Modeling the Universe

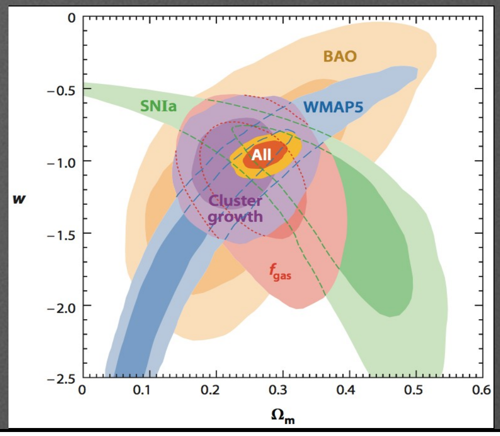
Interfacing Theory, Simulations, Statistical Methods, and Observations

Tim Eifler
(JPL/Caltech, University of Arizona)

The Challenge

→ reduced data and catalogs → summary statistics →

→



1) Independent probes
CMB, SN1a as priors

2) Large model vector
Self-consistent modeling of all observables as a function of
1) cosmological parameters (~10)
2) nuisance parameters (XXX)

large data vector

posterior probability

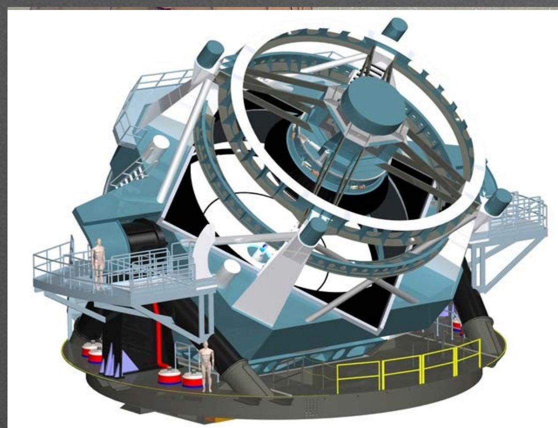
$$p\left(\pi|\hat{\mathrm{d}}\right) \propto p\left(\pi\right)\mathcal{L}\left(\hat{\mathrm{d}}|\mathbf{m}\left(\pi\right), C\right)$$

3) Enhanced modeling via
- Observations
- Simulations
- Theory

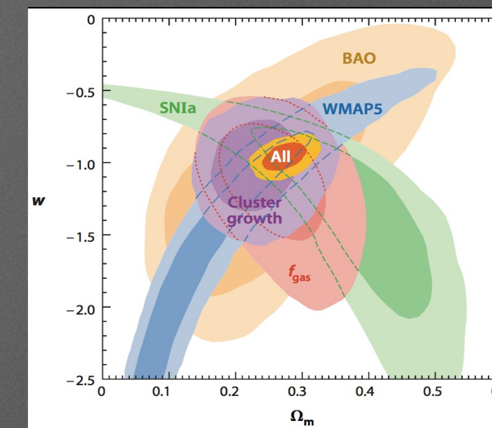
4) Statistics I - Likelihood function
- Multivariate Gaussian vs other parameterizations
- Non-parametric forms

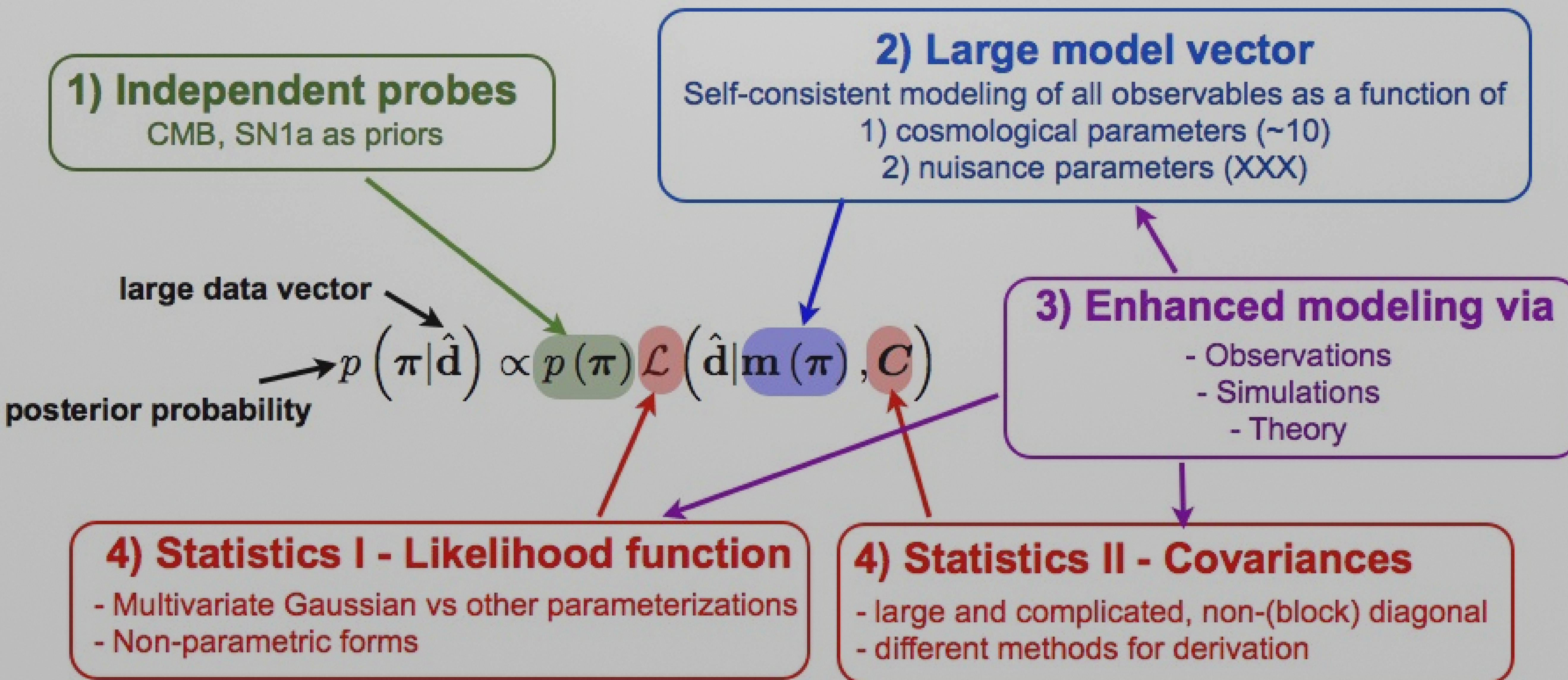
4) Statistics II - Covariances
- large and complicated, non-(block) diagonal
- different methods for derivation

The Challenge



reduced data
and catalogs





Introducing CosmoLike

Idea: consistent, multi-probe likelihood analysis software framework including

- Realistic statistical error bars (cross-probe covariances)
- Cross-correlations of observables/systematics
- Efficient treatment of nuisance parameters

Numerical Simulations/ Emulators

Weak Lensing, Galaxy Clustering, Clusters, CMB, CMB-LSS correlations

Systematics (photo-z, shape uncertainties)

Astrophysics (Intrinsic alignment, Baryonic Physics)

Explore fundamental physics (cosmic acceleration, neutrinos, tests of gravity)

Likelihood free inference Gaussianization of summary statistics

Galaxy bias models (linear, quadratic, HOD)

Multi-Probe Covariances/Hybrid Estimators

Project 1: Simulate a Multi-Probe Likelihood Analysis for LSST

Theory+Sims+Stats -> Obs

cosmolike - cosmological likelihood analyses for photometric galaxy surveys

CosmoLike release paper (www.cosmolike.info)
Krause & TE 2017

Example Data Vector and Systematics

- Weak Lensing (cosmic shear)
  - 10 tomography bins
  - 25 l bins, 25 < l < 5000
- Galaxy clustering
  - 4 redshift bins (0.2-0.4,0.4-0.6,0.6-0.8,0.8-1.0)
  - compare two samples: σ_z <0.04, redMaGiC
  - linear + quadratic bias only : l bins restricted to R> 10 Mpc/h
  - HOD modeling going to R>0.1 MPC/h
- Galaxy-galaxy lensing
  - galaxies from clustering (as lenses) with shear sources
- Clusters - number counts + shear profile
  - so far, 8 richness, 4 z-bins (same as clustering)
  - tomographic cluster lensing (500 < l < 10000)

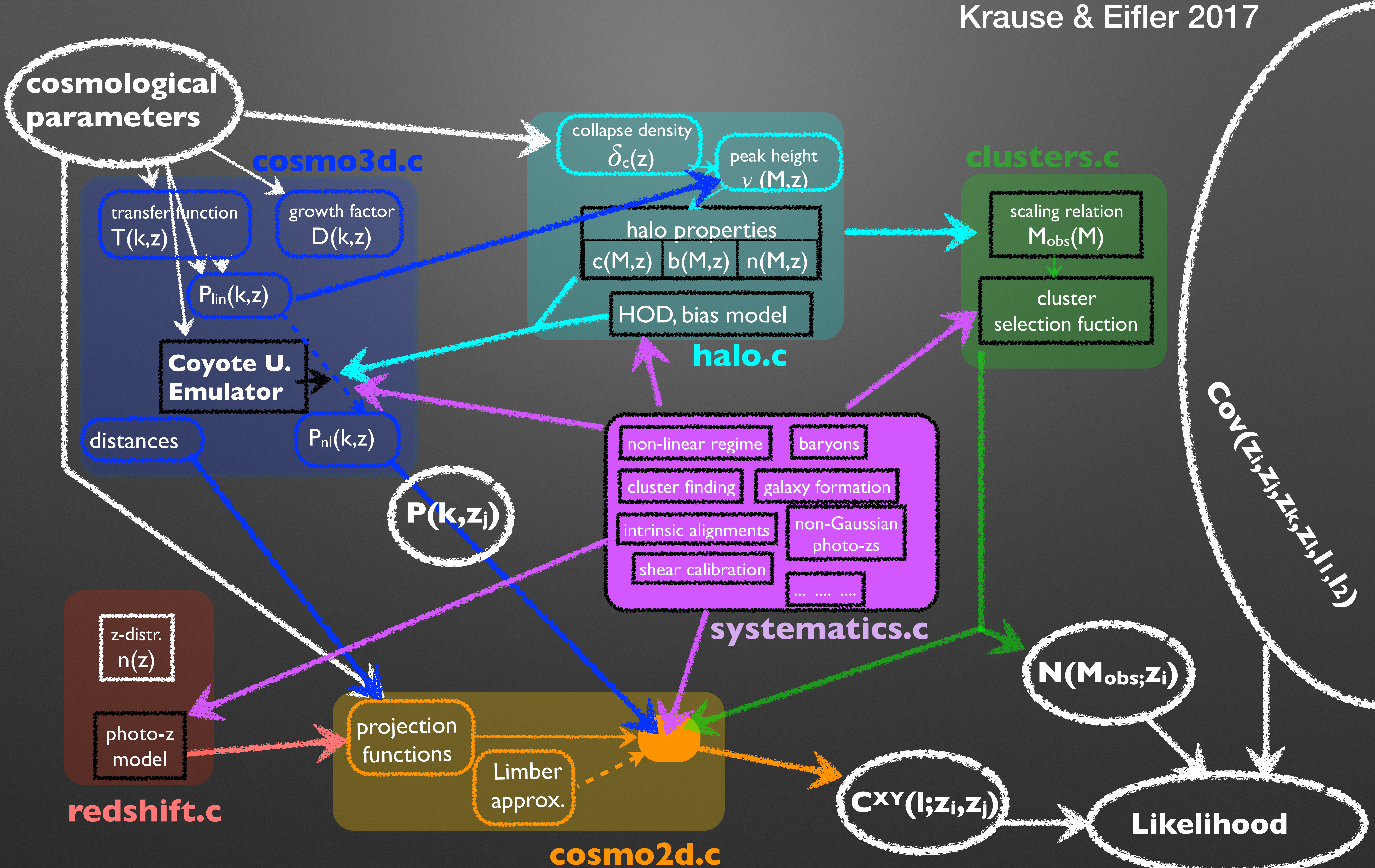
shear calibration,
photo-z (sources)
IA, baryons

b_1, b_2,...
photo-z (lenses)

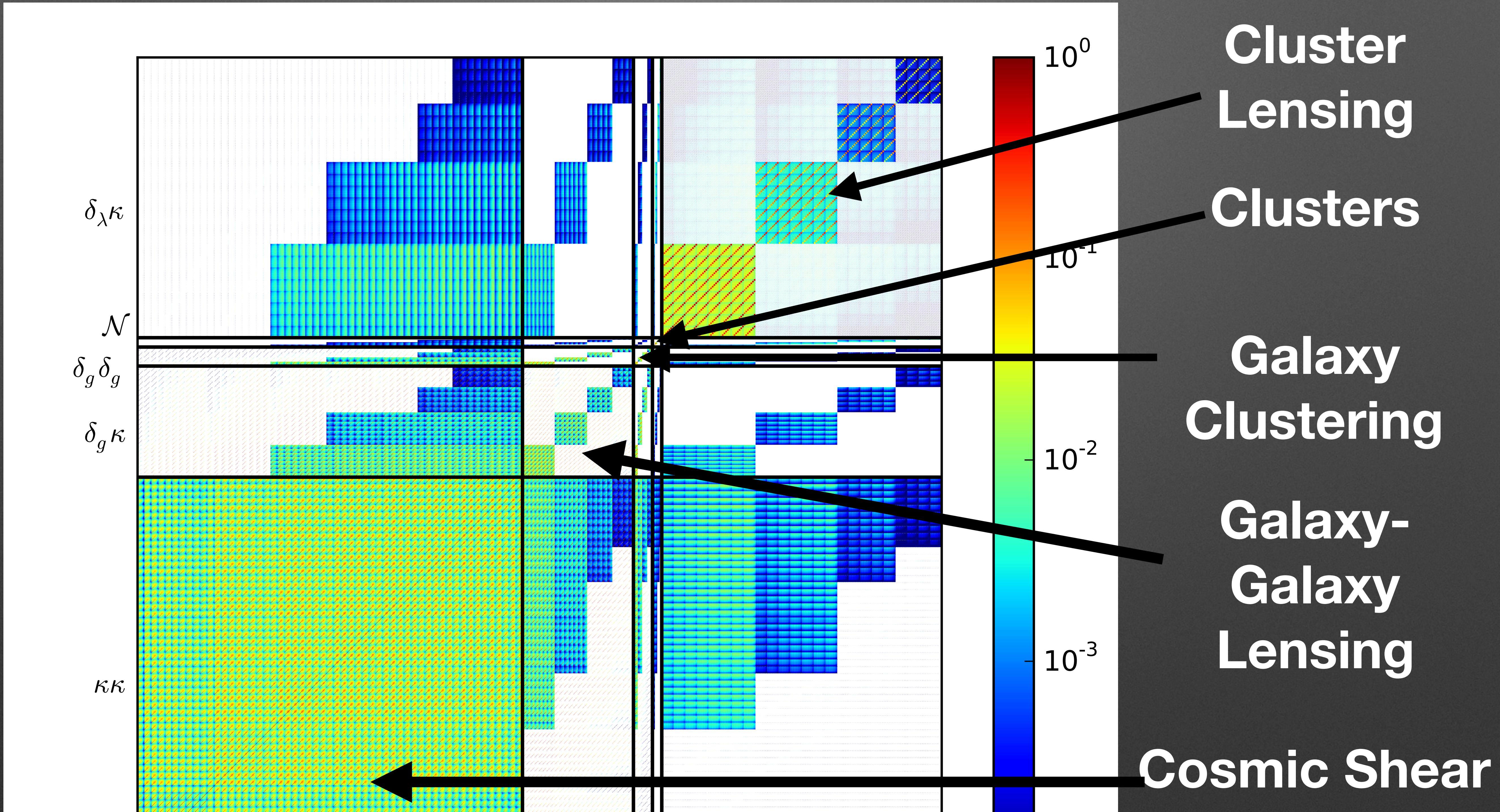
N-M relation
c-M relation
off-centering

CosmoLike - "Inner Workings"

Krause & Eifler 2017



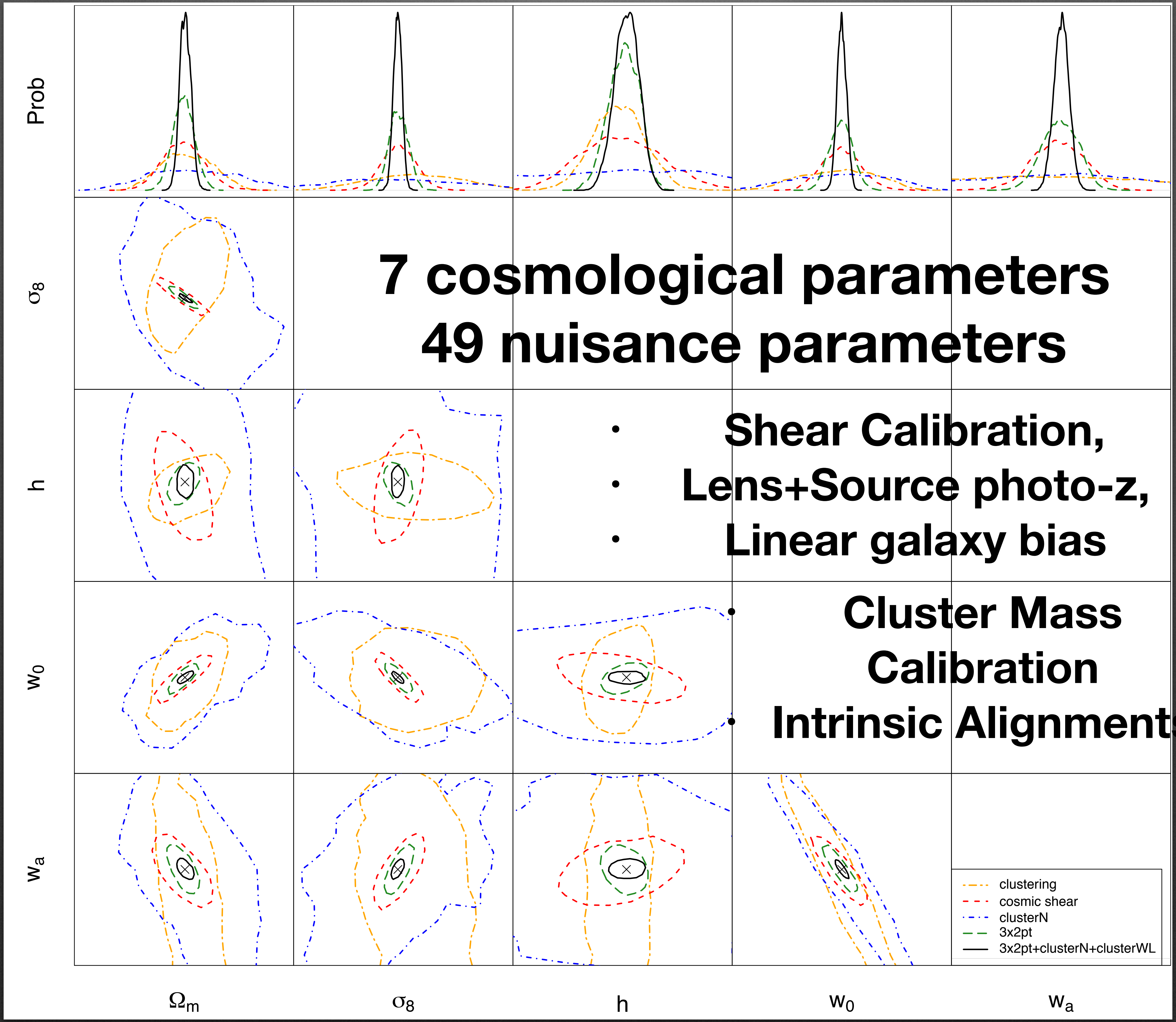
Multi-Probes Forecasts:



7+ million elements

details: Krause&TE '17

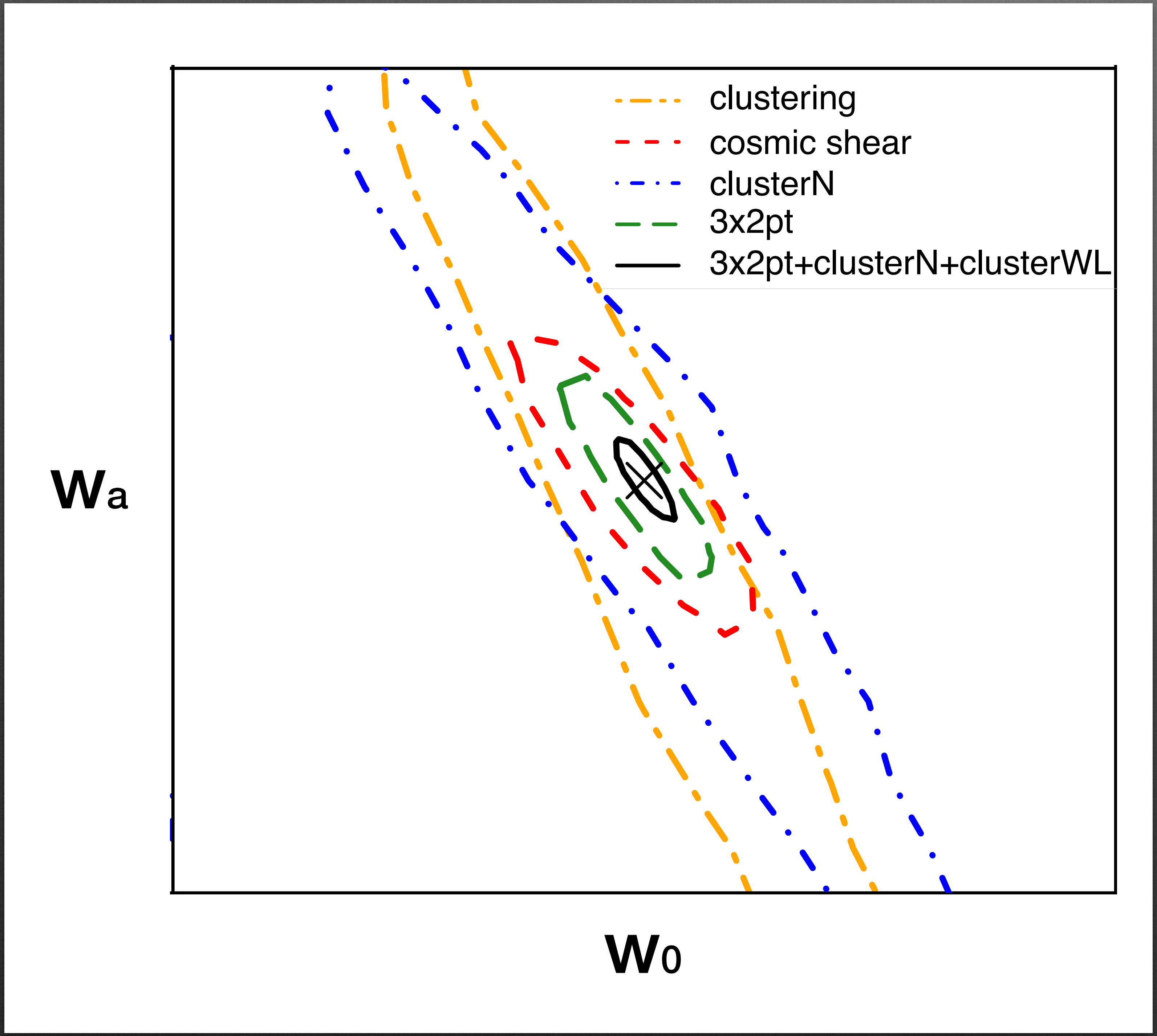
The Power of Combining Probes



7 cosmological parameters
49 nuisance parameters

- **Shear Calibration,**
- **Lens+Source photo-z,**
- **Linear galaxy bias**
- **Cluster Mass Calibration**
- **Intrinsic Alignments**

Zoom into w0-wa plane



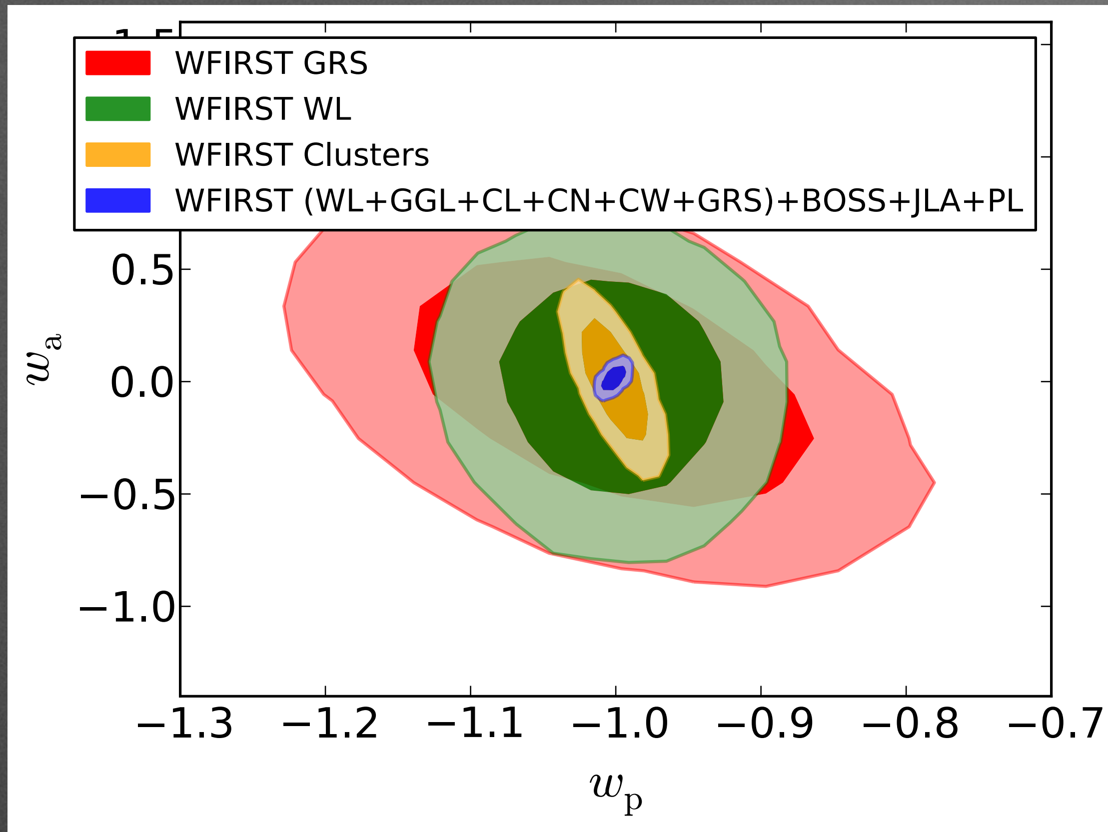
- Very non-linear gain in constraining power
- Most stringent requirements on numerical simulations, photo-z, shear calibration, etc flow from Multi-Probe statistical limits

Project 2: Exploring WFIRST survey strategies

Theory+Sims+Stats -> Obs

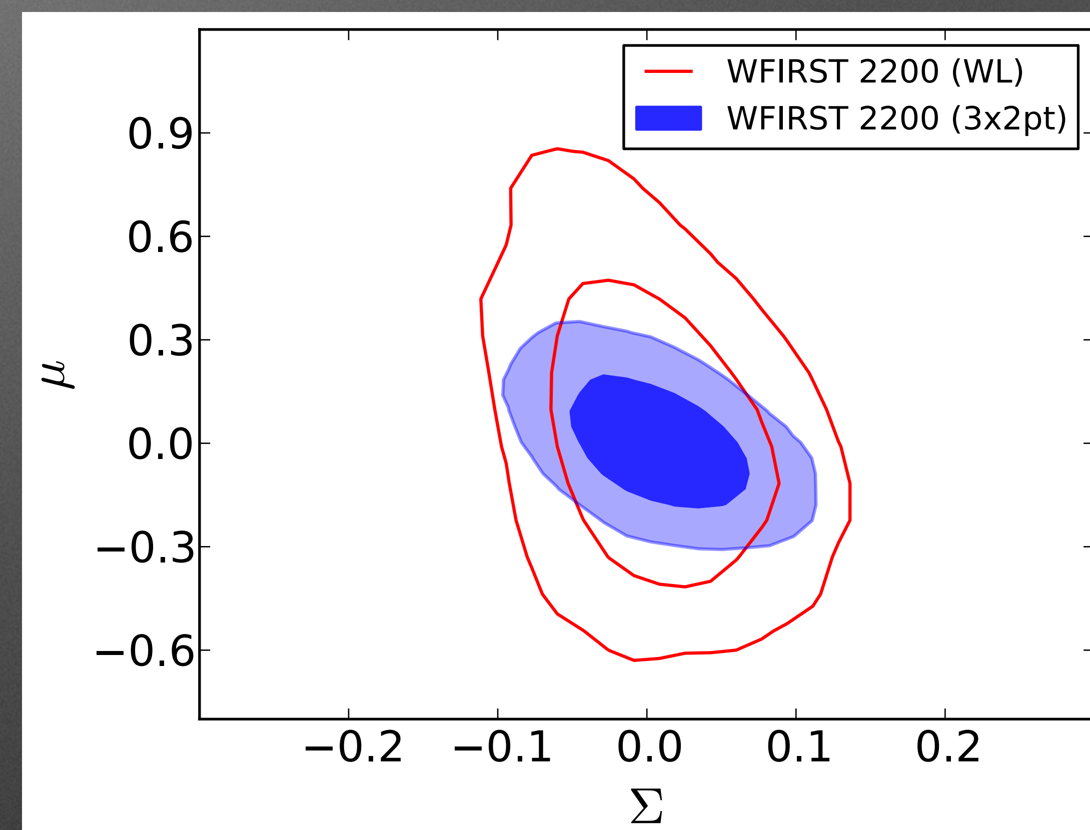
Project within the WFIRST *Cosmology with the High Latitude Survey* Science Investigation Team

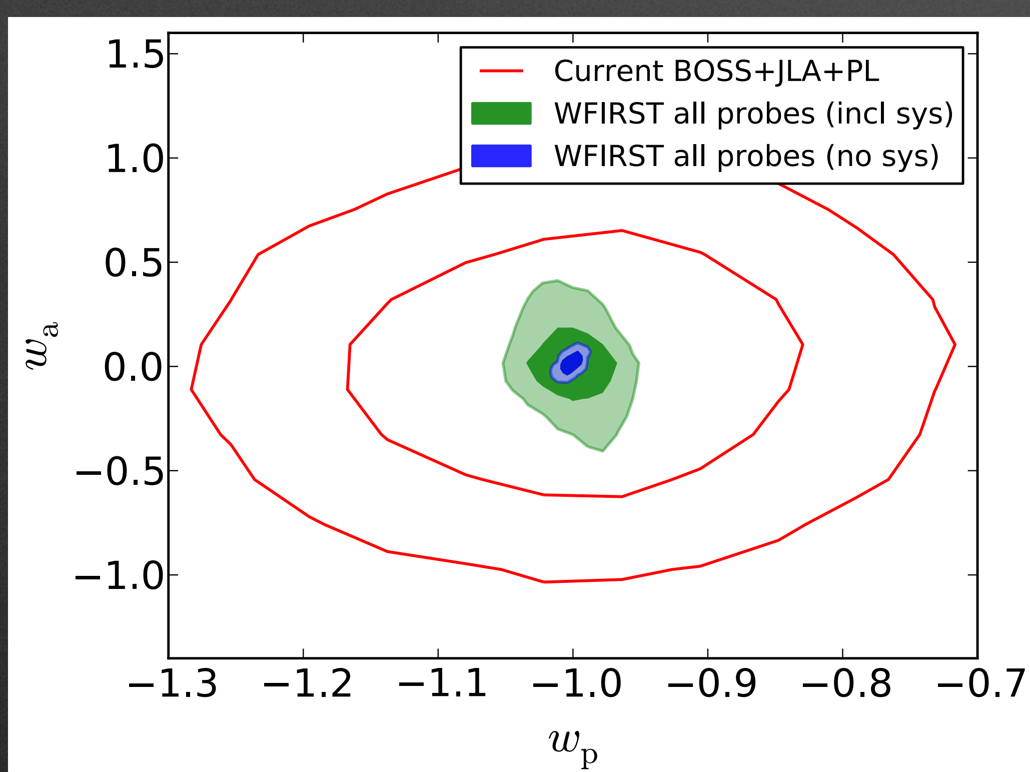
TE et al in prep



Individual vs multi-probe WFIRST analysis

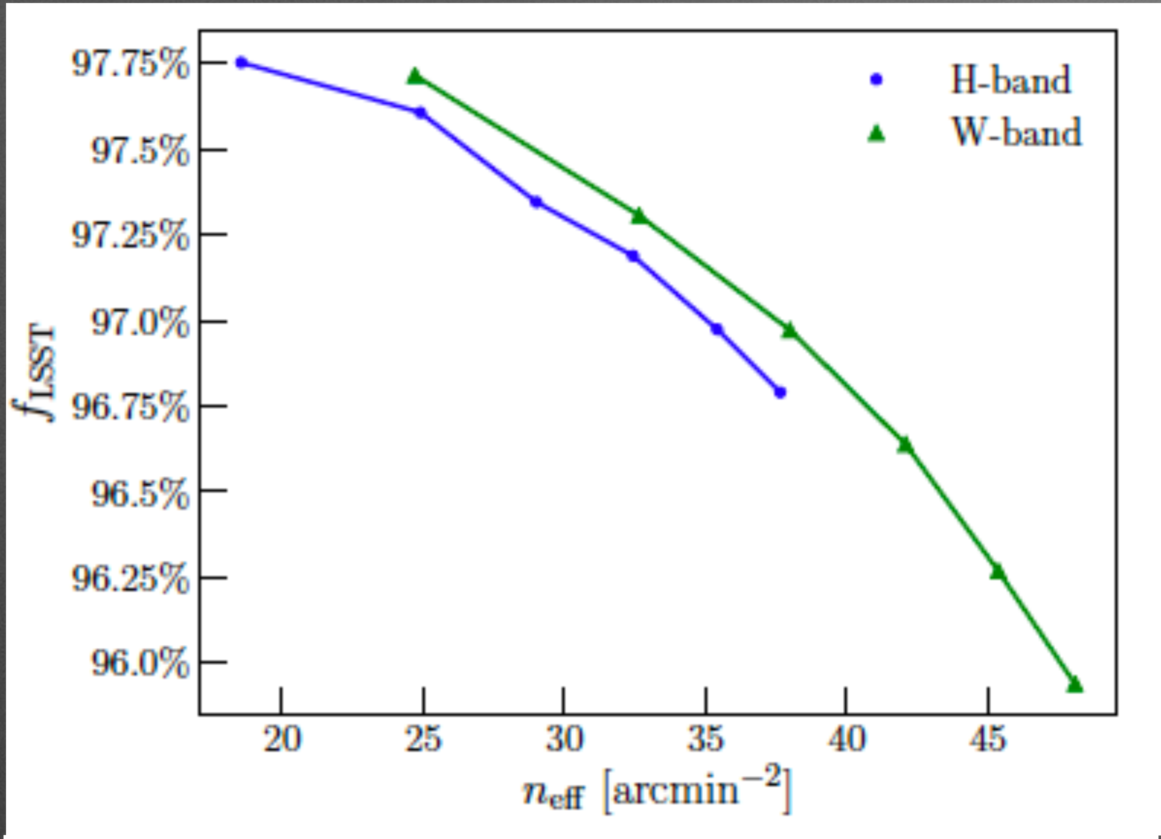
Modified Gravity

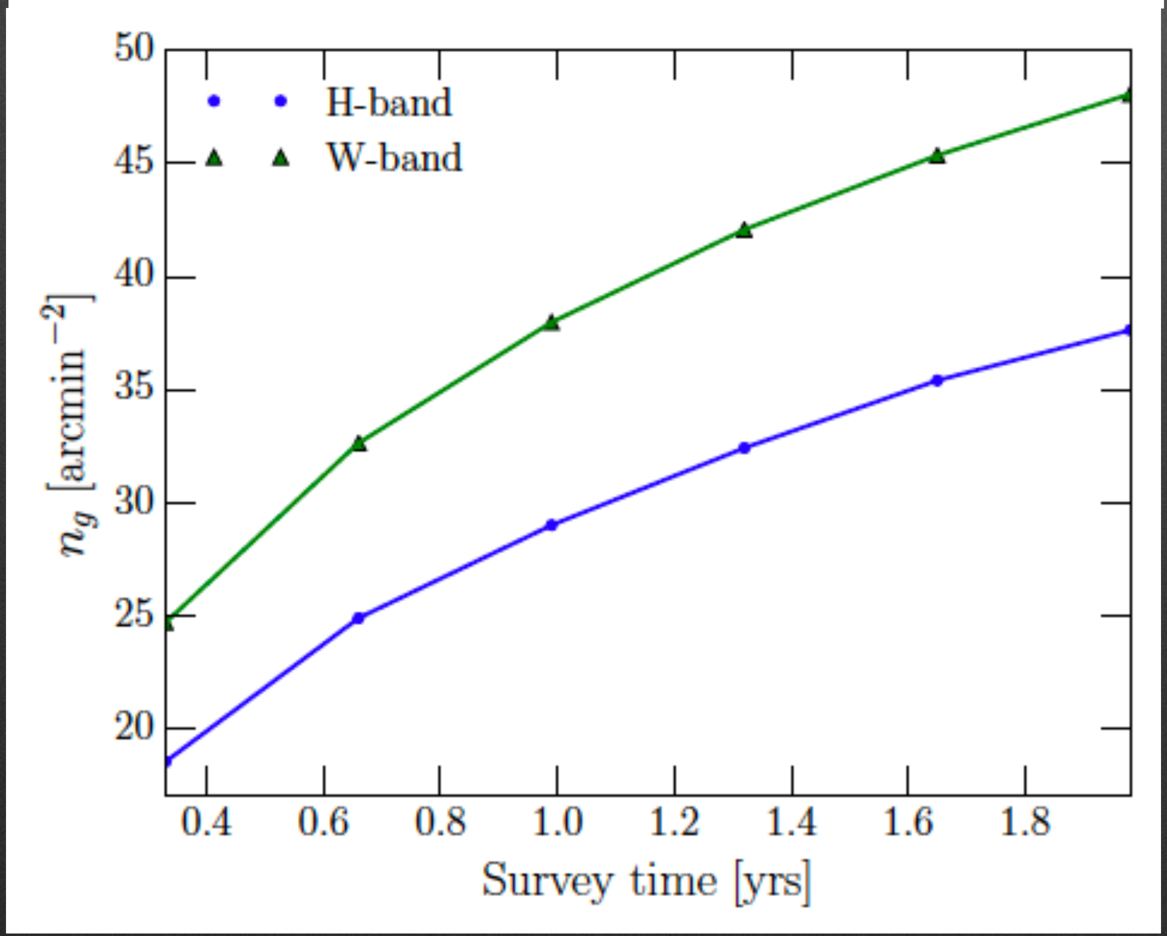




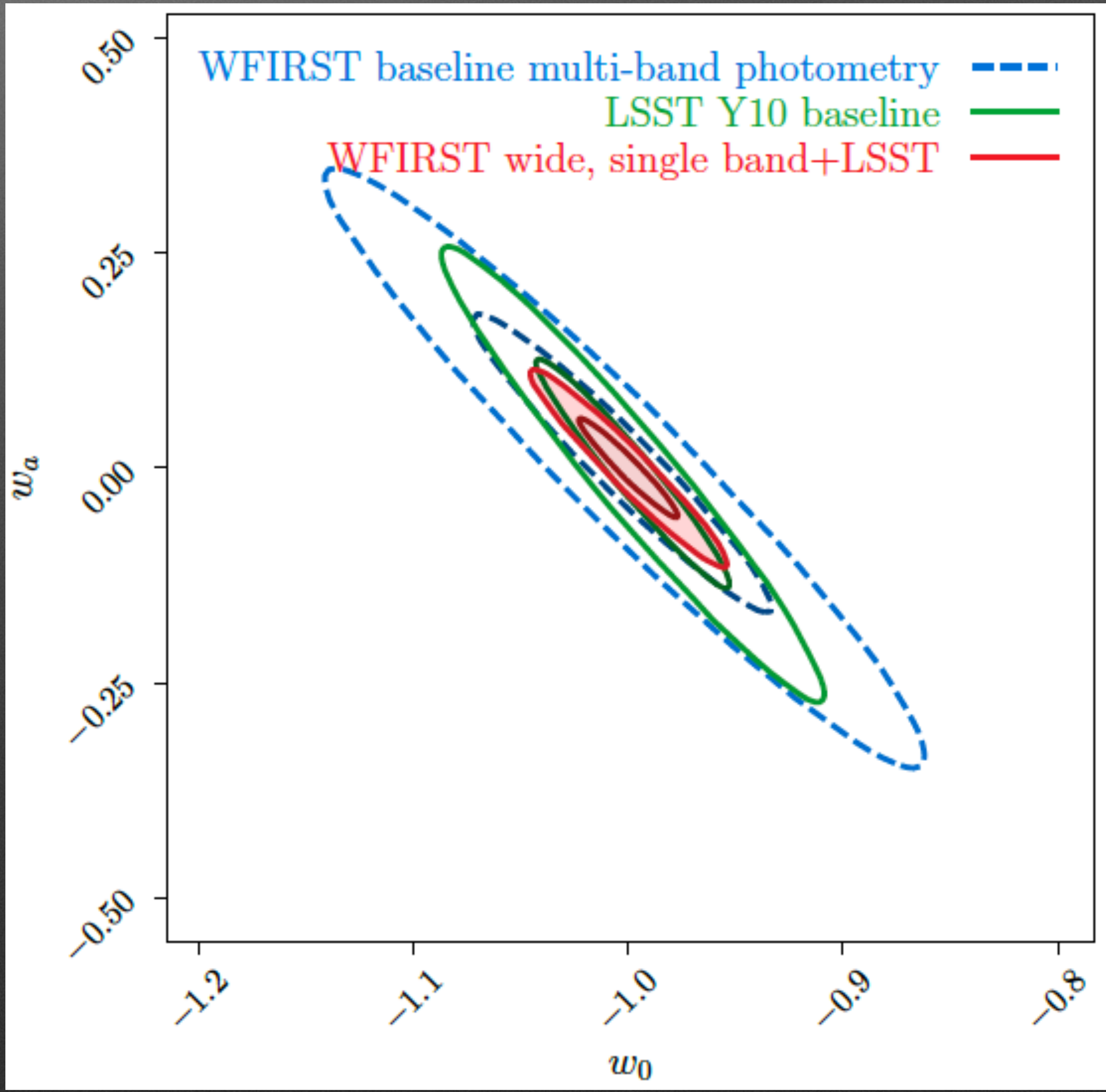
All-In Systematics
76 dimensions (7 cosmology, 69 systematics)

WFIRST - LSST synergies





Possible WFIRST extension of 1.6 years overlapping with LSST



Project 3:
New statistical methods to reduce Super-Computing needs

Theory+Stats -> Sims

Precision matrix expansion - efficient use of numerical simulations in estimating errors on cosmological parameters

Friedrich & TE 2018

The Problem: Inverse Covariance Estimation

- Analytical covariance model relies on approximations that might be too imprecise for an LSST Y10 data set
- Estimation the covariance from numerical simulations (brute force), requires 10^5-10^6 realizations of an LSST Year 10 like survey to shield against noise in the estimator
- Why?
    - The estimated inverse covariance is not the inverse of the estimated covariance
    - High-dimensionality of the data vector -> many elements in the covariance

Idea: Estimate the inverse directly

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{\xi}}) \sim \exp\left(-\frac{1}{2}\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right]\right) p(\boldsymbol{\pi})$$

$$\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right] = \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)^T \mathbf{C}^{-1}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)$$

Standard Estimator

$$\hat{\boldsymbol{\Psi}} = \frac{\nu - N_d - 1}{\nu}\hat{\mathbf{C}}^{-1}$$

New idea: Include theory information into estimator

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

$\mathbf{M} = \mathbf{A} + \mathbf{B}_m$

$$\mathbf{C} = \mathbf{M} + (\mathbf{B} - \mathbf{B}_m)$$

$\mathbf{X} := (\mathbf{B} - \mathbf{B}_m)\,\mathbf{M}^{-1}$

$$\mathbf{C} = (\mathbb{1} + \mathbf{X})\ \mathbf{M}$$

Invert and expand as power series

$$\begin{aligned}\boldsymbol{\Psi} &= \mathbf{M}^{-1}\left(\sum_{k=0}^{\infty}(-1)^k\mathbf{X}^k\right)\\ &= \mathbf{M}^{-1}\left(\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}\left[\mathbf{X}^3\right]\right)\end{aligned}$$

Build Estimator

$$\begin{aligned}\hat{\boldsymbol{\Psi}}_{\text{2nd}} = \ & \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\\ & +\mathbf{M}^{-1}\frac{\nu^2\hat{\mathbf{B}}\mathbf{M}^{-1}\hat{\mathbf{B}} - \nu\hat{\mathbf{B}}\ \mathrm{tr}\left(\mathbf{M}^{-1}\hat{\mathbf{B}}\right)}{\nu^2 + \nu - 2}\mathbf{M}^{-1}\end{aligned}$$

Only matrix multiplication, no inversion of estimated quantities

Idea: Estimate the inverse directly

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{\xi}}) \sim \exp\left(-\frac{1}{2}\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right]\right)\, p(\boldsymbol{\pi})$$

$$\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right] = \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)^T \mathbf{C}^{-1}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)$$

Standard Estimator

$$\hat{\boldsymbol{\Psi}} = \frac{\nu - N_d - 1}{\nu}\hat{\mathbf{C}}^{-1}$$

New idea: Include theory information into estimator

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

$\mathbf{M} = \mathbf{A} + \mathbf{B}_m$ →

$$\mathbf{C} = \mathbf{M} + (\mathbf{B} - \mathbf{B}_m)$$

$\mathbf{X} := (\mathbf{B} - \mathbf{B}_m)\,\mathbf{M}^{-1}$ →

$$\mathbf{C} = (\mathbb{1} + \mathbf{X})\;\mathbf{M}$$

Invert and expand as power series

$$\begin{aligned}\boldsymbol{\Psi} &= \mathbf{M}^{-1}\left(\sum_{k=0}^{\infty}(-1)^k\mathbf{X}^k\right)\\ &= \mathbf{M}^{-1}\left(\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}\left[\mathbf{X}^3\right]\right)\end{aligned}$$

Build Estimator

$$\begin{aligned}\hat{\boldsymbol{\Psi}}_{\mathrm{2nd}} = \;& \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\\ & +\mathbf{M}^{-1}\frac{\nu^2\hat{\mathbf{B}}\mathbf{M}^{-1}\hat{\mathbf{B}} - \nu\hat{\mathbf{B}}\,\mathrm{tr}\left(\mathbf{M}^{-1}\hat{\mathbf{B}}\right)}{\nu^2 + \nu - 2}\mathbf{M}^{-1}\end{aligned}$$

Only matrix multiplication, no inversion of estimated quantities

Standard estimator

$$\hat{\boldsymbol{\Psi}} = \frac{\nu - N_d - 1}{\nu} \hat{\mathbf{C}}^{-1}$$

$$\hat{\mathbf{C}} := \frac{1}{\nu} \sum_{i=1}^{N_\mathrm{s}} \left(\hat{\boldsymbol{\xi}}_i - \bar{\boldsymbol{\xi}} \right) \left(\hat{\boldsymbol{\xi}}_i - \bar{\boldsymbol{\xi}} \right)^T$$

Inverting quantities with “hats” is dangerous

Idea: Estimate the inverse directly

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{\xi}}) \sim \exp\left(-\frac{1}{2}\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right]\right)\, p(\boldsymbol{\pi})$$

$$\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right] = \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)^T \mathbf{C}^{-1}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)$$

Standard Estimator

$$\hat{\boldsymbol{\Psi}} = \frac{\nu - N_d - 1}{\nu}\hat{\mathbf{C}}^{-1}$$

New idea: Include theory information into estimator

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

$\mathbf{M} = \mathbf{A} + \mathbf{B}_m$

$$\mathbf{C} = \mathbf{M} + (\mathbf{B} - \mathbf{B}_m)$$

$\mathbf{X} := (\mathbf{B} - \mathbf{B}_m)\,\mathbf{M}^{-1}$

$$\mathbf{C} = (\mathbb{1} + \mathbf{X})\ \mathbf{M}$$

Invert and expand as power series

$$\begin{aligned}\boldsymbol{\Psi} &= \mathbf{M}^{-1}\left(\sum_{k=0}^{\infty}(-1)^k\mathbf{X}^k\right)\\ &= \mathbf{M}^{-1}\left(\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}\left[\mathbf{X}^3\right]\right)\end{aligned}$$

Build Estimator

$$\begin{aligned}\hat{\boldsymbol{\Psi}}_{\text{2nd}} = \ & \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\\ & -\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\\ & +\mathbf{M}^{-1}\frac{\nu^2\hat{\mathbf{B}}\mathbf{M}^{-1}\hat{\mathbf{B}} - \nu\hat{\mathbf{B}}\,\mathrm{tr}\left(\mathbf{M}^{-1}\hat{\mathbf{B}}\right)}{\nu^2 + \nu - 2}\mathbf{M}^{-1}\end{aligned}$$

Only matrix multiplication, no inversion of estimated quantities

Idea: Estimate the inverse directly

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{\xi}}) \sim \exp\left(-\frac{1}{2}\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right]\right) p(\boldsymbol{\pi})$$

$$\chi^2\left[\boldsymbol{\pi} \mid \hat{\boldsymbol{\xi}}, \mathbf{C}\right] = \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)^T \mathbf{C}^{-1} \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\boldsymbol{\pi}]\right)$$

Standard Estimator $\hat{\boldsymbol{\Psi}} = \frac{\nu - N_d - 1}{\nu} \hat{\mathbf{C}}^{-1}$

New idea: Include theory information into estimator

$\mathbf{C} = \mathbf{A} + \mathbf{B}$ → ($\mathbf{M} = \mathbf{A} + \mathbf{B}_m$) → $\mathbf{C} = \mathbf{M} + (\mathbf{B} - \mathbf{B}_m)$ → ($\mathbf{X} := (\mathbf{B} - \mathbf{B}_m)\, \mathbf{M}^{-1}$) → $\mathbf{C} = (\mathbb{1} + \mathbf{X})\, \mathbf{M}$

Invert and expand as power series

$$\begin{aligned}\boldsymbol{\Psi} &= \mathbf{M}^{-1}\left(\sum_{k=0}^{\infty}(-1)^k \mathbf{X}^k\right) \\ &= \mathbf{M}^{-1}\left(\mathbb{1} - \mathbf{X} + \mathbf{X}^2 + \mathcal{O}\left[\mathbf{X}^3\right]\right)\end{aligned}$$

Build Estimator

$$\begin{aligned}\hat{\boldsymbol{\Psi}}_{\text{2nd}} = \; & \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1} \\ & -\mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1} \\ & -\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1} \\ & -\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1} \\ & +\mathbf{M}^{-1}\frac{\nu^2\hat{\mathbf{B}}\mathbf{M}^{-1}\hat{\mathbf{B}} - \nu\hat{\mathbf{B}}\,\mathrm{tr}\left(\mathbf{M}^{-1}\hat{\mathbf{B}}\right)}{\nu^2 + \nu - 2}\mathbf{M}^{-1}\end{aligned}$$

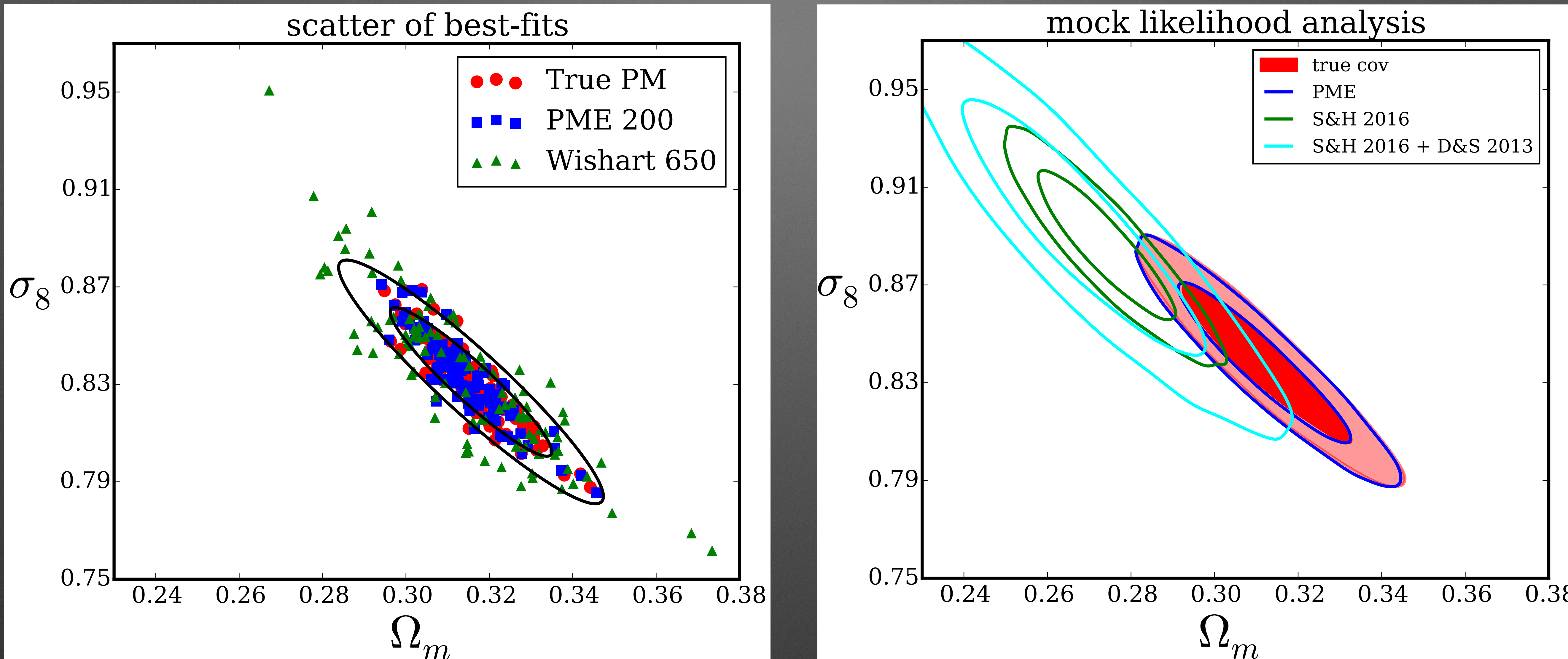
Only matrix multiplication, no inversion of estimated quantities

New estimator

$$
\begin{aligned}
\hat{\boldsymbol{\Psi}}_{\mathrm{2nd}} \quad = \quad & \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1} \\
& - \mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1} \\
& - \mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1} \\
& - \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1} \\
& + \mathbf{M}^{-1}\,\frac{\nu^2\hat{\mathbf{B}}\mathbf{M}^{-1}\hat{\mathbf{B}} - \nu\hat{\mathbf{B}}\,\mathrm{tr}\left(\mathbf{M}^{-1}\hat{\mathbf{B}}\right)}{\nu^2 + \nu - 2}\mathbf{M}^{-1}
\end{aligned}
$$

No more inversion of “hat quantities”…

New estimator performance



- Instead of >10^5 our new estimator only requires ~2000 numerical simulations (LSST case)
- Given that 1 sim is 1M CPUh, at 1c/CPUh
- New method reduces cost $1B to $20M (-> fund theorists!)
- Next step: data compression

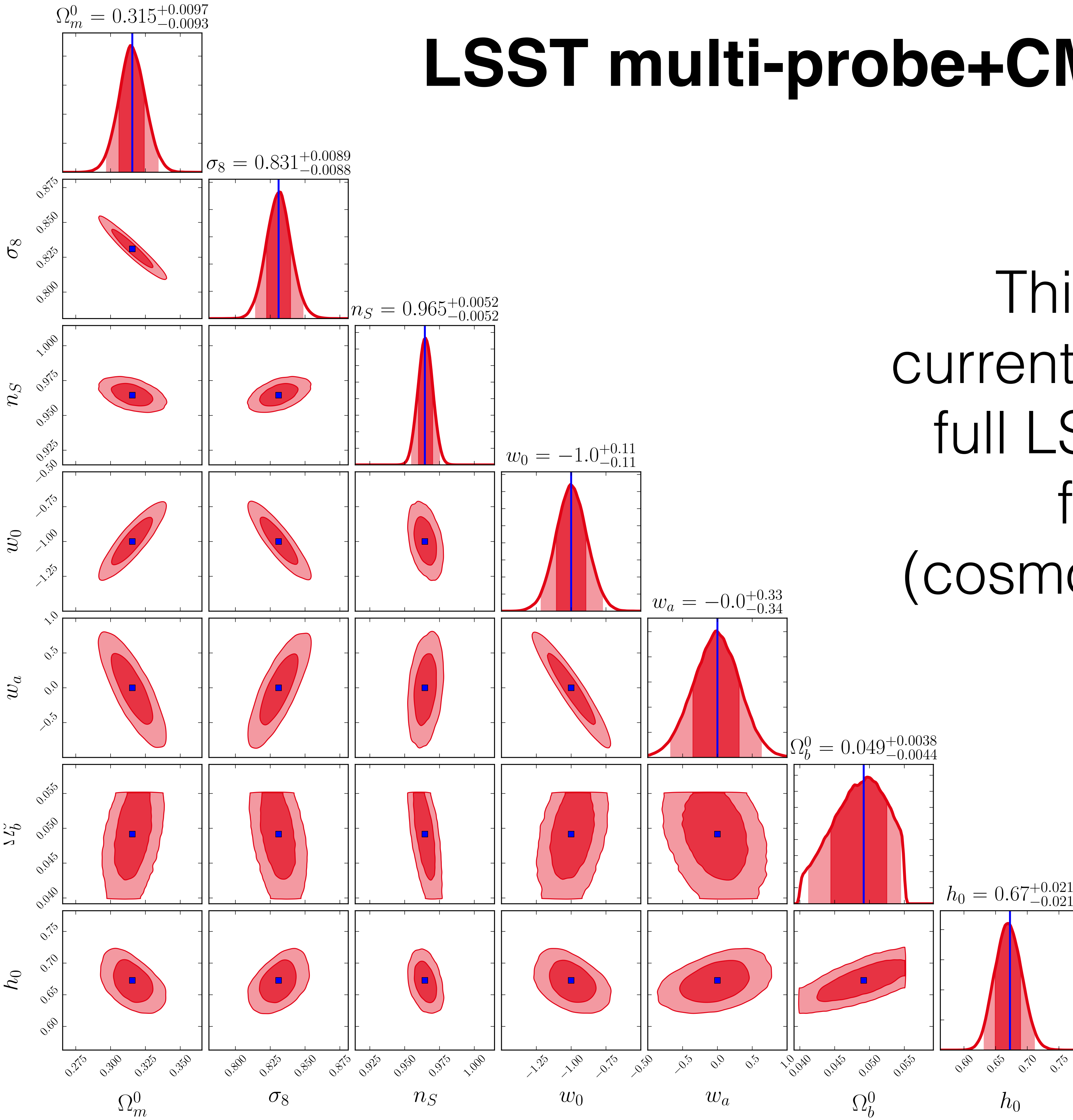
Project 4: Synergies of CMB-S4 and LSST

Obs -> Theory/Sims

Looking through the same lens: Shear calibration for LSST, Euclid, and WFIRST with stage 4 CMB lensing

Schaan, Krause, TE et al 2017

LSST multi-probe+CMB-S4 Lensing

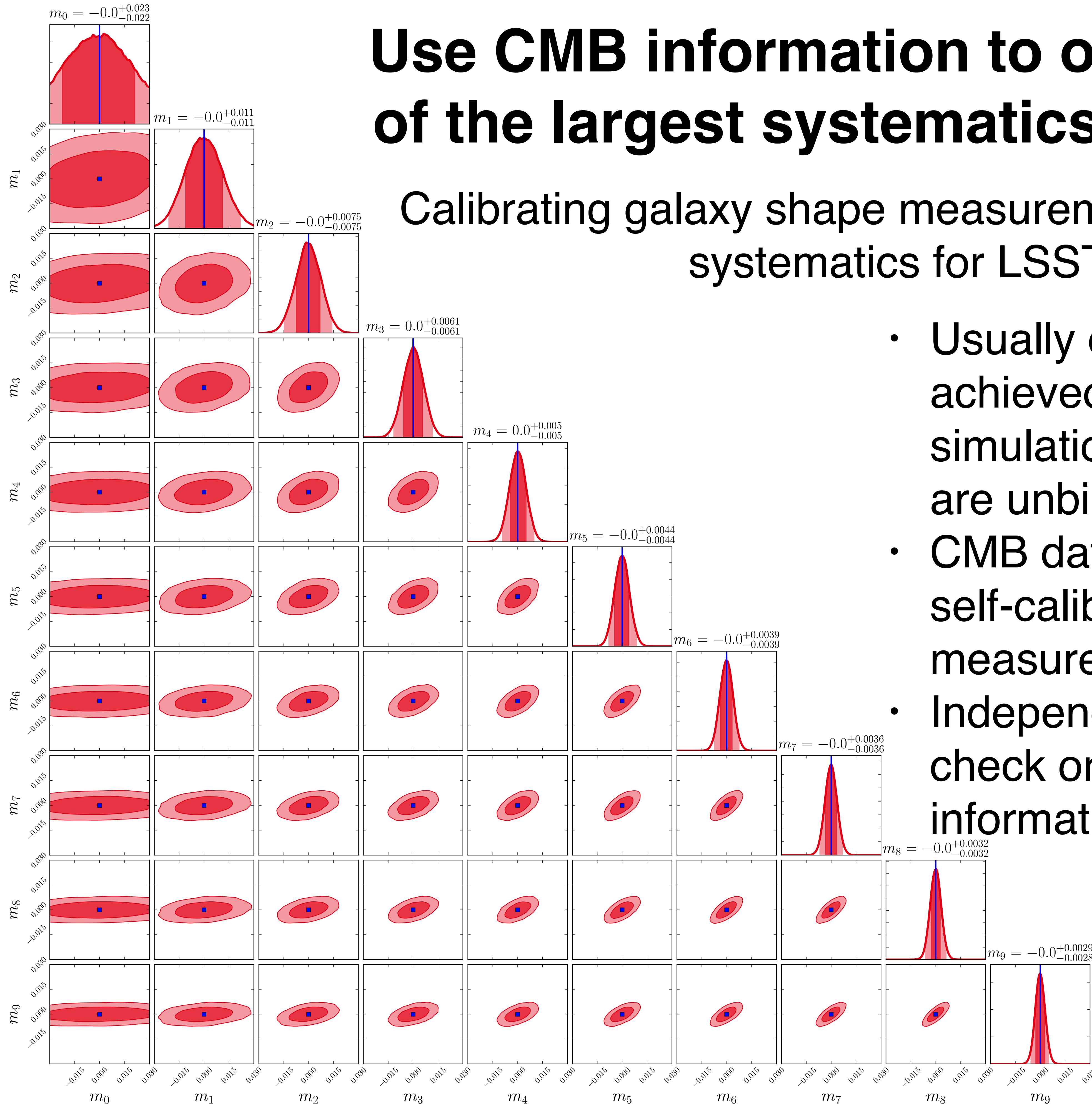


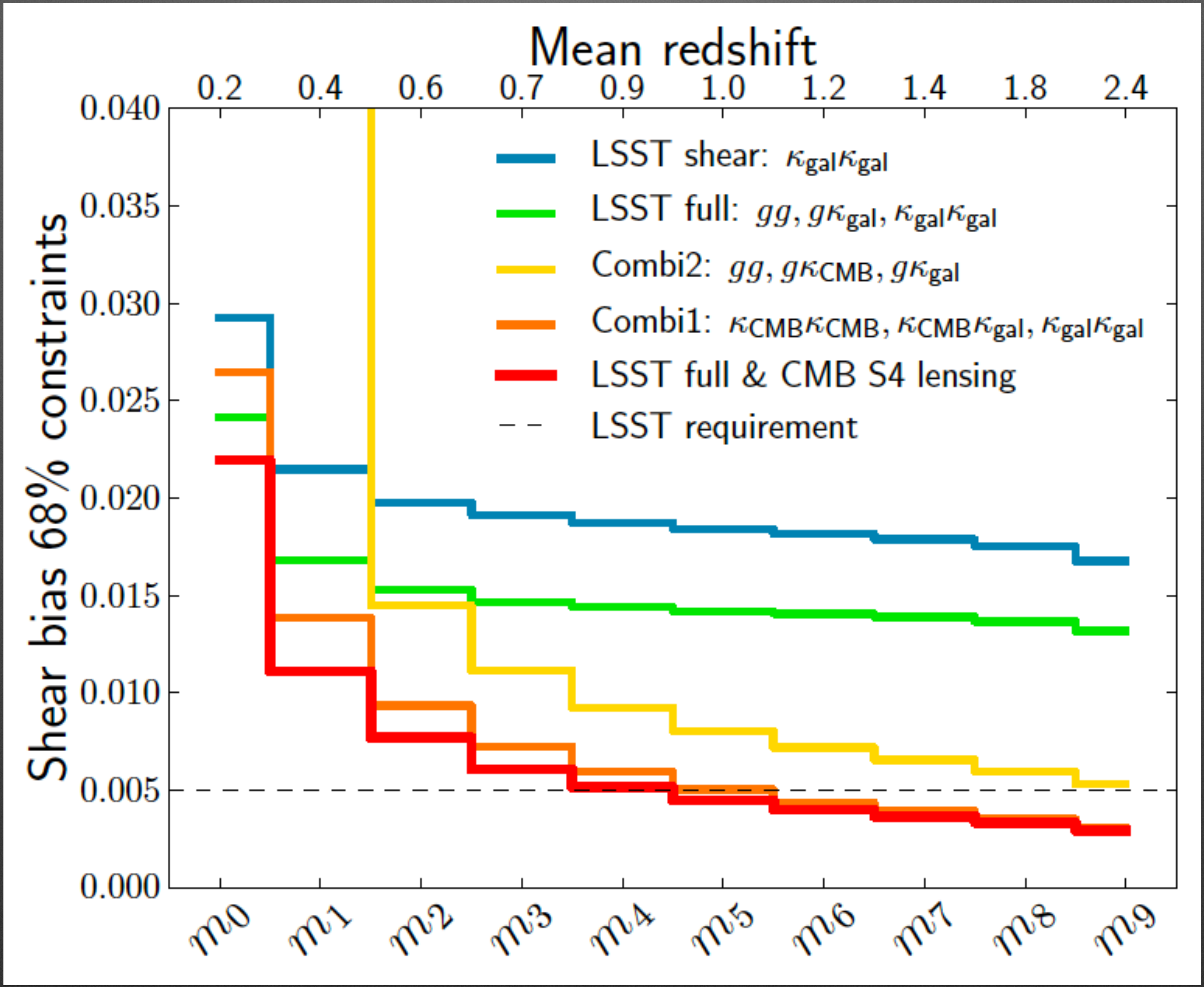
This project is currently evolving into full LSST+CMB-S4 forecasts (cosmology+inflation models)

Use CMB information to offset one of the largest systematics in LSST

Calibrating galaxy shape measurements is a major systematics for LSST

- Usually calibration is achieved through costly simulations, which we hope are unbiased
- CMB data can be used to self-calibrate LSST shape measurements
- Independent consistency check or additional information





Allows for independent LSST shear calibration at level of LSST requirements in highest z-bins (hard to achieve otherwise)

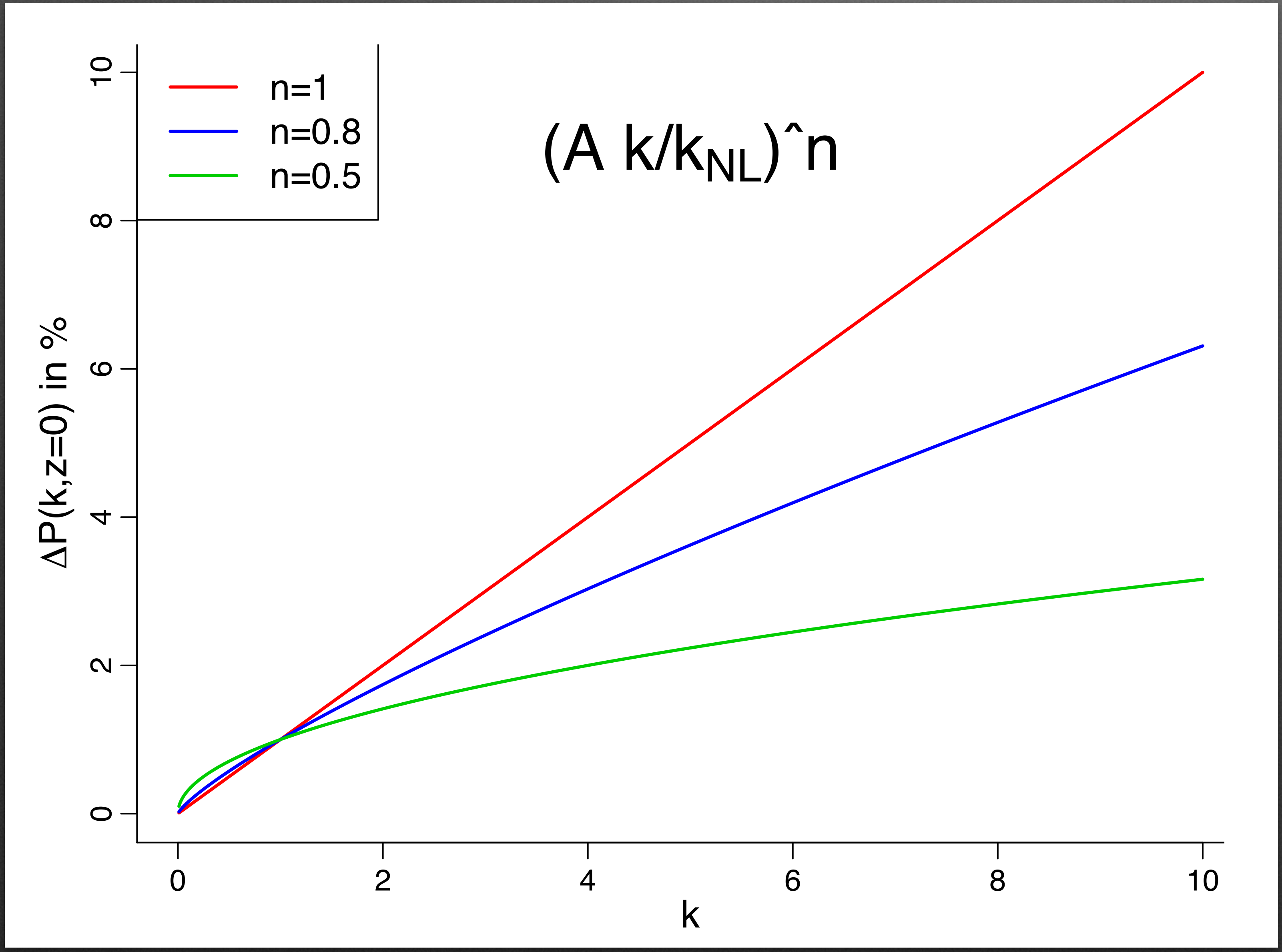
Project 5: Test Accuracy in Numerical Simulations

Theory -> Sims

Project in some shelf… might never see daylight…

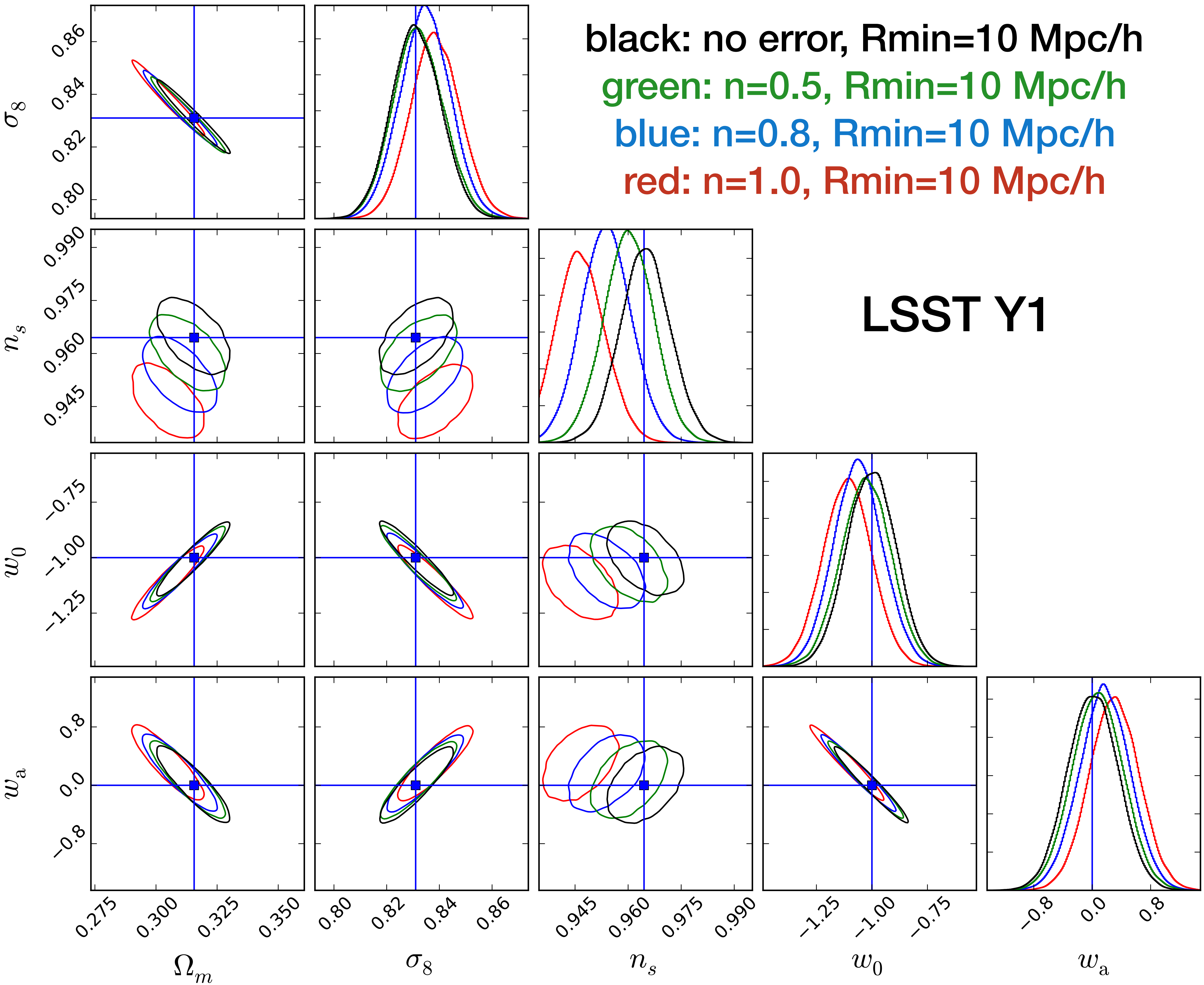
Numerical simulations have uncertainties

Matter Power Spectrum Error Models



Joint data vector details (LSST Y1, 18000 deg^2)

- **WL Source Sample:**
  - 5 tomographic bins [0:2.5]
  - 25 l-bins [30:5000]
  - n_gal=13 gal/arcmin^2
- **Clustering Lens Sample:**
  - 4 tomographic bins [0:1.0]
  - 25 l-bins [30:5000]
  - red sequence sample
  - k_max cut-off, R=[2,5,10] Mpc/h to justify linear galaxy bias models
- **Galaxy galaxy lensing using lens and source sample**


black: no error, Rmin=10 Mpc/h
green: n=0.5, Rmin=10 Mpc/h
blue: n=0.8, Rmin=10 Mpc/h
red: n=1.0, Rmin=10 Mpc/h
LSST Y1
σ_8
n_s
w_0
w_a
Ω_m
0.275
0.300
0.325
0.350
0.80
0.82
0.84
0.86
0.945
0.960
0.975
0.990
−1.25
−1.00
−0.75
−0.8
0.0
0.8

Conclusions

Future of cosmology is very exciting and very complex

- Exciting because of the enormous amount of cosmological data from a variety of surveys
- Complex because smart+precise multi-probe and multi-data set analyses are hard
- We need creative research on systematics mitigation, precise error calculation, model building, data inference
- Critical to interface expertise in simulations, observations, analytical modeling, statistical methods