Combining Probes of Large-Scale Structure

Analysis Strategies for the Precision Cosmology Era

Elisabeth Krause
University of Pennsylvania

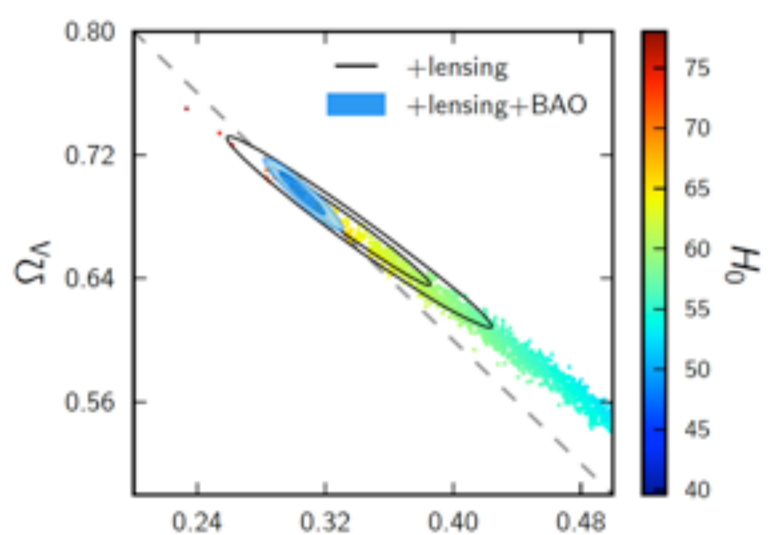
Overview

- **Cosmological constraints from combining probes of large-scale structure (LSS)**
  - focus on challenges of joint analysis
    (modeling of individual probes material for many individual talks)
  - CosmoLike software package
  - systematics limited analysis
  - blind analysis strategies

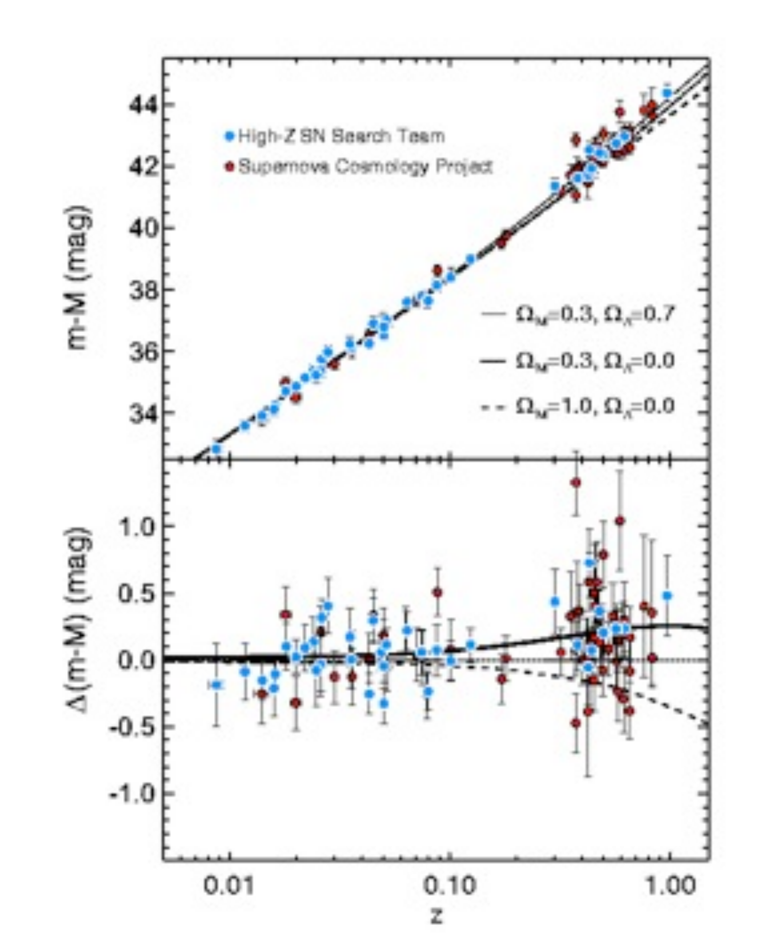
with Tim Eifler, Carlos Cunha, Scott Dodelson, DES collaboration

Dark Energy

- first detected with SN surveys
- current contraints (Planck ++): $\Omega_\Lambda \sim 0.7$, equation of state poorly constrained



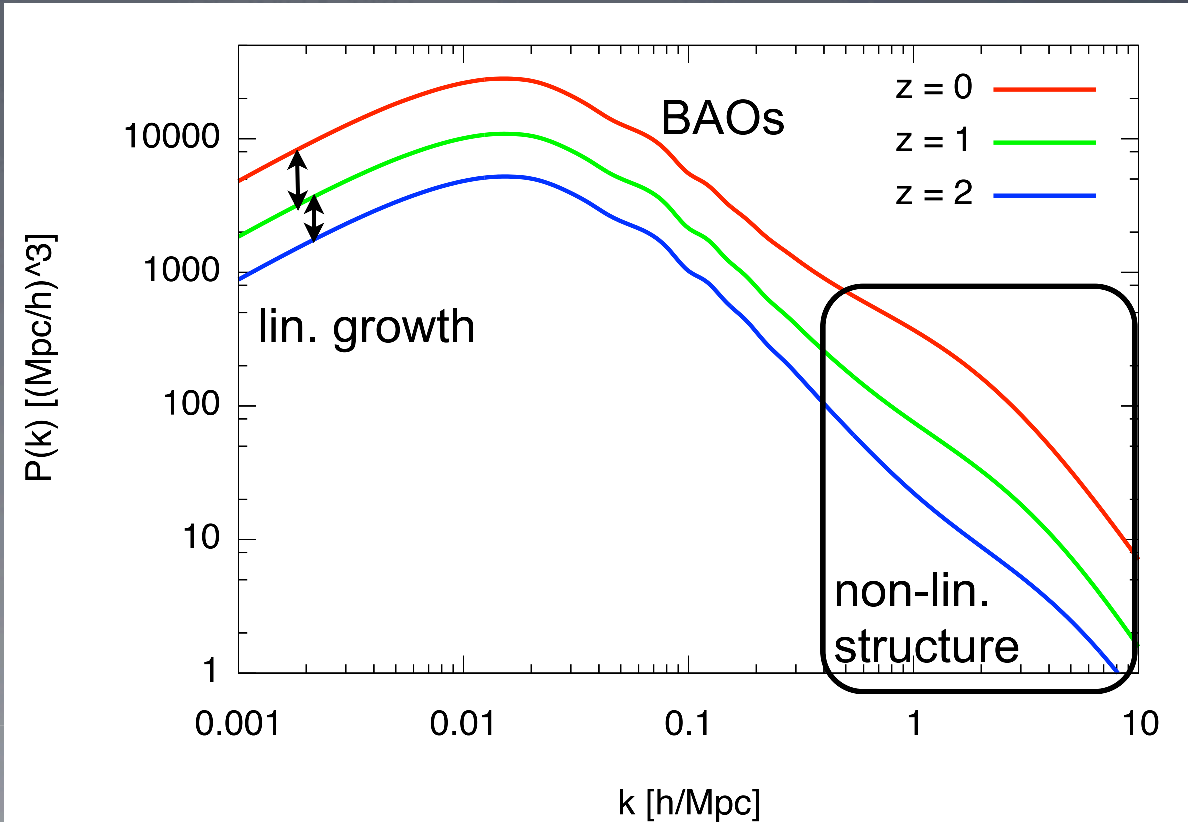
Planck XVI ’13



Riess et al. ’99 ff.

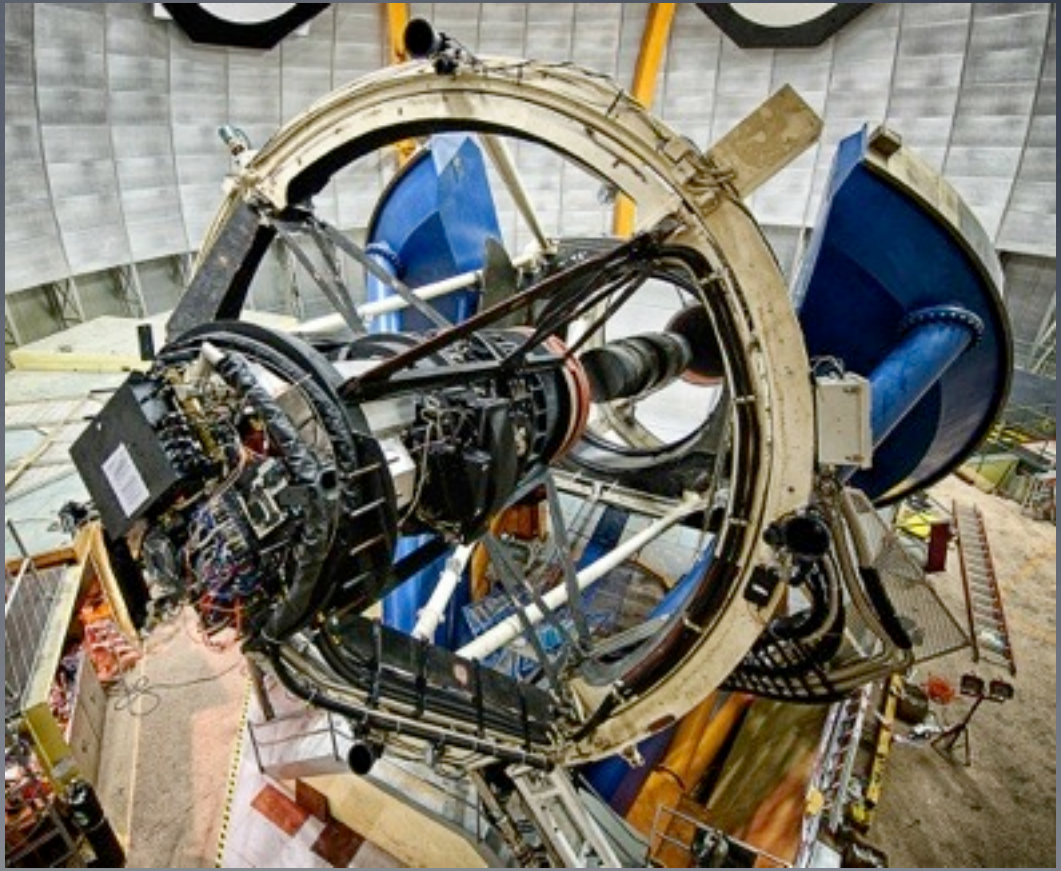
Dark Energy with LSS

- Dark Energy (modified gravity?) changes expansion rate
- dominates energy density at low redshift, affects
    - distances (standard rulers, - candles),
    - volume elements
    - growth of structure
- LSS ideal probe for late-time evolution



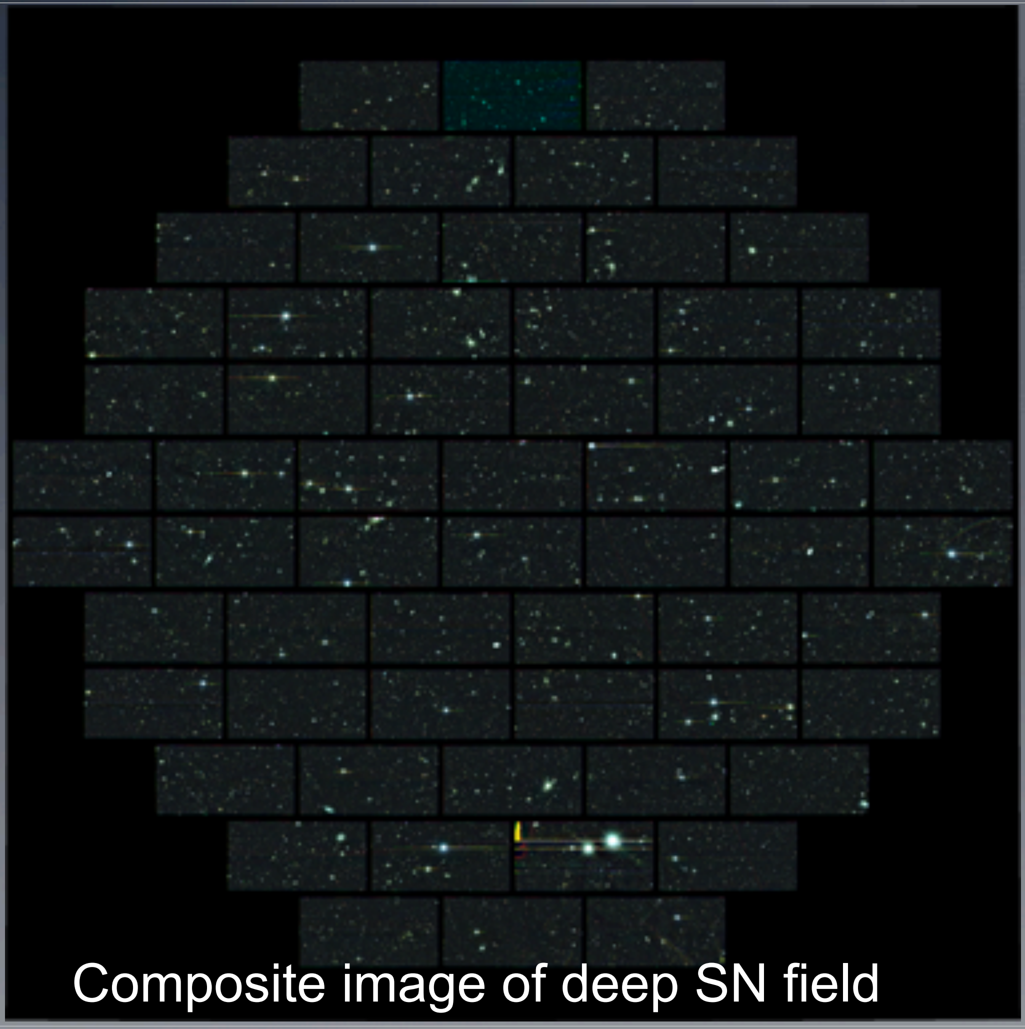
Dark Energy Survey

- Two multiband imaging surveys:
  - 300 million galaxies over 1/8 sky
  - 4000 supernovae (time-domain)
- New 570 Megapixel Dark Energy Camera on the Blanco 4-meter
  - Science Verification complete
  - Survey 2013-2018 (525 nights)
  - **Survey mode started Sept 1!**



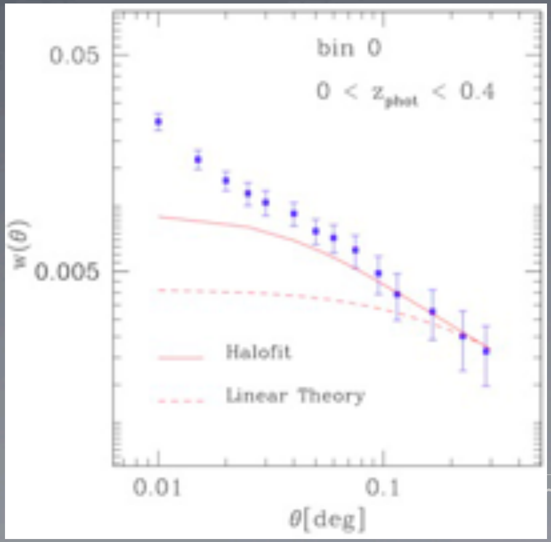
DECam on the Blanco 4m at NOAO Cerro Tololo InterAmerican Observatory

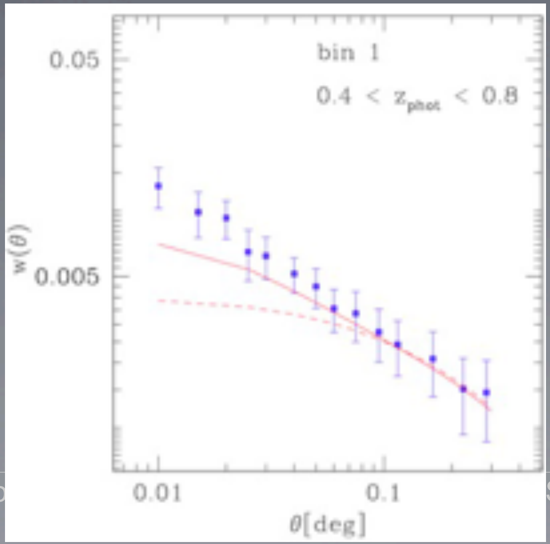
DECam

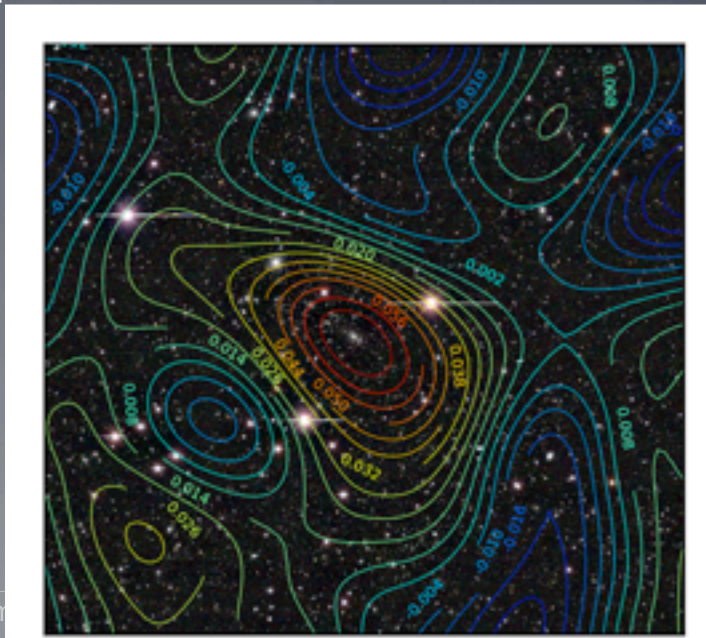


Science Verification Data

- 100 sq deg to full depth from science verification
  - detected and confirmed super novae (SN working group)
  - discovered many high-z clusters already (Rozo, Rykoff, et al.)
  - detected galaxy-galaxy lensing (Clampitt, EK + WL group)
  - early clustering measurements (Bauer + LSS group)
- working hard to understand instrument
  - cosmology can wait to year 1 data...



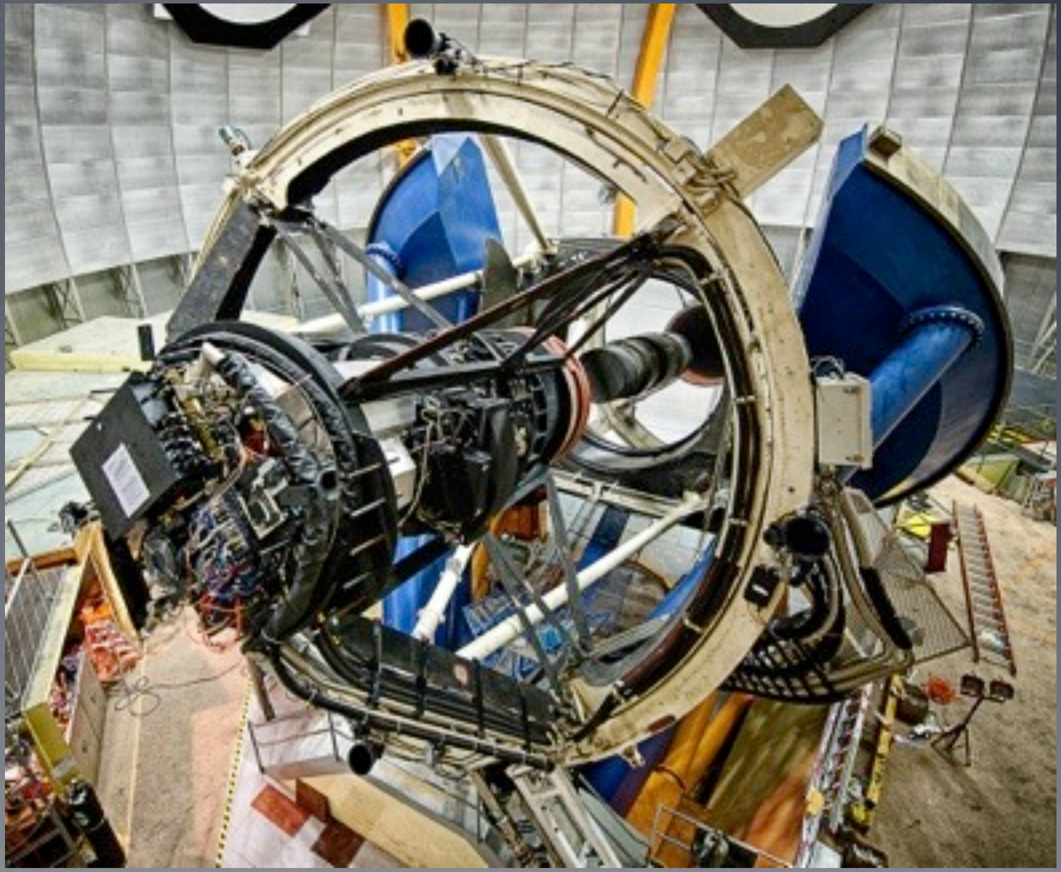




RXJ2238, Melchior et al.

Dark Energy Survey

- Two multiband imaging surveys:
  - 300 million galaxies over 1/8 sky
  - 4000 supernovae (time-domain)
- New 570 Megapixel Dark Energy Camera on the Blanco 4-meter
  - Science Verification complete
  - Survey 2013-2018 (525 nights)
- Stage III Survey using 4 complementary techniques:
  - I. Galaxy Clusters
  - II. Weak Gravitational Lensing
  - III. Galaxy Clustering
  - IV. Supernovae

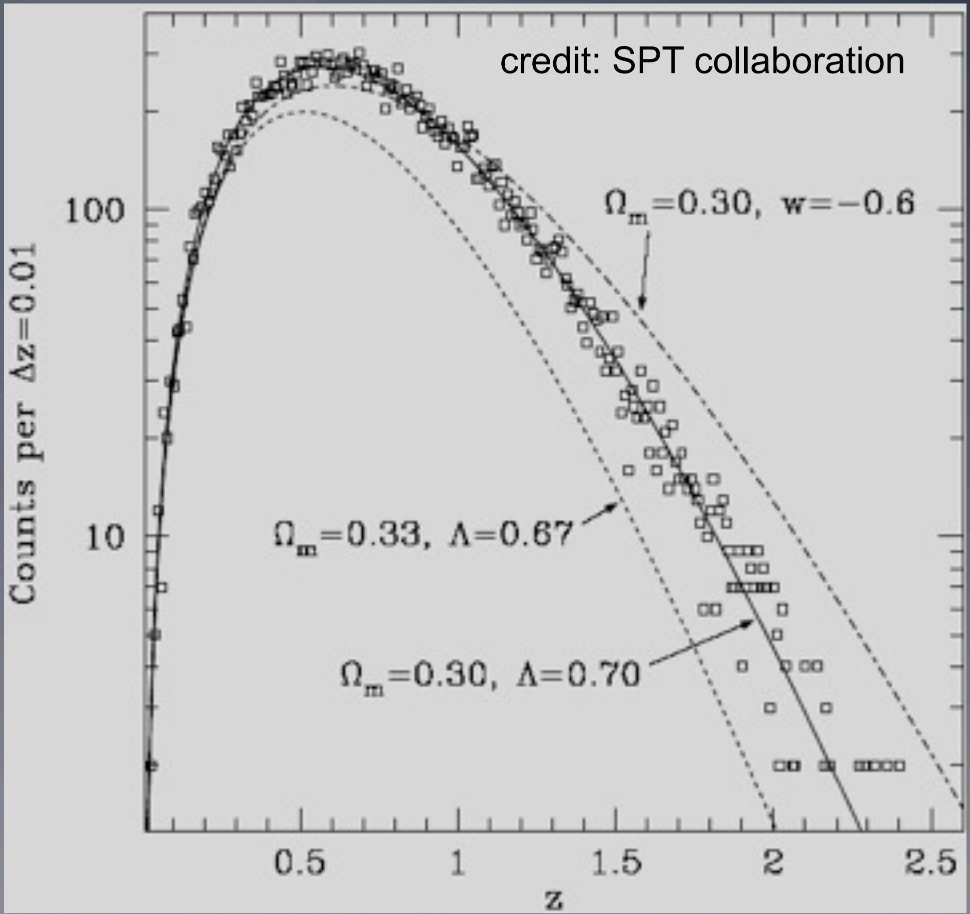


DECam on the Blanco 4m at NOAO Cerro Tololo InterAmerican Observatory

LSS Probes of Dark Energy I

Galaxy Clusters

- measure number counts (+ clustering)
  → growth of structure, expansion history
  - DES: 100,000 clusters to z ~ 1
  - overlap with SPT



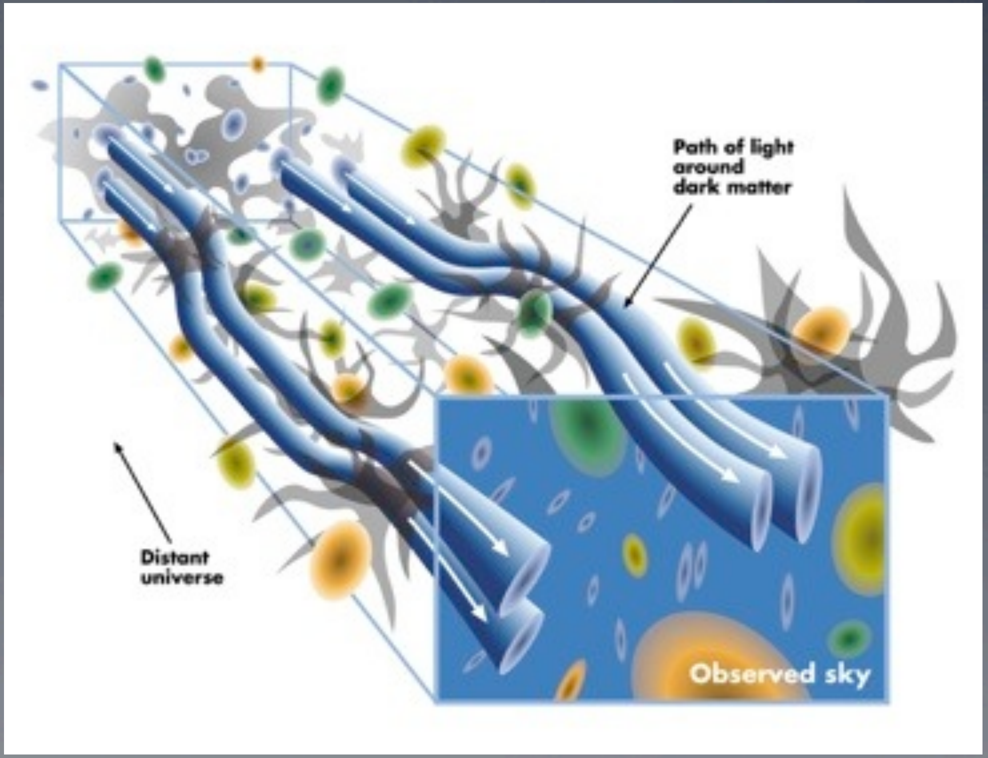
LSS Probes of Dark Energy II

Weak Gravitational Lensing

- Light deflected by tidal field of LSS
- Coherent distortion of galaxy shapes (“shear”)
- Shear statistically related to (projected) matter distribution

→ growth of structure, expansion history

- DES: shapes of 200 million galaxies



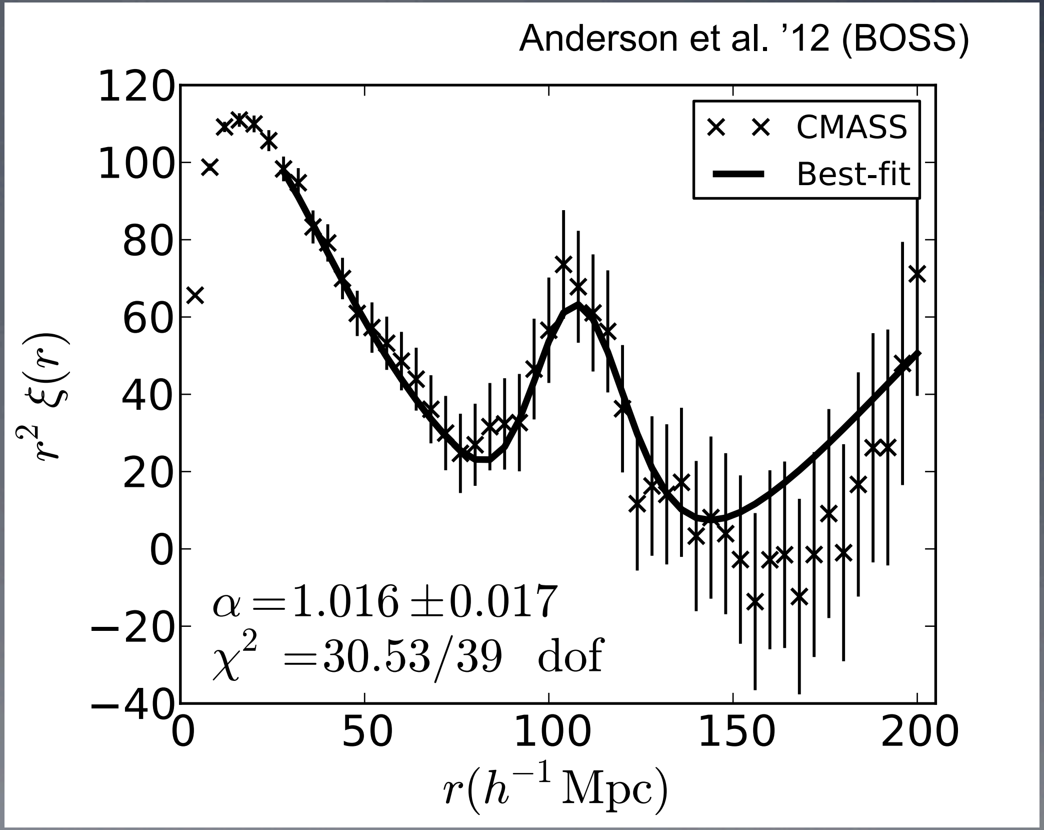
LSS Probes of Dark Energy III

Galaxy Clustering

- measure BAOs + shape of correlation function

  → growth of structure, expansion history

  - DES: 300 million galaxies to z ~ 1.5

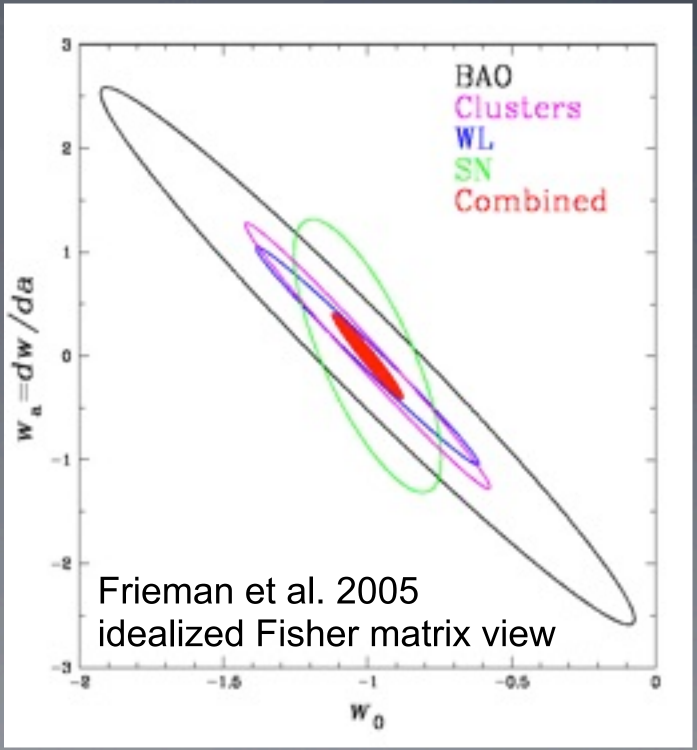


Combining Probes

- Clusters, WL, BAO probe different aspects of structure formation, subject to different systematics
- Combining them improves constraints

But...

- Probe same underlying density field, correlated information
- Shared systematics
- Combining posterior distributions not sufficient, **joint analysis** required



Joint Likelihood Analysis

model data vector

self-consistent modeling of all observables

matter power spectrum, halo mass + bias function
+ *selection function, systematics*

likelihood function

number counts: Poisson

2PCF: ~ Gaussian (?)

improvements needed for stage IV

joint covariance

large and complicated,
non-(block) diagonal matrix

$$p\left(\boldsymbol{\pi}|\hat{\mathbf{d}}\right) = p\left(\boldsymbol{\pi}\right)\mathcal{L}\left(\hat{\mathbf{d}}|\mathbf{d}\left(\boldsymbol{\pi}\right), C\right)$$

combined data vector

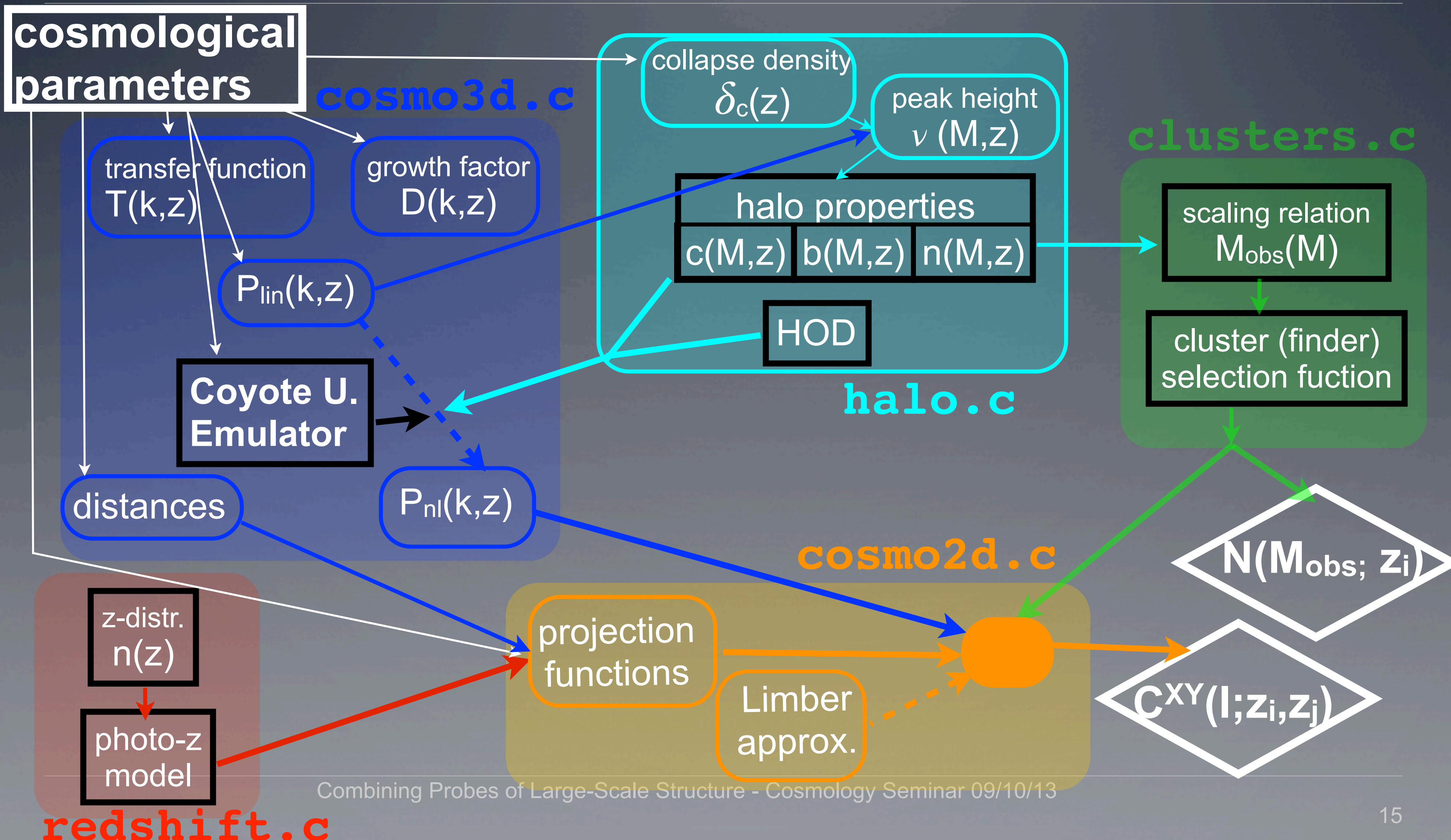
priors

parameters

Introducing CosmoLike

- Likelihood library for combined probes analyses
  - cosmic shear, galaxy-galaxy lensing, galaxy clustering, cluster abundance + mass self-calibration, CMB cross-correlations
  - covariances, model data vectors including systematic effects
  - optimized for high-dimensional likelihood analyses
- Integrated in DES pipeline, public release in late 2013

CosmoLike: Model Data Vector



Combined Data Vector

- Consider number counts (*N*), two-point statistics (cosmic shear <κκ>, g-g lensing <δκ>,clustering <δδ>, etc.)
- Number of observables:

Cluster counts	5 richness bins x 5 z-bins	= 25
Cluster lensing	5 x 5 x(source-z bins)x 20 radial bins	> 1000
Cluster clustering	5 x 5 x 20 radial bins	= 250
Galaxy clustering	5 z-bins x 20 radial bins x type bins	> 200
Shear/g-g lensing	2 x 20 radial bins x tomography	~ 1000

→ n > 2000!

Joint *Inverse* Covariance

- correlated noise makes inversion difficult
- covariance estimate from data/simulations requires thousands of representative patches N_p

$\widehat{C^{-1}} \approx \left[1 - \frac{n}{N_p}\right] \left(\hat{C}\right)^{-1}$, inflates contours

  - at least n + 2 *independent* realizations

General Covariance Structure

- Consider number counts (*N*), two-point statistics (e.g.,shear <κκ>, g-g lensing <δκ>,clustering <δδ>)

	N	<δδ>	<δκ>	<κκ>
N	Cov (*N*, *N*)	Cov (<δδ>, *N*)	Cov (<δκ>, *N*)	Cov (<κκ>, *N*)
<δδ>	Cov (<δδ>, *N*)	Cov (<δδ>, <δδ>)	Cov (<δδ>, <δκ>)	Cov (<δδ>, <κκ>)
<δκ>	Cov (<δκ>, *N*)	Cov (<δκ>, <δδ>)	Cov (<δκ>, <δκ>)	Cov (<δκ>, <κκ>)
<κκ>	Cov (<κκ>, *N*)	Cov (<κκ>, <δδ>)	Cov (<κκ>, <δκ>)	Cov (<κκ>, <κκ>)

General Covariance Structure

- Cov (N,N): Cluster number counts covariance

  Poisson error + sample variance → power spectrum

  (e.g., Lima & Hu '04, Takada & Bridle '07)

	N	<δδ>	<δκ>	<κκ>
N	Cov (N, N)	Cov (<δδ>, N)	Cov (<δκ>, N)	Cov (<κκ>, N)
<δδ>	Cov (<δδ>, N)	Cov (<δδ>, <δδ>)	Cov (<δδ>, <δκ>)	Cov (<δδ>, <κκ>)
<δκ>	Cov (<δκ>, N)	Cov (<δκ>, <δδ>)	Cov (<δκ>, <δκ>)	Cov (<δκ>, <κκ>)
<κκ>	Cov (<κκ>, N)	Cov (<κκ>, <δδ>)	Cov (<κκ>, <δκ>)	Cov (<κκ>, <κκ>)

General Covariance Structure

- Cov (<δδ>, N): Covariance of cluster abundance + 2pt statistic of (projected) density field → power spectrum, bispectrum (Takada & Bridle ‘07)
- Cov (N,N): power spectrum

	N	<δδ>	<δκ>	<κκ>
N	Cov (N, N)	Cov (<δδ>, N)	Cov (<δκ>, N)	Cov (<κκ>, N)
<δδ>	Cov (<δδ>, N)	Cov (<δδ>, <δδ>)	Cov (<δδ>, <δκ>)	Cov (<δδ>, <κκ>)
<δκ>	Cov (<δκ>, N)	Cov (<δκ>, <δδ>)	Cov (<δκ>, <δκ>)	Cov (<δκ>, <κκ>)
<κκ>	Cov (<κκ>, N)	Cov (<κκ>, <δδ>)	Cov (<κκ>, <δκ>)	Cov (<κκ>, <κκ>)

General Covariance Structure

- Cov (<δδ>, <δδ>), etc.: Covariance of (different) 2pt statistics of (projected) density field

$$\mathrm{Cov}(P(\mathbf{k}_1), P(\mathbf{k}_2)) \approx \underbrace{\frac{2\delta_D(\mathbf{k}_1+\mathbf{k}_2)}{N_{k_1}} P^2(k_1)}_{\text{Gaussian cosmic variance}} + \underbrace{\frac{\bar{T}(k_1,k_2)}{V_s}}_{\text{non}-\text{Gaussian c.v.}} + \underbrace{\frac{\partial P(k_1)}{\partial \rho_L}\frac{\partial P(k_2)}{\partial \rho_L}\sigma^2(\rho_L)}_{\text{sample variance}}$$

- Cov (<δδ>, N): bispectrum, power spectrum
- Cov (N,N): power spectrum

	N	<δδ>	<δκ>	<κκ>
N	Cov (N, N)	Cov (<δδ>, N)	Cov (<δκ>, N)	Cov (<κκ>, N)
<δδ>	Cov (<δδ>, N)	Cov (<δδ>, <δδ>)	Cov (<δδ>, <δκ>)	Cov (<δδ>, <κκ>)
<δκ>	Cov (<δκ>, N)	Cov (<δκ>, <δδ>)	Cov (<δκ>, <δκ>)	Cov (<δκ>, <κκ>)
<κκ>	Cov (<κκ>, N)	Cov (<κκ>, <δδ>)	Cov (<κκ>, <δκ>)	Cov (<κκ>, <κκ>)

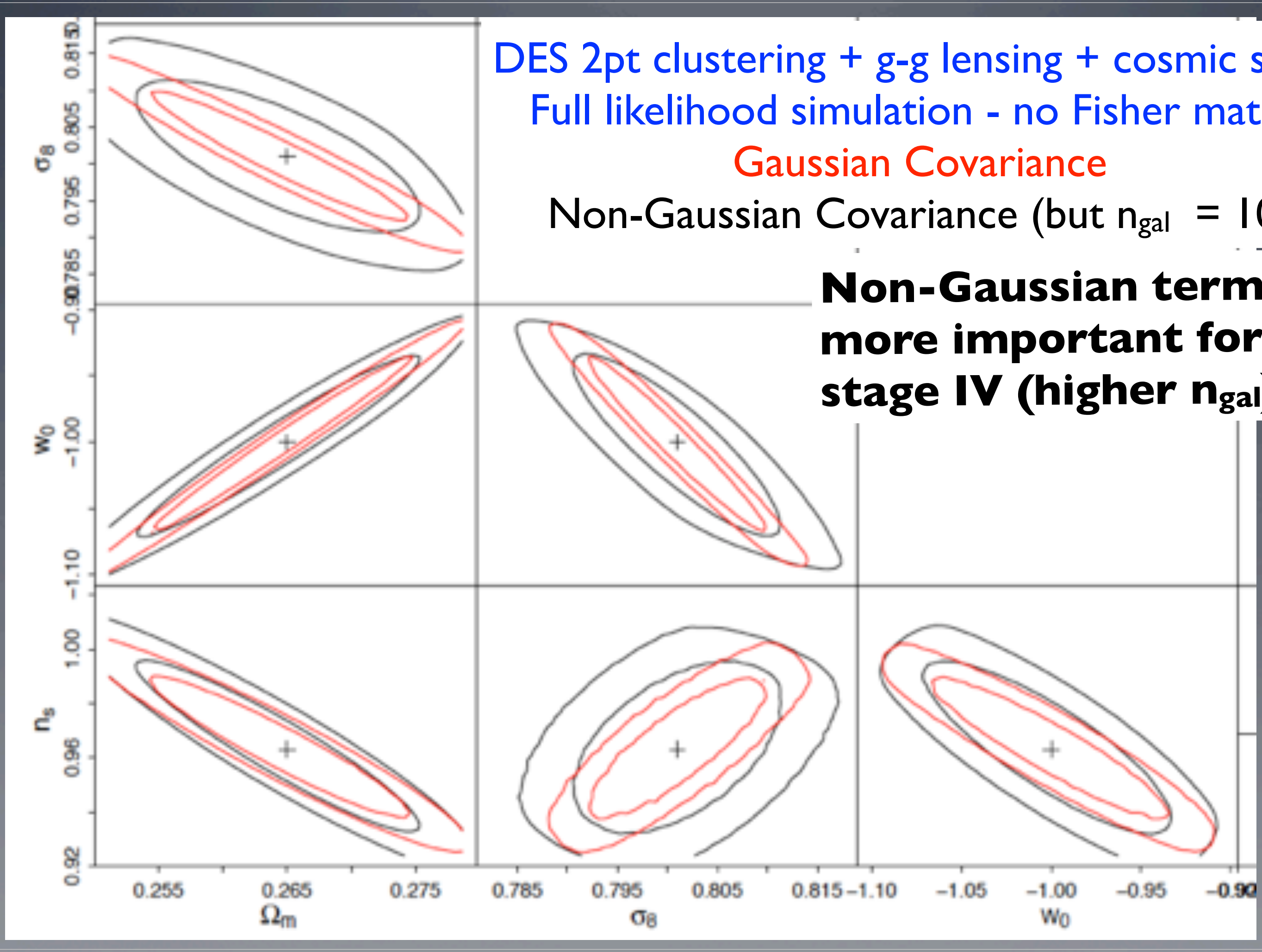
Joint *Inverse* Covariance

- correlated noise makes inversion difficult
- covariance estimate from data/simulations requires thousands of representative patches N_p

  $\widehat{C^{-1}} \approx \left[1 - \frac{n}{N_p}\right] \left(\hat{C}\right)^{-1}$, inflates contours

  - at least n + 2 *independent* realizations
- in practice, combine theoretical templates with matrix regularization
  - for DES, $\mathcal{O}(25)$ simulated realizations sufficient for combined probes covariance

Impact of Non-G. Covariance

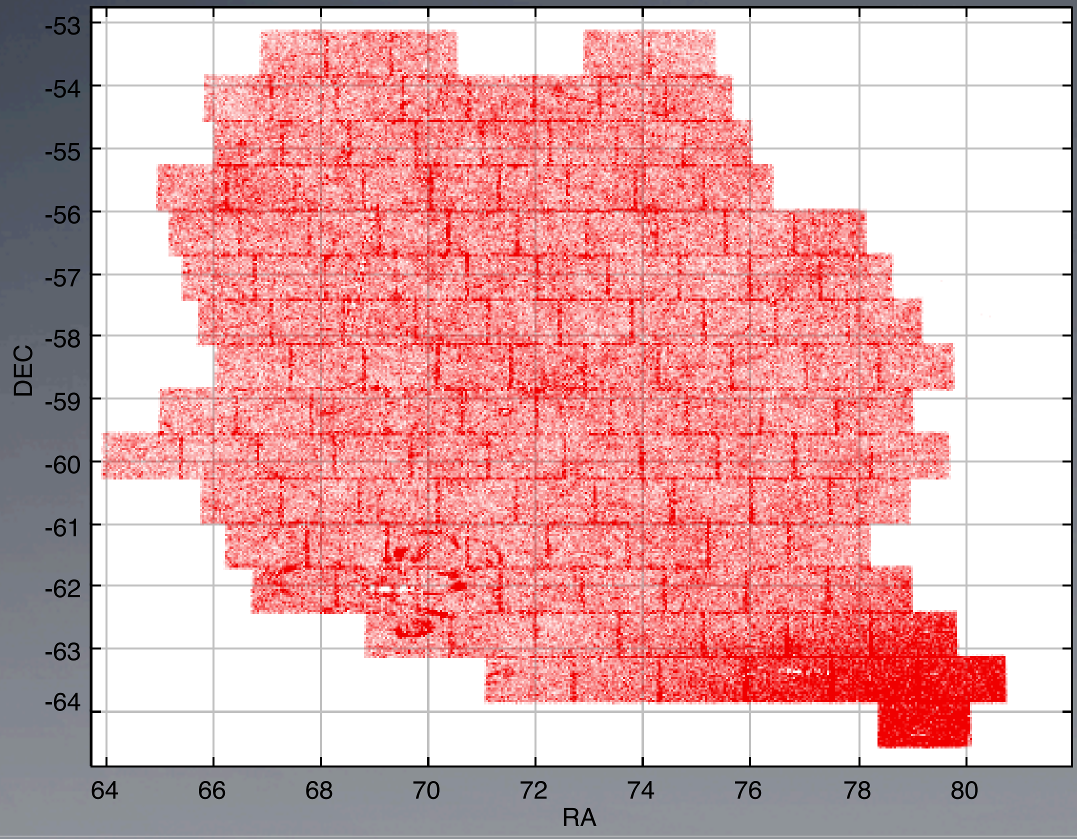


DES 2pt clustering + g-g lensing + cosmic shear
Full likelihood simulation - no Fisher matrix
Gaussian Covariance
Non-Gaussian Covariance (but n_{gal} = 10)

Non-Gaussian terms more important for stage IV (higher n_{gal})

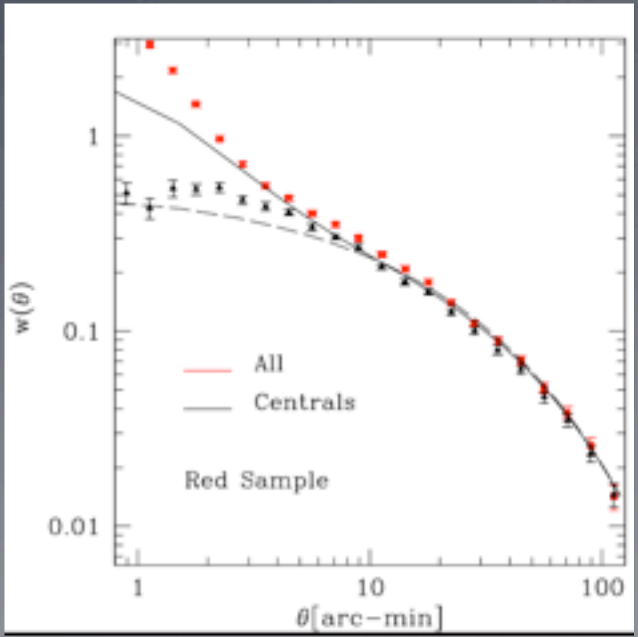
CosmoLike: Applications

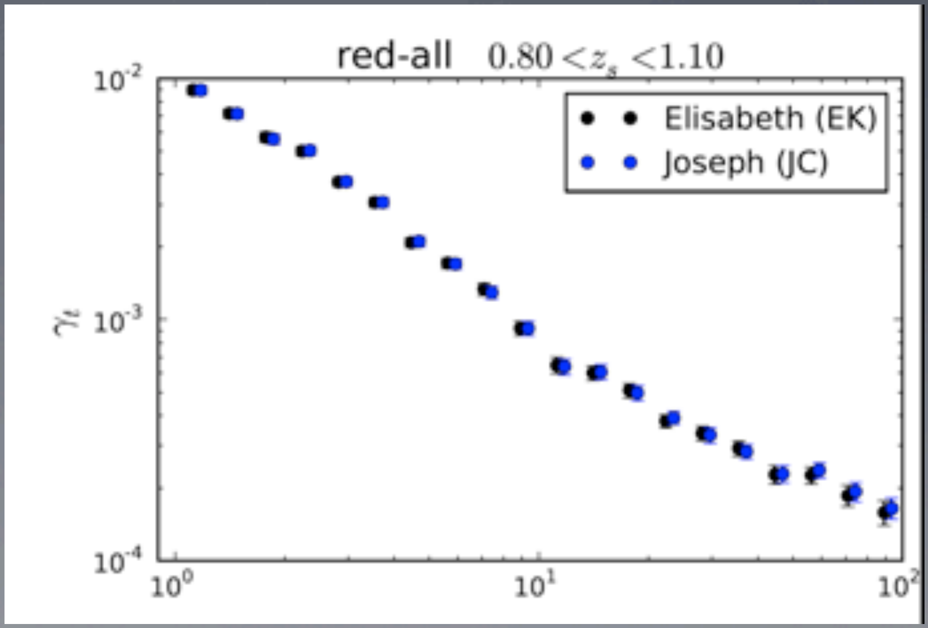
- model SPT lensing x DES (galaxies, shear)
  - expect detection in SV data already - catalog cleaning underway...



CosmoLike: Applications

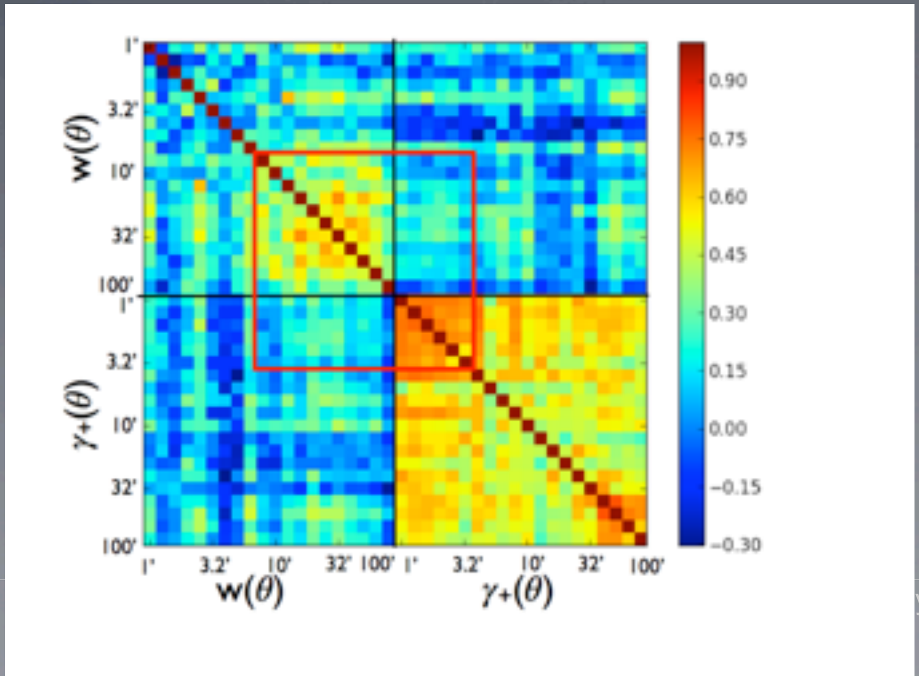
- model SPT lensing x DES (galaxies, shear)
  - expect detection in SV data already - catalog cleaning underway...
- combining galaxy-galaxy lensing + clustering
  - tests on DES mocks (EK, Clampitt, Crocce, et al.)





CosmoLike: Applications

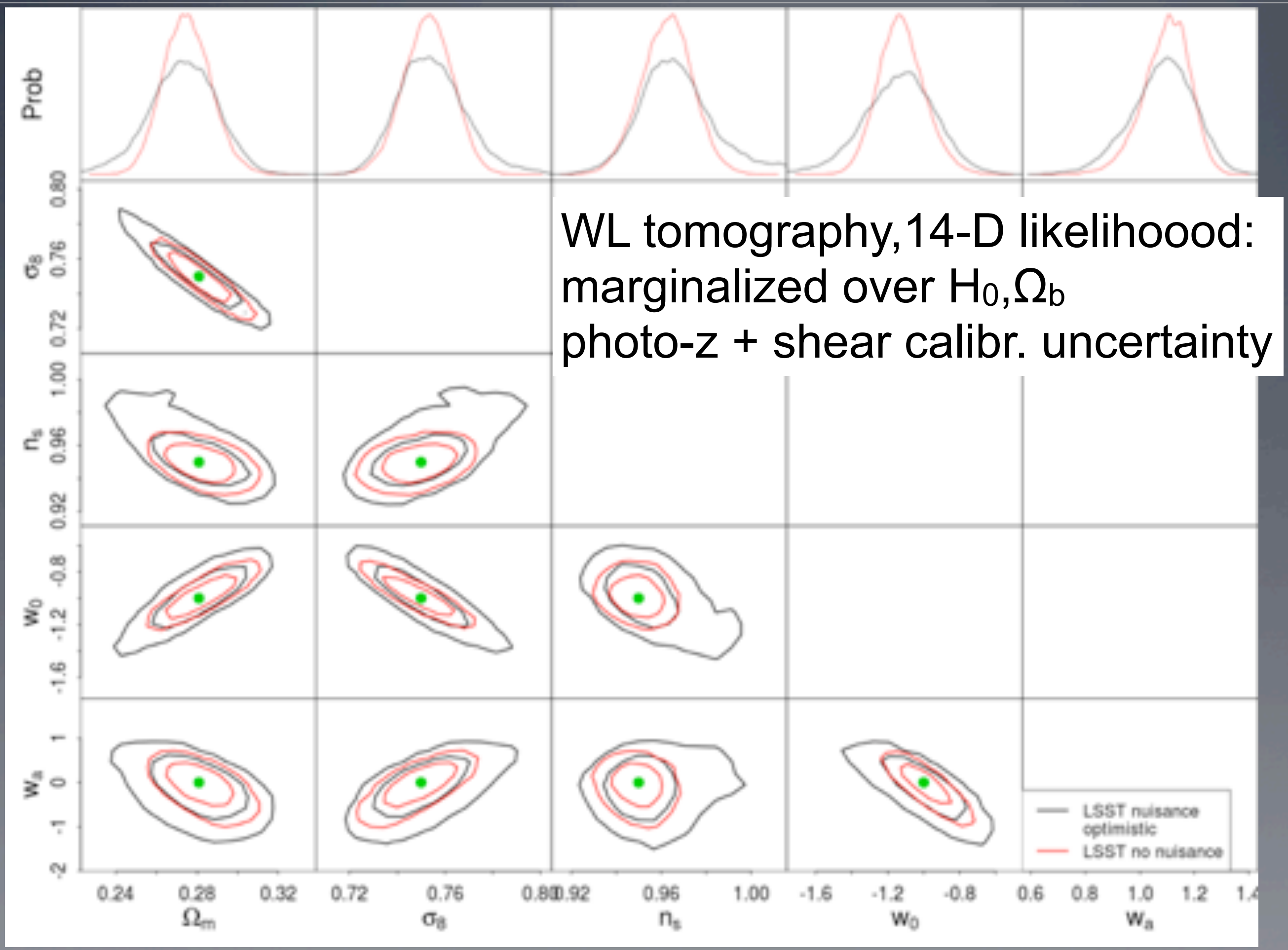
- model SPT lensing x DES (galaxies, shear)
  - expect detection in SV data already - catalog cleaning underway...
- combining galaxy-galaxy lensing + clustering
  - tests on DES mocks (EK, Clampitt, Crocce, et al.)
  - which scales to include? (Yoo & Seljak ’12, everything,...?)



CosmoLike: Applications

- model SPT lensing x DES (galaxies, shear)
  - expect detection in SV data already - catalog cleaning underway...
- combining galaxy-galaxy lensing + clustering
  - tests on DES mocks (EK, Clampitt, Crocce, et al.)
  - which scales to include? (Yoo & Seljak ’12, everything,...?)
- explore parameter space of systematic effects:
  - large survey volume, “precision cosmology”
    - limited by systematic uncertainties

Impact of Systematics



EK, Eifler et al. in prep.

Nuisance Parameters

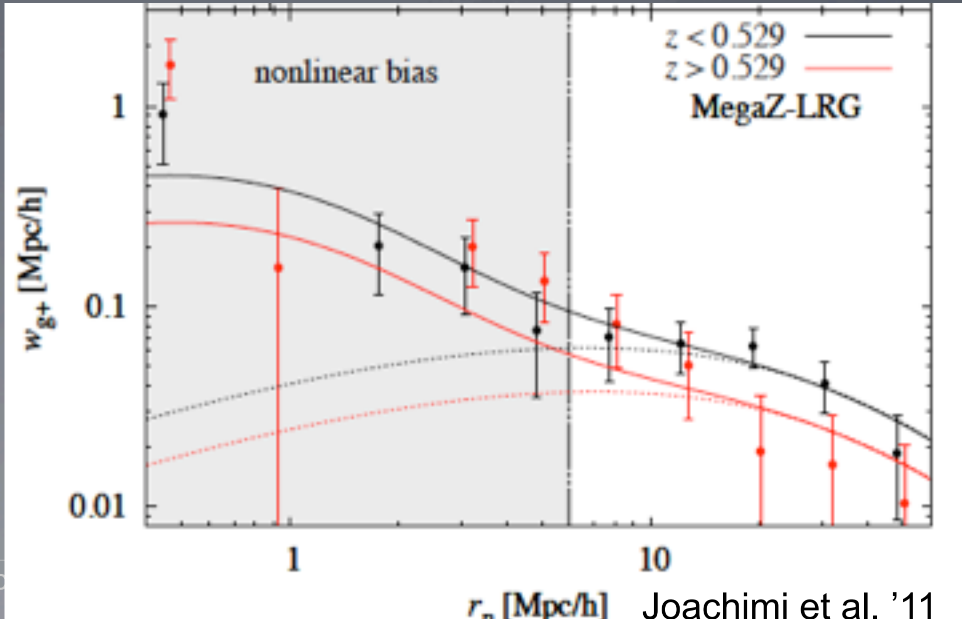
- large survey volume, “precision cosmology”
  - limited by systematic uncertainties
- easy to come up with many parameters
  - galaxy bias: e.g., 5 HOD + b_2 per bin
  - mass - observable relation for clusters
    - 2-3 parameters (mean relation) + scatter parameters, per bin
  - shear calibration, photo-z uncertainties, ...
  - Σ(poll among DES working groups) ~ 500-1000 parameters
- self-calibration + marginalization
  - can be costly (computationally, constraining power)

Combined Probes Approach

- specify data vector (probes included)
- identify (+prioritize) systematic effects
  - find suitable (minimal sets of) parameterizations + limits
  - joint systematics hardest
- possible constraints on nuisance parameters
  - independent observations
  - other observables from same data set
  - split data set
- combine theory, simulations & observations to reduce limits (priors) on nuisance parameters

Intrinsic Alignment Calibration

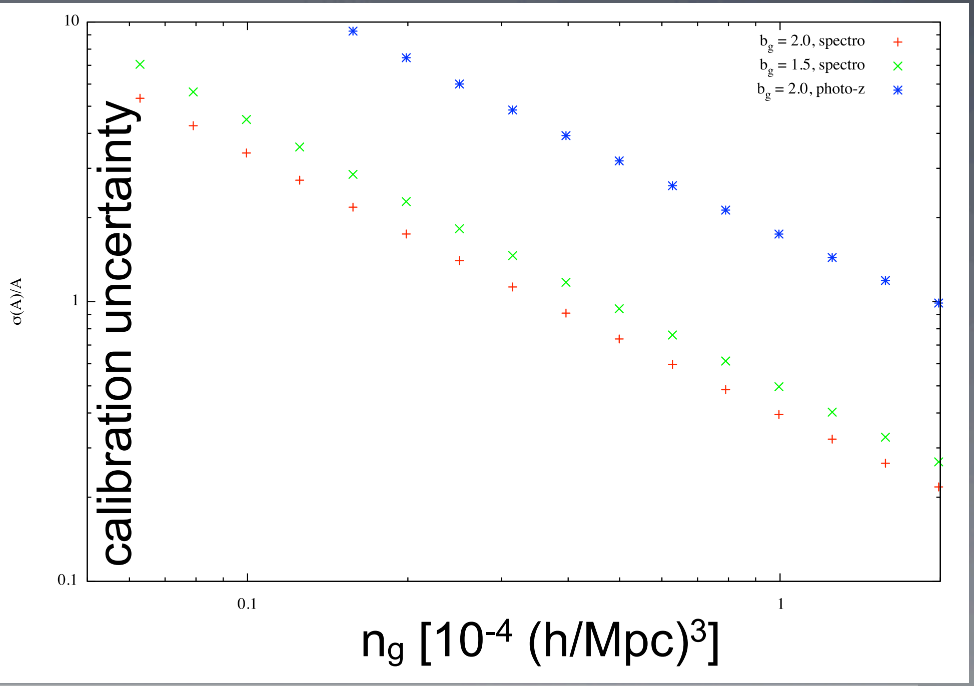
- tidal alignment of galaxies key contaminant for WL
- measured directly with position - shape correlation of low-z galaxies
  - use these observations to calibrate models (Hirata et al. ’07, Mandelbaum et al. ’11, Joachimi et al. ’12)



Joachimi et al. ’11

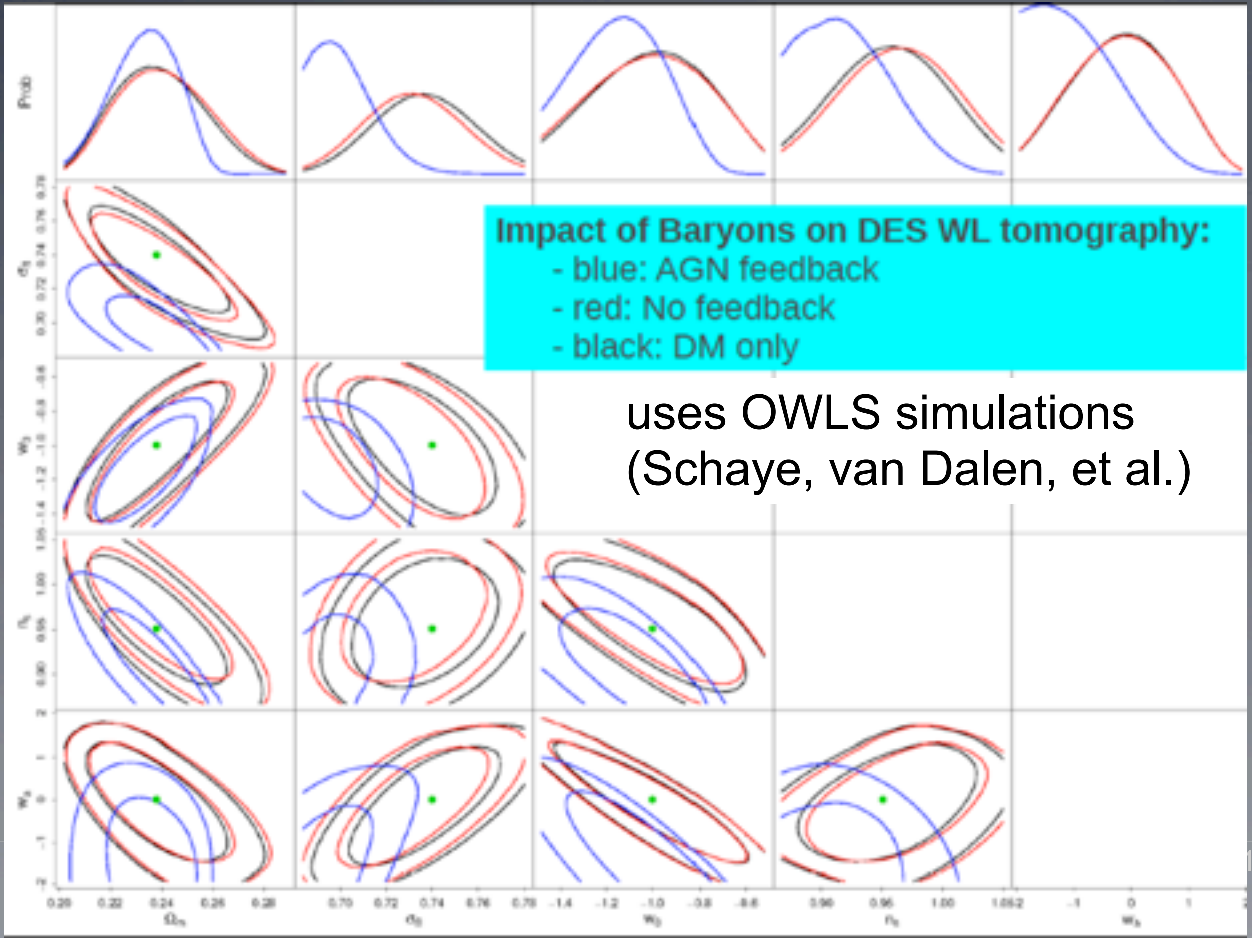
Intrinsic Alignment Calibration

- calibration depends on galaxy sample
  - need to calibrate DES galaxies
- calibration with spectroscopic sample more precise
- optimize follow-up strategy for IA calibration
- currently folding in galaxy properties from SV data, propagation through likelihood (EK, Eifler, Dodelson et al., in prep.)



Baryonic Effects

- baryonic effects (adiabatic contraction, feedback, ...) key uncertainties on small angular scales



uses OWLS simulations (Schaye, van Dalen, et al.)

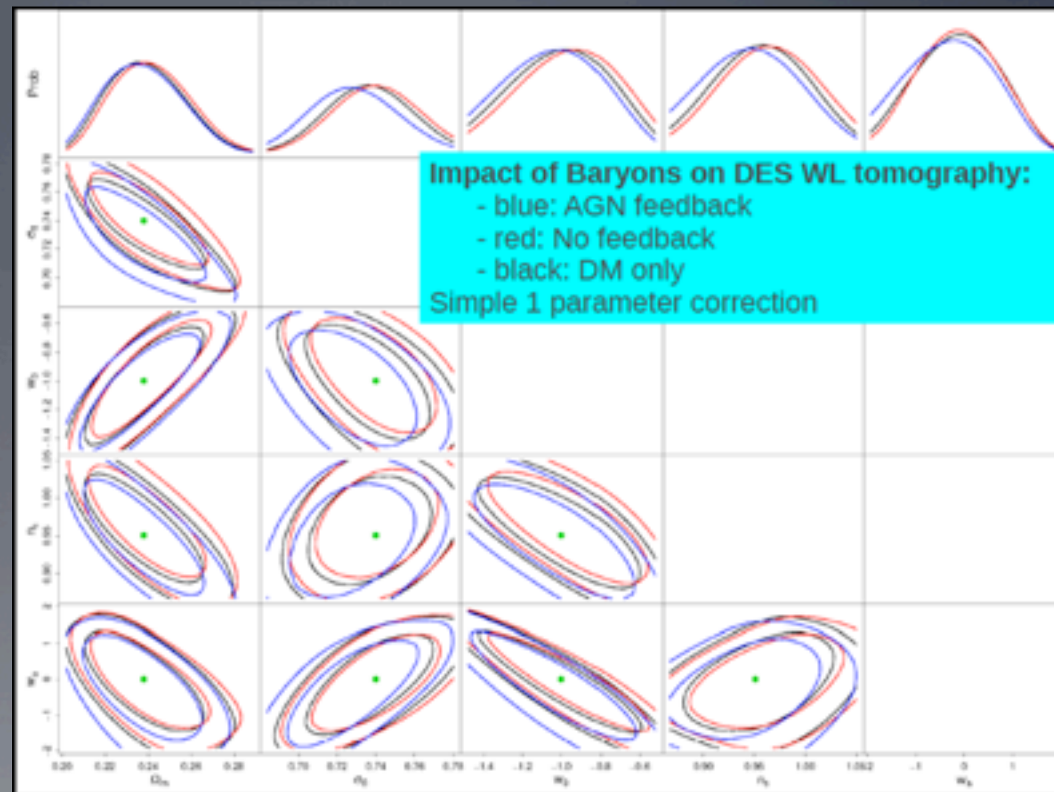
Eifler, EK, Dodelson, et al., in prep.

Baryonic Effects

marginalizing over simple parameterization based on range of simulations reduces bias

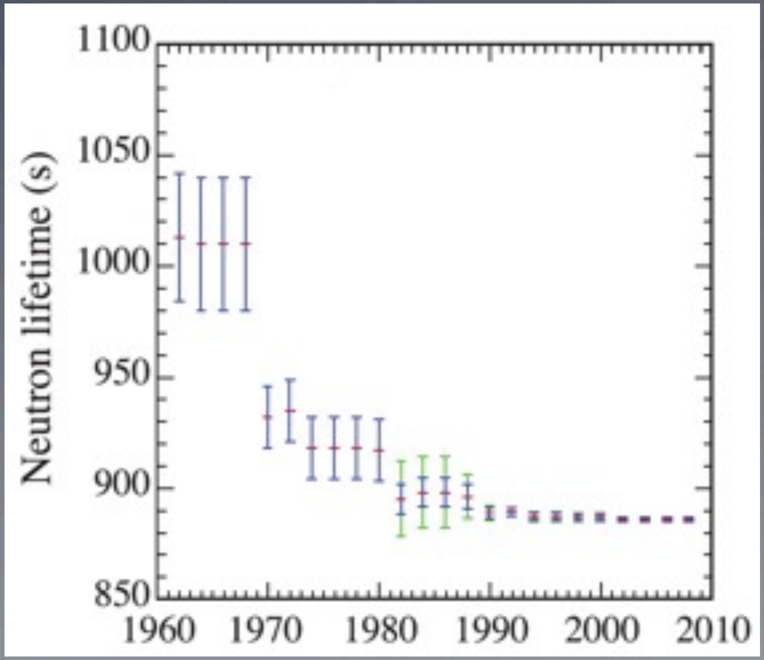
Future work:

- measure stacked halo profiles from SPT data (w/ L. Bleem)
- update parameter range for simulations, run new sims (w/ K. Heitmann & S. Habib)
- feed these into updated marginalization scheme



Experimenter Bias?

- nuisance parameters will outnumber cosmological parameters by far
  - what models + priors to adapt? when is the analysis done?
  - *don't use w = -1 to constrain intrinsic alignment model*



a warning from particle physics

Credit: A. Roodman, R. Kessler, Particle Data Group

Blind Analysis

- Experimenter's bias
  - choice of data samples + selections
  - choice of priors + evaluation of systematics
  - decision to stop work + publish
- Blind Analysis: Method to prevent experimenter's bias
  - hide the answer
  - must be customize for measurement
  - blind, but not dumb
- for DES
  - need to decide on cosmology parameters now
  - steps in unblinding, requirement to publish?

Summary I

- best constraints on cosmology from joint analysis of different probes
  - but individual constraints critical for consistency tests
- conceptually + computationally challenging
  - covariance, joint systematics require advanced modeling
  - nuisance parameters complicate analysis strategy

Outlook

- stage III surveys will improve constraints on DE
  - but, have to learn *a lot* about systematics first...
- for DES, 6-8 cosmological parameters in 5 years
- I’m excited to include detailed DE models, modified gravity parameterizations into the pipeline
  - realistically, in a few years
- with stage III as a training set, much to learn in stage IV (Euclid, LSST, etc.)

Conclusion

likelihood function

number counts: Poisson

2PCF: ~ Gaussian (?)

improvements needed for stage IV

model data vector

self-consistent modeling of all observables

matter power spectrum, halo mass + bias function + *selection function, systematics*

joint covariance

large and complicated, non-(block) diagonal matrix

use template + regularization

$$p\left(\boldsymbol{\pi}|\hat{\mathrm{d}}\right) = p\left(\boldsymbol{\pi}\right)\mathcal{L}\left(\hat{\mathrm{d}}|\mathrm{d}\left(\boldsymbol{\pi}\right), C\right)$$

large data vector

external data
blind analysis

validate

priors

parameters

~10 for cosmology
XXX for systematics