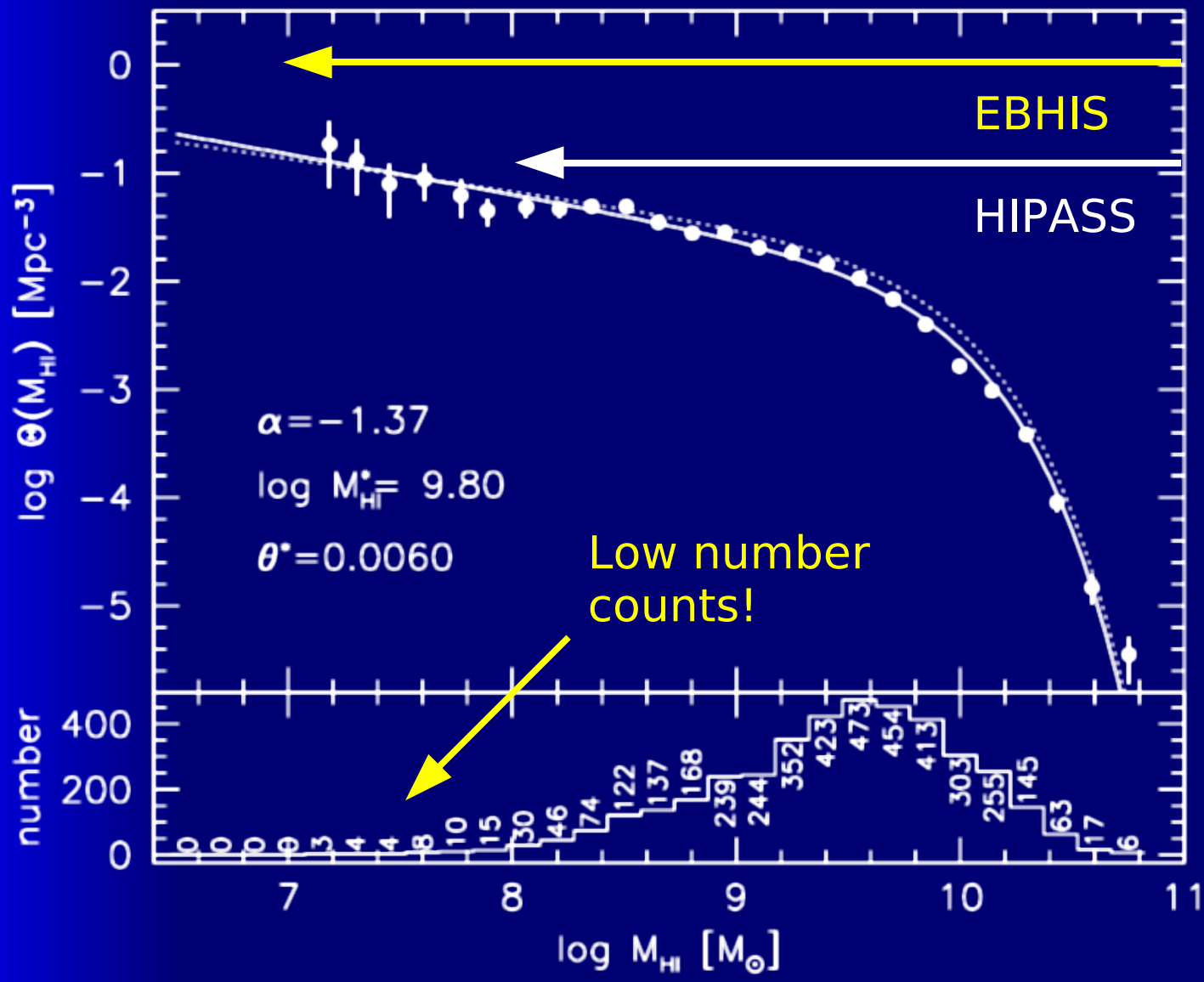
Data reduction scheme



Simulations

- Test the data reduction software and algorithms
- What is the impact of RFI on the results?
- Investigate possible bias/selection-effects
- Determine completeness function
- Other statistical tests

The HI mass function

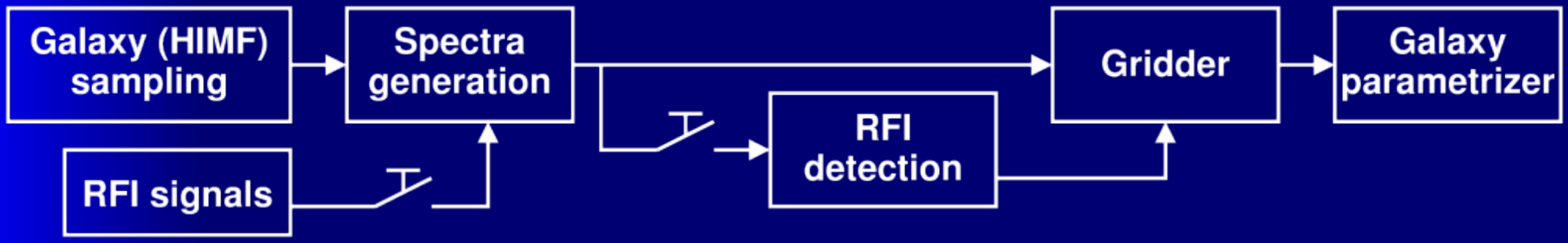


Local universe:

$\Omega_{HI}=3.5 \bullet 10^{-4}\, h^{-1}$

Zwaan et al. 2005

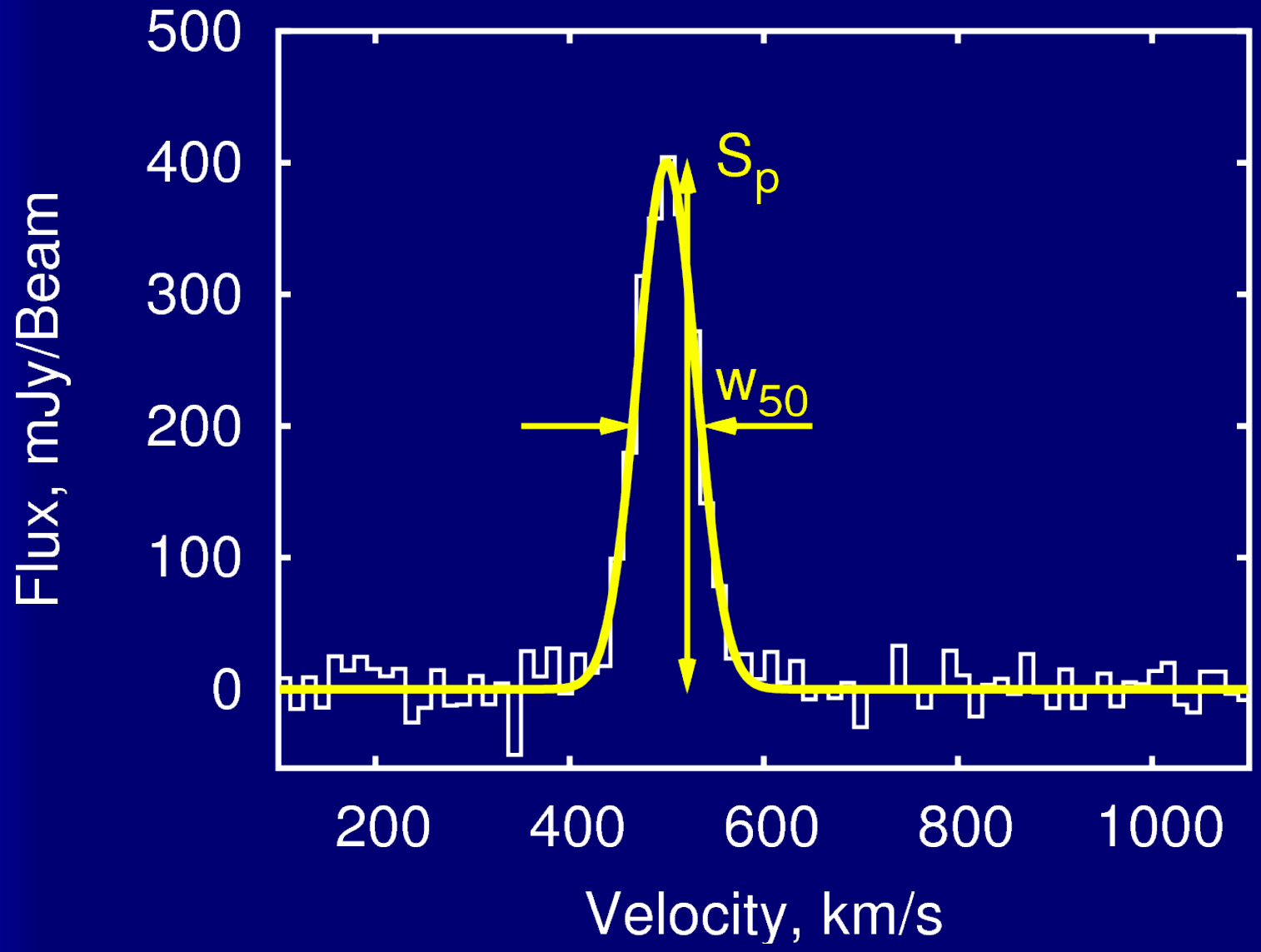
Simulations - Setup



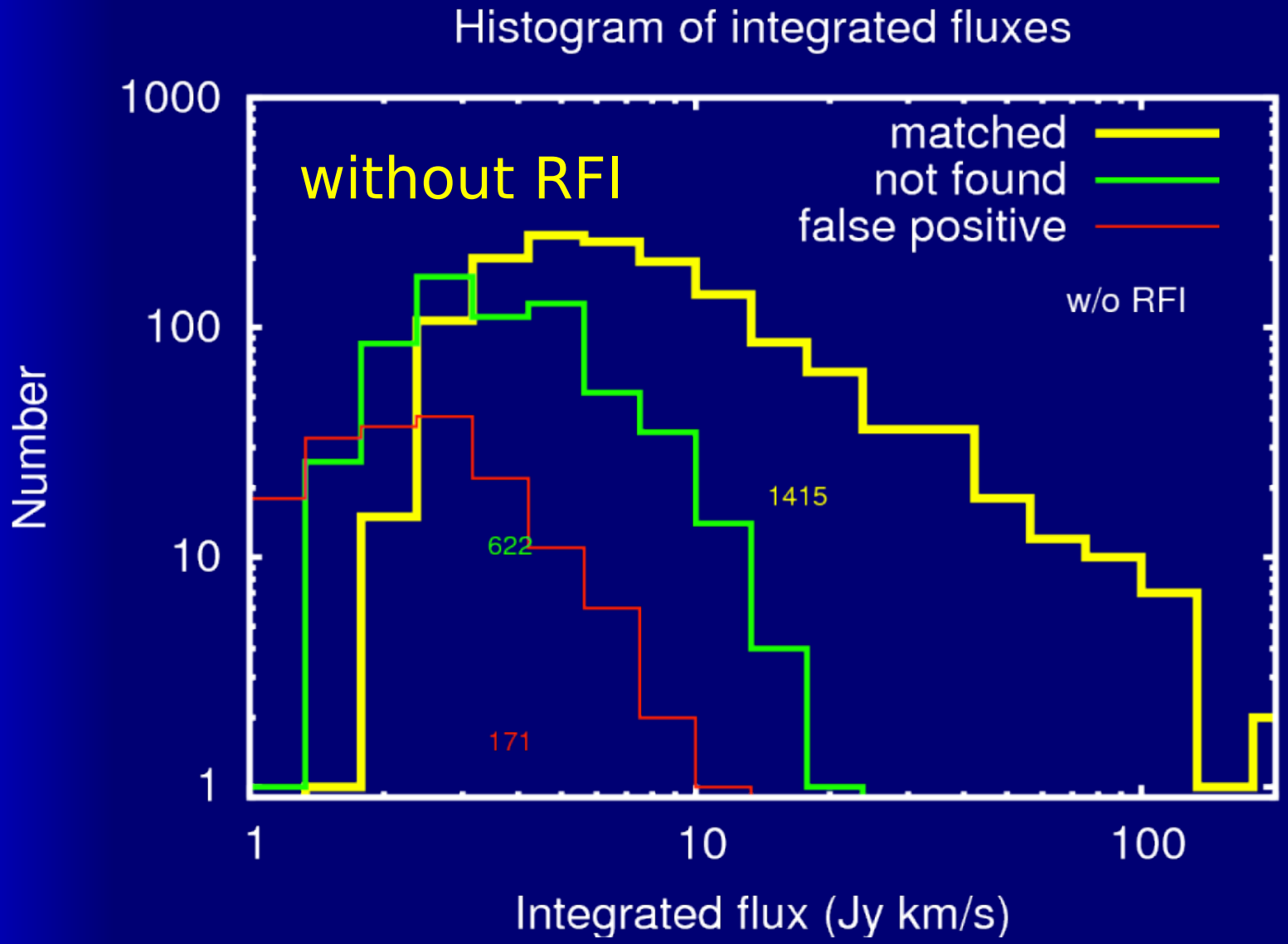
Some numbers:

- ~120 data cubes: 10°x10°x160 Mpc (150x150x1024)
- 1 Mio raw spectra per cube (8 GB)
- Three runs:
  - w/o RFI
  - with RFI, w/o mitigation
  - with RFI, with mitigation

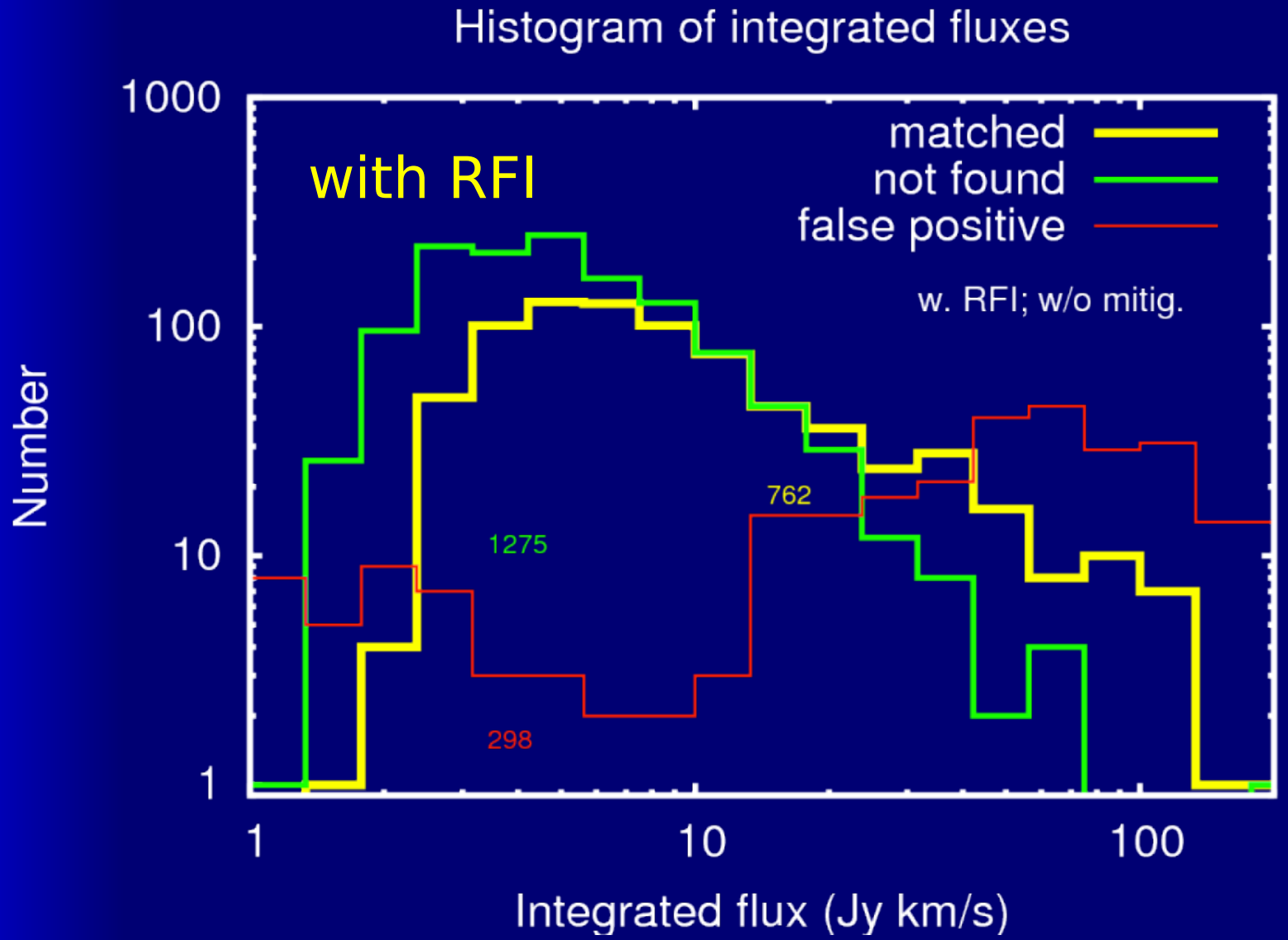
Source Parametrization (HIPASS)



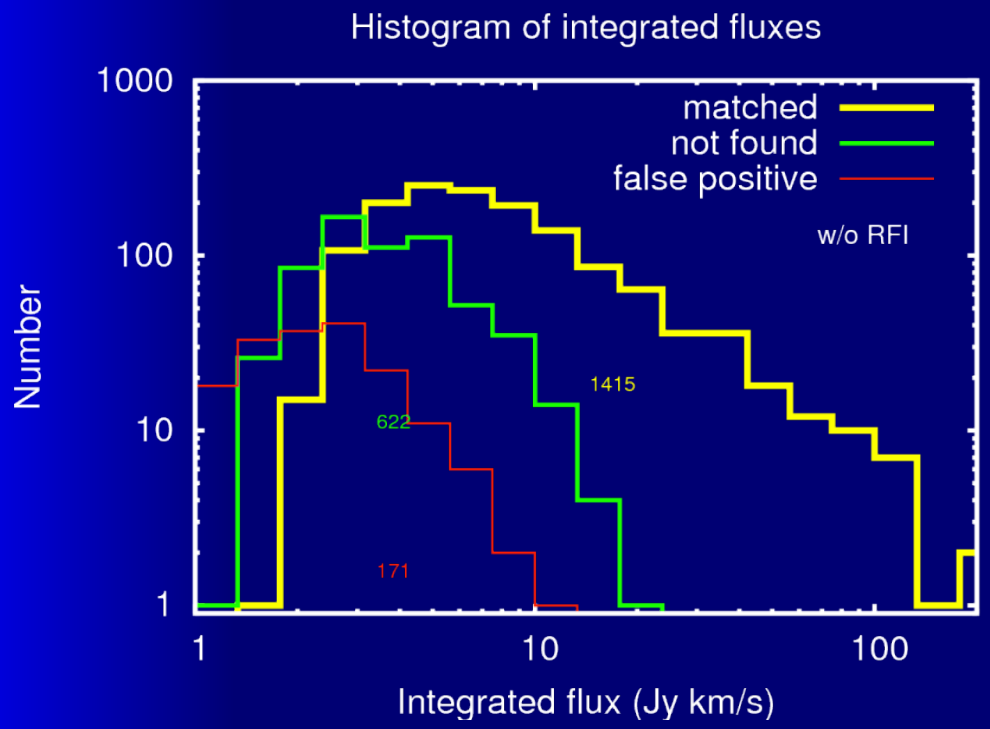
Simulations - Results



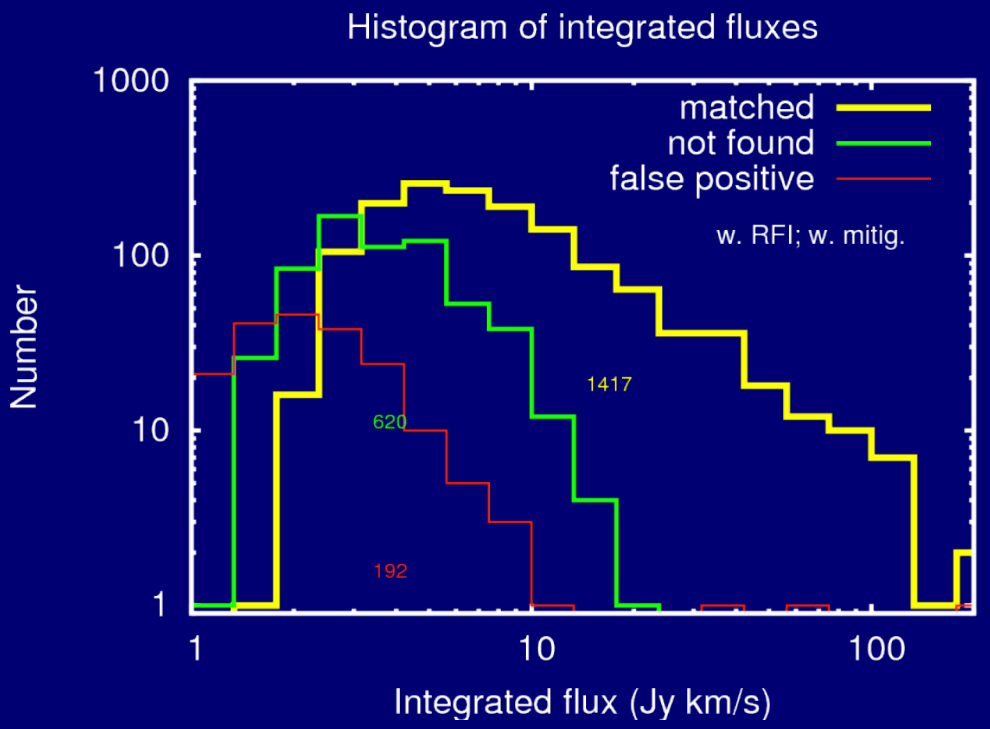
RFI destroys HIMF!



RFI mitigation works!



without RFI



with RFI mitigation

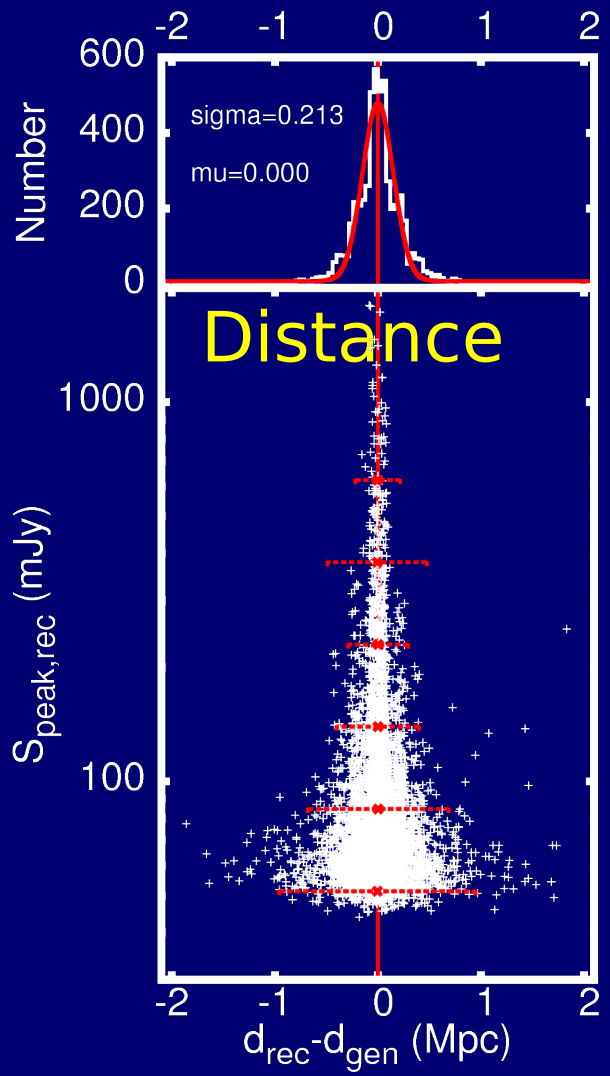
Systematic Effects

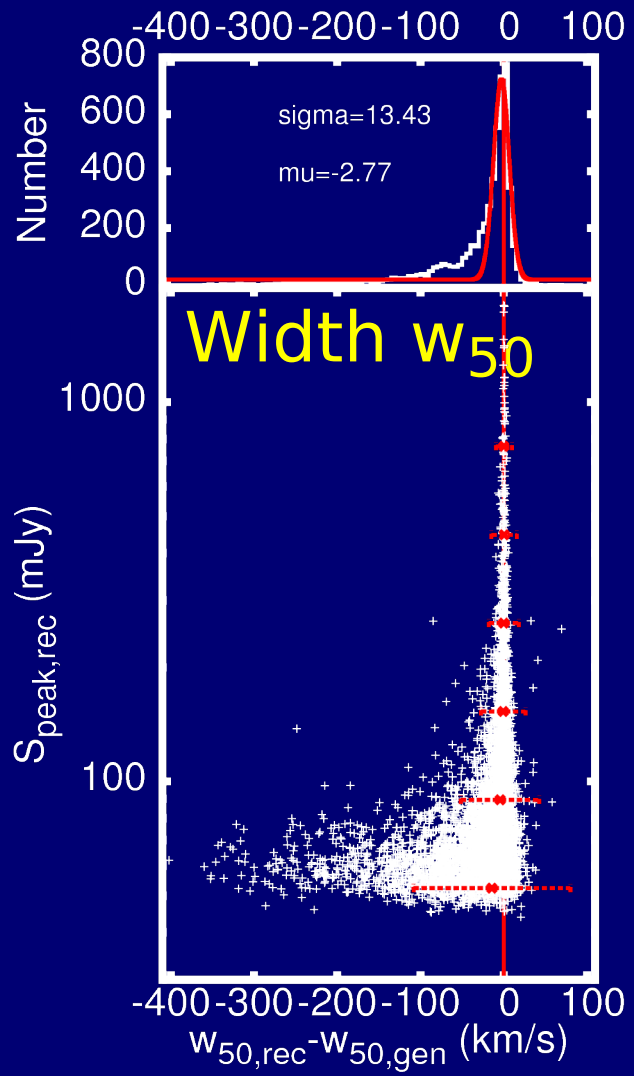
Distance d:

- No bias

Velocity profile width w_{50}:

- Strong under-estimation for faint sources





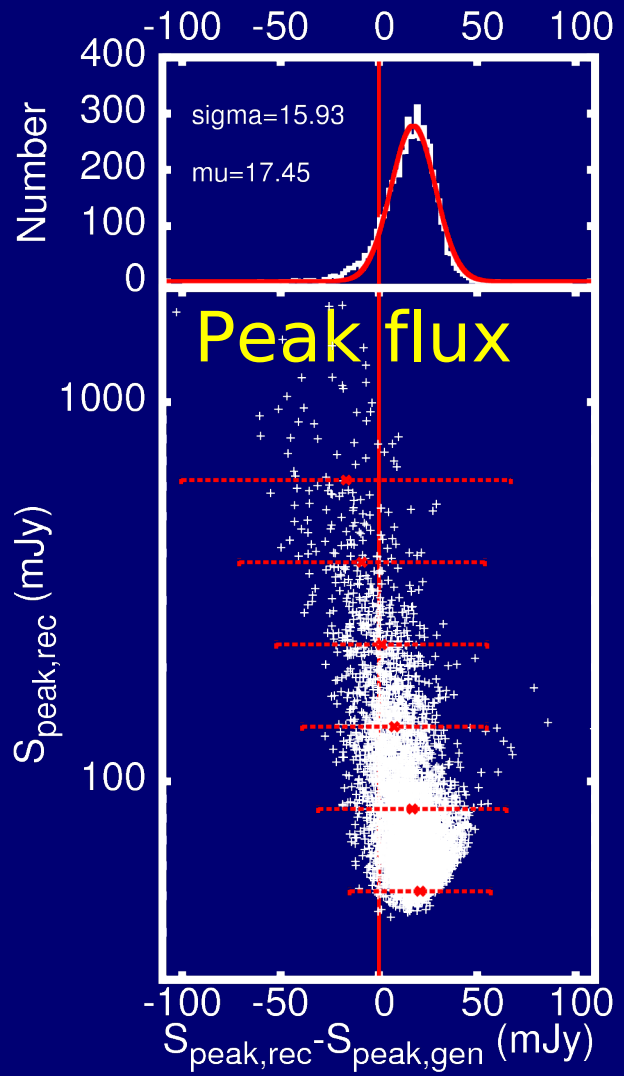
Systematic Effects

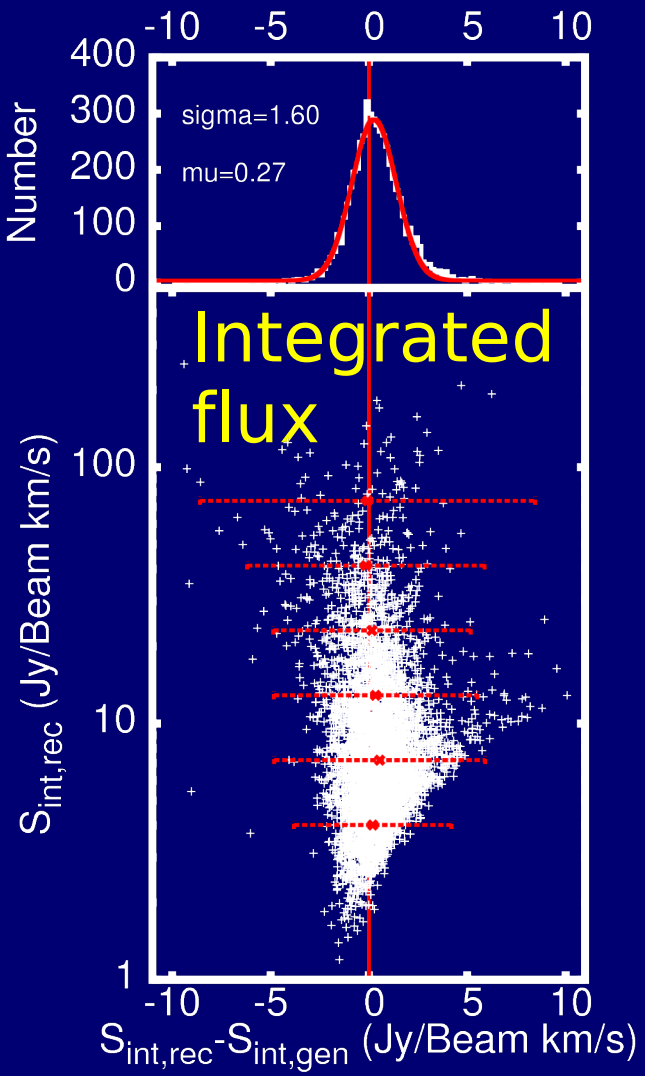
Peak flux:

- Overestimation for low fluxes
- Underestimation for high fluxes

Integrated flux:

- Small excess for intermediate fluxes

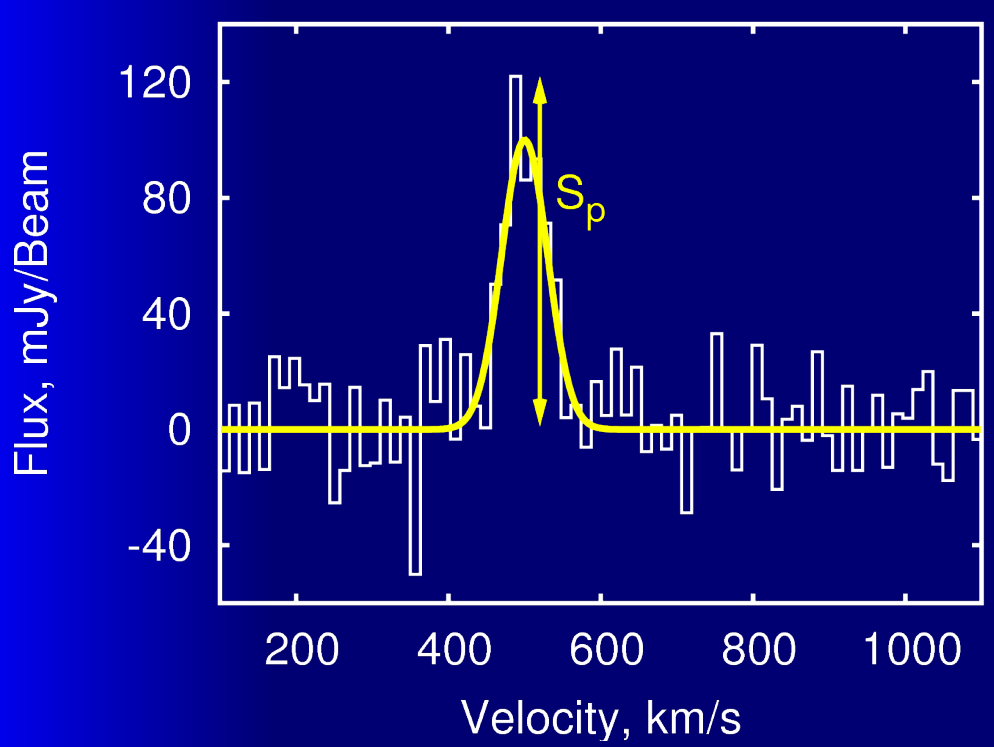


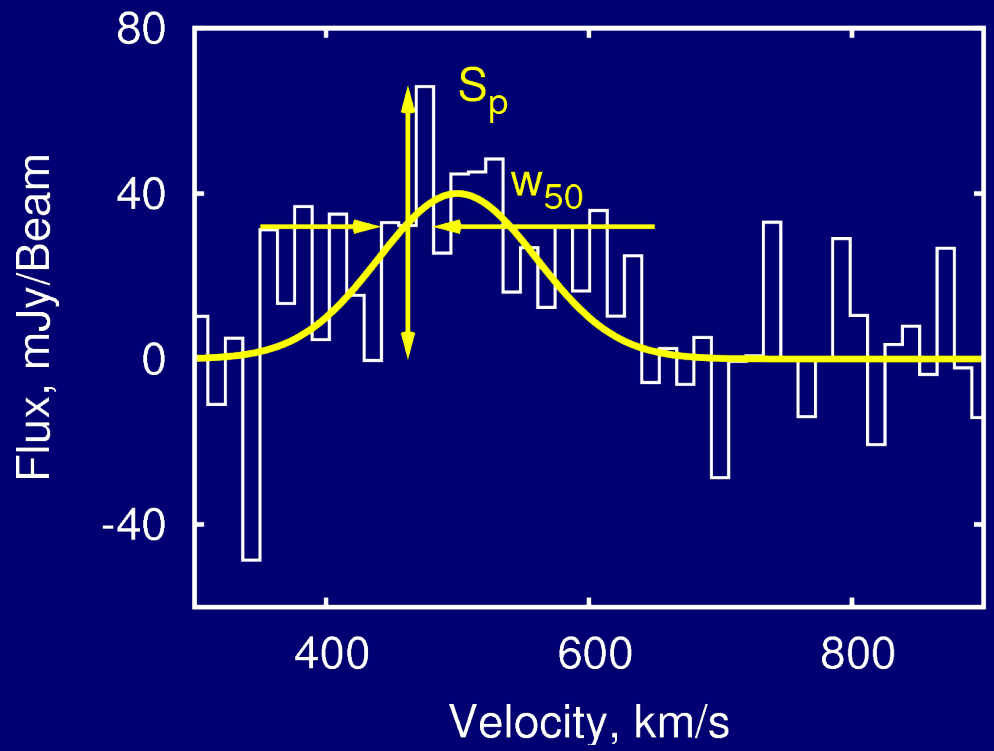


Systematic Effects

- Parametrization issues
  - The peak value of a spectrum is a bad estimator for the peak flux
  - For small peak fluxes the velocity width is often heavily underestimated
- Selection effects
  - Incompleteness of the survey (correction possible)
  - Integrated flux can be biased to higher values
- Other
  - Aliasing
  - ...

Systematic Effects – Parametrization

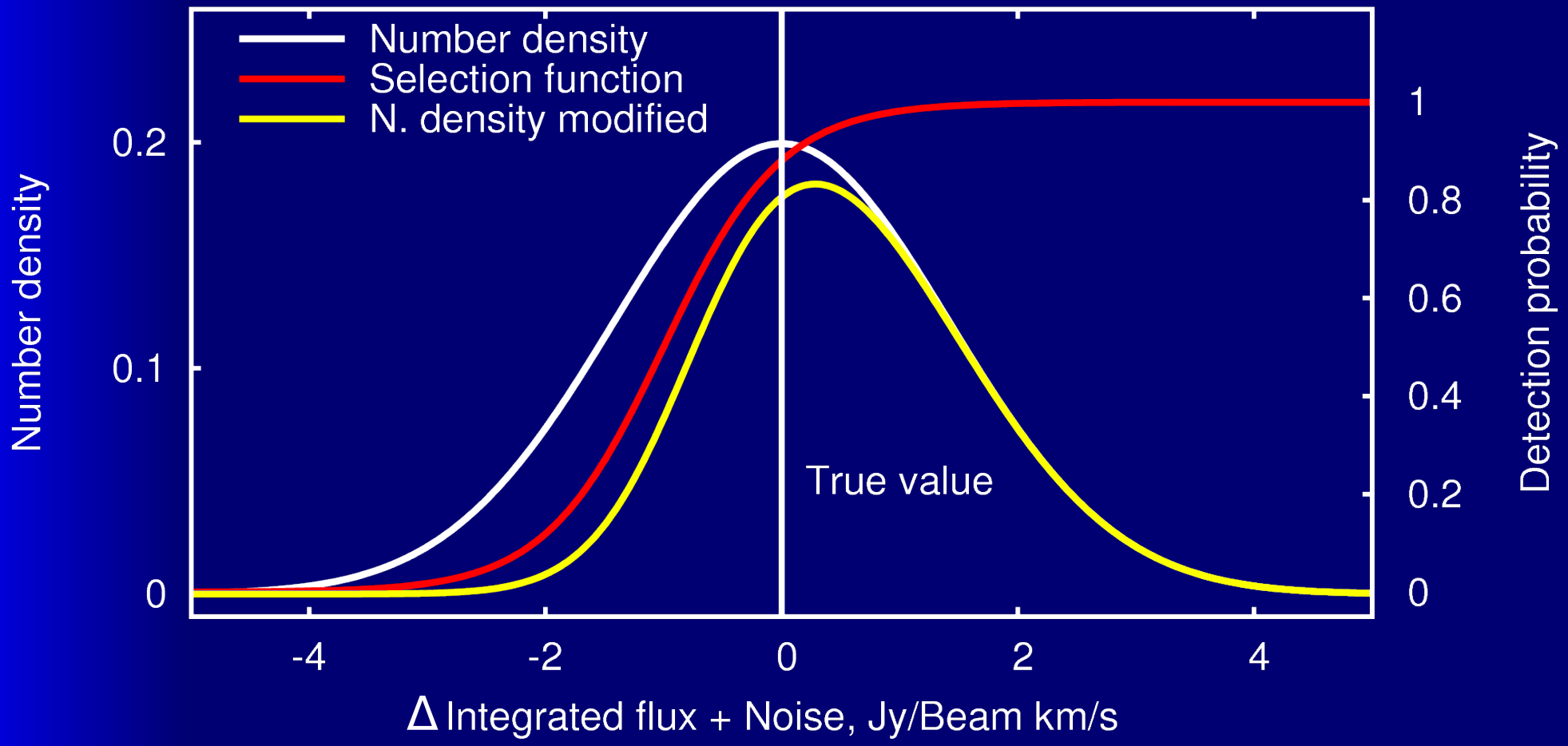




Systematic Effects

- Parametrization issues
  - The peak value of a spectrum is a bad estimator for the peak flux
  - For small peak fluxes the velocity width is often heavily underestimated
- Selection effects
  - Incompleteness of the survey (correction possible)
  - Integrated flux can be biased to higher values
- Other
  - Aliasing
  - ...

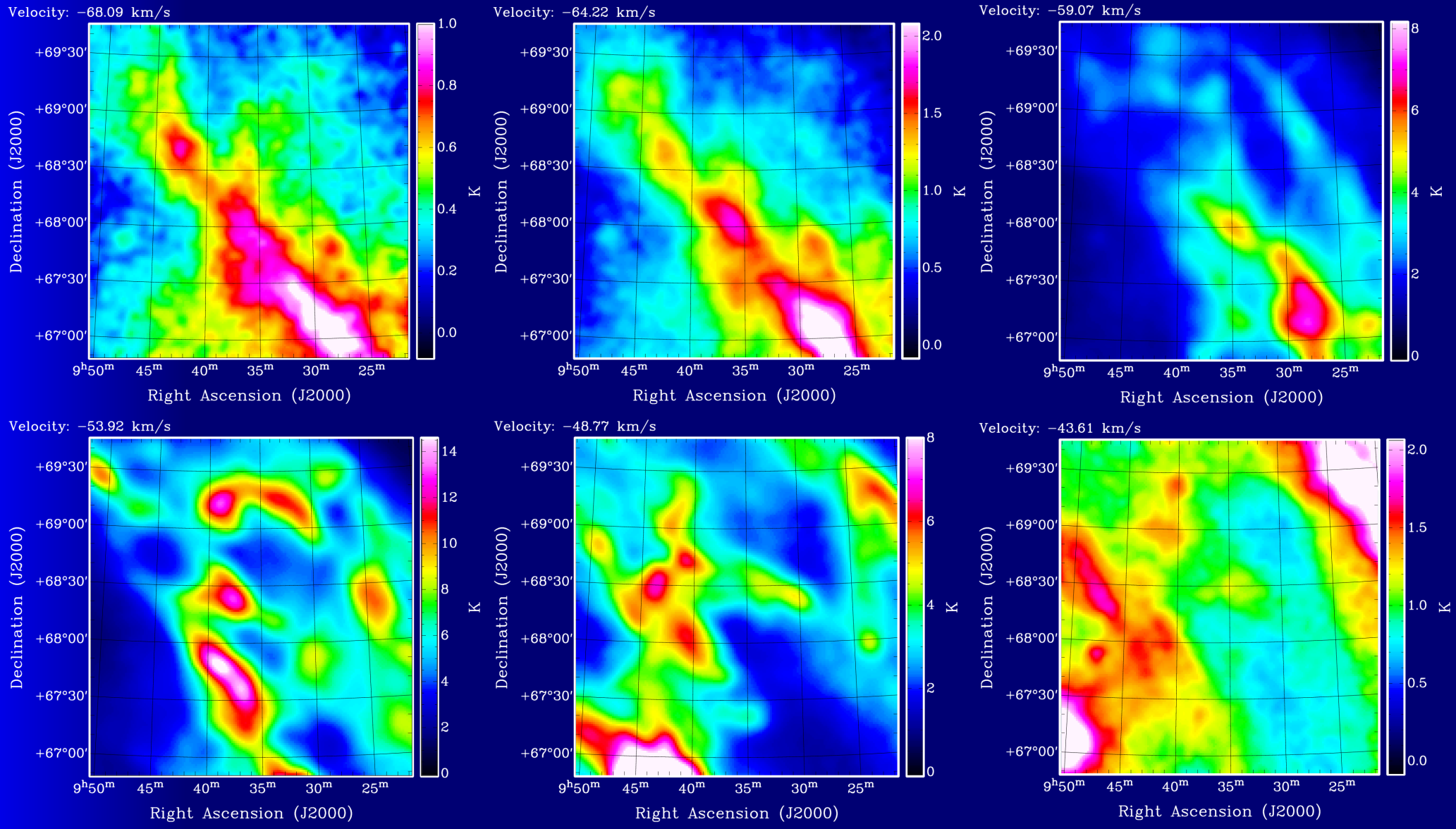
Systematic Effects – Selection

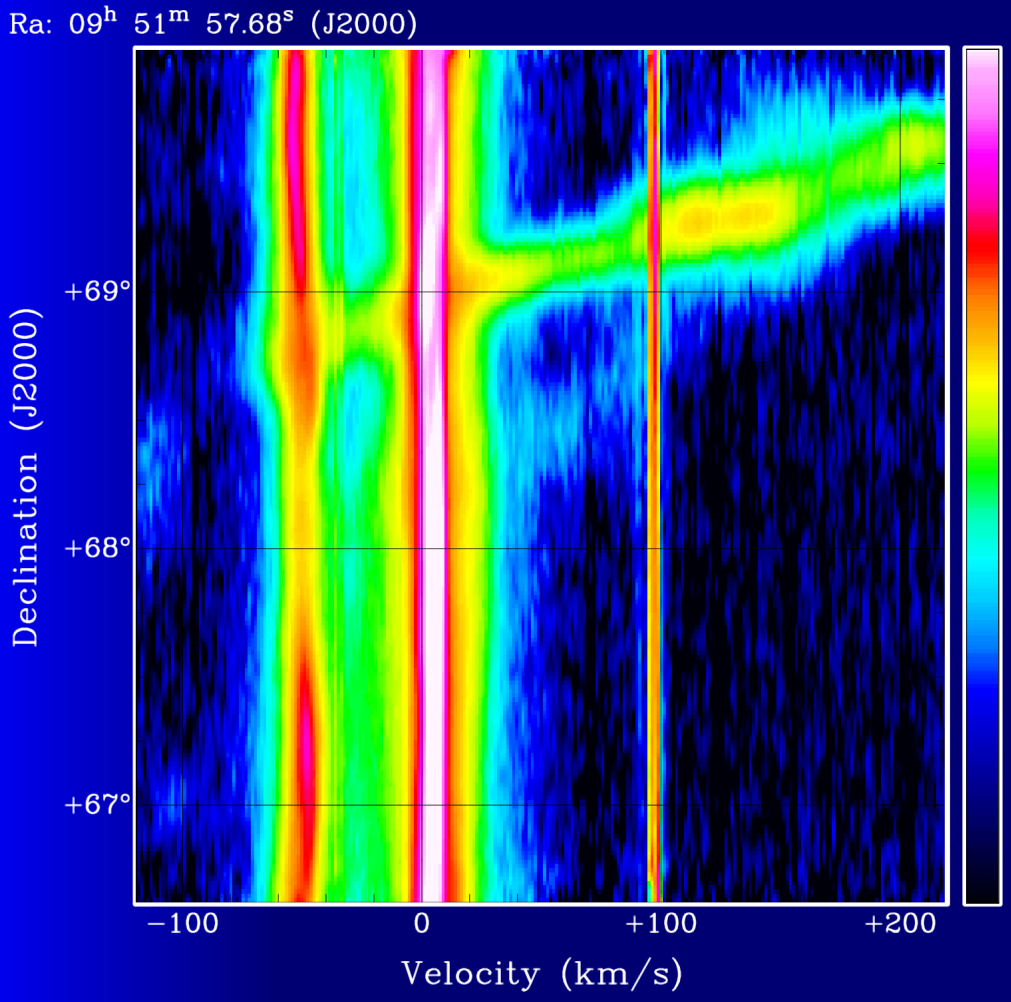


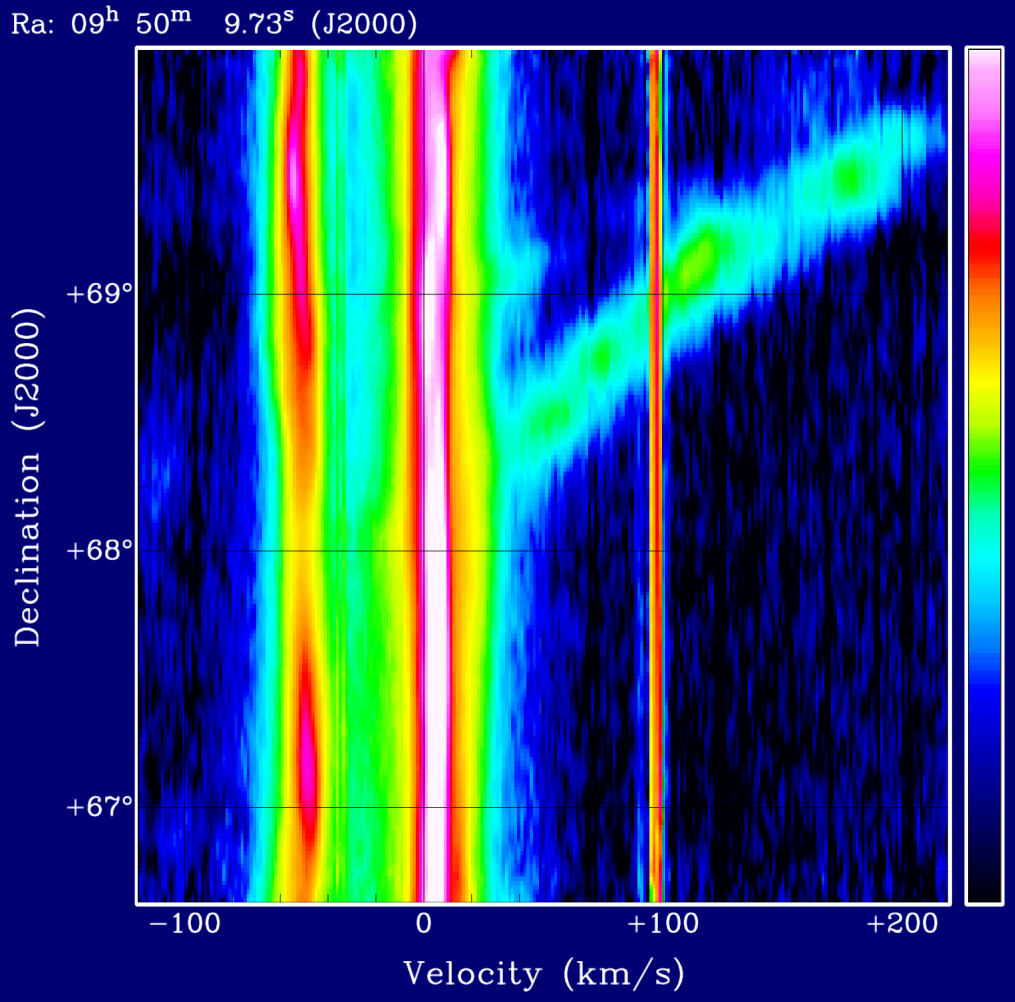
Systematic Effects

- Parametrization issues
  - The peak value of a spectrum is a bad estimator for the peak flux
  - For small peak fluxes the velocity width is often heavily underestimated
- Selection effects
  - Incompleteness of the survey (correction possible)
  - Integrated flux can be biased to higher values
- Other
  - Aliasing
  - Gridding
  - ...

First observations




Ra: 09h 51m 57.68s (J2000)
Declination (J2000)
+69°
+68°
+67°
−100
0
+100
+200
Velocity (km/s)

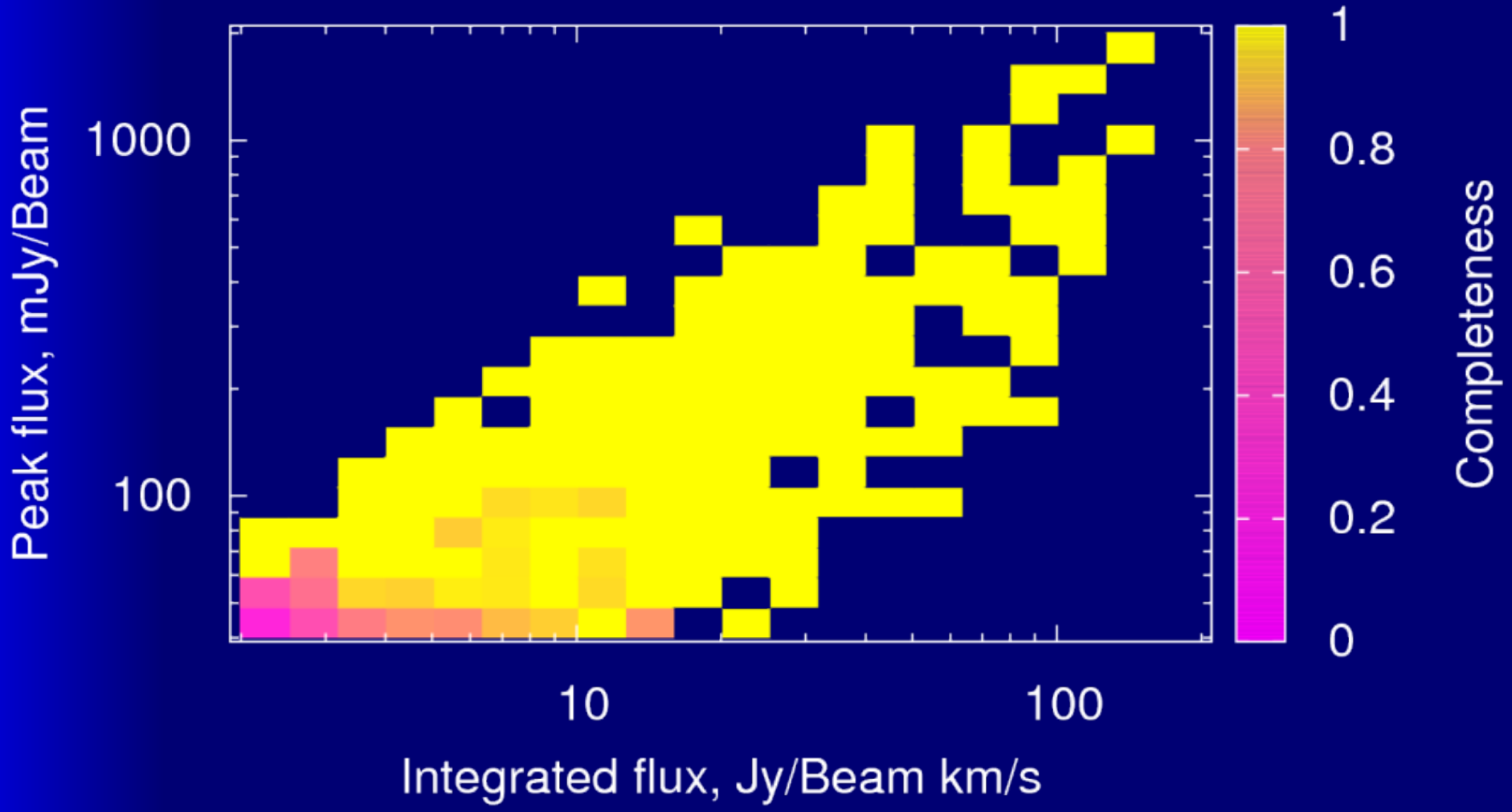

Ra: 09h 50m 9.73s (J2000)
Declination (J2000)
+69°
+68°
+67°
−100
0
+100
+200
Velocity (km/s)

Conclusions

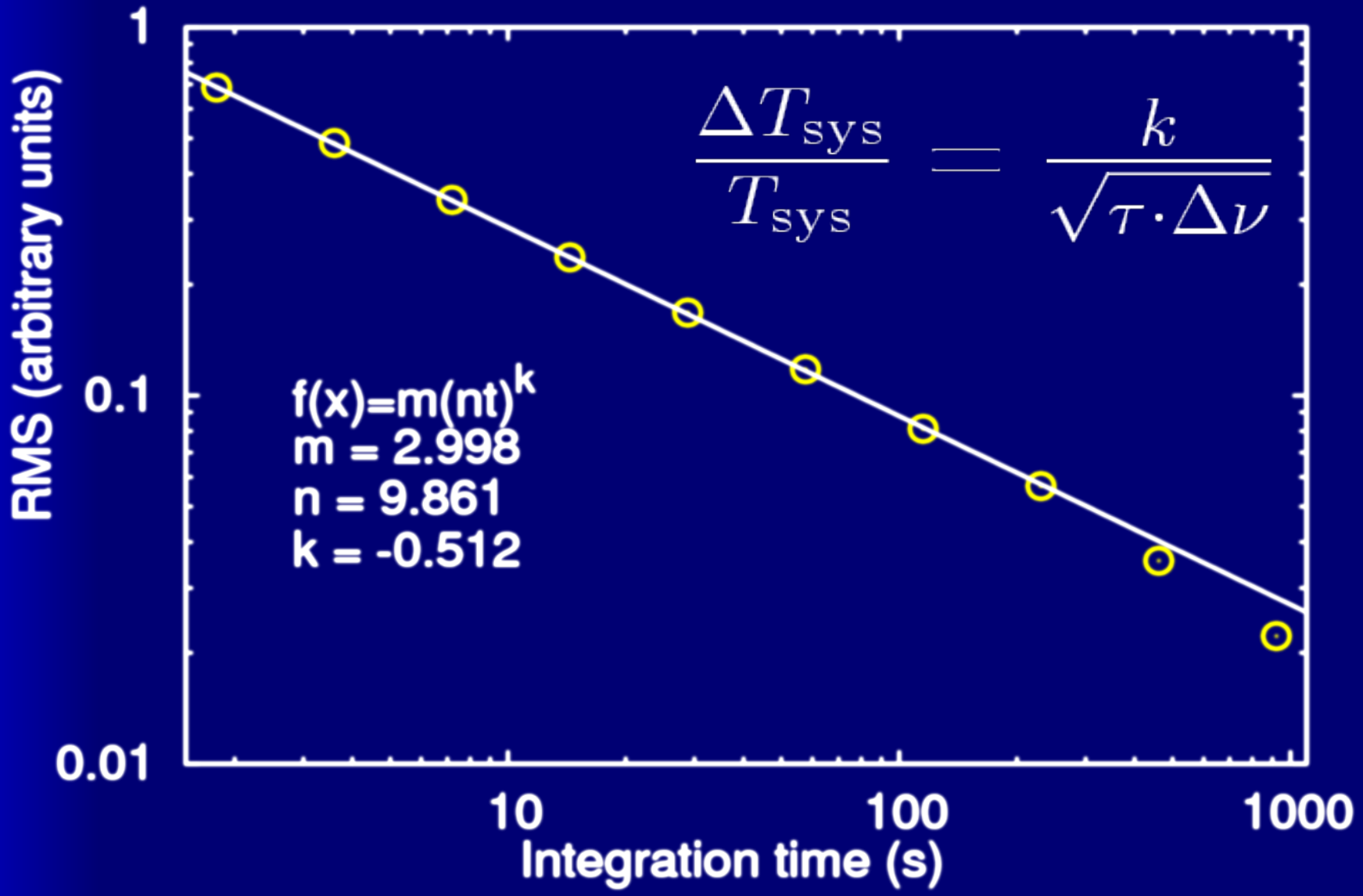
- Data reduction software is working
- Simulations
  - RFI must be mitigated
  - Systematic effects could be explained
- First observations successful

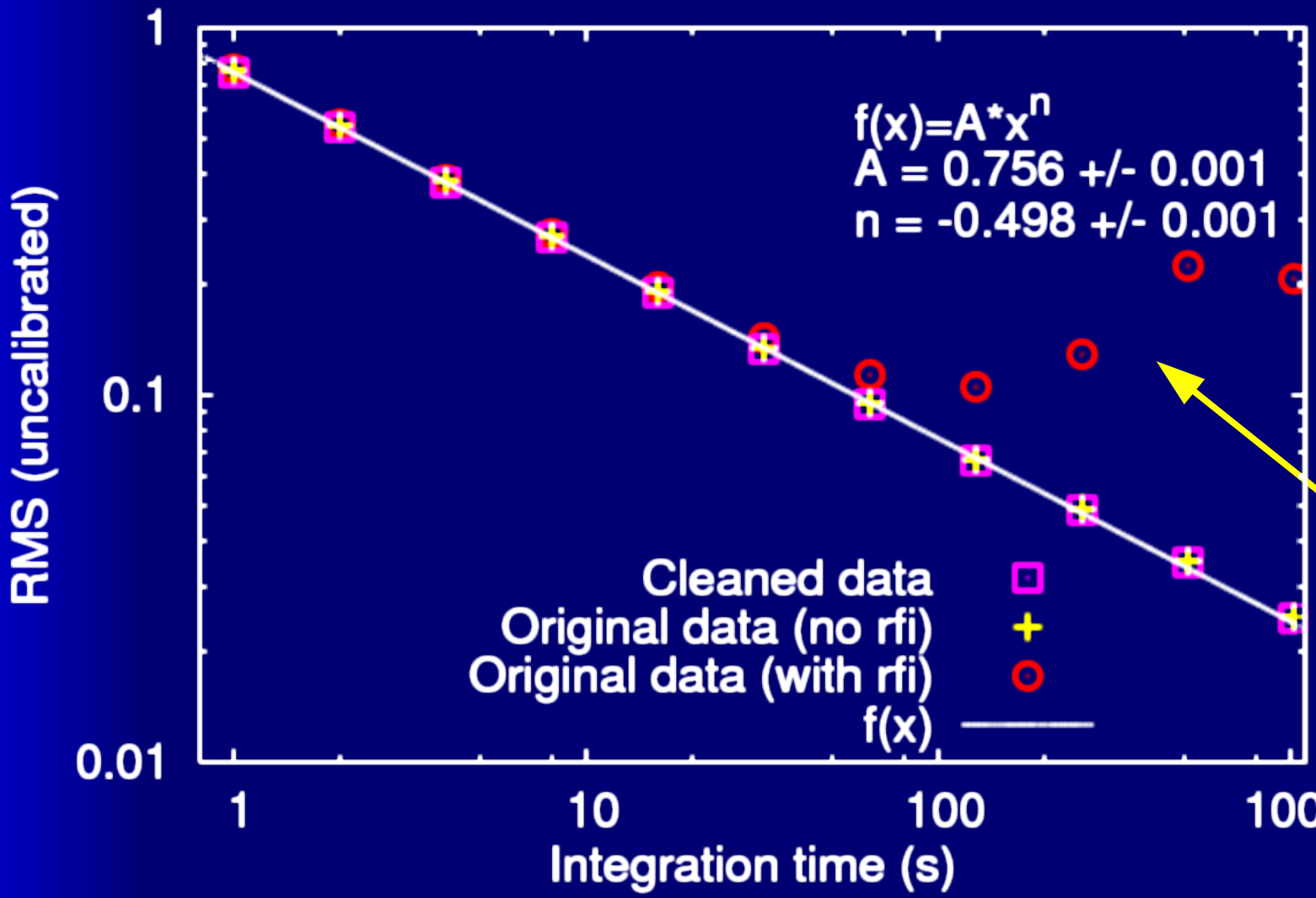
The end

Completeness 2D



Allan plot

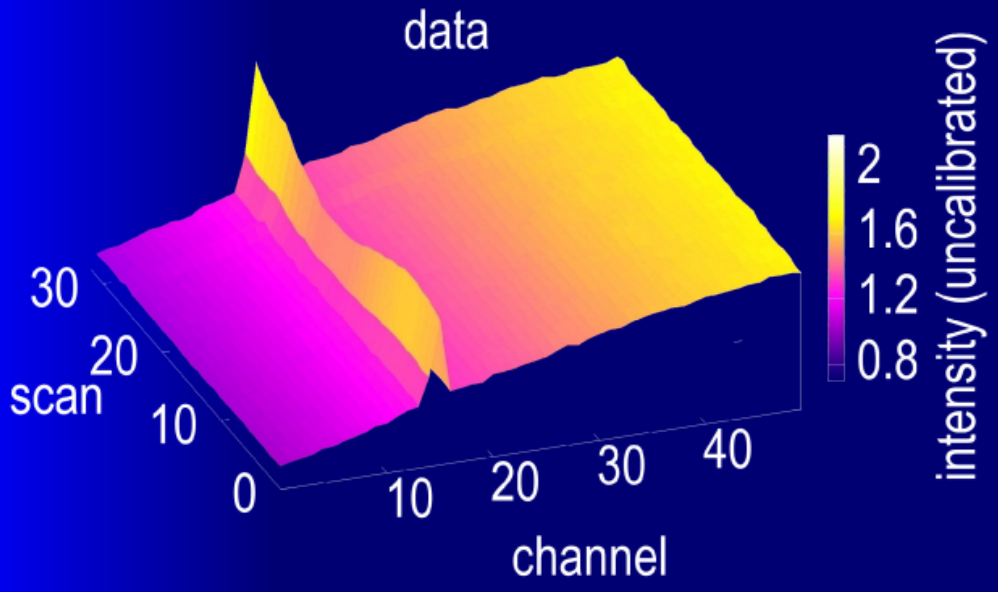

1
0.1
0.01
RMS (arbitrary units)
$\frac{\Delta T_{\mathrm{sys}}}{T_{\mathrm{sys}}} = \frac{k}{\sqrt{\tau \cdot \Delta\nu}}$
f(x)=m(nt)^k
m = 2.998
n = 9.861
k = -0.512
10
100
1000
Integration time (s)

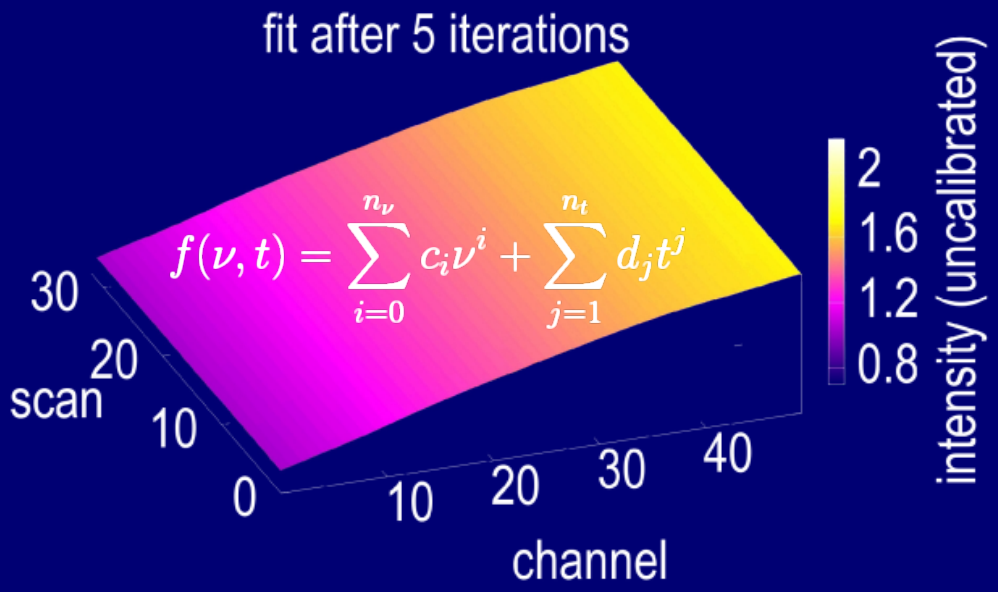


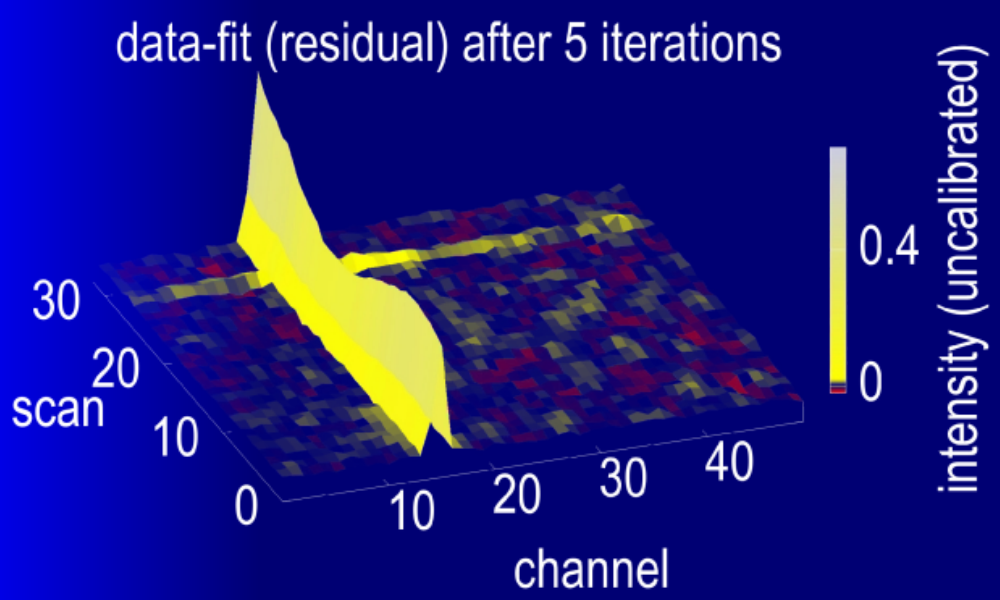
Sensitivity drops!

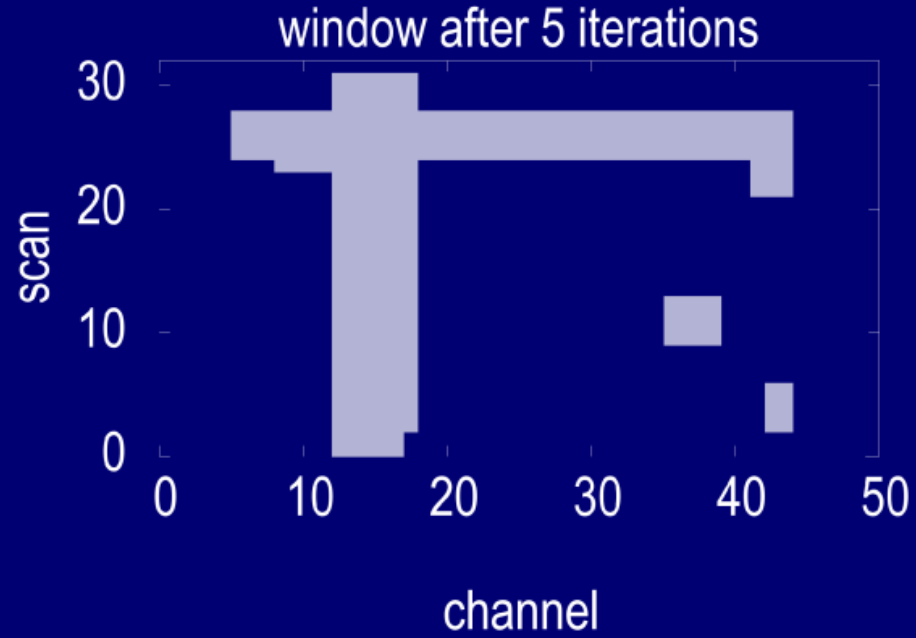
Winkel et al. (2007)

Detection algorithm









Winkel et al. (2007)

Introduction

Why HI surveys? – Science drivers

- 3D distribution of galaxies
- HI Mass Function (low mass end!)
- Environmental/evolutionary effects

Extra-galactic

- (Super-)shells in the Milky Way
- IVC/HVCs: statistical properties, Ultra CHVCs, connection to Halo/Disk
- Accretion of gas

Galactic

- Multiple wavelength studies (Next Generation XRay!)

Data reduction scheme

- High flexibility
- Memory efficient
- Multiple workstations can work independently
- Most functions are multithreaded
- Raw data can be kept
- Easy backup of the database