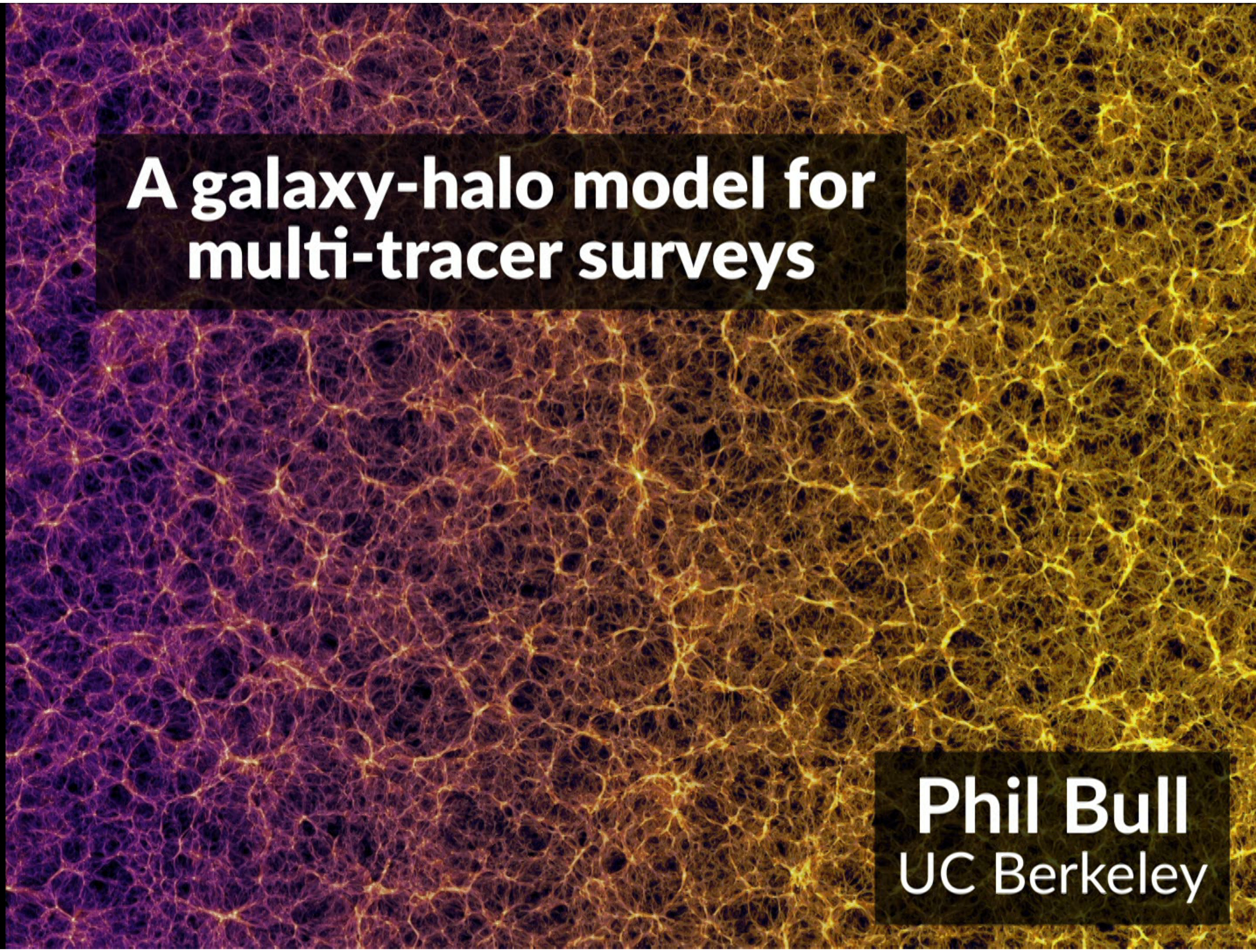
Title: A galaxy-halo model for multi-tracer surveys

Date: Mar 20, 2018 11:00 AM

URL: http://pirsa.org/18030116

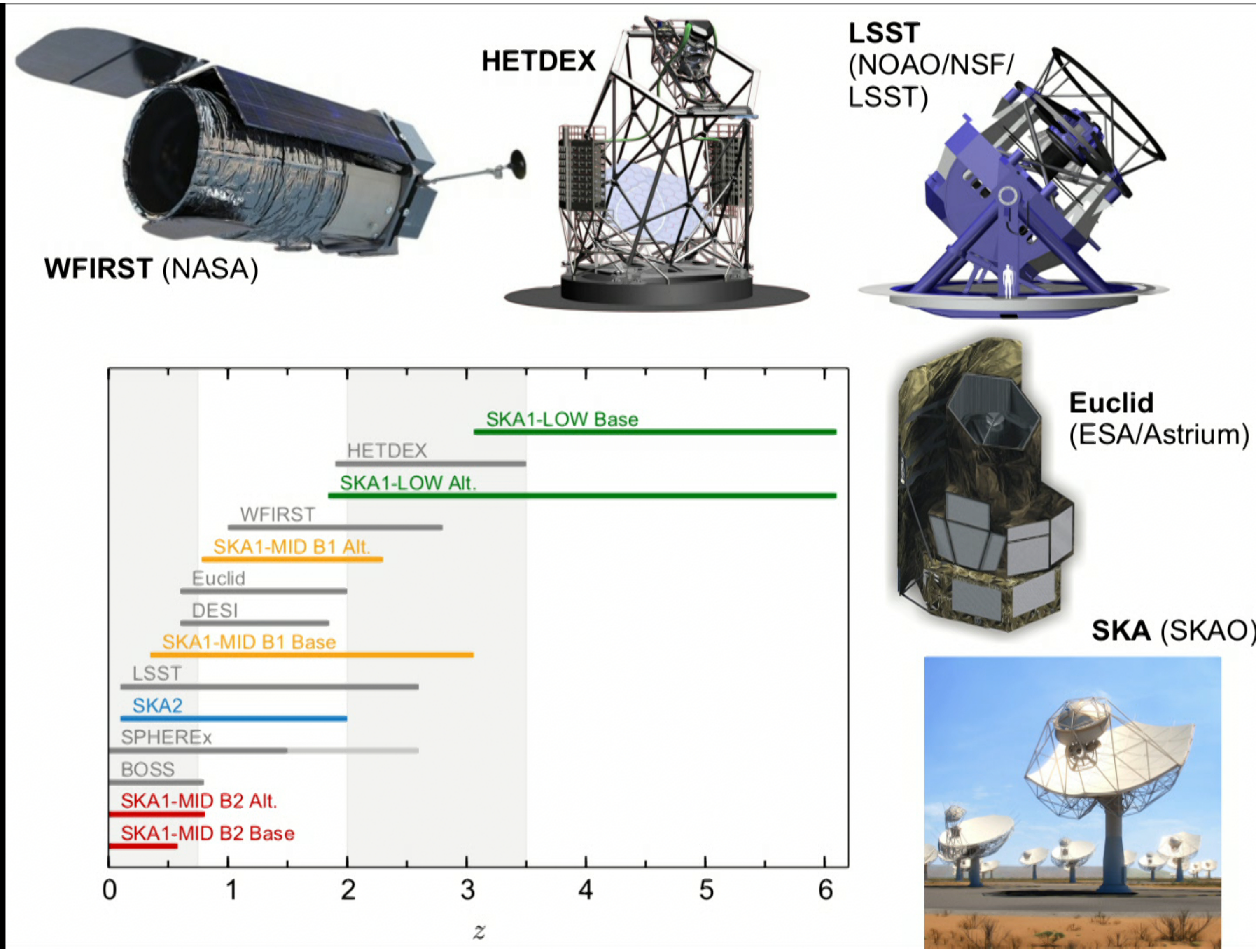
Abstract: <p>Cosmic variance puts hard limits on what we can learn about fundamental physical processes on the largest cosmological scales. A neat trick allows this limit to be circumvented in some cases though, by cross-correlating multiple tracers of the cosmic matter distribution. After giving a few examples of where the multi-tracer technique can be useful, I will outline a new analytic statistical model to describe how the galaxies seen in different surveys (and at different wavelengths) jointly populate the dark matter halo distribution. While the model is necessarily simplified, its flexibility and analytic nature allow rapid exploration of the parameter space relevant to forthcoming surveys in the optical, IR, microwave, and radio. I will show the results of an MCMC fit to low-redshift multi-frequency (radio and optical) luminosity function data, and then discuss several successful (and not-so-successful) consistency tests of the model.</p>


A galaxy-halo model for multi-tracer surveys
Phil Bull
UC Berkeley

There will be a proliferation of large, overlapping LSS surveys over the next ~10 years

By **combining surveys**, we can tease out information that was hidden in the individual surveys

(It's also a good way of controlling systematics)


WFIRST (NASA)
HETDEX
LSST
(NOAO/NSF/
LSST)
Euclid
(ESA/Astrium)
SKA (SKAO)
SKA1-LOW Base
HETDEX
SKA1-LOW Alt.
WFIRST
SKA1-MID B1 Alt.
Euclid
DESI
SKA1-MID B1 Base
LSST
SKA2
SPHEREx
BOSS
SKA1-MID B2 Alt.
SKA1-MID B2 Base
0
1
2
3
4
5
6
z

Ways of combining surveys

(in order of sophistication)

- Independently (combine parameter constraints only)
- Template subtraction (using data from A to clean B)
- Cross-correlation (measure cross-spectrum of A x B)
- Joint analysis (include auto *and* cross-spectra)

Ways of combining surveys

(in order of sophistication)

- Independently (combine parameter constraints only)
- Template subtraction (using data from A to clean B)
- Cross-correlation (measure cross-spectrum of A x B)
- Joint analysis (include auto *and* cross-spectra)

An important application of joint analysis is **evading cosmic variance**

Multi-tracer cosmology

Cosmic variance: fundamental limit to measurement precision

- Only see a finite number of Fourier modes
- Biased populations probe the **same** DM field → <u>deterministic</u>

Multi-tracer cosmology

Cosmic variance: fundamental limit to measurement precision

- Only see a finite number of Fourier modes
- Biased populations probe the **same** DM field → deterministic
- Tracer-dependent quantities not CV-limited (Seljak 2008)

Tracer 1 $\delta_1 = (b_1 + f\mu^2)\boxed{\delta_M}$

Tracer 2 $\delta_2 = (b_2 + f\mu^2)\boxed{\delta_M}$

Tracers are stochastic

$$\implies \delta_1/\delta_2 = \boxed{\frac{b_1 + f\mu^2}{b_2 + f\mu^2}}$$

Ratio is deterministic

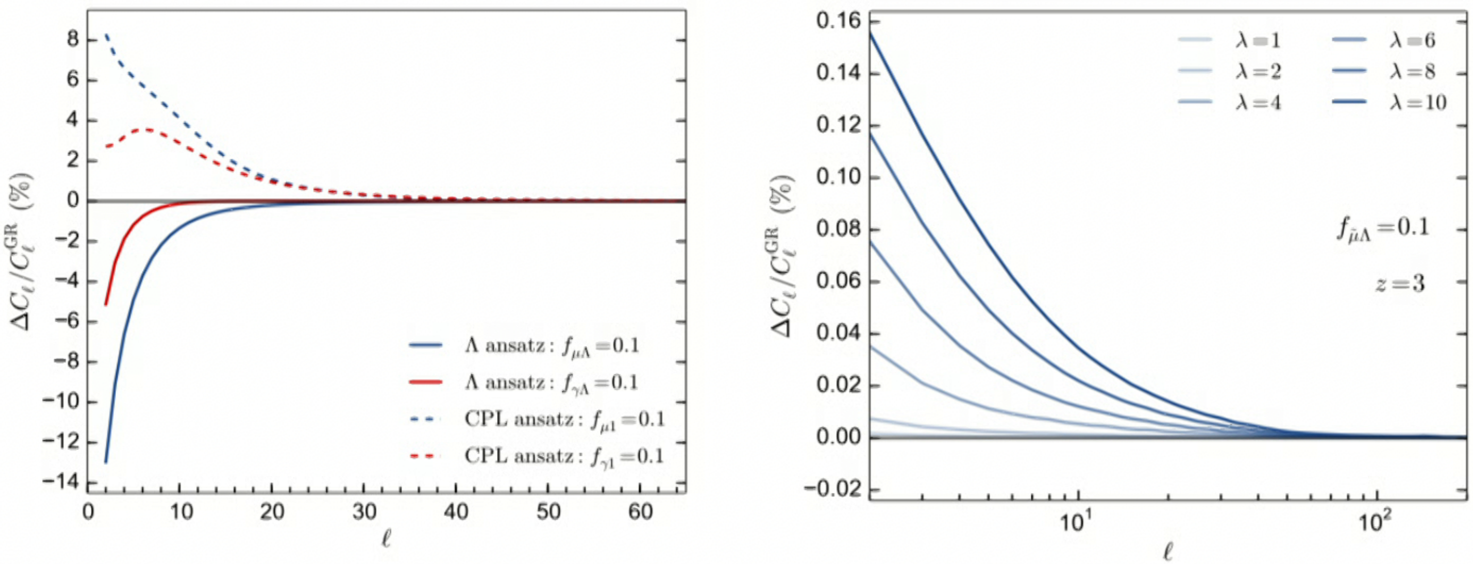
→ Only shot noise limits measurement precision on this quantity

Applications of joint analysis

Physics on the Hubble scale

Cosmic acceleration ~ Hubble scale

Natural place to look for new physics due to DE/MG



Baker & **PB** (2015)

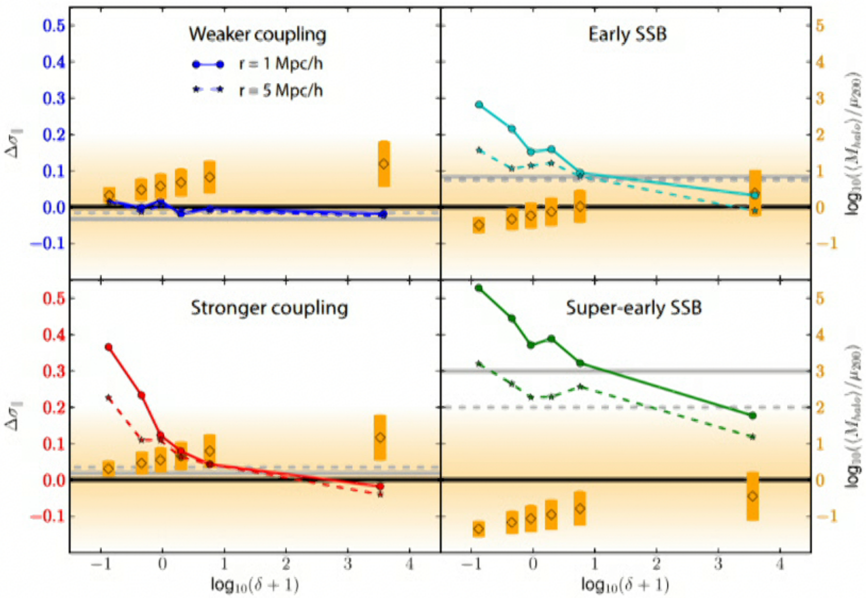
Applications of joint analysis

Marked correlation functions

Modifications to GR are often screened

Screening causes environment-dependent clustering

Need clustering + “mark” (tag) for environment



Velocity dispersion correlation function

vs

Density of local environment

Ivarsen, **PB**, Llinares, Mota (2016)

Applications of joint analysis

Measuring primordial non-Gaussianity

Scale-dependent bias on extremely large scales

Can only reach sigma(fNL) < 1 with multi-tracer

Alonso, PB et al. [1505.07596]; Alonso & Ferreira (2015)

In all cases, we are making use of differences in the clustering properties of multiple tracers

Crucial to understand the connection between tracers (e.g. galaxies) and the underlying dark matter

This requires a galaxy model (but galaxies are complicated...)

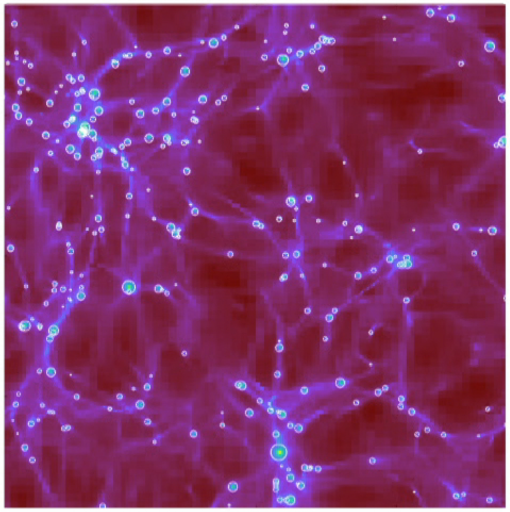
Galaxy models

Halo models

- Fast, simple, analytic. Can marginalise over model uncertainty.
- Predict only a few properties. Tied to empirically-defined "populations".

Empirical models

- Reproduce observables by construction. Can be fast and detailed.
- Strongly reliant on the quality of training data. Not really predictive.



yt project

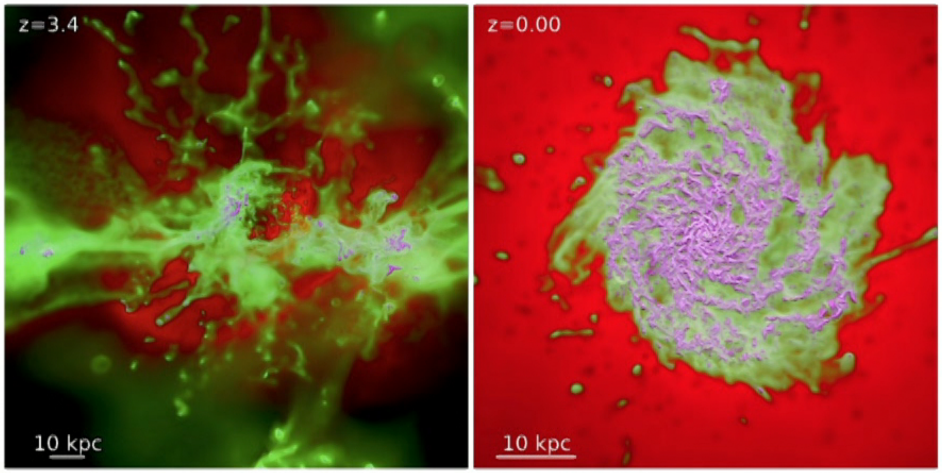
Galaxy models

Semi-analytic models

- Take into account lots of physics. Detailed, multi-observable output.
- Expensive. Many input parameters, which are expensive to change.

Detailed hydro sims

- “Full physics” approach, provides very detailed picture.
- Very expensive (small volumes only); need to choose sub-grid params.



FIRE project

Extension to multiple populations

How are the galaxies detected by Survey A related to the ones from Survey B?

→ e.g. Want to predict LSST x SKA clustering signal.

Extension to multiple populations

How are the galaxies detected by Survey A related to the ones from Survey B?

→ e.g. Want to predict LSST x SKA clustering signal.

- **Empirical approaches:** Not (yet) enough multi-wavelength data of sufficient depth to calibrate the models!
- **Halo model:** Difficult to ensure consistency – can't just calibrate two populations separately.
- **SAMs + hydro:** Consistent physics-based predictions of multiple populations, but expensive to change parameters, generate multiple mocks etc.

A statistical galaxy-halo model

Presenting an analytic statistical model that offers a “middle ground” between these approaches:

- Based on modelling the galaxy-halo connection using simple analytic expressions (scaling relations) → **Fast, cheap**
- Includes multiple physical properties of galaxies (e.g. SFR, stellar mass) → **(Somewhat) Detailed**

A statistical galaxy-halo model

Approach:

- Given halo properties, predict galaxy properties (e.g. SFR)
- Given galaxy properties, predict observables (e.g. $L_{h\alpha}$)

→ Need to predict multiple properties *and correlations between them*

The **joint distribution** of properties $p(M_h, M_\star, SFR \ldots)$ is complicated! Not obvious how to model it.

A statistical galaxy-halo model

Approach:

- Given halo properties, predict galaxy properties (e.g. SFR)
- Given galaxy properties, predict observables (e.g. $L_{h\alpha}$)

→ Need to predict multiple properties *and correlations between them*

The **joint distribution** of properties $p(M_h, M_\star, SFR \ldots)$ is complicated! Not obvious how to model it.

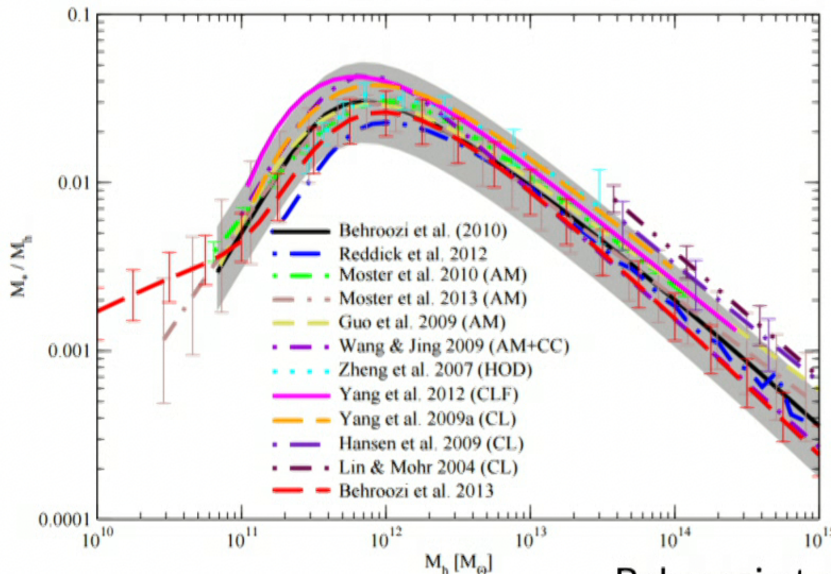
Solution: Decompose into simpler **conditional distributions**:

$$p(a, b, c) = p(a, b|c)\, p(c)$$
$$= p(a|b, c)\, p(b|c)\, p(c)$$

Example: Stellar mass vs. halo mass

Low-scatter relation observed between DM halo mass and stellar mass

$p(M_\star | M_h)$



Behroozi et al. (2013)

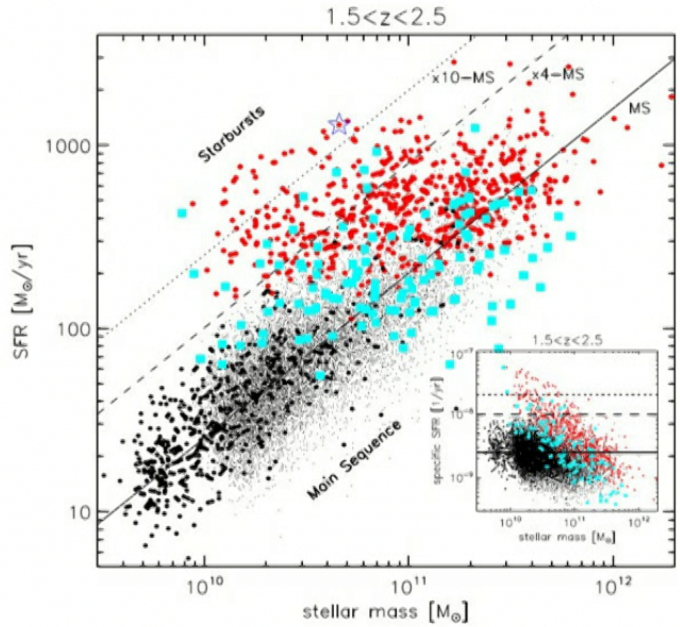
Simple model: double power law with some scatter (Moster+ 2010; Behroozi+ 2010)

Example: Star-forming main sequence

Star-forming galaxies seem to be confined to a narrow "main sequence" in the SFR-$M_\star$ plane

Rodighiero et al. (2011)

$p(SFR|M_\star)$



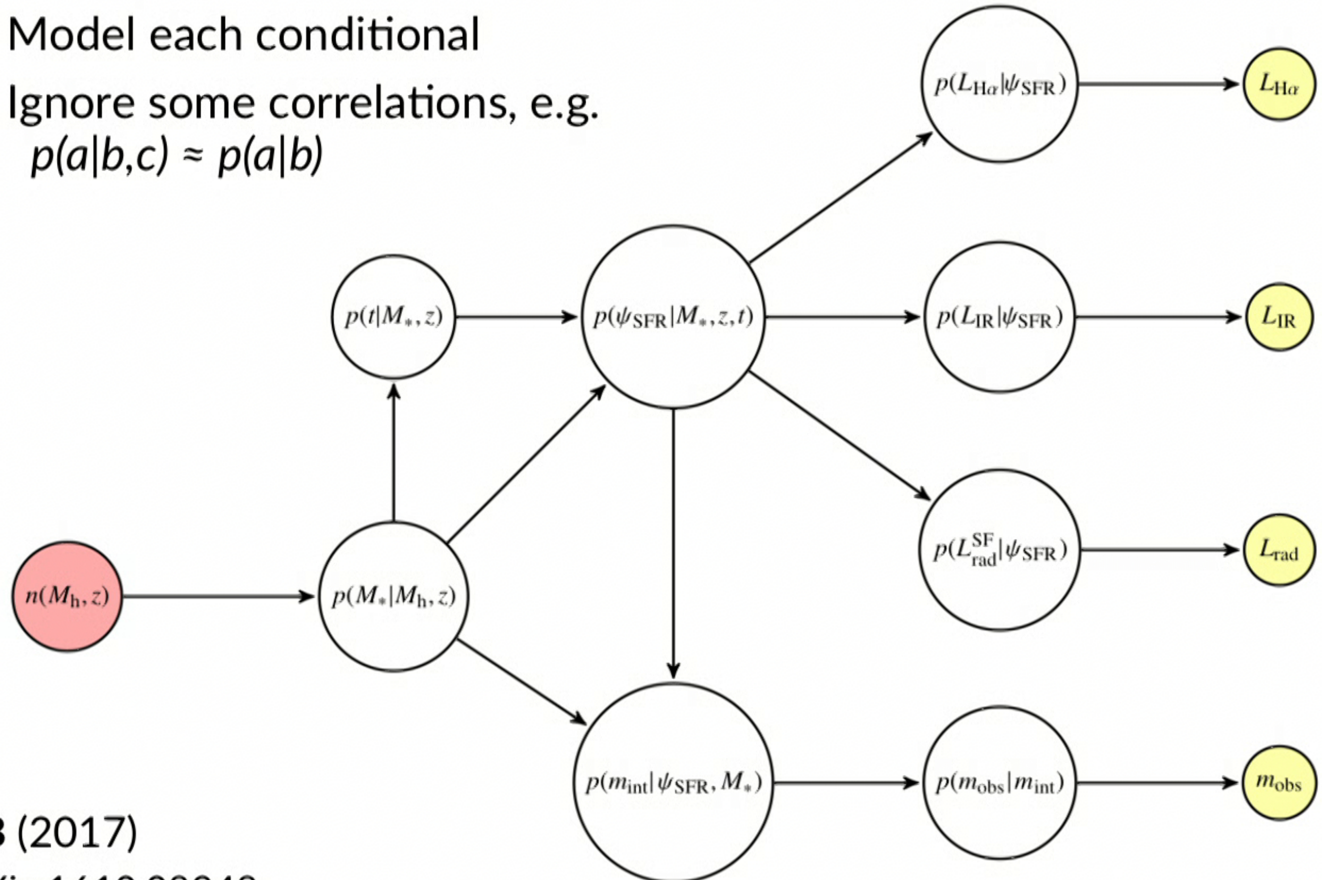
Simple model: power law with (log-normal) scatter (Wang+ '13)

Statistical model

- Break joint dist. into chain of conditionals
- Model each conditional
- Ignore some correlations, e.g.
  $p(a|b,c) \approx p(a|b)$

Statistical model

- Break joint dist. into chain of conditionals
- Model each conditional
- Ignore some correlations, e.g.
  p(a|b,c) ≈ p(a|b)



PB (2017)
arXiv:1610.08948

Optical magnitudes

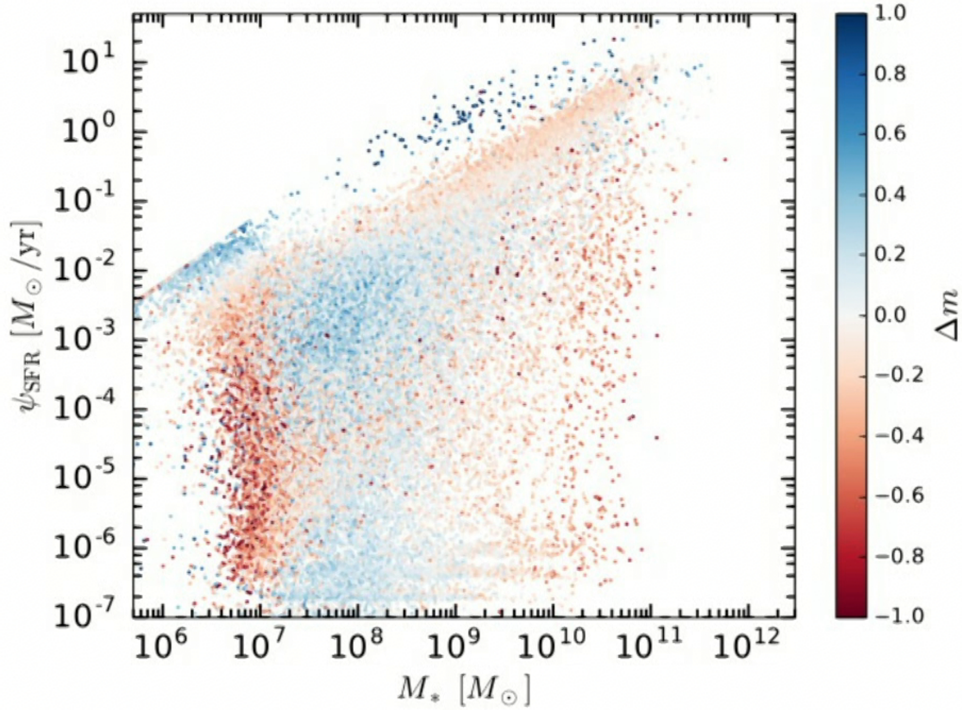
Ansatz:

1) Stellar mass is a proxy for the number of long-lived main sequence stars
2) SFR is a proxy for the number of young O/B stars (but SFR already has a dependence on $M_\star$)
3) Characteristic galaxy magnitude in each band

$$\overline{m}_\nu(\psi_{\rm SFR}, M_*) = \boxed{-c_0^{(\nu)}} + \boxed{A_*\left(c_* + \left(\frac{M_*}{10^9 M_\odot}\right)^{\beta_*}\right)} \boxed{+A_\times^{(\nu)}\left(\frac{M_*}{10^9 M_\odot}\right)^{\beta_\times}\left(\psi_{\rm SFR}\right)^{\gamma_\times}}.$$

Optical magnitudes

Difference between Guo et al. (2011) SAM *u*-band magnitude and best-fit analytic model



Model the remaining scatter to get: *p(mag|M★, SFR)*

Constraints with multi-wavelength data

- Resulting model has 17 free parameters. (Some other parameters fixed by fitting to SAM output.)

- Manually constructed likelihoods for:
  - GAMA g-band **optical** luminosity function
  - GAMA z-band **optical** luminosity function
  - NVSS/6dFGS 1.4 GHz **radio** luminosity function

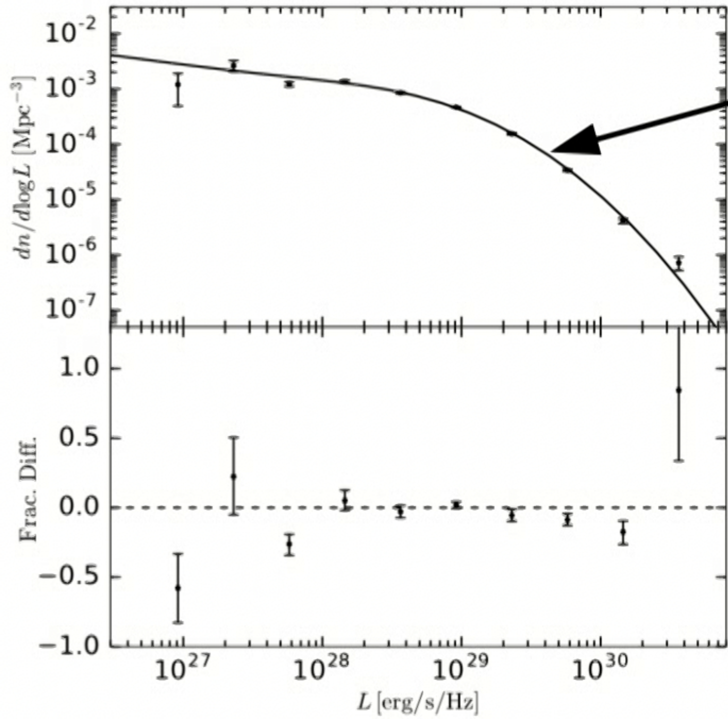
$$
\begin{aligned}
\frac{dn}{d\log L} = & \int d\psi_{\mathrm{SFR}} \int dM_\star \int d\log M_{\mathrm{h}} \\
& \times n(M_{\mathrm{h}})\, p(M_\star|M_{\mathrm{h}})\, p_{\mathrm{SFMS}}(\psi_{\mathrm{SFR}}|M_\star) \\
& \times L_{\mathrm{rad}}\, p(L_{\mathrm{rad}}|\psi_{\mathrm{SFR}})
\end{aligned}
$$

- Used *emcee* (128 walkers, 2000 samples each) to derive model constraints

Radio luminosity function

Radio (1.4 GHz) luminosity fn. at z~0, from NVSS+6dFGS
(Mauch + Sadler 2007)

MCMC fits used only radio data + g,z bands



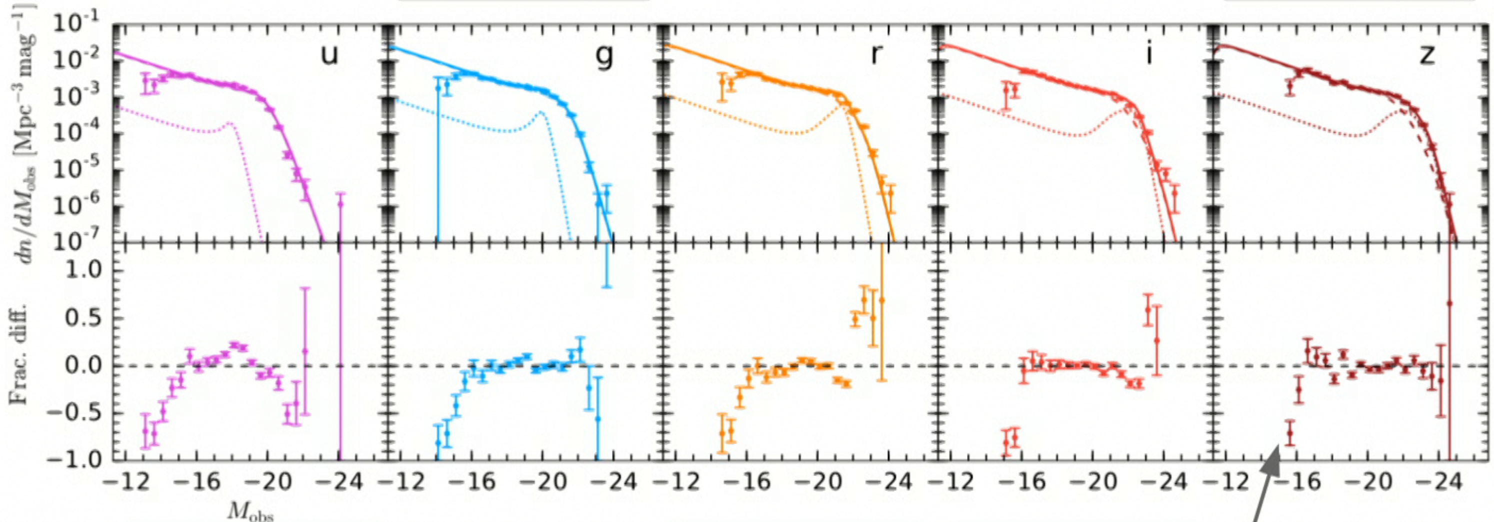
Best-fit model from MCMC

Residuals (fractional)

Optical luminosity functions

Optical (*ugriz*) luminosity functions at z~0, from GAMA
(Loveday et al. 2011)

MCMC fits used only radio data + **g,z bands**

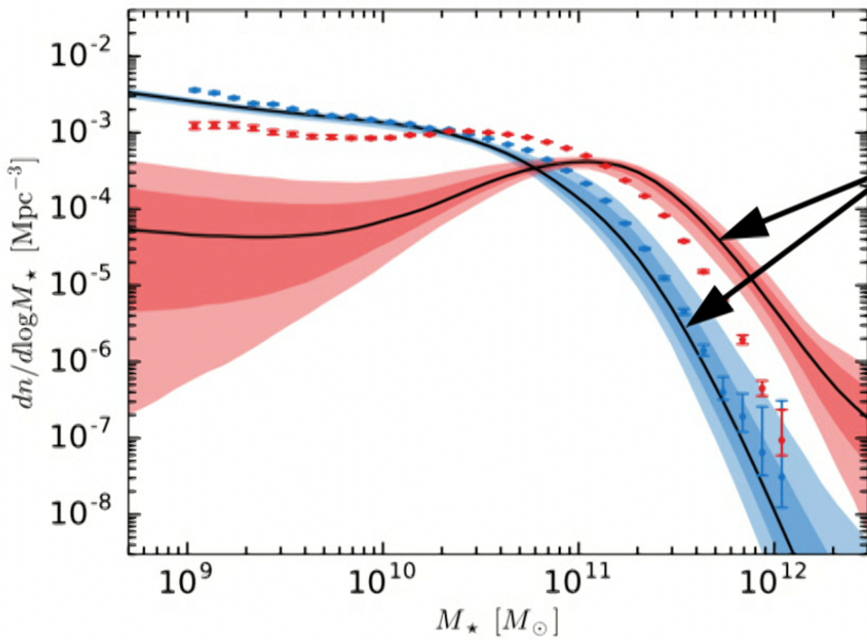


Predicted LFs (not used in fits)

Completeness
effects (not used)

Stellar mass function

Model-predicted SMF for star-forming and passive galaxies
(data points from SDSS/GALEX; Moustakas+ 2013)



Predicted by model fit to GAMA+NVSS LFs

Applications

- Model is **tractable for MCMC** → marginalise over galaxy-halo nuisance parameters in LSS studies
  - Modular – easy to try alternative models / physics
  - Analytic – useful for theoretical studies; analytic marginalisation

Conclusions

Subtle signatures of fundamental physics might only be detectable with a multi-tracer approach

→ need a consistent multi-tracer galaxy-halo model

Conclusions

Subtle signatures of fundamental physics might only be detectable with a multi-tracer approach

→ need a consistent multi-tracer galaxy-halo model

ghost is a statistical model of galaxy properties at multiple wavelengths, using empirical/theoretical scaling relations

- It's fast and analytic; can do MCMC etc.
- Approximate, but modular → can be extended/improved
- (Provides a consistent framework for dealing with scaling relations that everyone is using anyway...)

See **https://github.com/philbull/ghost**