Persistent Homology for Constraining Reionization using Ly-α Emitters

Shifting Landscapes in Astrophysics: New Fronteirs to Explore with Roman

Michael M. Wyatt, Ph.D. Candidate, UCLA

Advisor:
Steven R. Furlanetto
Collaborators:
Andrei Mesinger, Nikolaos Triantafyllou,
Samuel Gagnon-Hartman,
Masami Ouchi, Hiroya Umeda

In this talk…

- Background: how do we constrain reionization using Ly-α emitters?
- Introduce **persistent homology** and the **Betti-0 curve**
- Compare Betti-0 curve and existing methods:
    - **simulate** reionization and Ly-α emitters
    - create **mock observations** possible with Roman
    - **compare constraining power** of methods

How (and why) we use LAEs to constrain reionization

Reionization

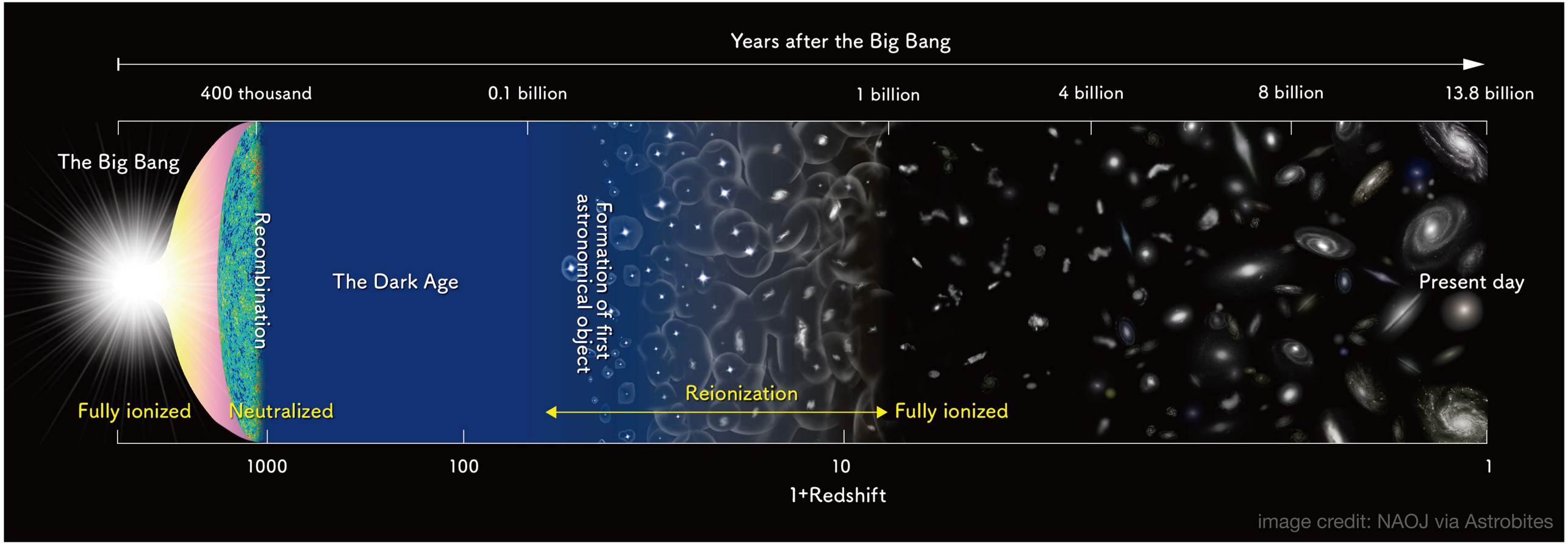


image credit: NAOJ via Astrobites

Reionization

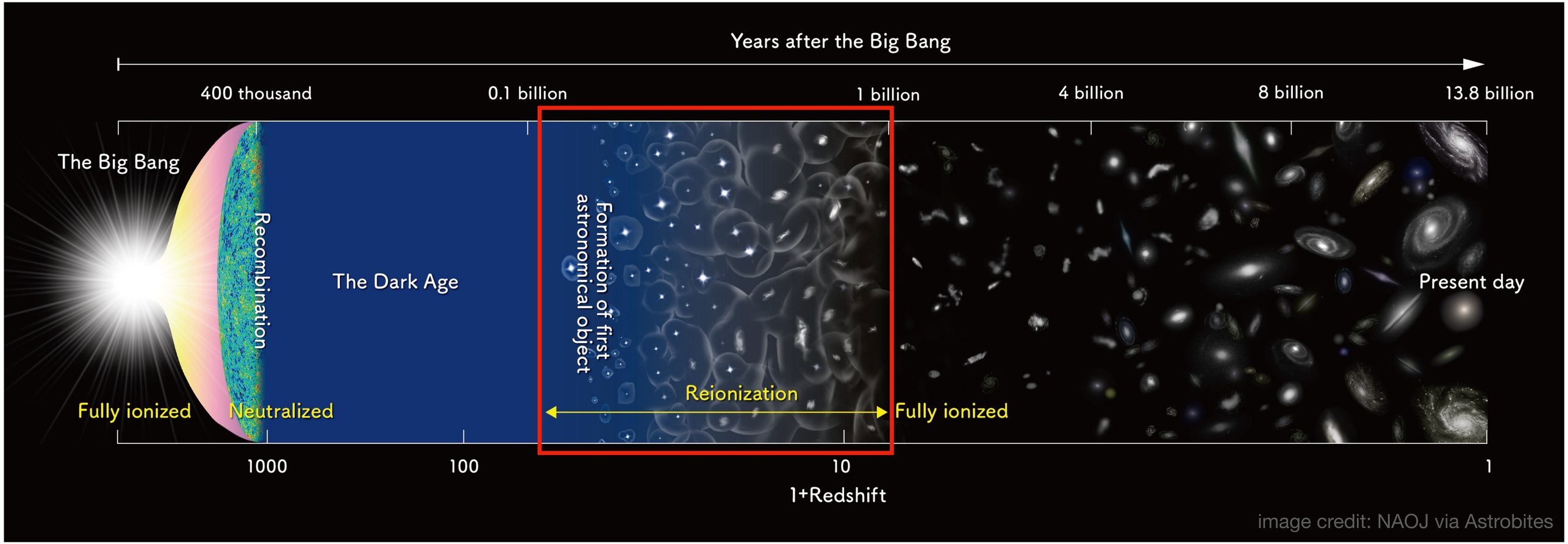
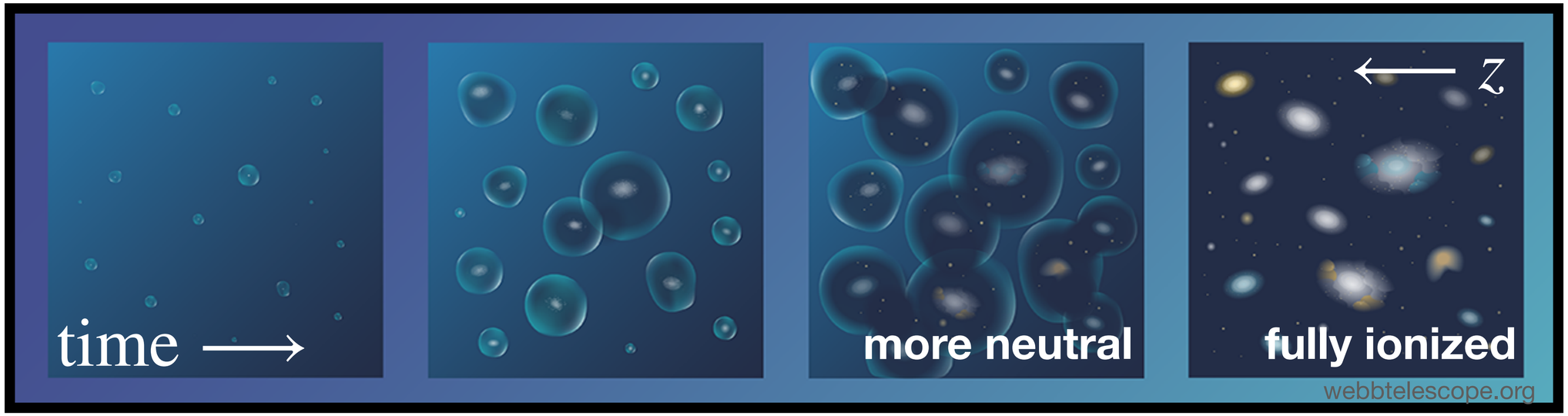


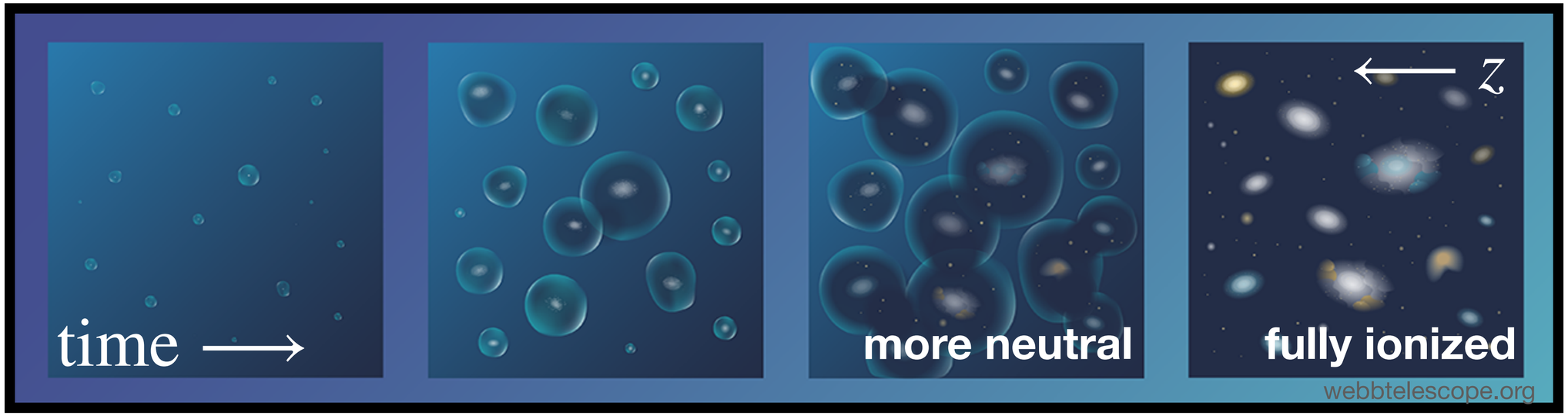
image credit: NAOJ via Astrobites

Ionized “bubble” morphology reveals information about sources

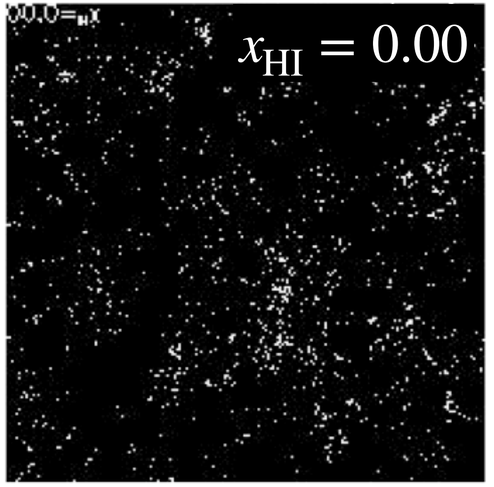


Reionization Enhances Clustering of LAEs

- LAEs visible only inside large ionized bubbles



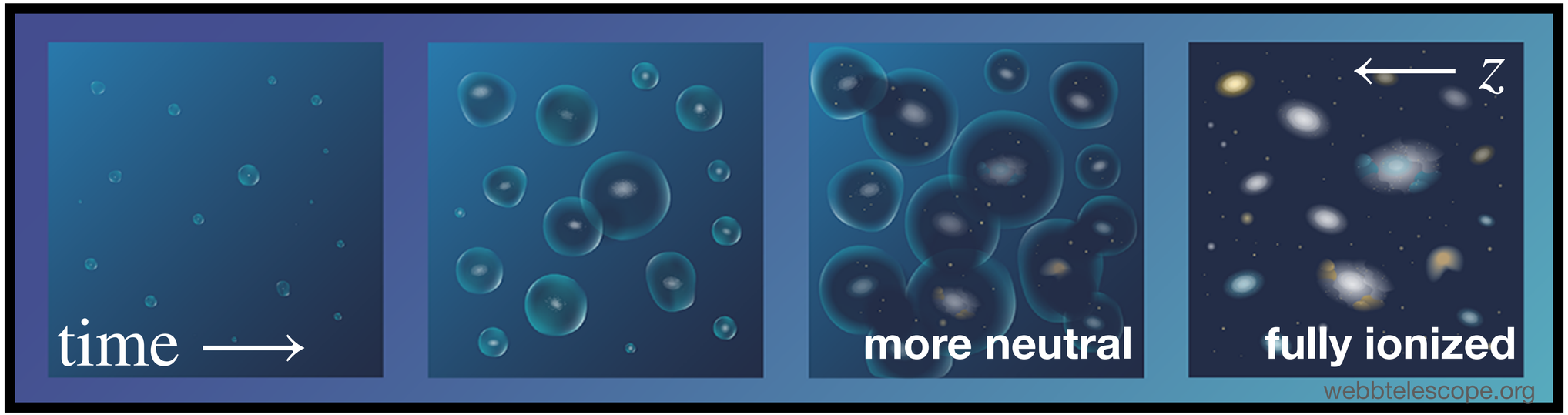
neutral fraction



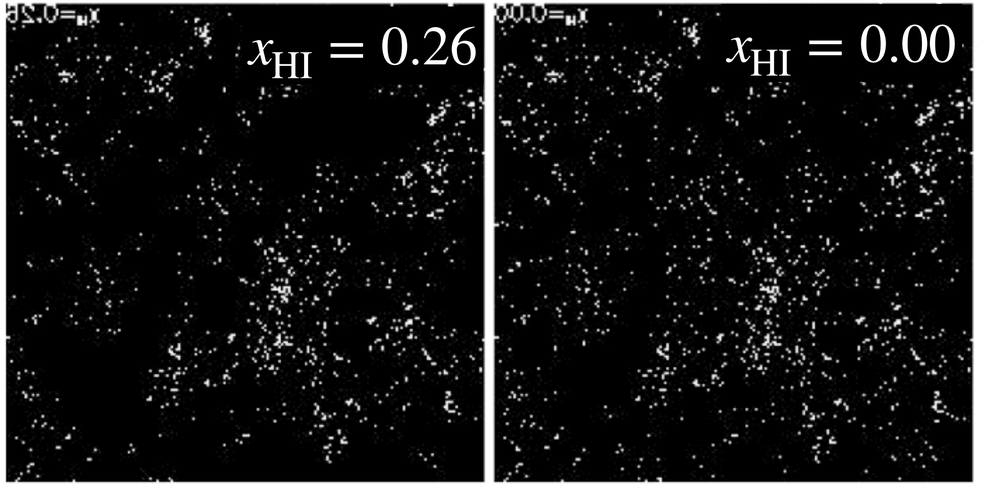
Mesinger+08

Reionization Enhances Clustering of LAEs

- LAEs visible only inside large ionized bubbles



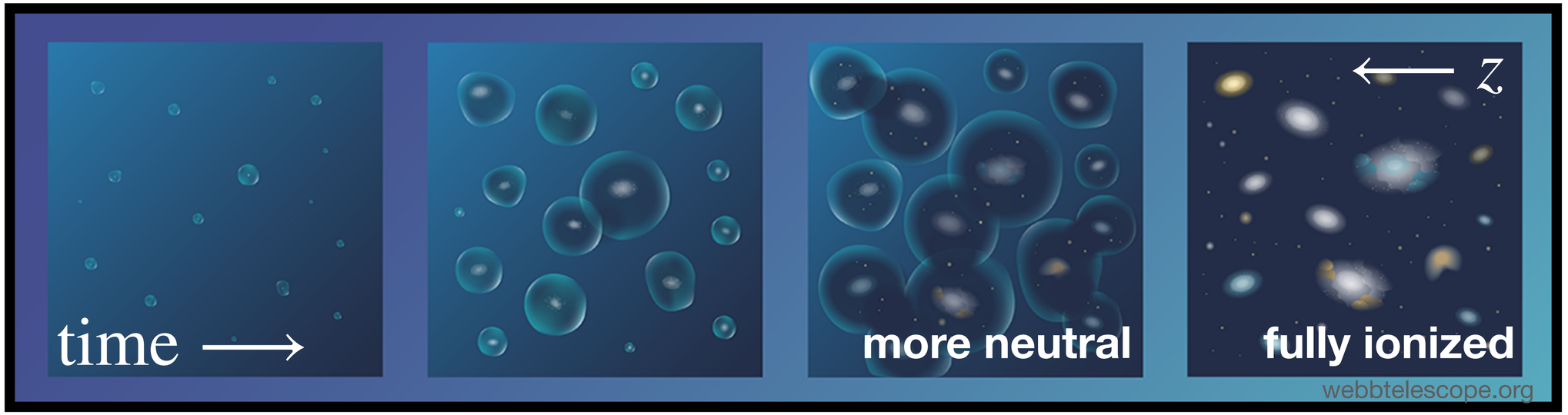
neutral fraction



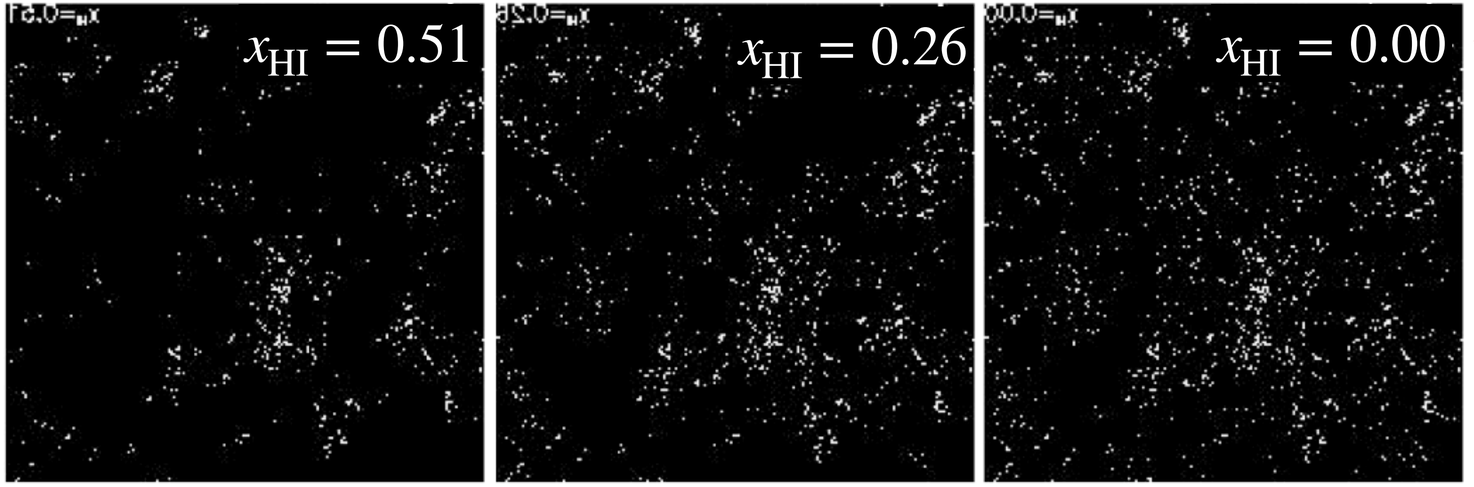
Mesinger+08

Reionization Enhances Clustering of LAEs

- LAEs visible only inside large ionized bubbles



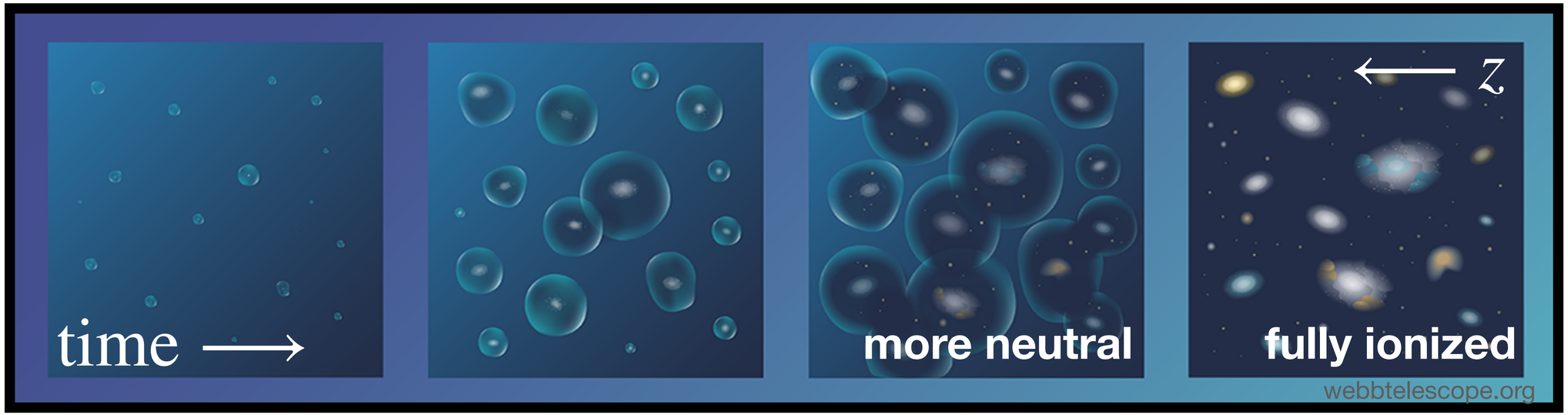
neutral fraction



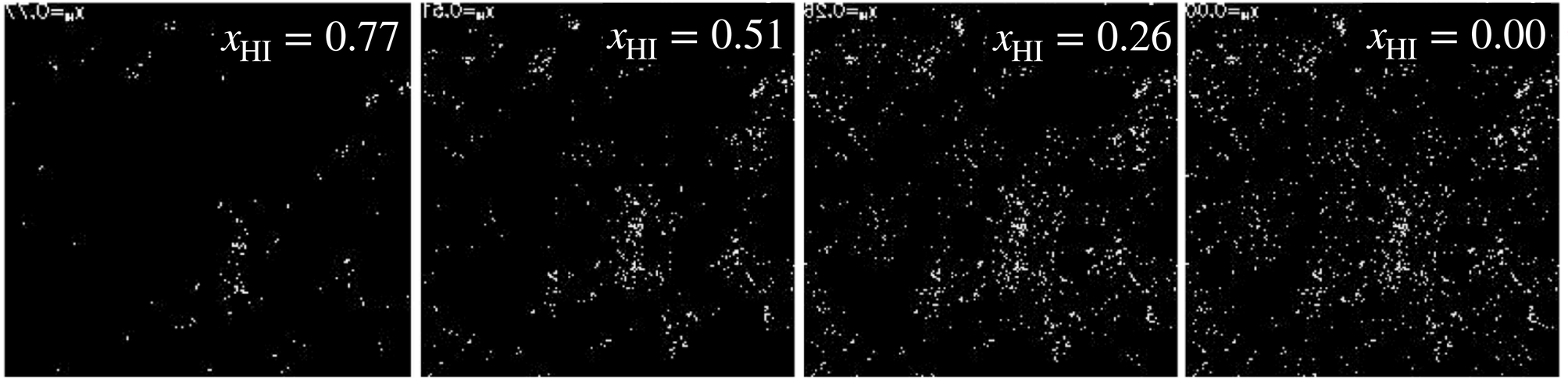
Mesinger+08

Reionization Enhances Clustering of LAEs

- LAEs visible only inside large ionized bubbles



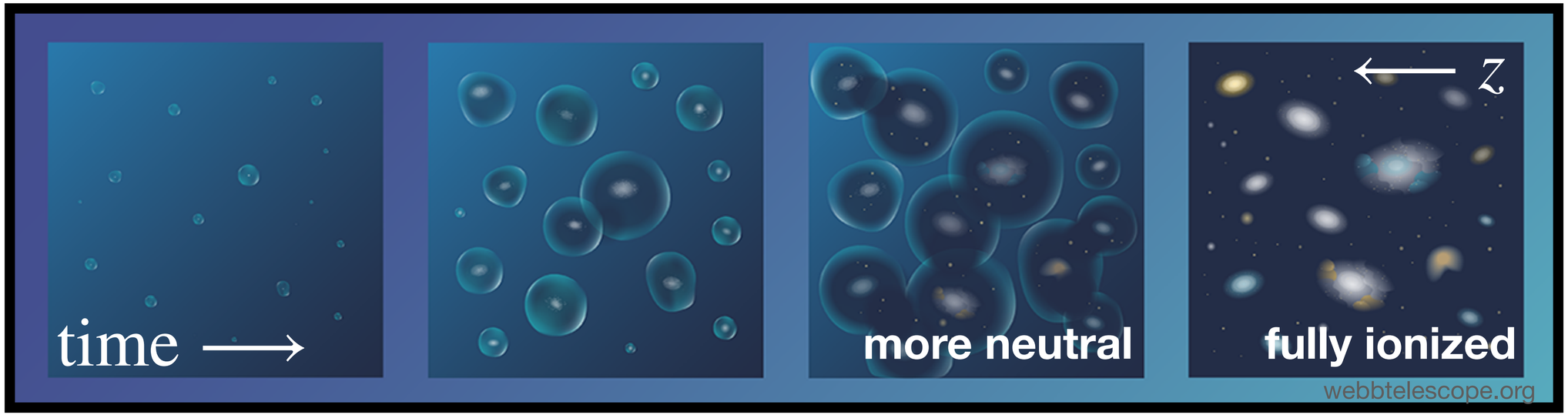
neutral fraction

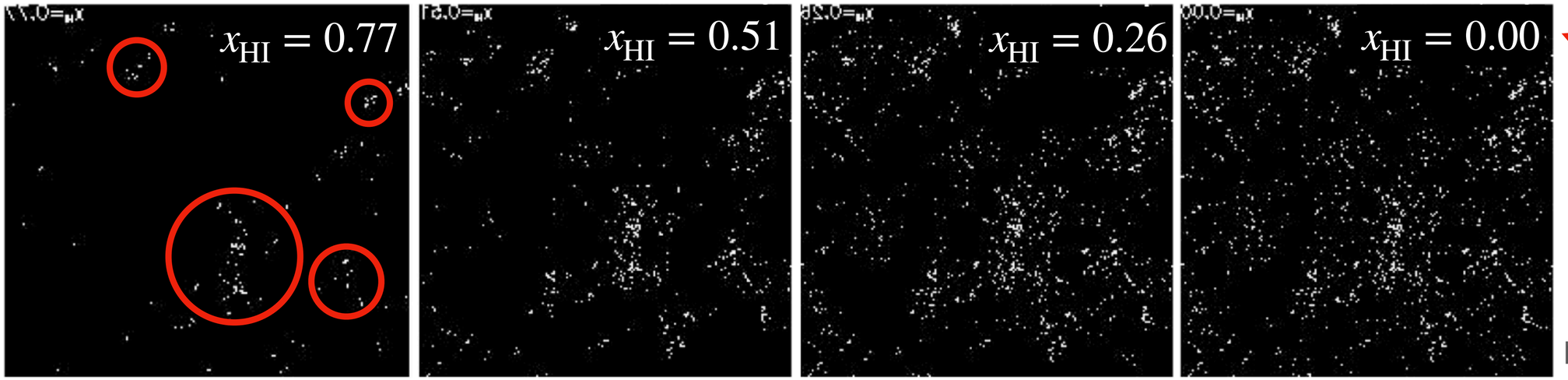


Mesinger+08

Reionization Enhances Clustering of LAEs

- LAEs visible only inside large ionized bubbles





neutral fraction

Mesinger+08

How do we quantify this clustering?

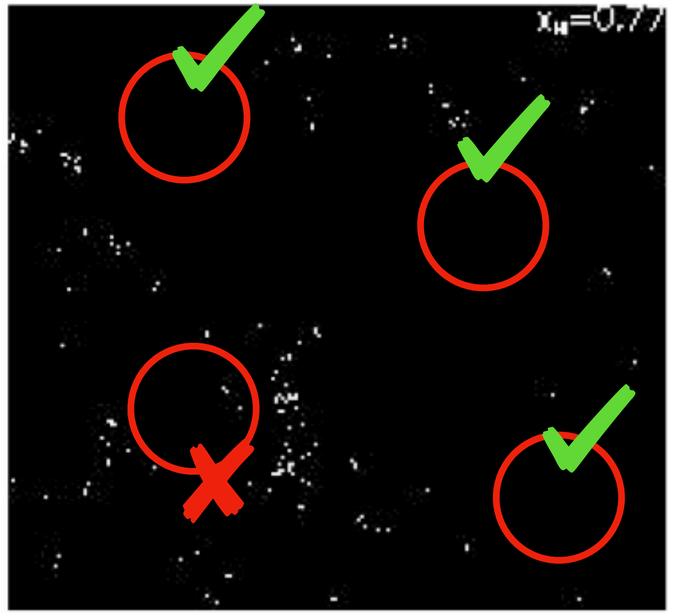
- There exist many methods…
- 2 point correlation function (or angular correlation function; **ACF**)
    - count number of pairs at a given separation

How do we quantify this clustering?

- There exist many methods…
- 2 point correlation function (or angular correlation function; **ACF**)
    - count number of pairs at a given separation
- Void probability function (**VPF**)
    - fraction of circles of some radius which are empty

How do we quantify this clustering?

- There exist many methods…
- 2 point correlation function (or angular correlation function; **ACF**)
    - count number of pairs at a given separation
- Void probability function (**VPF**)
    - fraction of circles of some radius which are empty

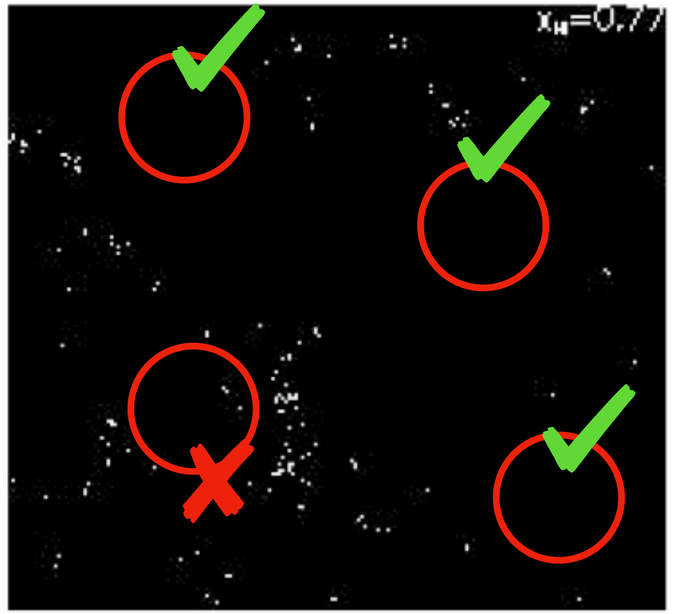


75% empty

How do we quantify this clustering?

- There exist many methods…
- 2 point correlation function (or angular correlation function; **ACF**)
    - count number of pairs at a given separation
- Void probability function (**VPF**)
    - fraction of circles of some radius which are empty
- **Persistent homology**
    - has been used in cosmology
    - not for LAEs

NEW!



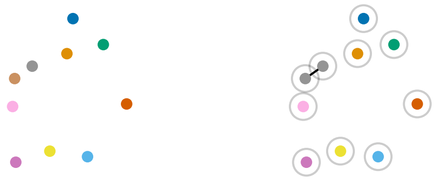
75% empty

Persistent Homology for Constraining Reionization

Persistent Homology

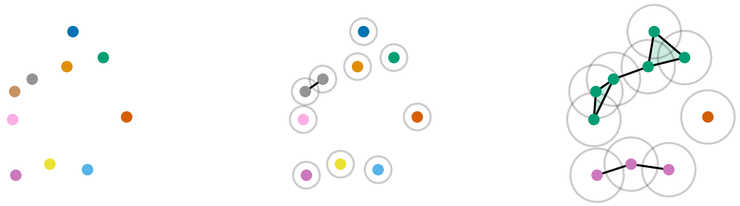
Persistent Homology

$\epsilon = 0.0$ $\epsilon = 0.25$



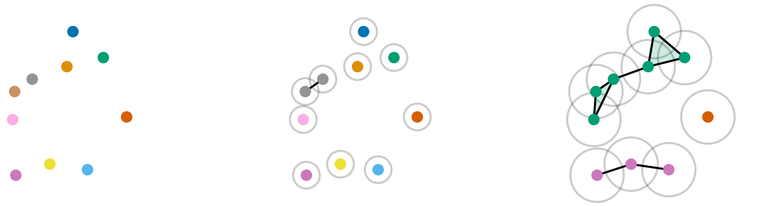
Persistent Homology

$\epsilon = 0.0$ $\epsilon = 0.25$ $\epsilon = 0.5$



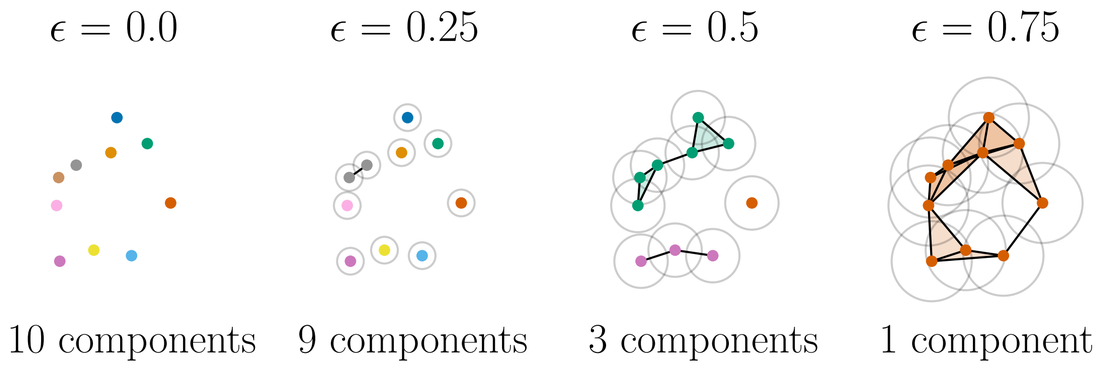
Persistent Homology

$\epsilon = 0.0$ $\epsilon = 0.25$ $\epsilon = 0.5$

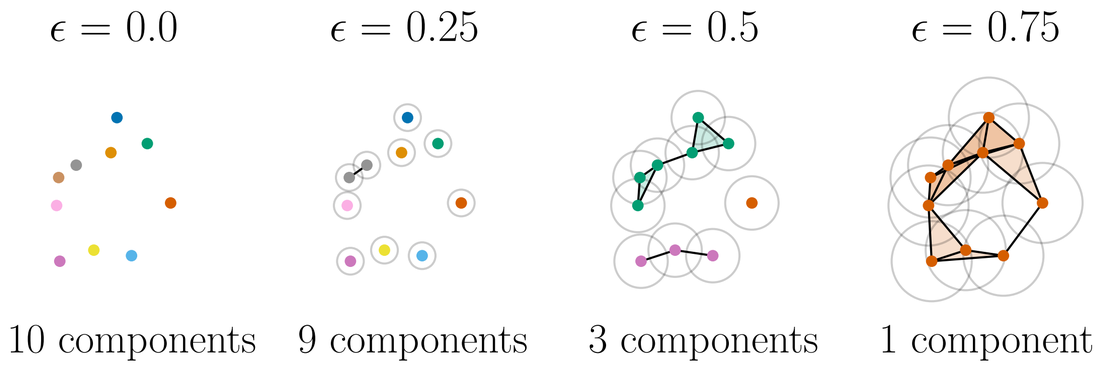


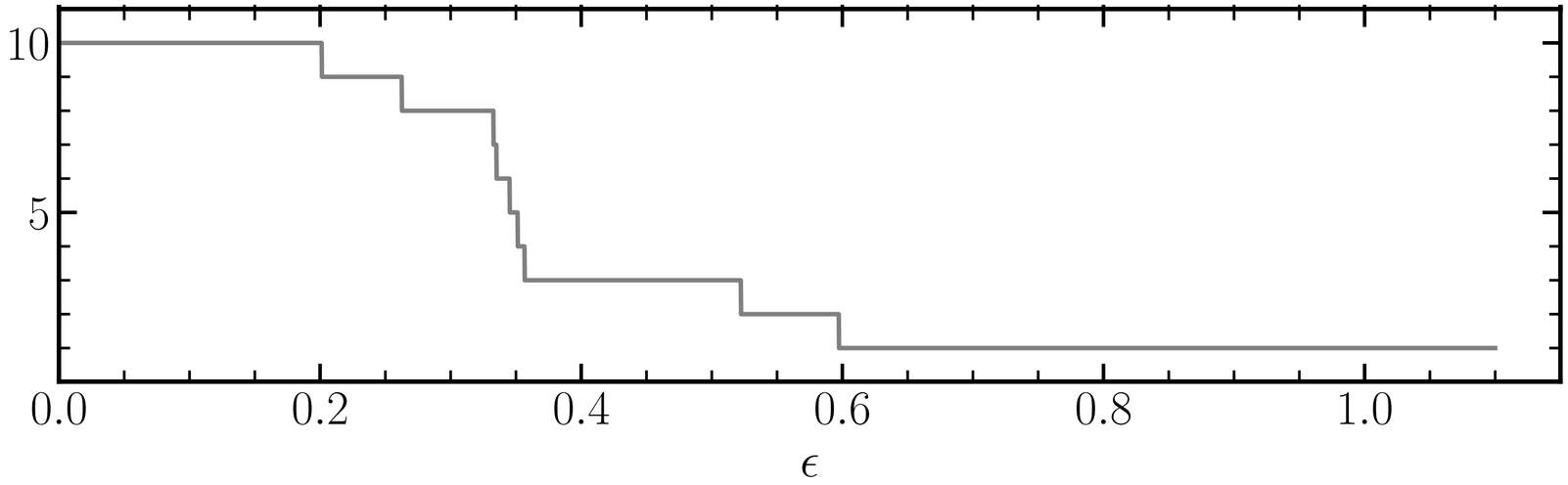
10 components 9 components 3 components

Persistent Homology



Persistent Homology





“Betti-0 curve”

Simulating Reionization and Mock Observations

Simulations of LAEs

- 21cmFAST
- 2,000 slices of the Universe, 300 x 300 cMpc^2, ~4 deg^2 at $z = 7.2$
    - Locations and observed fluxes of LAEs $f_{\alpha,\mathrm{obs}}$

Mock Observations

- Simple, uniform selection of LAEs with fluxes above limit $f_{\alpha,\mathrm{obs}} > f_{\mathrm{lim}}$
- We will focus on a global neutral fraction of $\left\langle x_{\mathrm{HI}} \right\rangle = 0.4$ at $z = 7.2$
    - $\Delta z = 0.25$ or ~ 80 cMpc thick
-

Mock Observations

- Simple, uniform selection of LAEs with fluxes above limit $f_{\alpha,\mathrm{obs}} > f_{\mathrm{lim}}$
- We will focus on a global neutral fraction of $\langle x_{\mathrm{HI}} \rangle = 0.4$ at $z = 7.2$
  - $\Delta z = 0.25$ or ~ 80 cMpc thick
- Three observations:

description	total survey area	total survey time	f_{lim} (50% completeness)
small area, deep (Wold et al., 2024)	1 deg^2	12 days	8.2×10^{-18} erg s^{-1} cm^{-2}
large area, shallow	4 deg^2	12 days	16.4×10^{-18} erg s^{-1} cm^{-2}
large area, medium (Wold et al., 2024)	4 deg^2	28 days	10.2×10^{-18} erg s^{-1} cm^{-2}

$f_{\mathrm{lim}} \propto t_{\mathrm{int}}^{-1/2}$

Mock Observations

- Simple, uniform selection of LAEs with fluxes above limit $f_{\alpha,\mathrm{obs}} > f_{\mathrm{lim}}$
- We will focus on a global neutral fraction of $\left\langle x_{\mathrm{HI}} \right\rangle = 0.4$ at $z = 7.2$
  - $\Delta z = 0.25$ or ~ 80 cMpc thick
- Three observations:

	description	total survey area	total survey time	f_{lim} (50% completeness)
Wold et al., 2024	small area, deep	1 deg^2	12 days	8.2×10^{-18} erg s^{-1} cm^{-2}
	large area, shallow	4 deg^2	12 days	16.4×10^{-18} erg s^{-1} cm^{-2}
Wold et al., 2024	large area, medium	4 deg^2	28 days	10.2×10^{-18} erg s^{-1} cm^{-2}

$f_{\mathrm{lim}} \propto t_{\mathrm{int}}^{-1/2}$

Mock Observations

- Simple, uniform selection of LAEs with fluxes above limit $f_{\alpha,\mathrm{obs}} > f_{\mathrm{lim}}$
- We will focus on a global neutral fraction of $\langle x_{\mathrm{HI}} \rangle = 0.4$ at $z = 7.2$
  - $\Delta z = 0.25$ or ~ 80 cMpc thick
- Three observations:

	description	total survey area	total survey time	f_{lim} (50% completeness)
Wold et al., 2024	small area, deep	1 deg^2	12 days	8.2×10^{-18} erg s^{-1} cm^{-2}
	large area, shallow	4 deg^2	12 days	16.4×10^{-18} erg s^{-1} cm^{-2}
Wold et al., 2024	large area, medium	4 deg^2	28 days	10.2×10^{-18} erg s^{-1} cm^{-2}

$f_{\mathrm{lim}} \propto t_{\mathrm{int}}^{-1/2}$

Mock Observations

- Simple, uniform selection of LAEs with fluxes above limit $f_{\alpha,\mathrm{obs}} > f_{\mathrm{lim}}$
- We will focus on a global neutral fraction of $\langle x_{\mathrm{HI}} \rangle = 0.4$ at $z = 7.2$
  - $\Delta z = 0.25$ or ~ 80 cMpc thick
- Three observations:

	description	total survey area	total survey time	f_{lim} (50% completeness)
Wold et al., 2024	small area, deep	1 deg^2	12 days	8.2×10^{-18} erg s^{-1} cm^{-2}
	large area, shallow	4 deg^2	12 days	16.4×10^{-18} erg s^{-1} cm^{-2}
Wold et al., 2024	large area, medium	4 deg^2	28 days	10.2×10^{-18} erg s^{-1} cm^{-2}

$f_{\mathrm{lim}} \propto t_{\mathrm{int}}^{-1/2}$

Mock Observations

- Simple, uniform selection of LAEs with fluxes above limit $f_{\alpha,\mathrm{obs}} > f_{\mathrm{lim}}$
- We will focus on a global neutral fraction of $\langle x_{\mathrm{HI}} \rangle = 0.4$ at $z = 7.2$
  - $\Delta z = 0.25$ or ~ 80 cMpc thick
- Three observations:

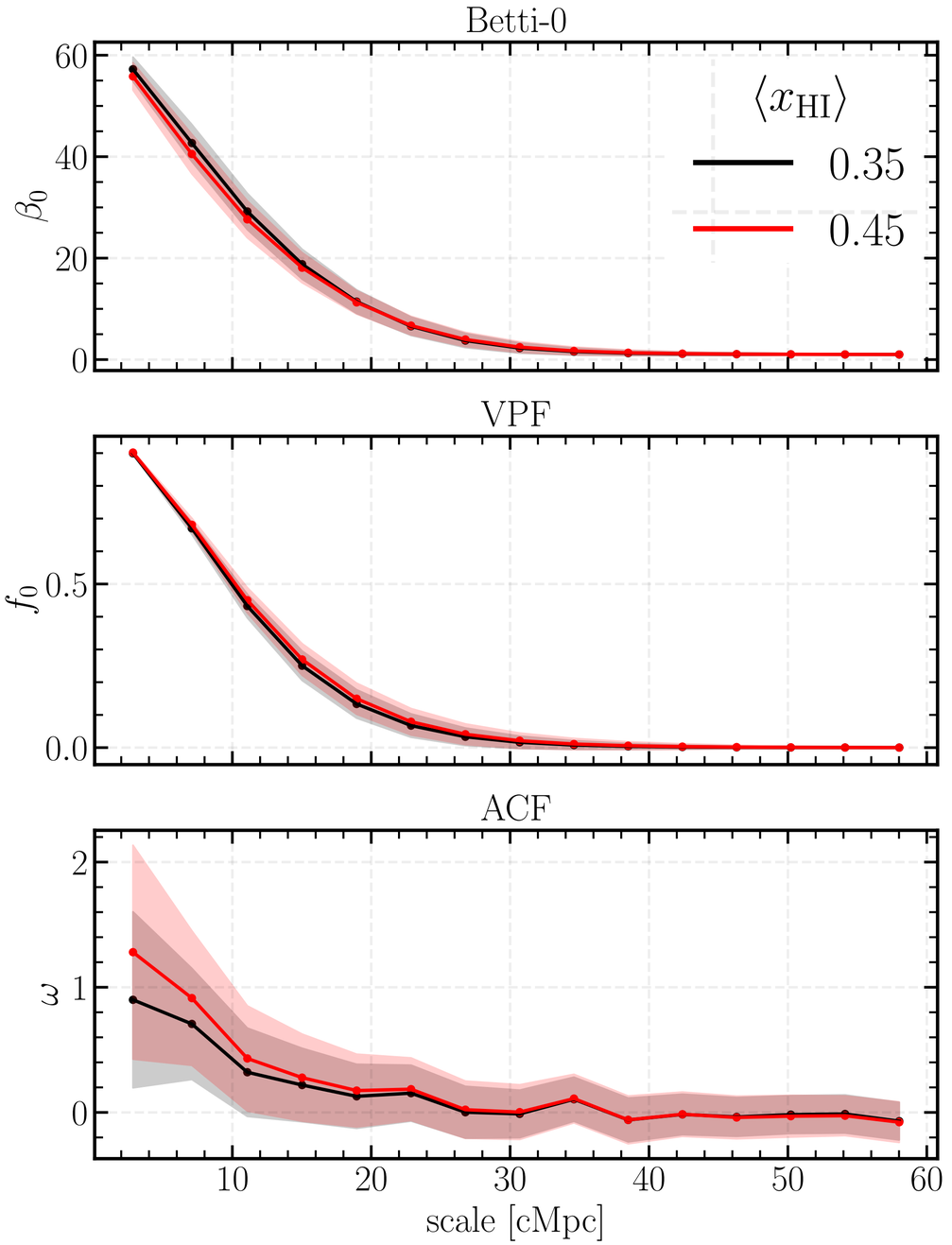
	description	total survey area	total survey time	f_{lim} (50% completeness)
Wold et al., 2024	small area, deep	1 deg^2	12 days	8.2×10^{-18} erg s^{-1} cm^{-2}
	large area, shallow	4 deg^2	12 days	16.4×10^{-18} erg s^{-1} cm^{-2}
Wold et al., 2024	large area, medium	4 deg^2	28 days	10.2×10^{-18} erg s^{-1} cm^{-2}

$f_{\mathrm{lim}} \propto t_{\mathrm{int}}^{-1/2}$

Results

Small area, deep survey

1 deg², 12 days

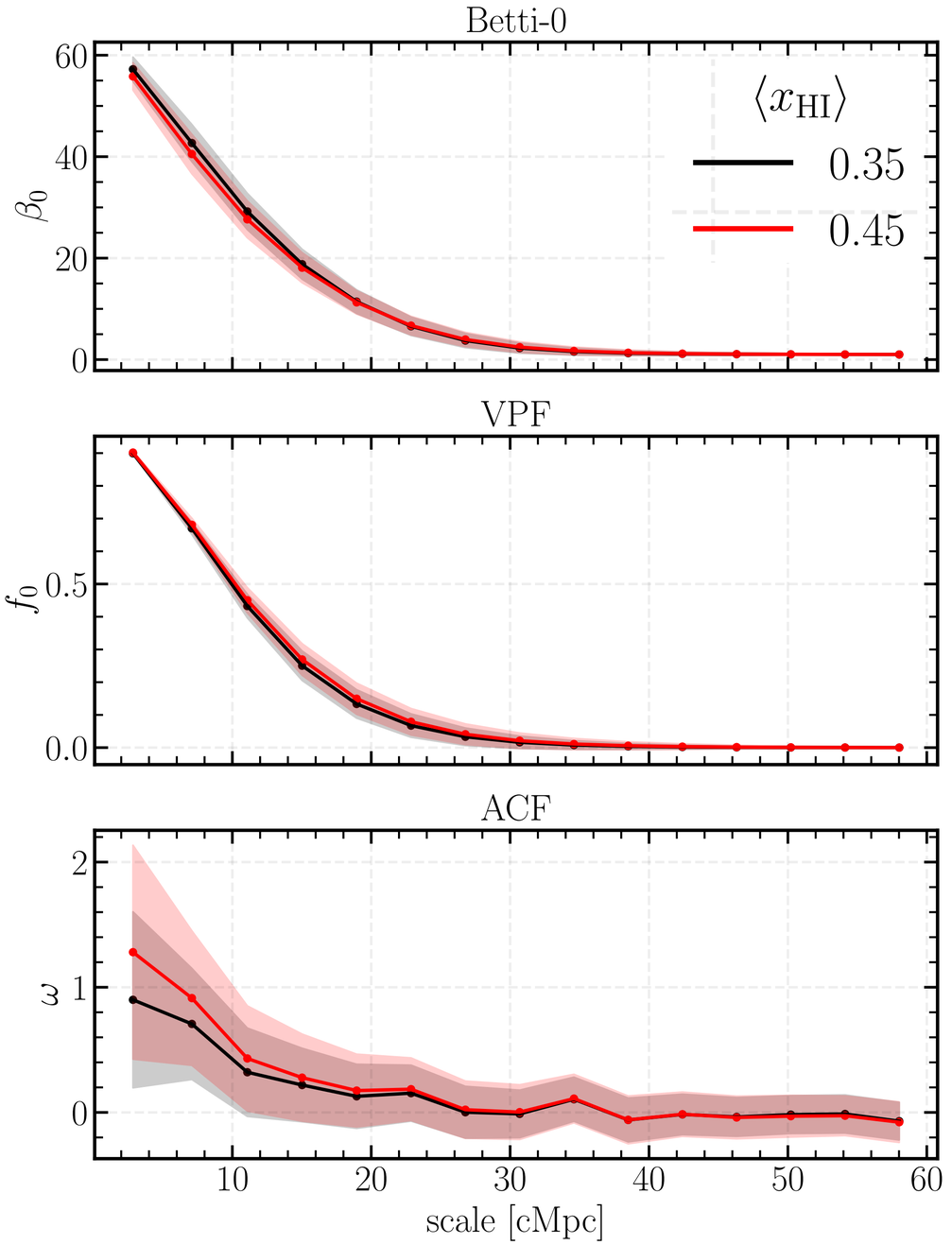


Small area, deep survey

1 deg², 12 days

- To compare quantitatively:
- Fisher information:

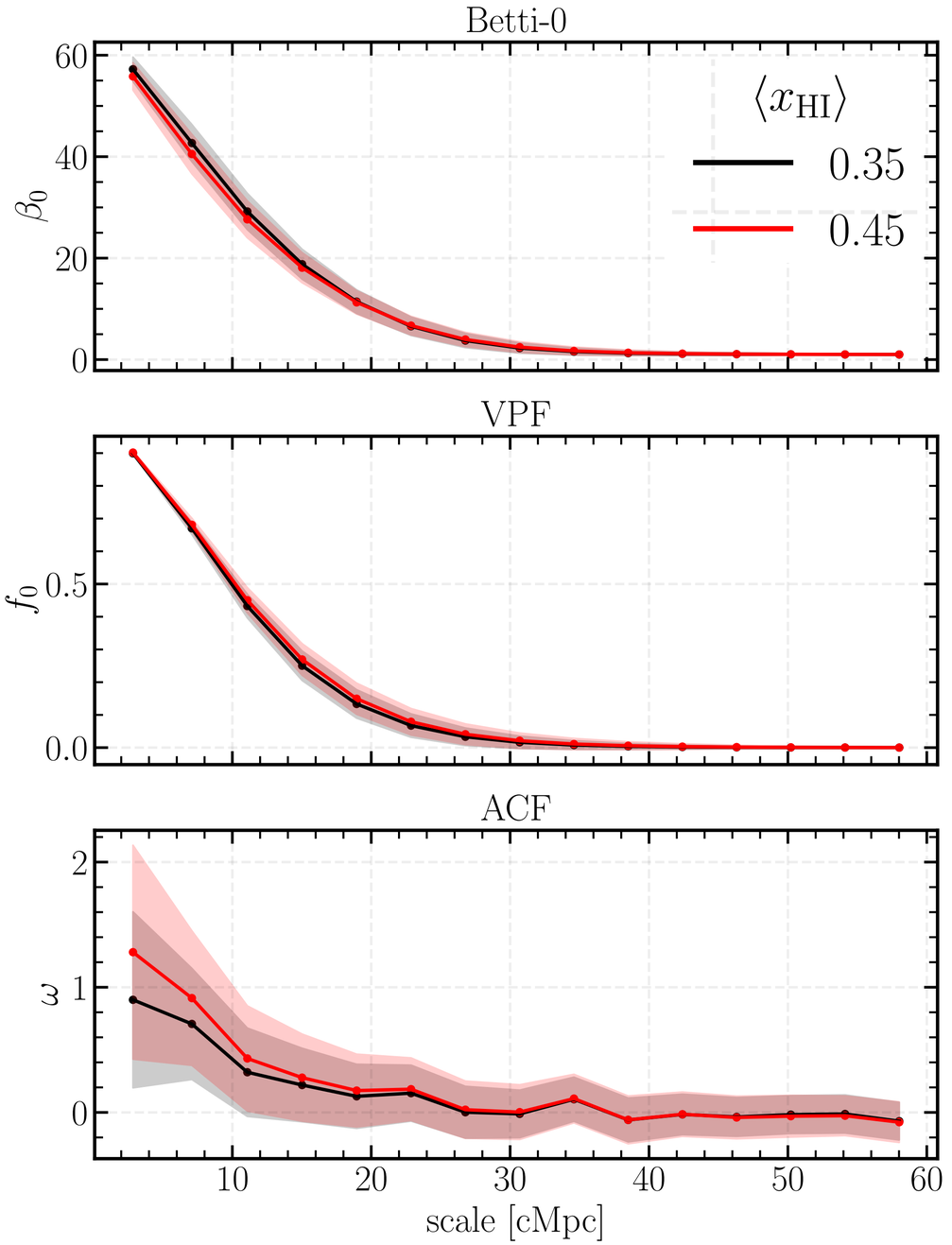
$$F = \frac{\partial \boldsymbol{\mu}^T}{dx_{\rm HI}} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{dx_{\rm HI}}$$



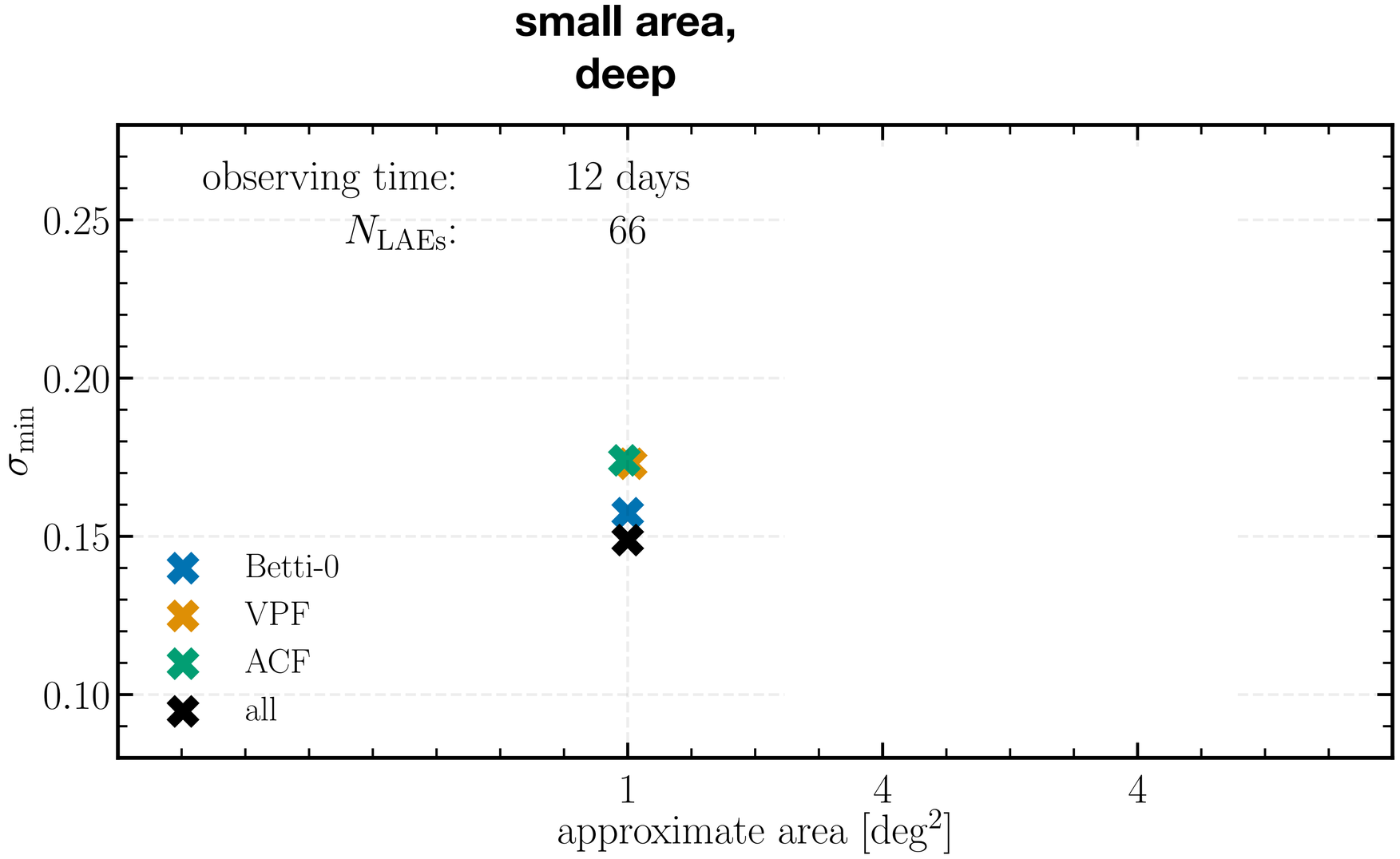
Small area, deep survey

1 deg², 12 days

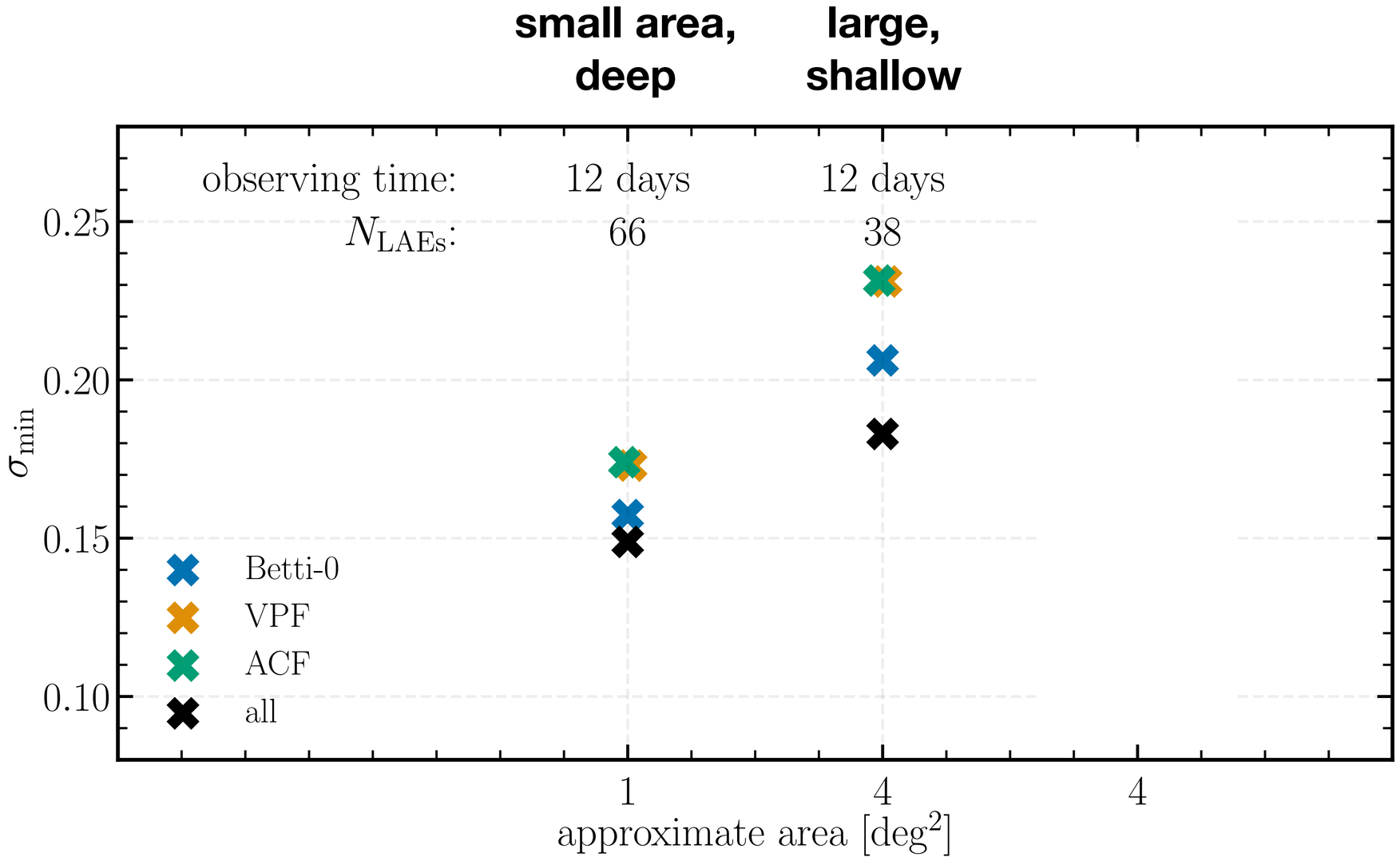
- To compare quantitatively:
- Fisher information:
  - $$F = \frac{\partial \boldsymbol{\mu}^T}{dx_{\rm HI}} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{dx_{\rm HI}}$$
- Cramér-Rao bound:
  - $$\sigma \geq 1/\sqrt{F}$$



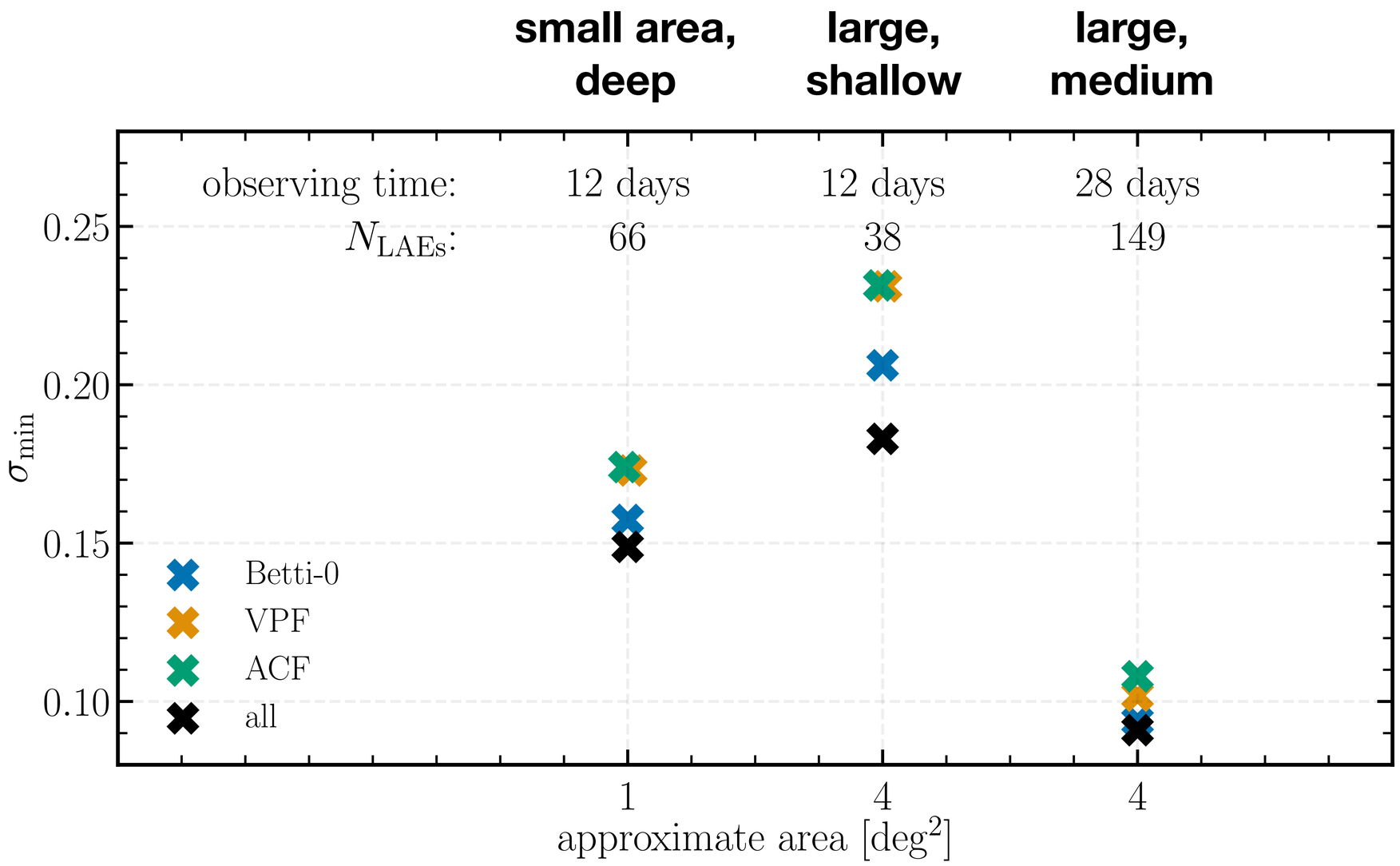
Results


small area,
deep
observing time: 12 days
N_{LAEs}: 66
0.25
0.20
0.15
0.10
σ_{min}
Betti-0
VPF
ACF
all
1
4
4
approximate area [deg^2]

Results


small area, deep
large, shallow
observing time:
12 days
12 days
N_{LAEs}:
66
38
0.25
0.20
0.15
0.10
σ_{min}
Betti-0
VPF
ACF
all
1
4
4
approximate area [deg^2]

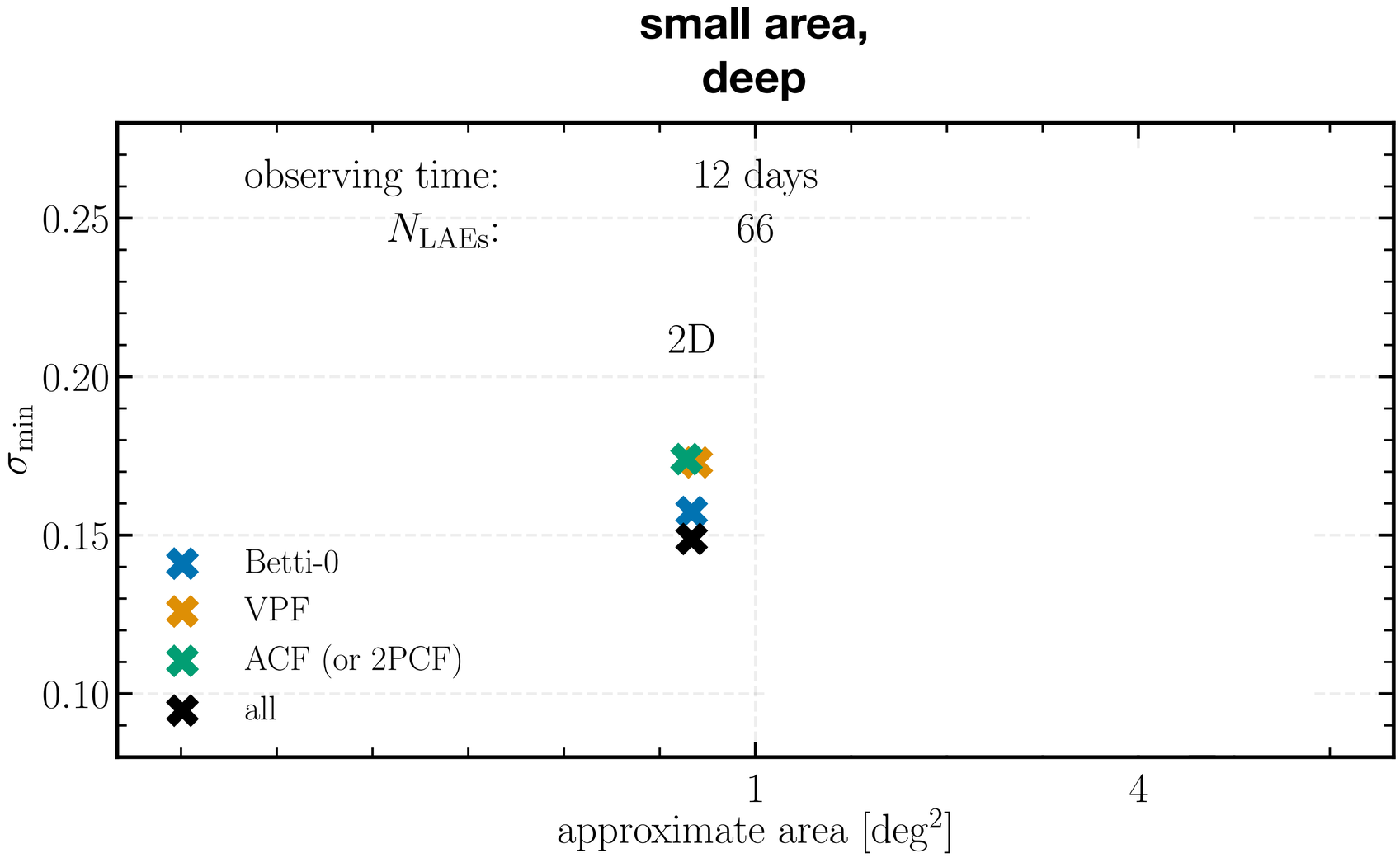
Results


small area, deep
large, shallow
large, medium
observing time:
12 days
12 days
28 days
N_{LAEs}:
66
38
149
0.25
0.20
0.15
0.10
σ_{min}
Betti-0
VPF
ACF
all
1
4
4
approximate area [deg^2]

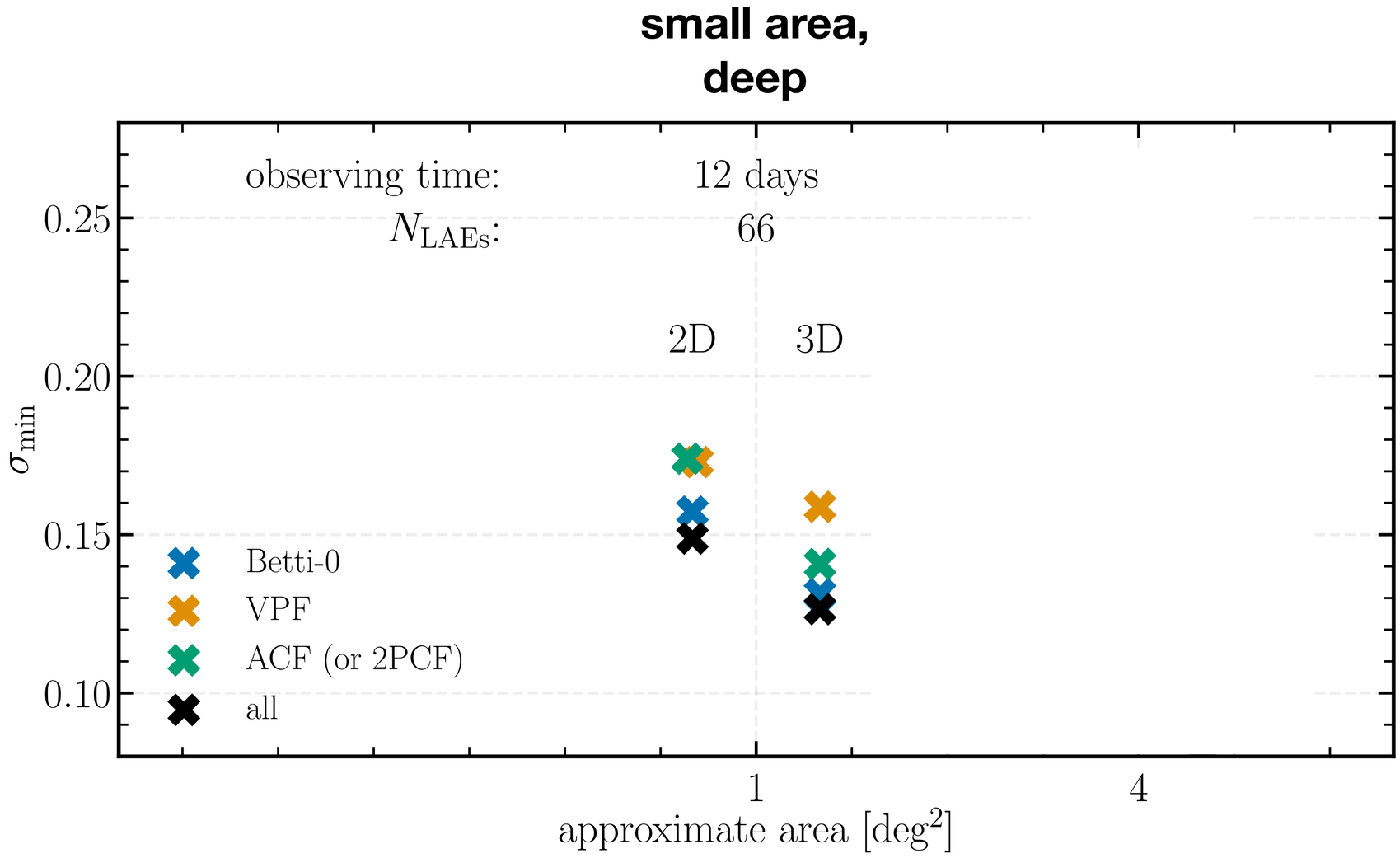
2D vs 3D statistics

- All statistics can be defined in 3 dimensions
- Include position along redshift dimension and recalculate statistics
    - ignore redshift uncertainty

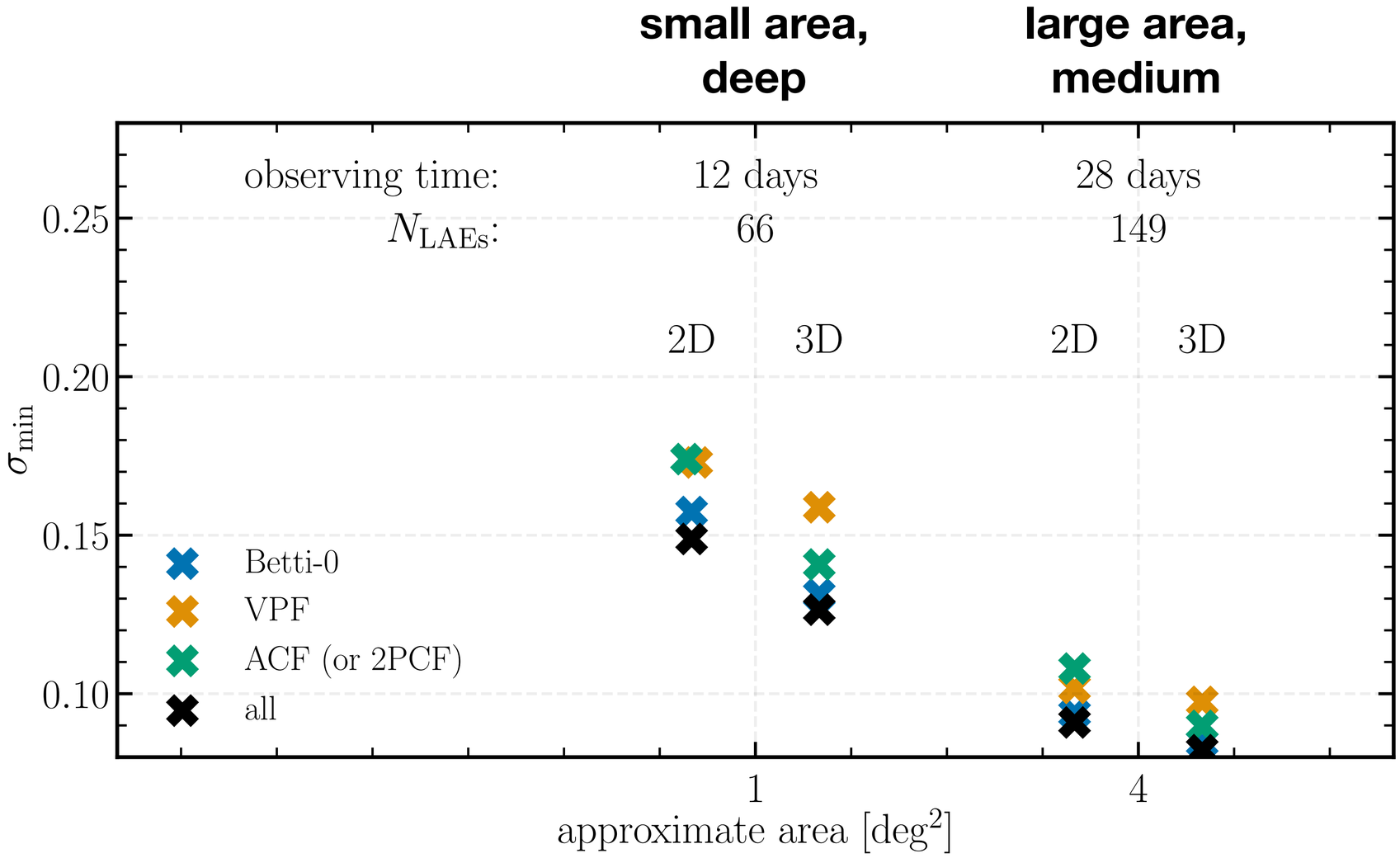
2D vs 3D statistics



2D vs 3D statistics



2D vs 3D statistics



Conclusions

- Persistent homology a promising new category of methods for constraining reionization using LAEs, and should be considered for upcoming observations
    - In this work, we consider only the simplest statistic (Betti-0)
- In observations possible with Roman, the Betti-0 curve performs as well as, or slightly better than, commonly used summary statistics
- The Betti-0 curve is not redundant with the ACF and VPF — fitting the statistics simultaneously provides more information and constraining power
- At the number of LAEs expected in the observations explored here, we find marginal improvements in x_{HI} constraints when using 3D statistics, particularly at lower number densities

Conclusions

Paper coming soon!

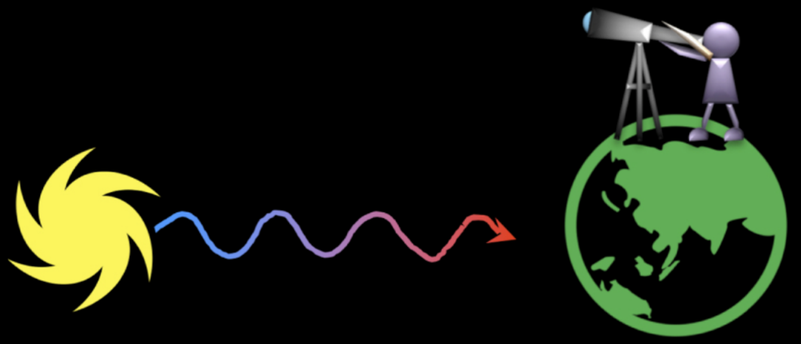
- Persistent homology a promising new category of methods for constraining reionization using LAEs, and should be considered for upcoming observations
    - In this work, we consider only the simplest statistic (Betti-0)
- In observations possible with Roman, the Betti-0 curve performs as well as, or slightly better than, commonly used summary statistics
- The Betti-0 curve is not redundant with the ACF and VPF — fitting the statistics simultaneously provides more information and constraining power
- At the number of LAEs expected in the observations explored here, we find marginal improvements in $x_{\rm HI}$ constraints when using 3D statistics, particularly at lower number densities

backups

Ly-α emitters as a tracer of bubble morphology

neutral IGM

ionized IGM



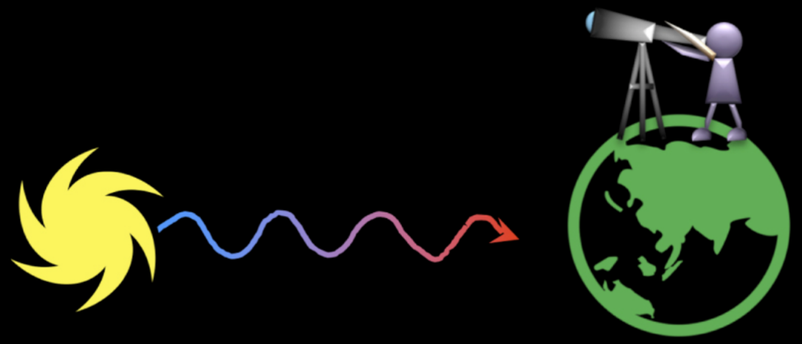
images adapted from
aasnova.org/2022/09/27/bright-galaxies-blowing-bubbles/

Ly-α emitters as a tracer of bubble morphology

Ly-α absorbed by neutral IGM

neutral IGM

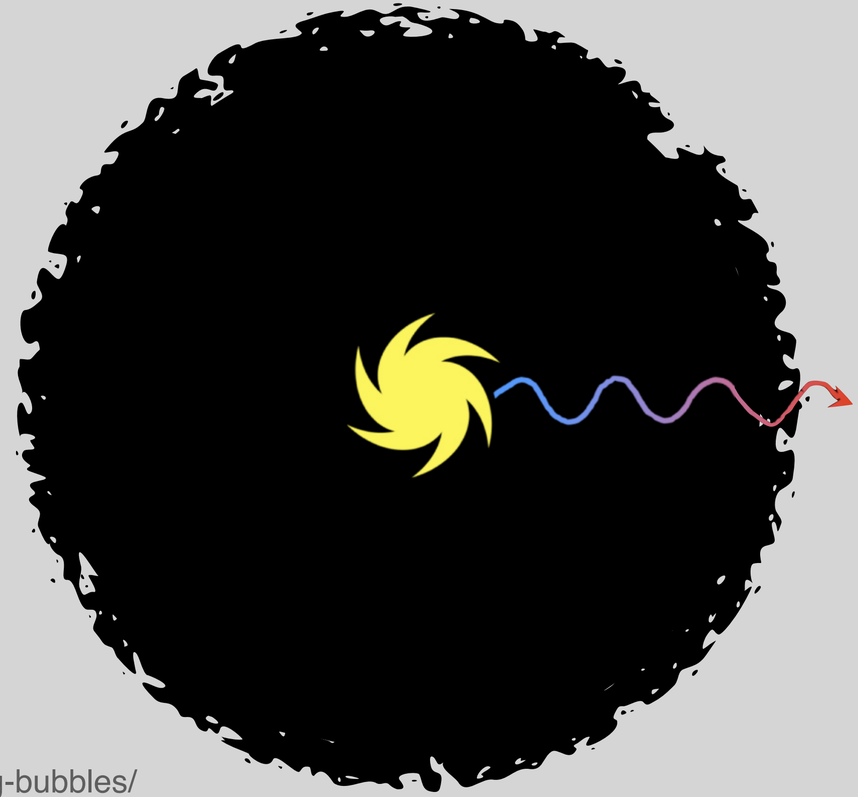
ionized IGM



images adapted from
aasnova.org/2022/09/27/bright-galaxies-blowing-bubbles/

Ly-α emitters as a tracer of bubble morphology

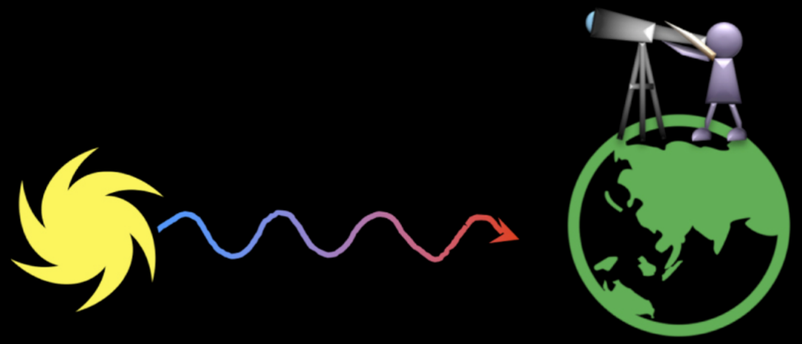
Ly-α absorbed by neutral IGM



Ly-α redshifts away from resonant frequency, passes through IGM

neutral IGM

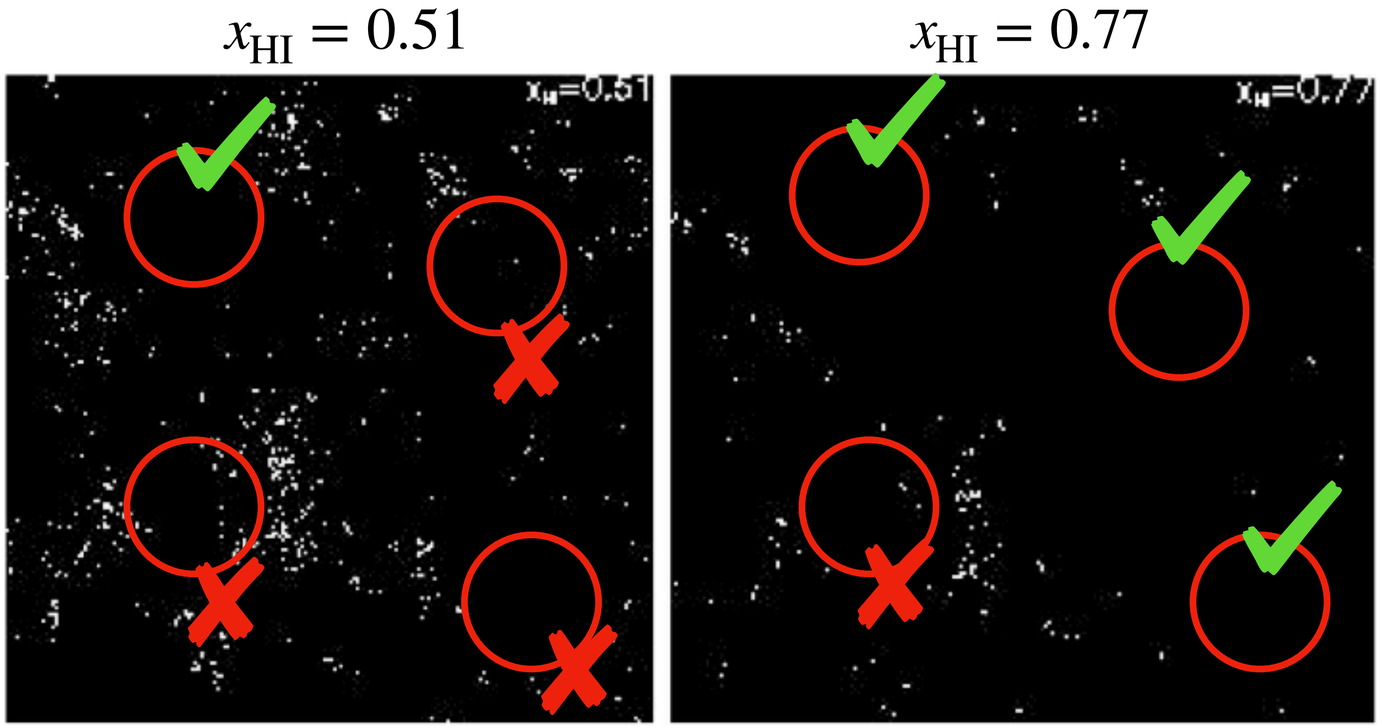
ionized IGM



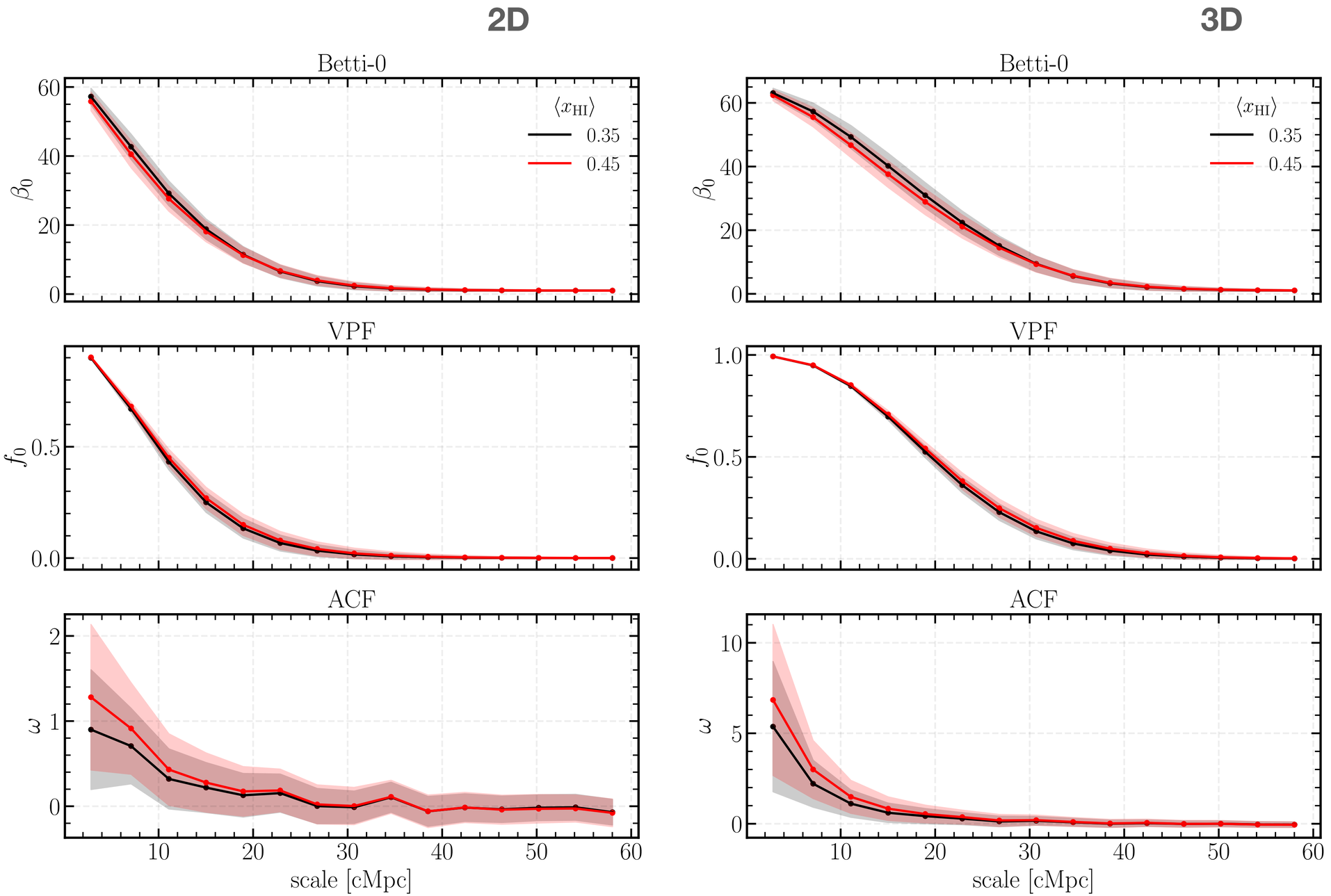
images adapted from
aasnova.org/2022/09/27/bright-galaxies-blowing-bubbles/

Void probability function

- Draw many circles of a given radius
- Count fraction of empty (void) circles
- Can repeat for multiple scales



2D vs 3D

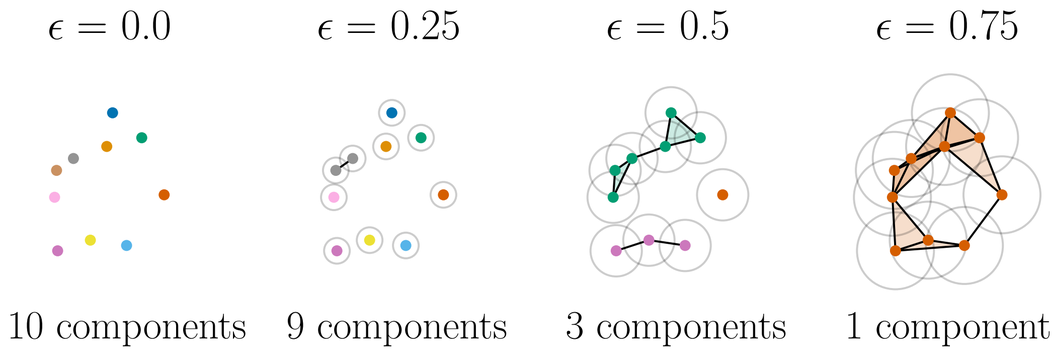

2D
3D
Betti-0
VPF
ACF
$\langle x_{\rm HI} \rangle$
0.35
0.45
β_0
f_0
ω
scale [cMpc]

Uv luminosities

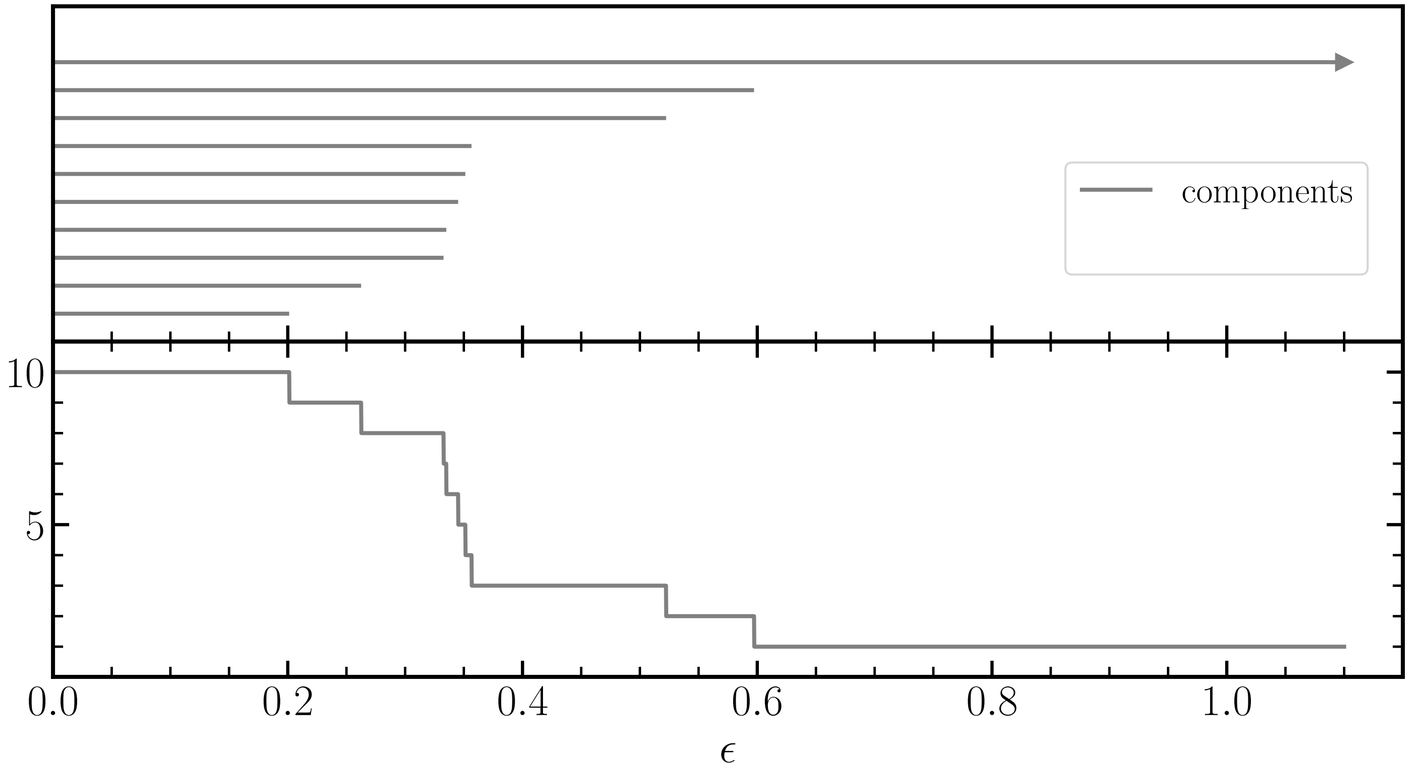
- Relationship from G. Sun and S.R. Furlanetto, *Constraints on the star formation efficiency of galaxies during the epoch of reionization*, MNRAS 460 (2016) 417
- assume that the star formation rate of a galaxy is proportional to its rest-frame UV continuum luminosity density at 1500 Å:
  - $L_{1500,\nu}(\dot{\rho}_\star) = \dot{\rho}_\star / K_{\rm UV}$
  - $K_{\rm UV} = 1.15 \times 10^{-28} M_\odot\,{\rm yr}^{-1}/\,{\rm erg\ s^{-1}\,Hz^{-1}}$
  - assumes a continuous mode star formation with a Salpeter IMF

Persistent Homology

“Vietoris–Rips filtration”

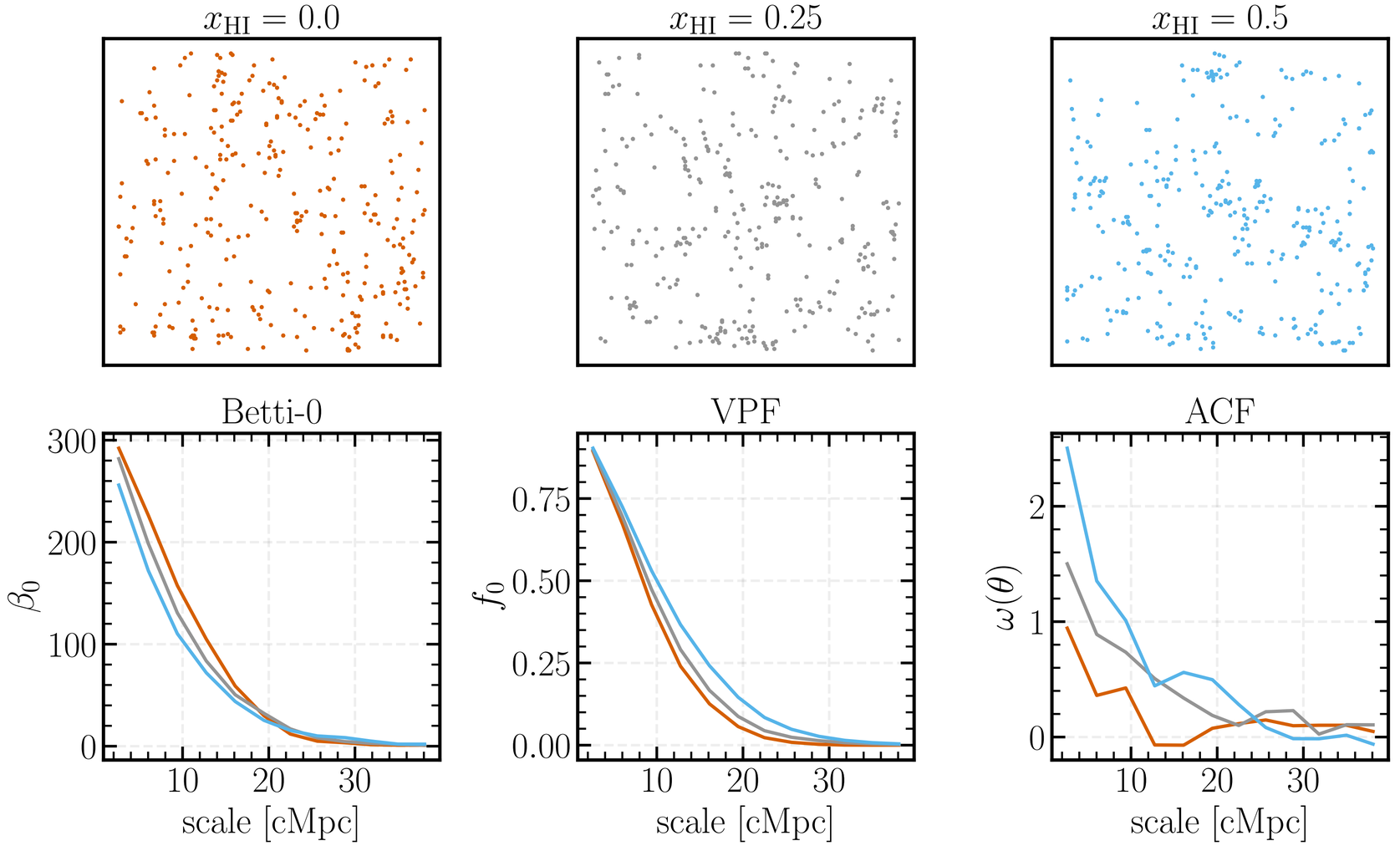


“Betti-0 curve”



“persistence barcode”

What does it look like for LAEs?



Persistent Homology

- From topological data analysis
- Allows us to define **topological features** for a discrete distribution
- Has been used in some cosmological contexts

Why persistent homology?

- established technique in topological data analysis
    - backed by a rich mathematical formalism
    - existing resources, studies
- uniquely defined on a given data set, no free parameters
- computation is straightforward
    - existing packages in Python (`Ripser`)

Simulations of LAEs

- 21cmFAST
- 2,000 lightcones
- ionized bubbles, halos

Simulations of LAEs

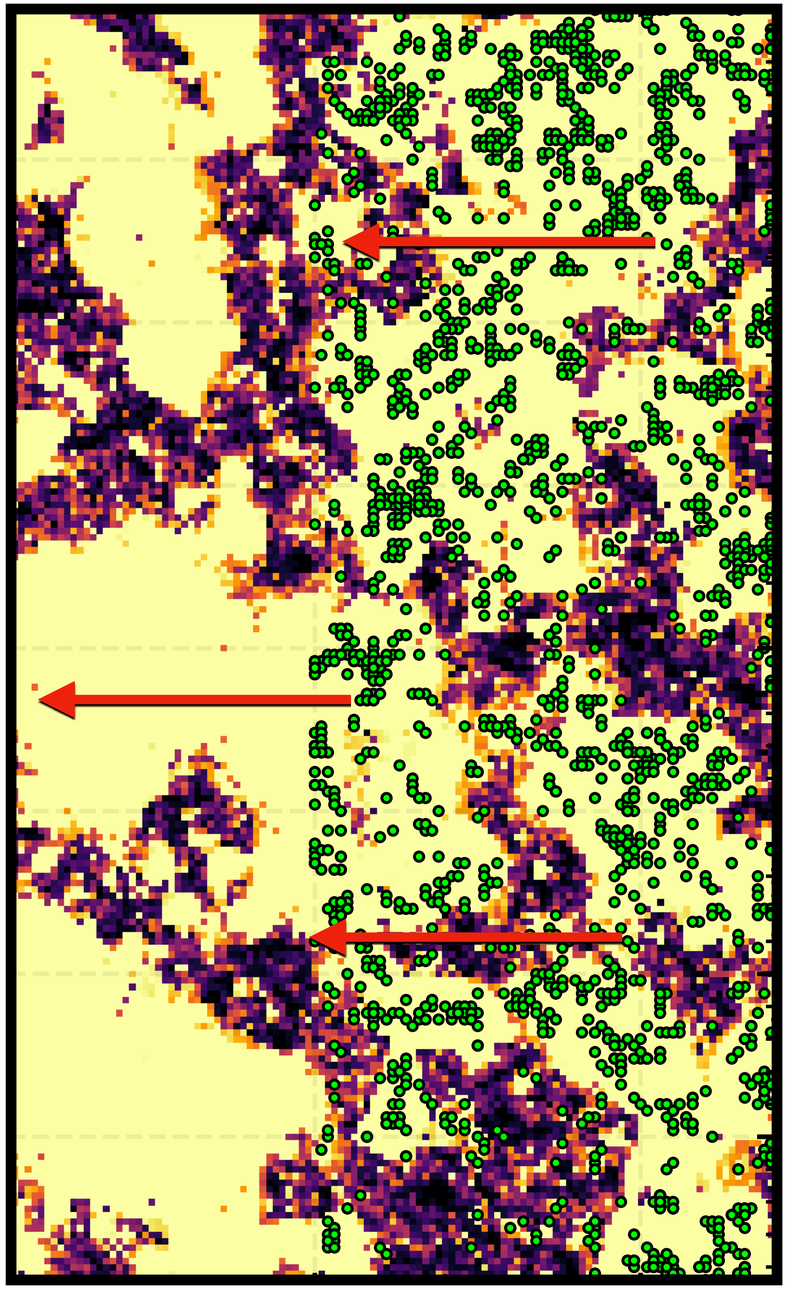
- 21cmFAST
- 2,000 lightcones
- ionized bubbles, halos
- assign halos intrinsic Ly-α luminosities
- account for absorption from IGM

Model of reionization

- Suite of simulations using `21cmFAST`
- 2,000 lightcones created with identical cosmological/astrophysical parameters — different initial conditions
- 300 x 300 cMpc^2 (~ 4 deg^2 at $z = 7.2$); 1.5 cMpc / pixel
- produces ionized bubble morphologies
- discrete halo locations
- star formation rates for each halo → UV luminosities
- Assign Ly-α luminosity and velocity offset probabilistically based on UV magnitude: $P(\{L_\alpha, \Delta v\} \,|\, M_{\rm UV})$; emperical fit to data from Gagnon-Hartman et al., 2026

IGM attenuation

- Draw lines of sight from halos
- In each pixel, sum contribution to optical depth τ based on density, neutral fraction, and redshift
- Attenuated Ly-α luminosity is:
  - $L_{\alpha,\mathrm{att}} = L_\alpha e^{-\tau}$



Roman observations

- using WFI's grism at 1.00-1.93 μm, we can detect LAEs at $z \gtrsim 7.2$ in a FOV of 0.281 deg^2
- obervation strategies from Wold et al., 2024
- Uses multiple exposures with rotated position angles in a single FOV
- Two basic strategies:

number of exposures	total time per FOV	50% completeness limit
25	70 hours	8.2×10^{-18} erg s^{-1} cm^{-2}
10	42 hours	10.2×10^{-18} erg s^{-1} cm^{-2}

LAE selection

- Convert 50% flux limits to luminosity limits $L_{\rm lim}$
- Select all LAEs with $L_{\alpha,\rm att} > L_{\rm lim}$
- Take appropriate survey area (1 or 4 deg²)
- undersample LAEs so that all slices have the same number
    - 3 cases (*1 deg², 12 day survey*):
        - ‘mean:’ 66 / deg²
        - ‘optimistic:’ 96 / deg²
        - ‘pessimistic:’ 36 / deg²