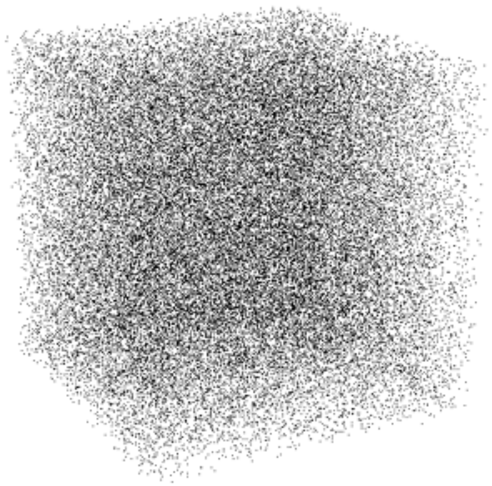
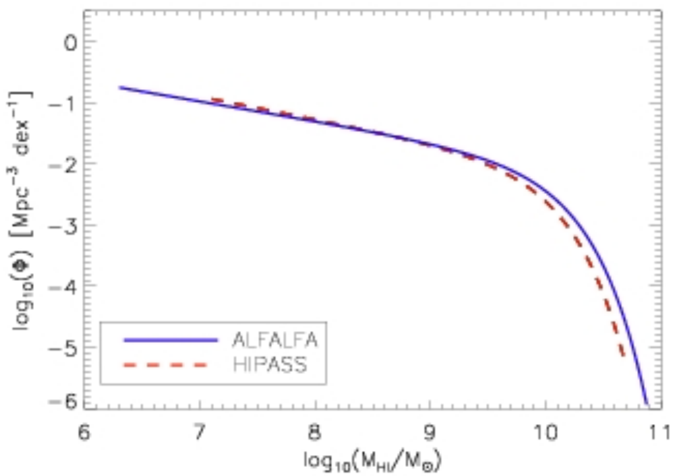
Challenges in calculating galaxy space densities from HI surveys

Martin Zwaan - ESO

HI mass functions…

The problem…





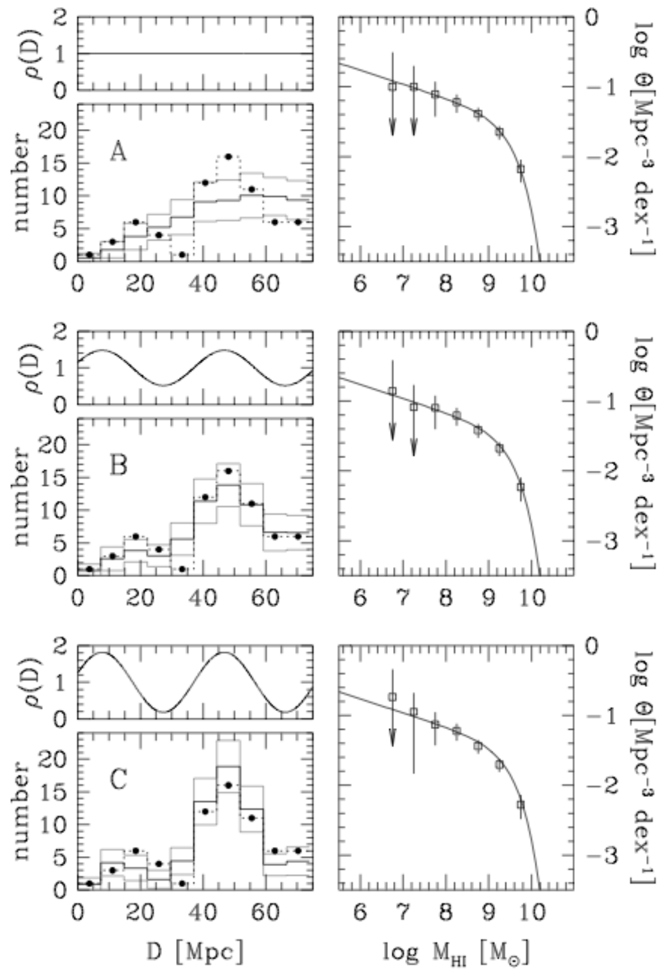
- Complications:
  - Complicated completeness limit (S_{peak}, W, profile shape, freq) $\otimes$
  - Large scale structure

The methods

1/V	ML
φ/Φ	C

The 1/V_{max} method

- The ‘classical’ Schmidt (1968) method
- Calculate maximum distance D_{max} out to which the galaxies can be detected
- Convert D_{max} into a V_{max}
- Used for early Arecibo surveys
- Advantages: simple and automatically normalised ✔
- Disadvantage: sensitive to large scale structure ✗



Zwaan et al 1997

Maximum likelihood methods

- Defined by Efstathiou et al 1988, Sandage et al 1979
- Find θ that yields maximal joint probability of detecting all sources in sample

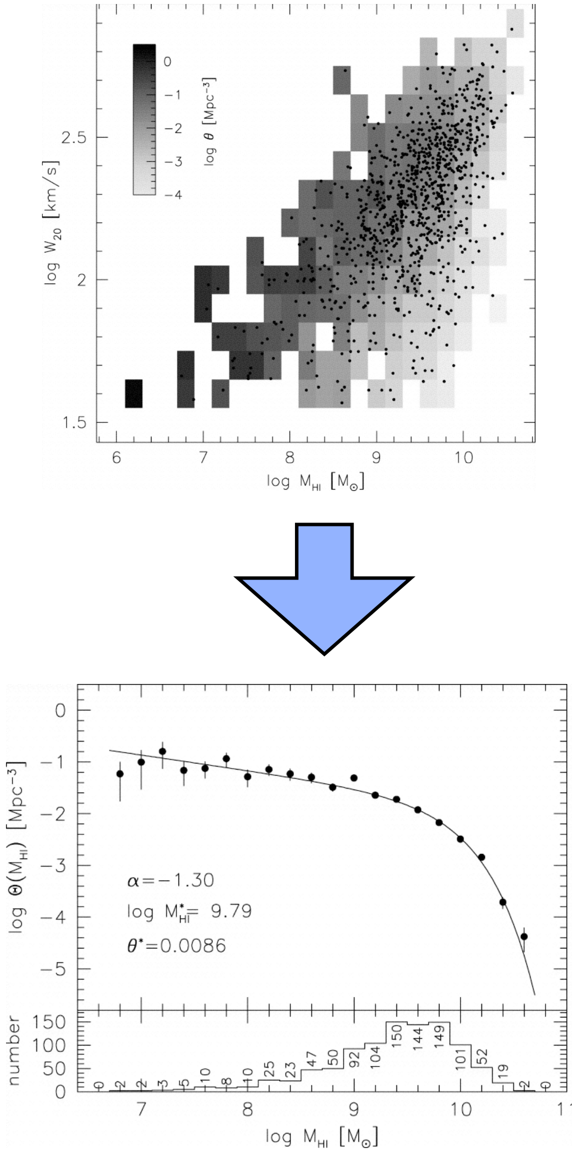
$$p(M_{\mathrm{HI},i}|D_i) = \frac{\theta(M_{\mathrm{HI},i})}{\int_{M_{\mathrm{HI,min}(D_i)}}^{\infty} \theta(M_{\mathrm{HI}}) dM_{\mathrm{HI}}}$$

minimal detectable HI mass at distance D_i

generally **not defined** for HI selected samples

2D Stepwise Maximum likelihood method

- Solution: multi-dimensional stepwise maximum likelihood methods
- Find $\theta(M_{HI},W)$
- Collapse to find HIMF
- (Or, find ML-based effective volume accessible to each galaxy individually)
- **Used for HIPASS and ALFALFA**
- Advantage: robust against LSS ✔
- Disadvantage: slow ✗



The Turner or φ/Φ-method

NEW!

- Introduced by Turner (1979) for 3C and 4C quasar catalogues
- Calculate the ratio of number of galaxies in interval dM_{HI} and number of galaxies brighter than M_{HI}

$$\frac{N(dM_{\mathrm{HI}})}{N(\geq M_{\mathrm{HI}})} = \frac{\theta(M_{\mathrm{HI}})\rho(D)dM_{\mathrm{HI}}dV}{\int_{M_{\mathrm{HI}}}^{\infty}\theta(M_{\mathrm{HI}})\rho(D)dM'_{\mathrm{HI}}dV} = \frac{\theta(M_{\mathrm{HI}})dM_{\mathrm{HI}}}{\Theta(M_{\mathrm{HI}})} = d\ln\Theta(M_{\mathrm{HI}})$$

- Advantage: fast and robust against LSS ✔
- Disadvantage: correlated errors ✗

The C^- method

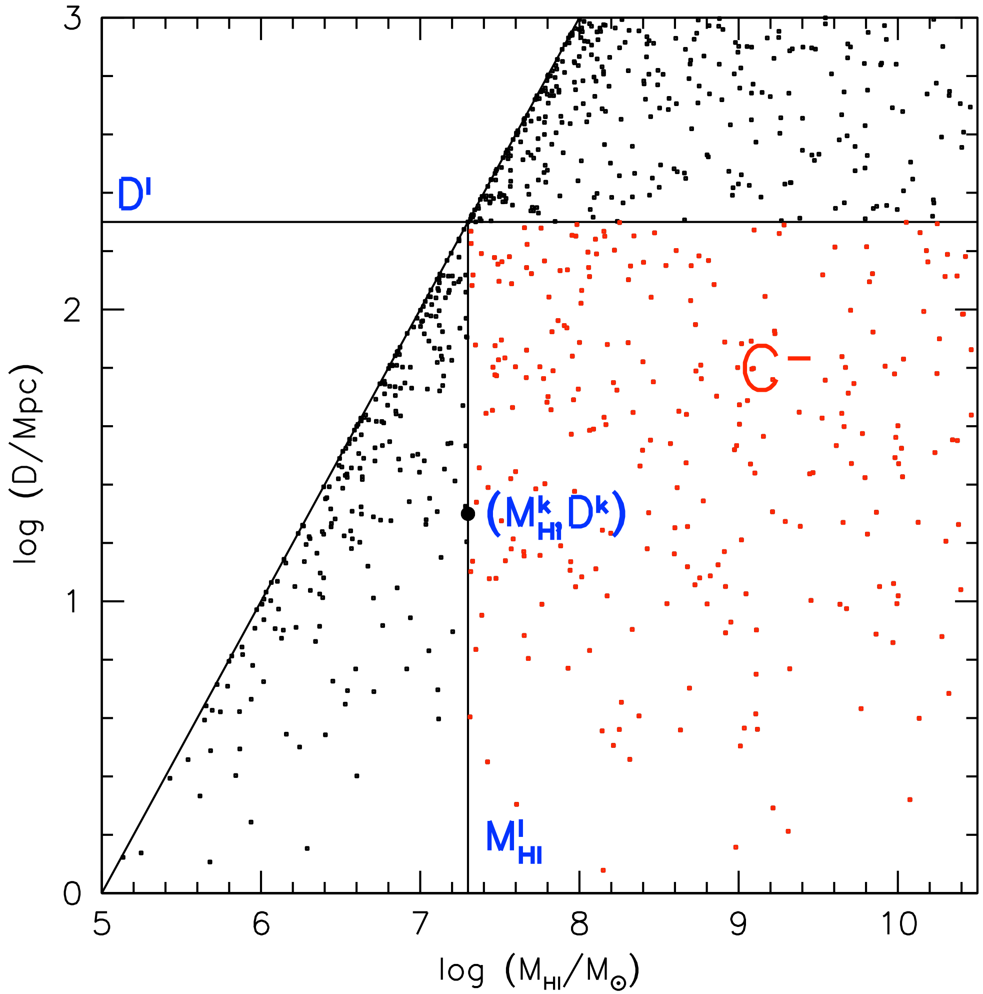
- Developed by Lynden-Bell (1971) for quasars
- Maximum likelihood procedure. Does not require any binning.
- Estimates the cumulative luminosity function (CLF).
- Does not require any assumptions about the distribution of objects within the data-set.

The C⁻ method

$$\psi_{k+1} = \psi_k \frac{C_k^-(M_{\rm HI}+1}{C_{k+1}^-(M_{\rm HI})}$$

$$\Theta(M_{\rm HI}) = \int_{M_{\rm HI}^{\rm min}}^{M_{\rm HI}} \theta(M_{\rm HI}) dM_{\rm HI}$$

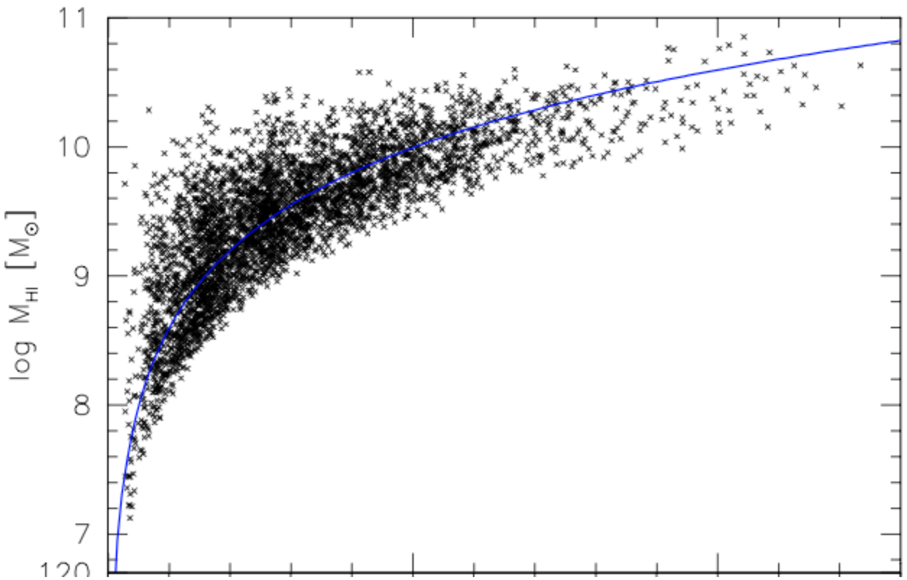
$$= \psi_1 \prod_i^{M_{\rm HI}^k < M_{\rm HI}} \frac{C_k^-(M_{\rm HI}+1}{C_k^-(M_{\rm HI})}$$



- then differentiate to obtain $\theta(M_{\rm HI})$
- advantage: independent of clustering effects ✔ and fast ✔

Dealing with gradual drop off in completeness (as opposed to sharp flux limits…)

- All these methods are designed for optical galaxy samples with sharp magnitude limits (m_{lim})
- The 2DSWML, Turner, and C- method are easily adaptable to work for complicated completeness limits



Normalising the HIMF

- Most methods (apart from $1/V_{max}$) lose the normalisation of the HIMF
- Various methods for recovering the normalisation
  - **n_3**: integral over selection function (ALFALFA)

$$n_3 = \frac{N_T}{\int_0^{z_{max}} s(z)dz}$$

  - **n_1**: calculating number of galaxies in redshift shells

$$n_1 = \frac{\int_0^{zmax} \frac{N(z)}{s(z)} dz}{\int_0^{zmax} \frac{dV}{dz} dz}.$$

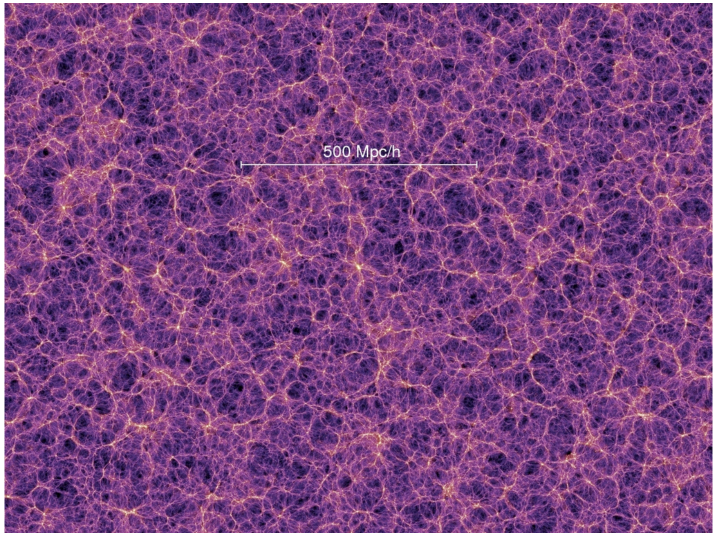
  - **$\bar{n}$**: "minimum-variance" - weighting by selection function and second moment of correlation function (HIPASS)

$$n = \frac{\sum_{i=1}^{N_g} N_i(z_i) w(z_i)}{\int_0^{z_{max}} s(z) w(z) \frac{dV}{dz} dz}$$

  - **counts**: compare real and expected number of galaxies

Simulations to test HIMF recovery

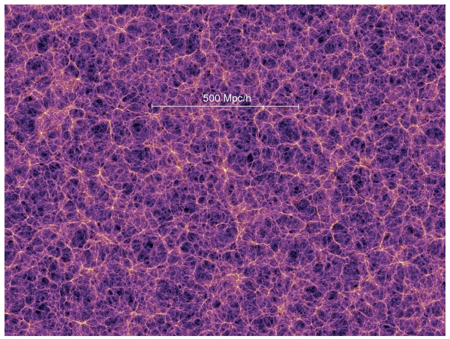
- Millennium Simulation (Springel et al 2005)
- 9 million galaxies in the full simulation box (500 Mpc/h on a side)
- Stitch several cubes together
- Assume a HIPASS HI mass function
- Low mass (log M_{HI}<8.5) cluster around larger ones



Simulations

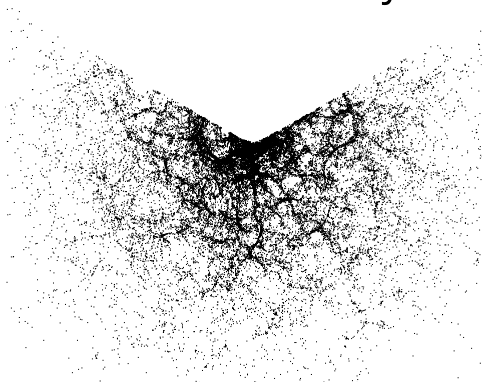
- Rotational velocity - HI mass relation from Obreschkow & Rawlings (2009)
- Random inclinations -> velocity widths
- Realistic scatter on all parameters
- Select galaxies from simulated boxes, assuming ‘optimal smoothing’

input catalogue



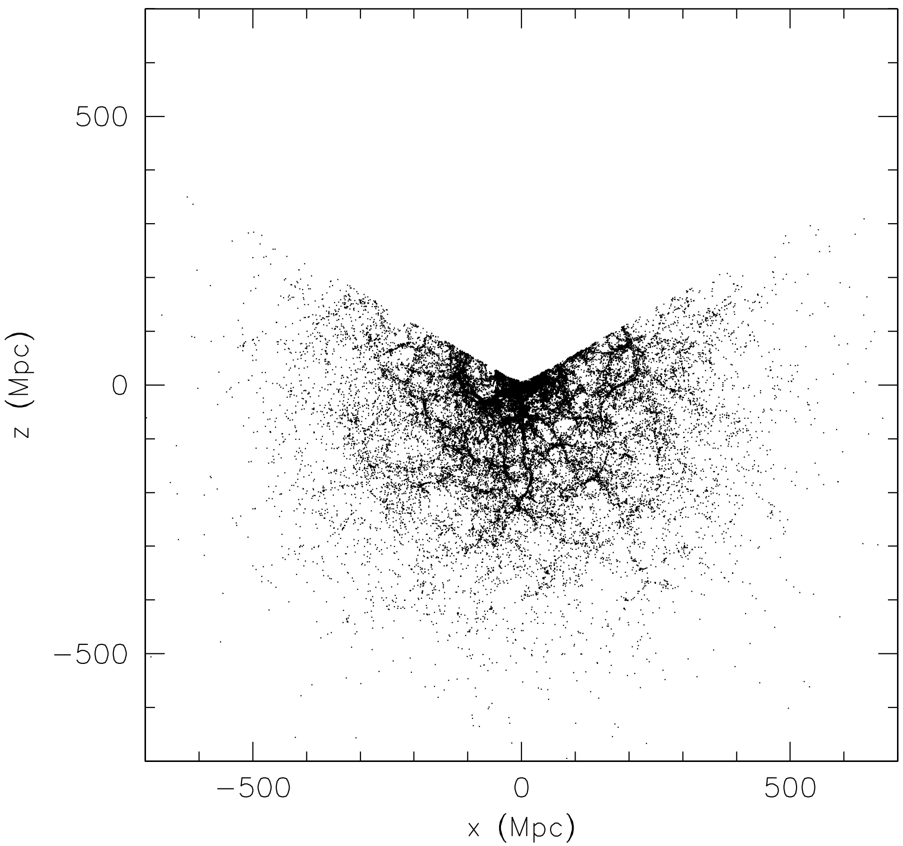
selection method

simulated sky

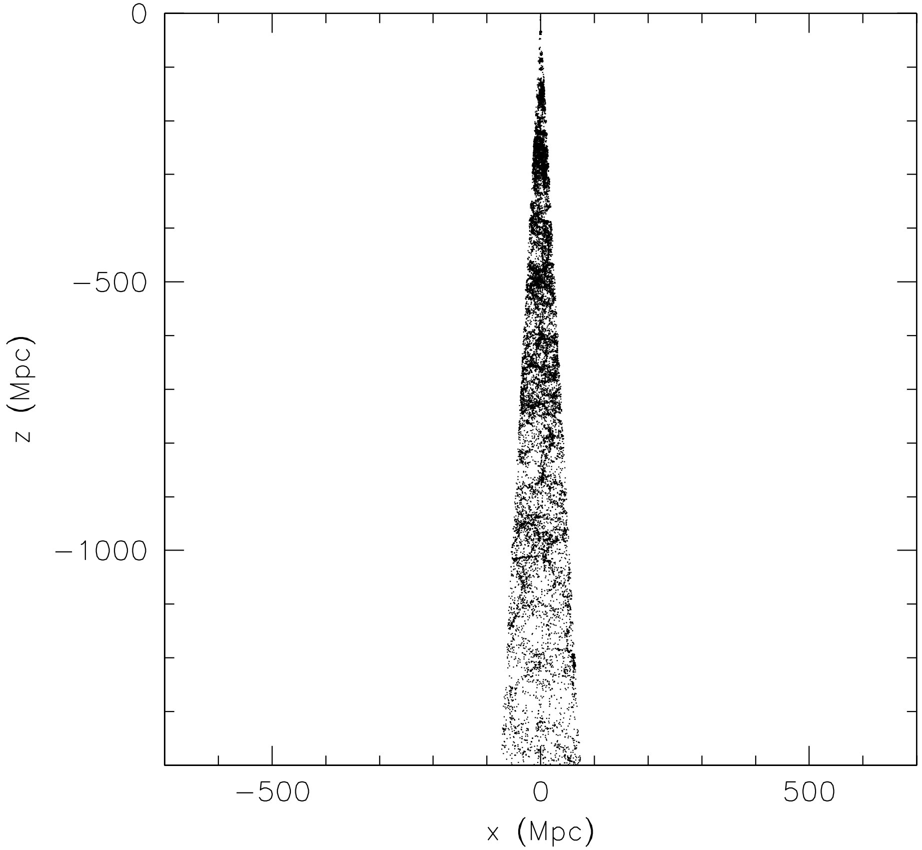


Simulated HI skies

"Wallaby"
all sky shallow

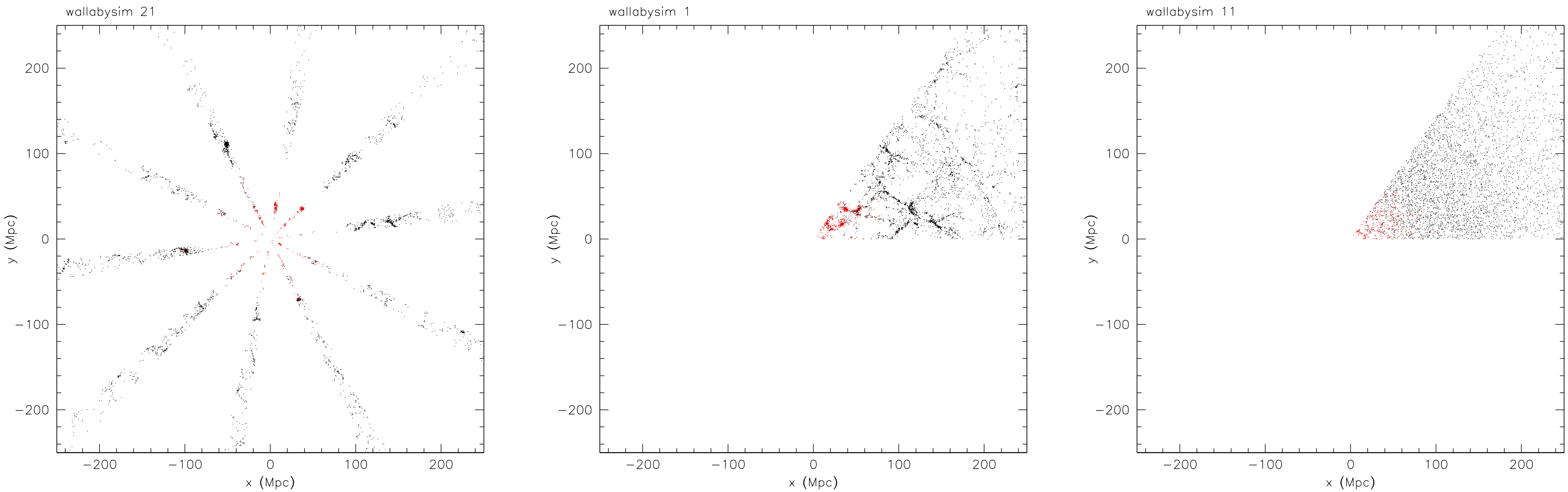


"Dingo"
30deg^2 deep



Testing the methods

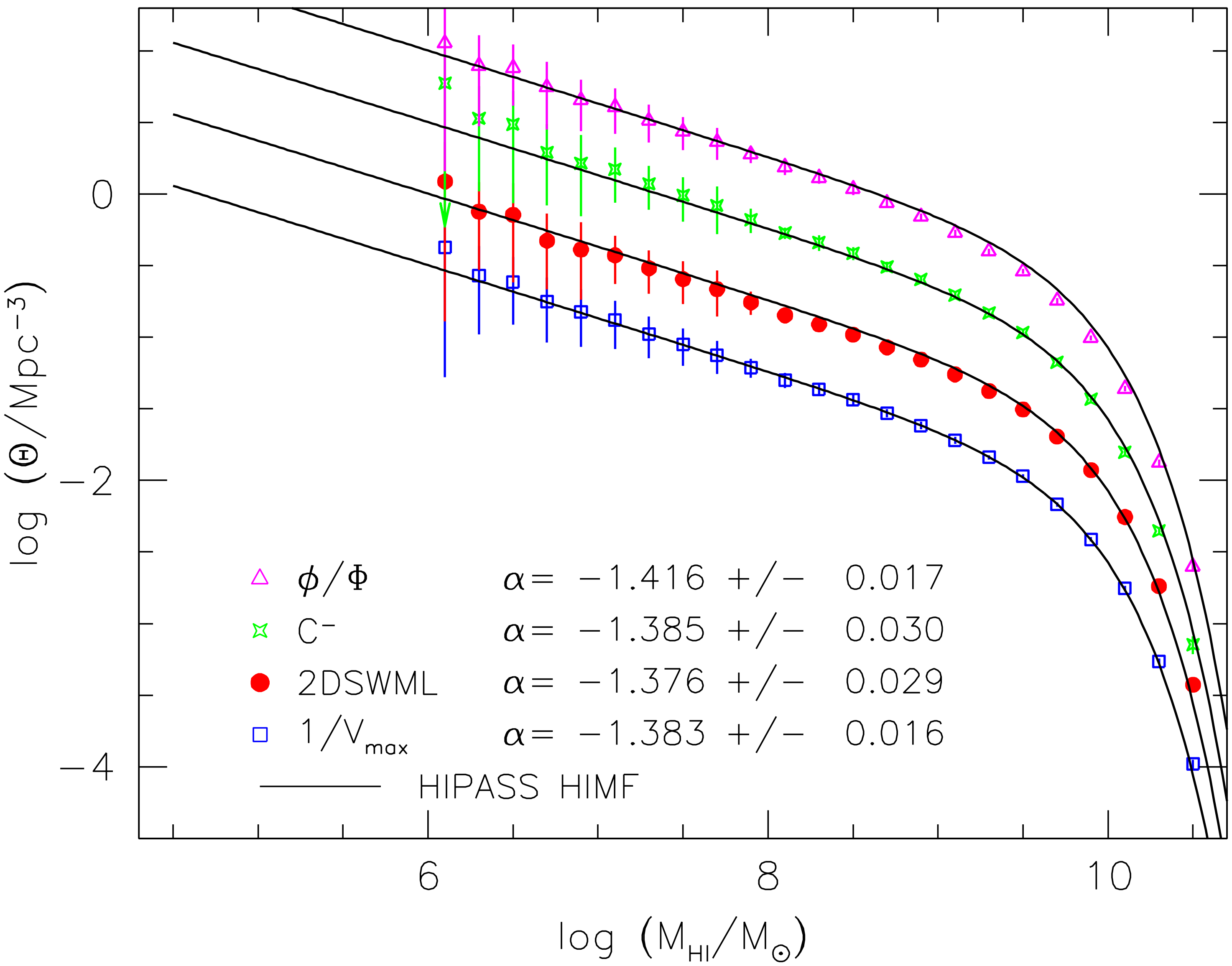
- Using Wallaby ‘early science’ fields



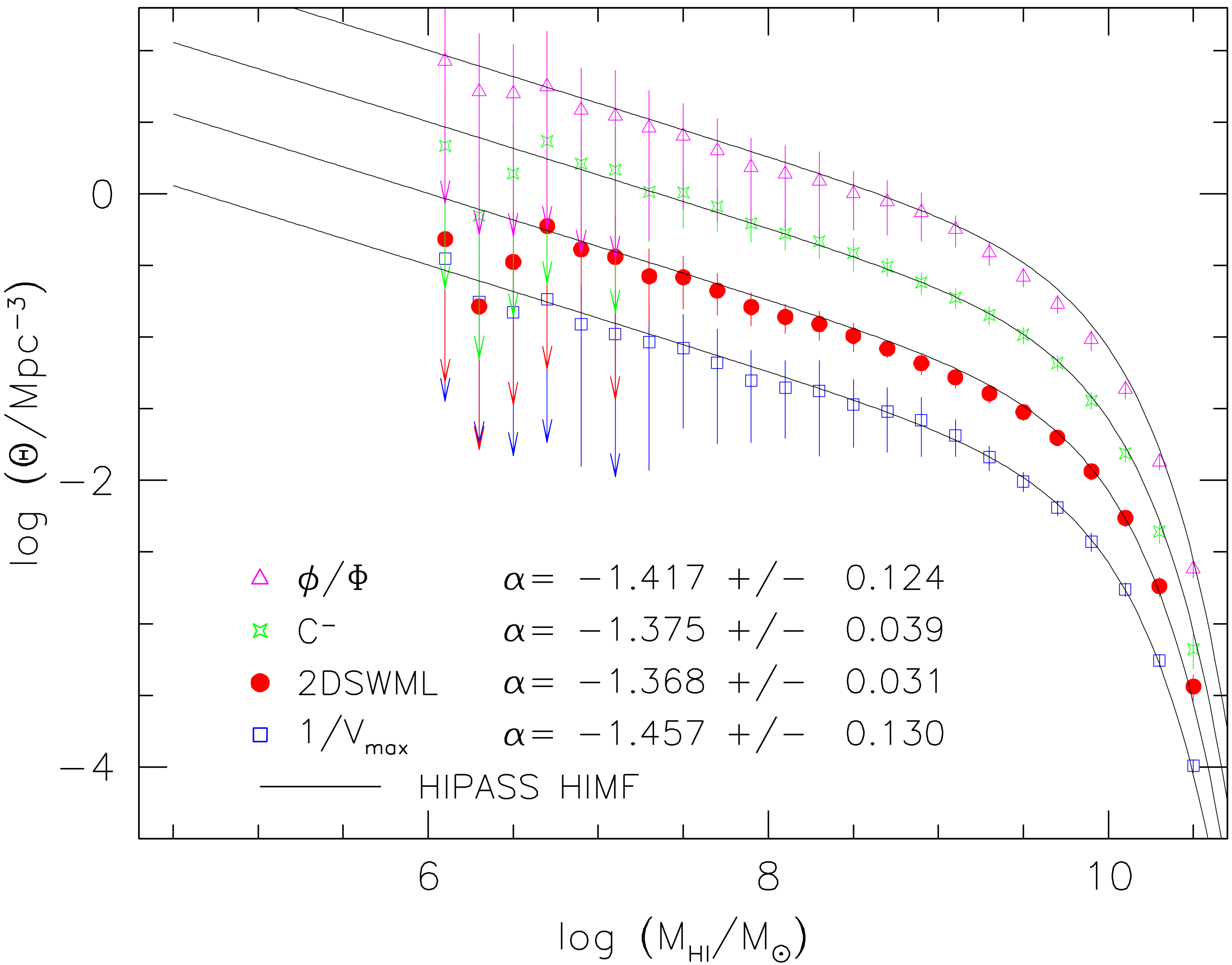
Simulate 10 times

$1/V_{max}$	ML
φ/Φ	C^-

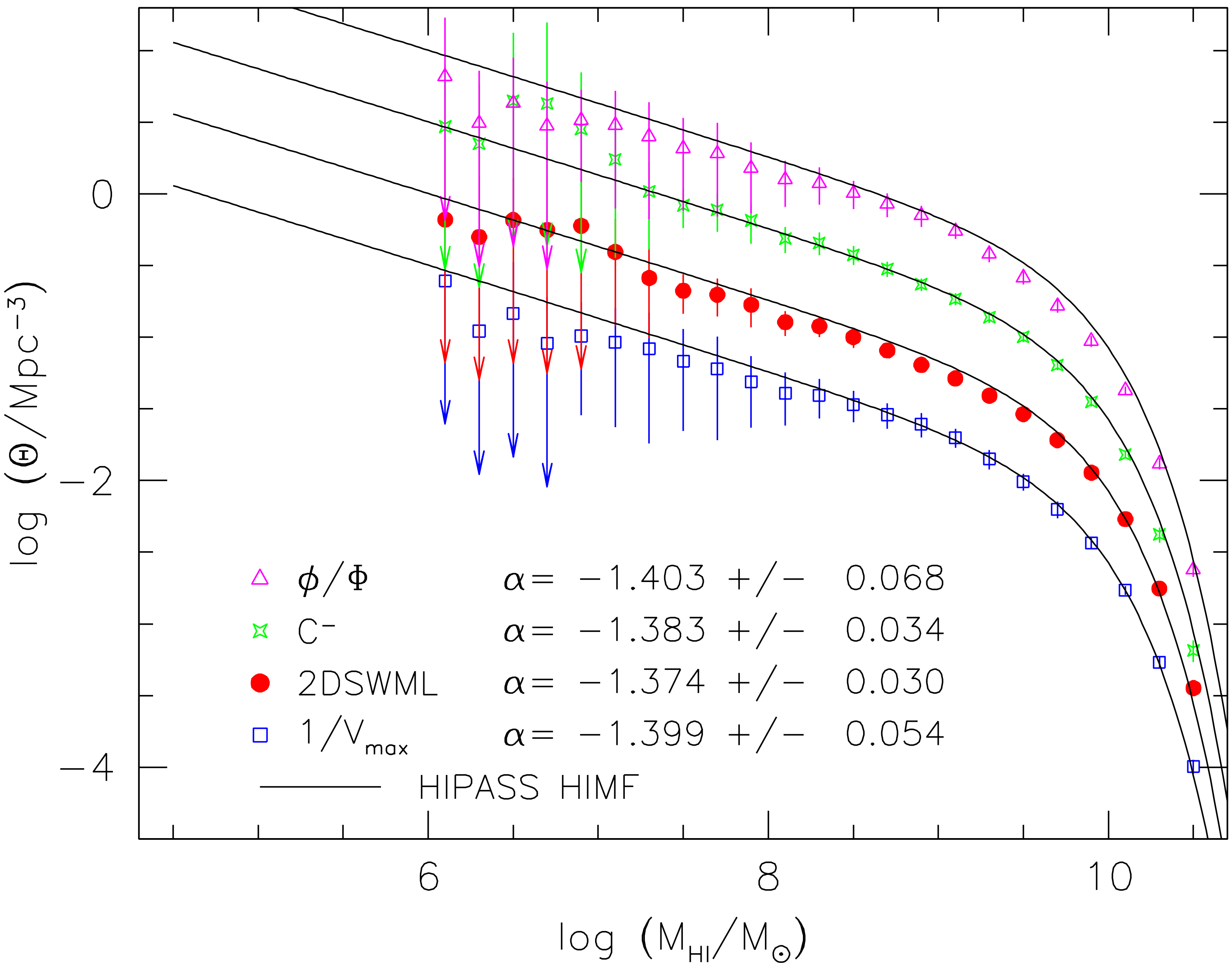
No large scale structure



10 ASKAP pointings - contiguous



10 ASKAP pointings - widely spaced



Results of testing methods on shallow HI survey

As expected. Sensitive to LSS

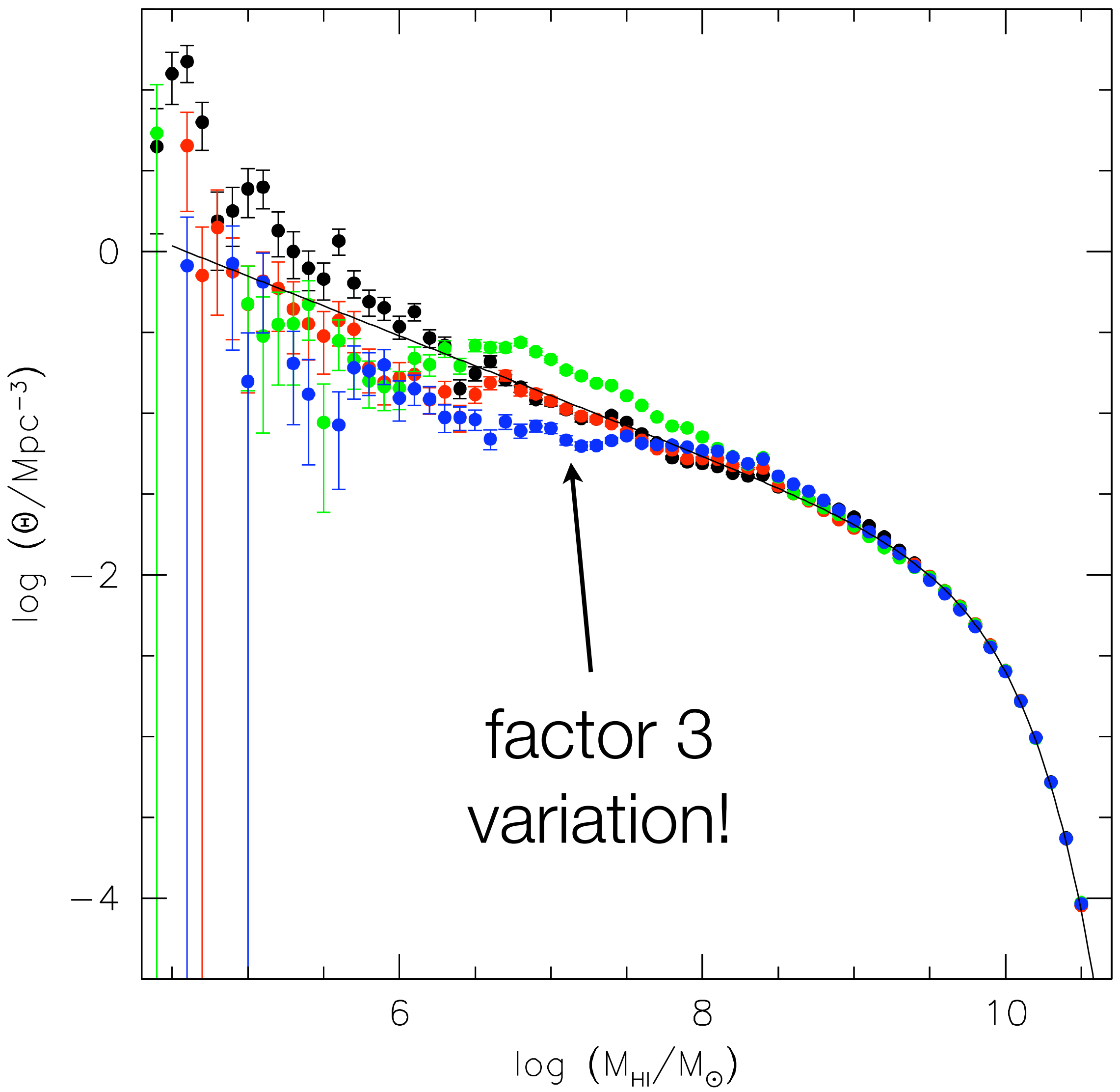
Not better than $1/V_{max}$

$1/V_{max}$	ML
φ/Φ	C^-

Performs well. Robust against LSS

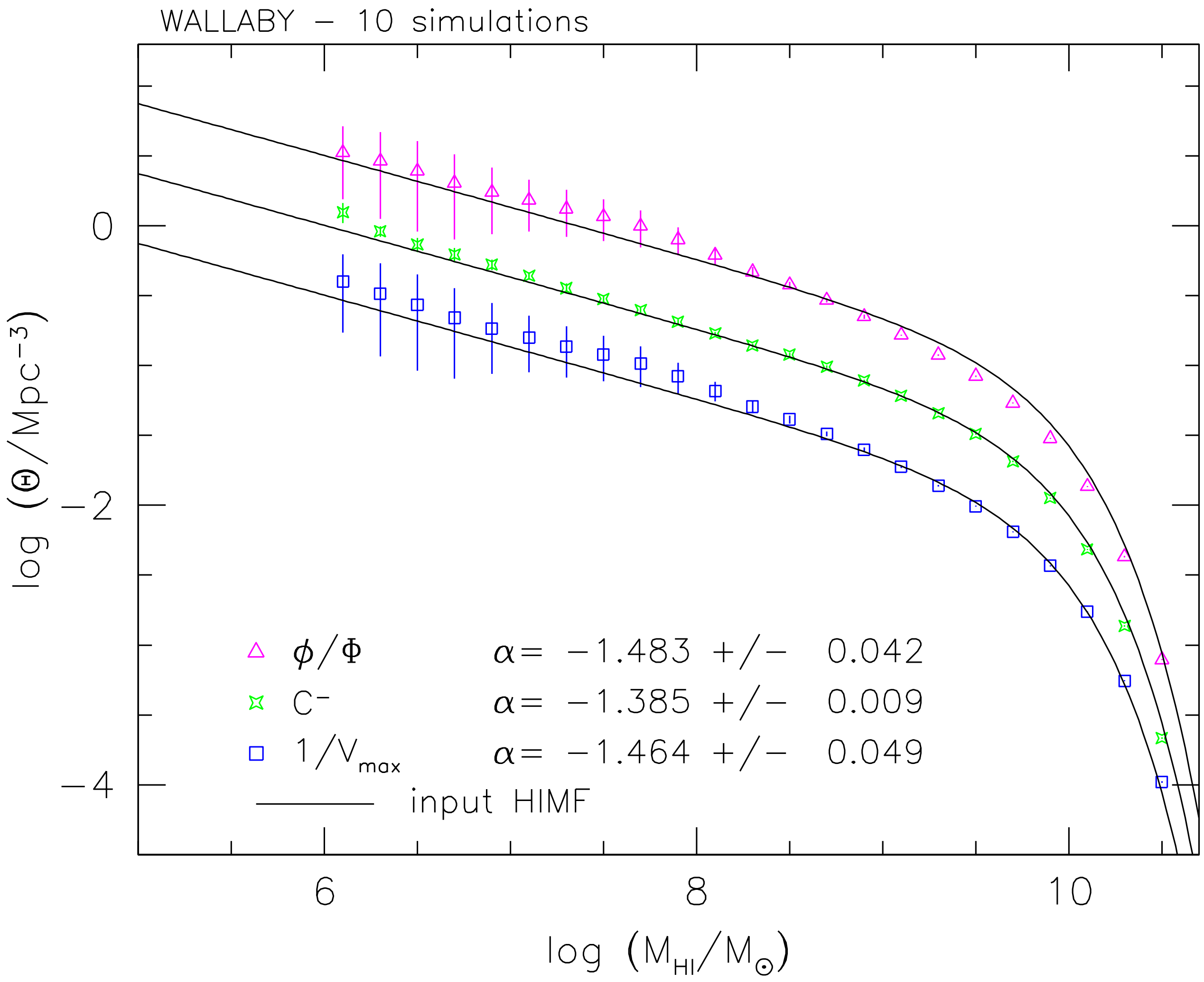
As good as 2DSWML. Faster!

Full Wallaby



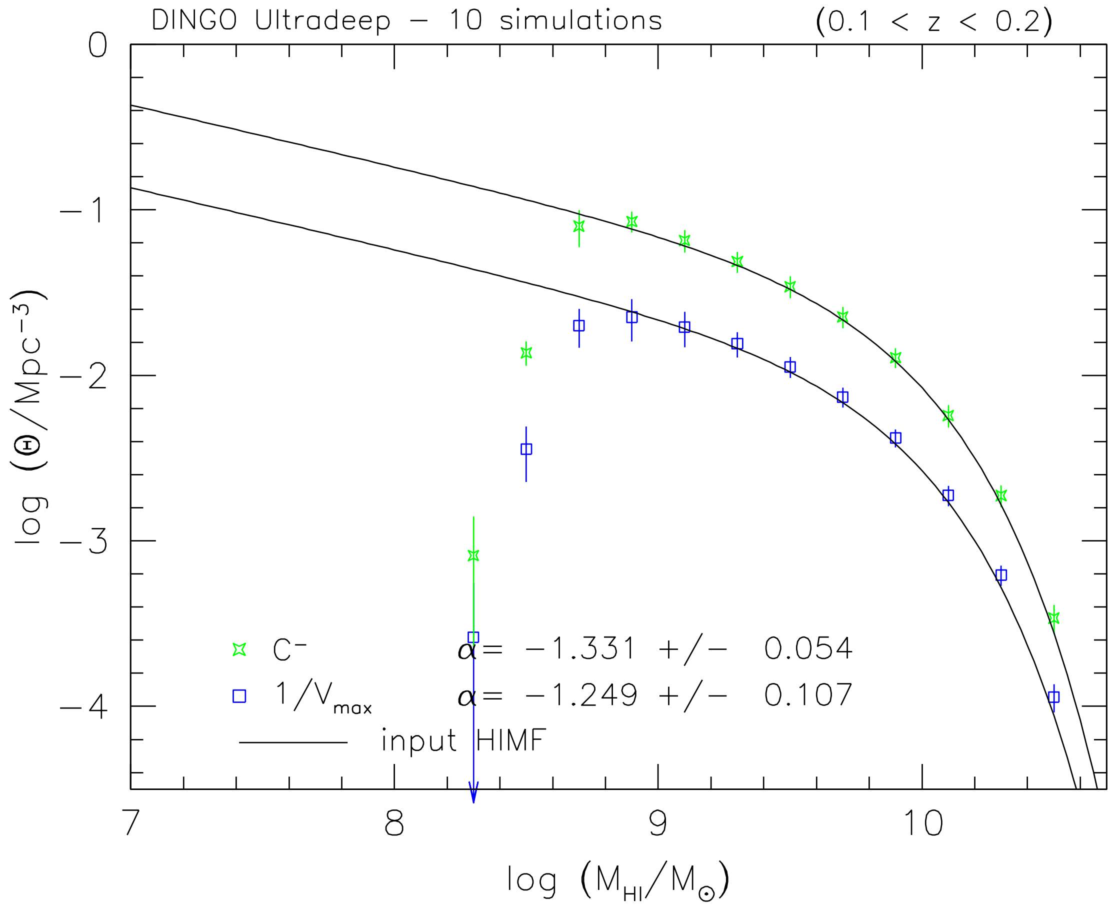
- HIMFs based on $1/V_{max}$ method

Full Wallaby

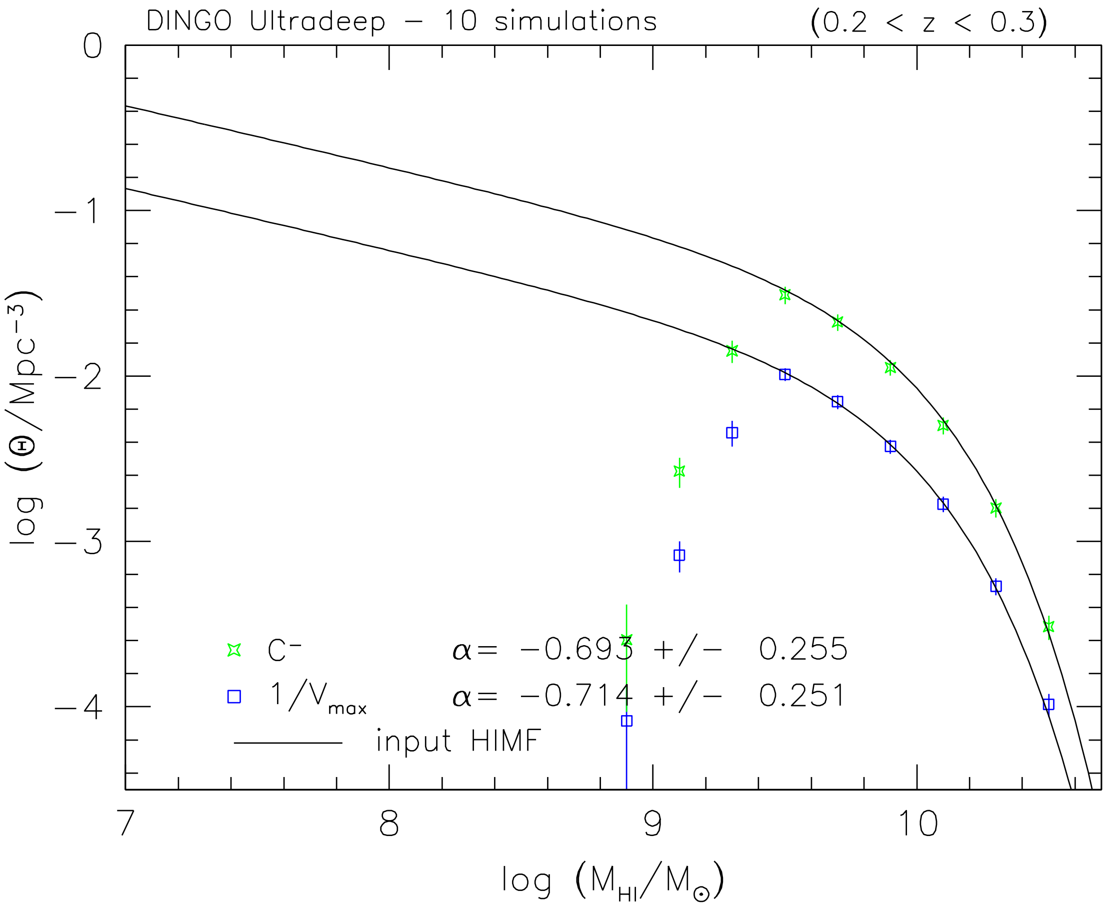


- C- performs very well
- SWML takes too much processing time
- HIMF slope accurate to ~0.01

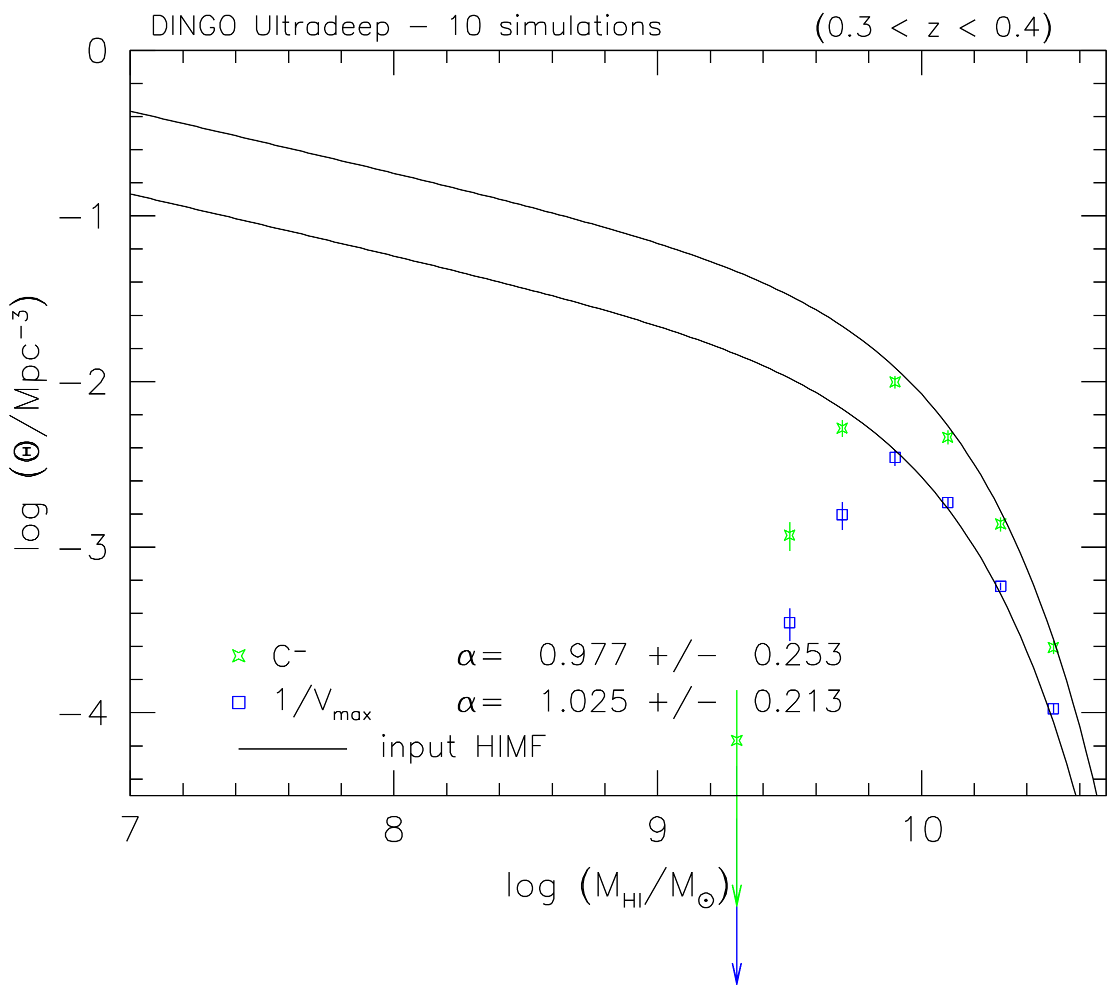
DINGO Ultradeep: 0.1 < z < 0.2



DINGO Ultradeep: 0.2 < z < 0.3



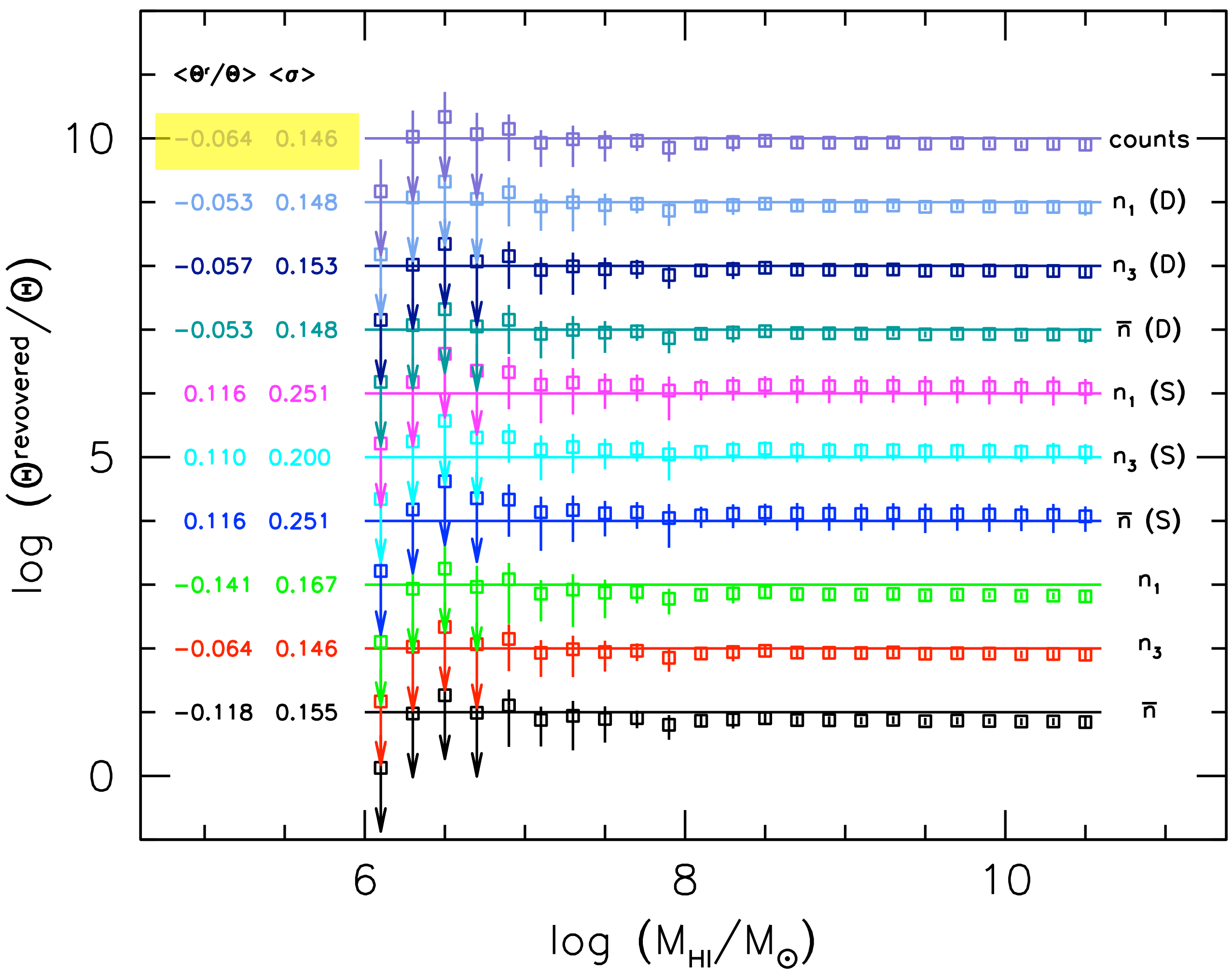
DINGO Ultradeep: 0.3 < z < 0.4



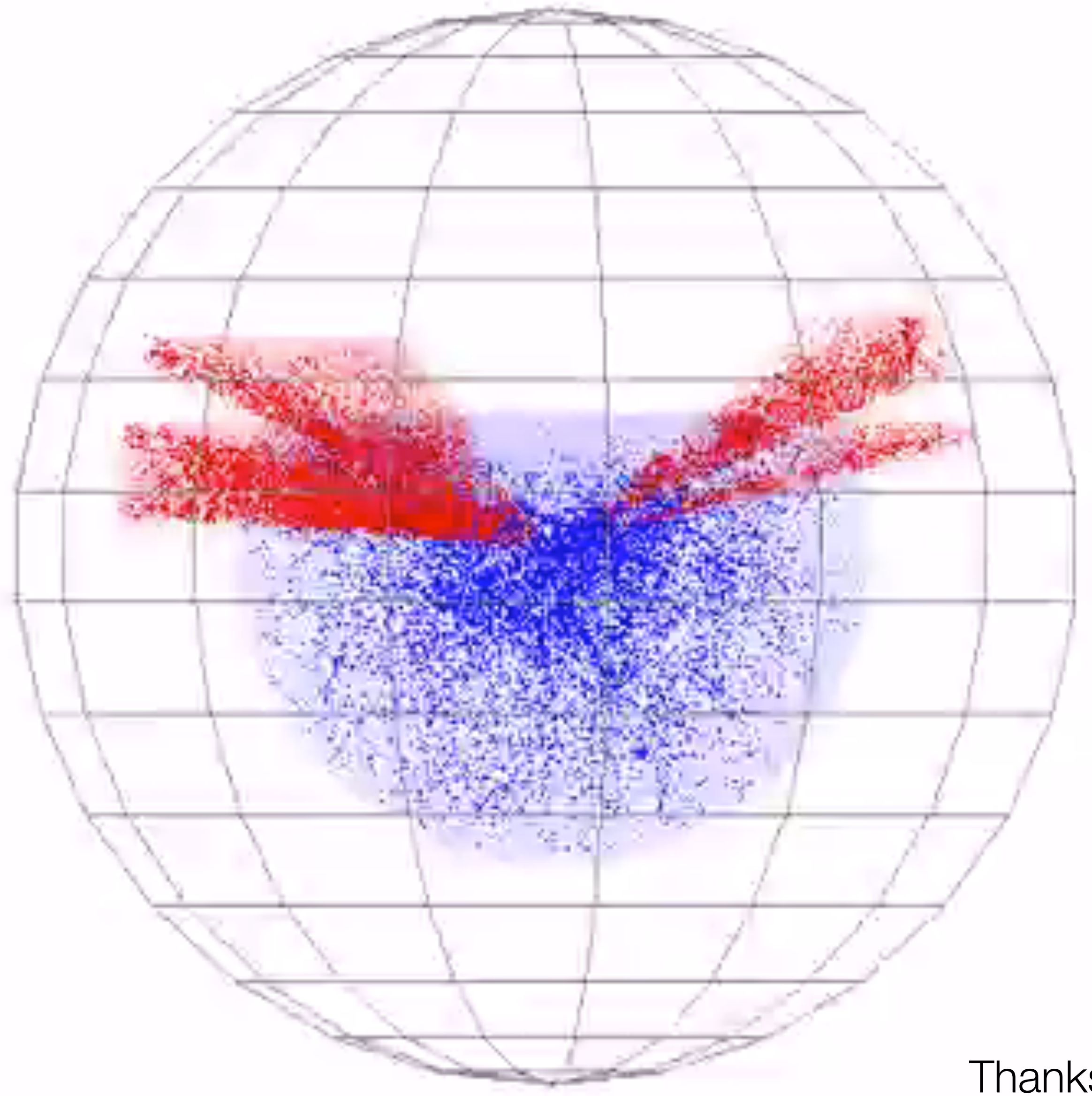
- Dingo:
  - measure HIMF above M_{HI}^* out to z~0.3
  - C- method performs well

Testing normalisations

- Normalising using ‘counts’ is very reliable

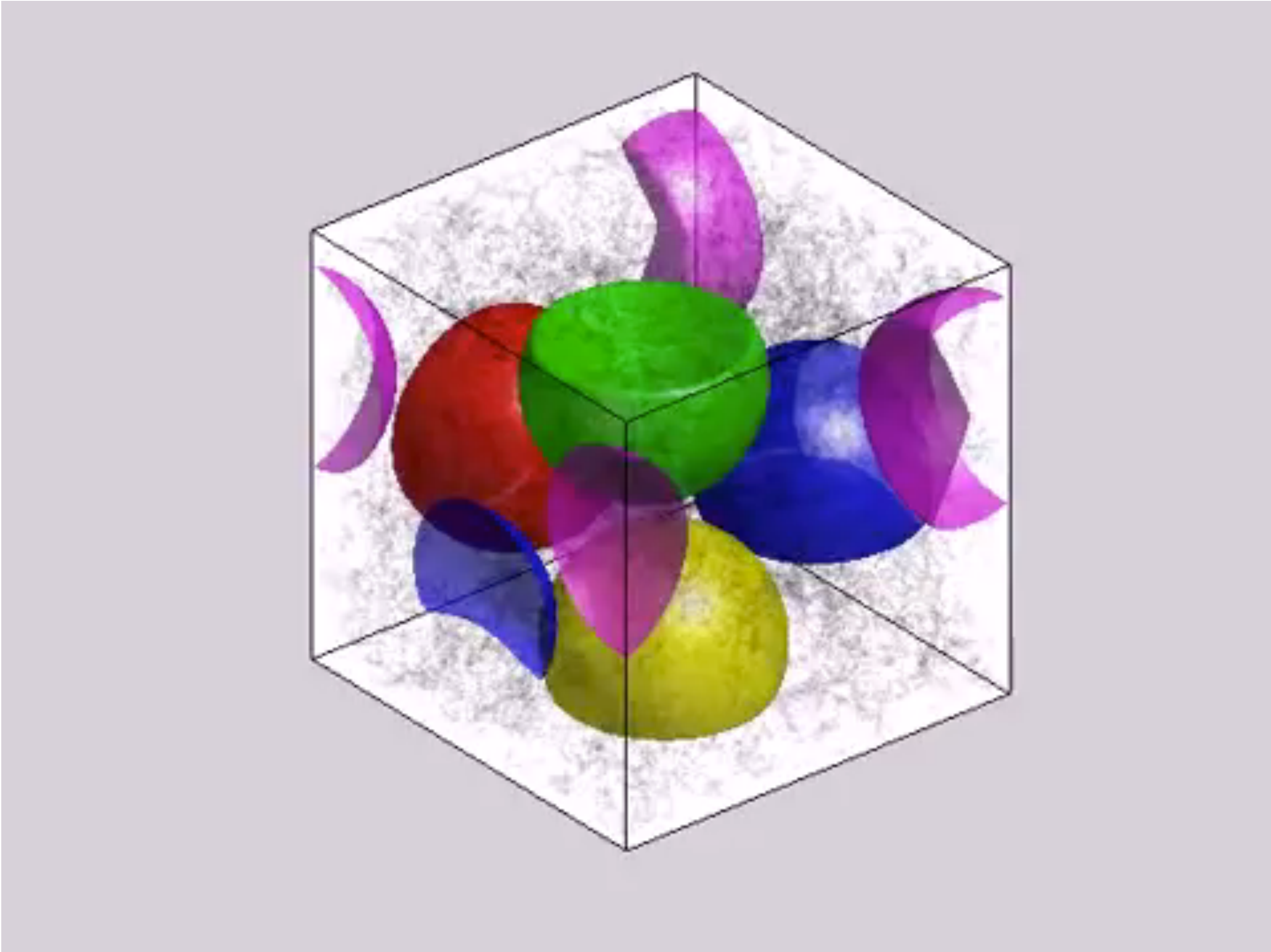


HIPASS and 40% ALFALFA in Millennium



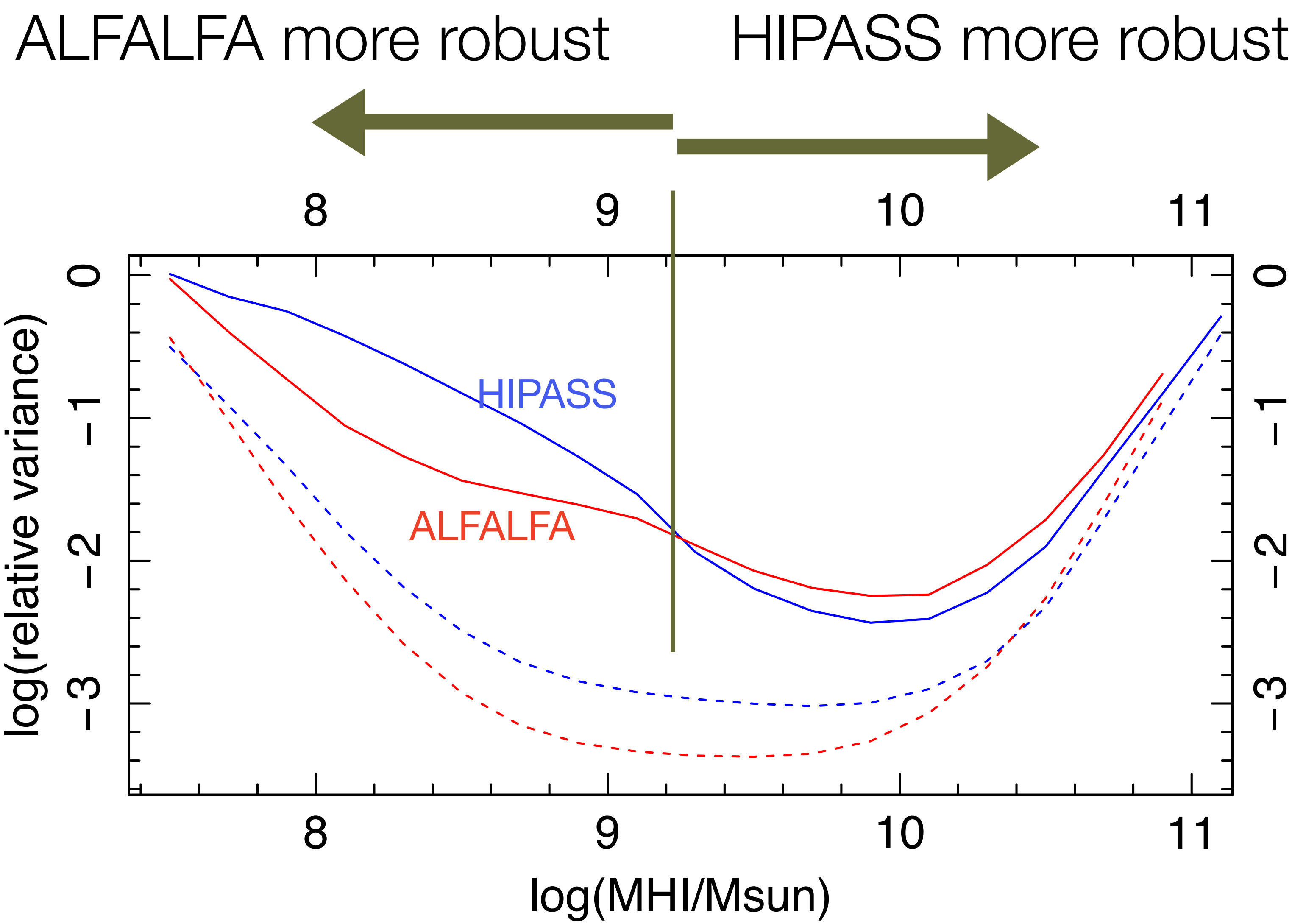
Thanks to Danail Obreschkow

HIPASS and 40%-ALFALFA inside Millennium simulation box (S3-SAX)



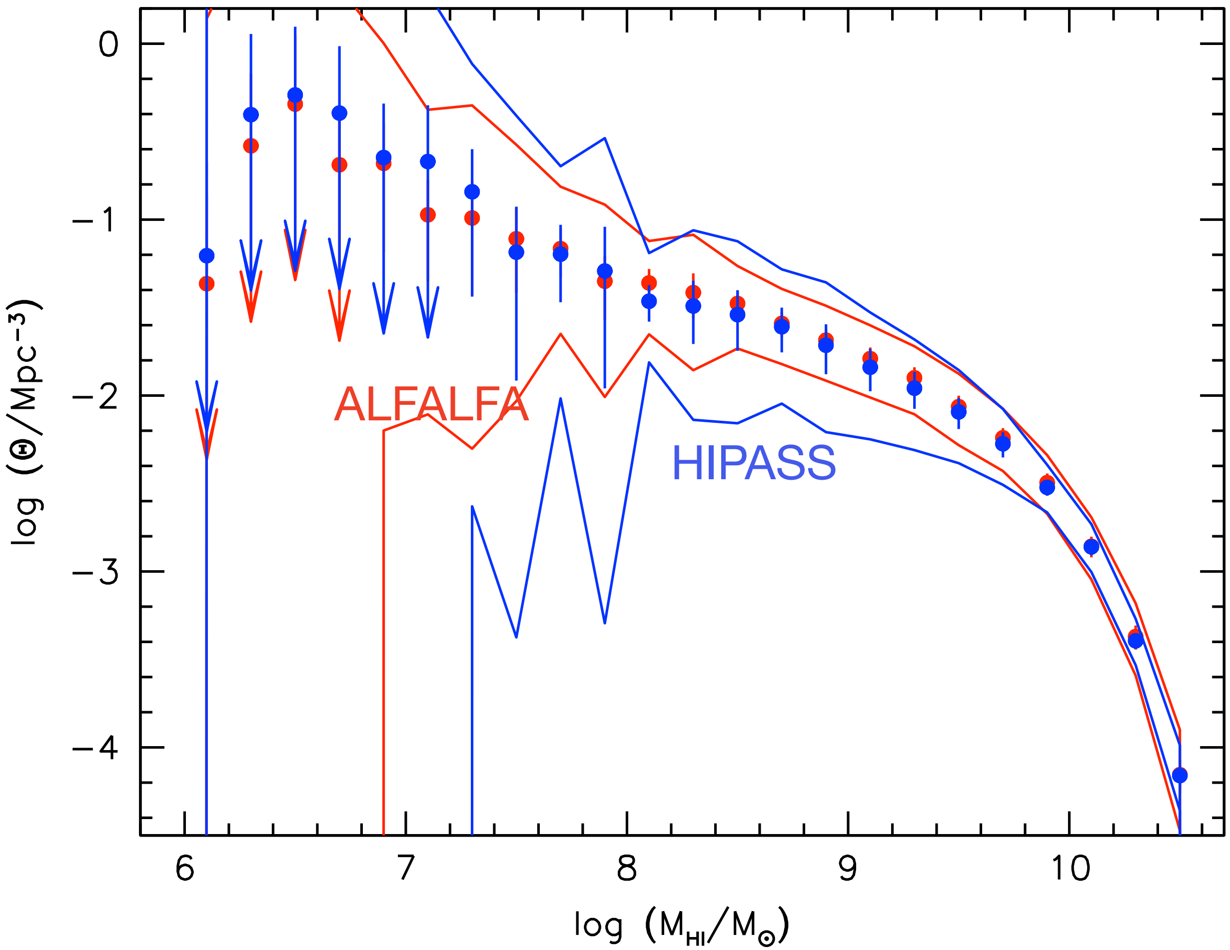
Thanks to Danail Obreschkow

Cosmic variance in HIPASS and ALFALFA (40%)



Compare HIPASS and ALFALFA (40%) HIMF

- Based on Millennium and C- method



- At high HI masses HIPASS and ALFALFA are equally accurate

Ignored…

- Peculiar motions
- Noise bias
- Confusion
- Inclination bias
- Eddington effect
- HI self-absorption

HIMFs from next generation HI surveys

- Use the C- method
  - Robust against LSS
  - Works with ‘soft’ completeness limits
  - Fast