Model selection for the spectrum of light baryons

Michael Doering

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Jefferson Lab

Uncertainty Quantification (UQ) at the Extremes (ISNET-6)

EMMI

8-12 October 2018
TU Darmstadt

Supported by

National Science Foundation

HPC support by JSC grant *jikp07*

JÜLICH FORSCHUNGSZENTRUM

DOE DE-AC05-06OR23177 & DE-SC0016582; NSF PIF 1415459 & CAREER PHY-1452055

Outline

- Brief physics motivation & definition of objective function
- Model selection for determination of
  - Partial waves
  - The baryon spectrum

- Questions:
  - Are the used penalties reasonable? How to choose?
  - Alternative methods?
  - How to deal with systematic errors?

Motivation
Defining the objective function

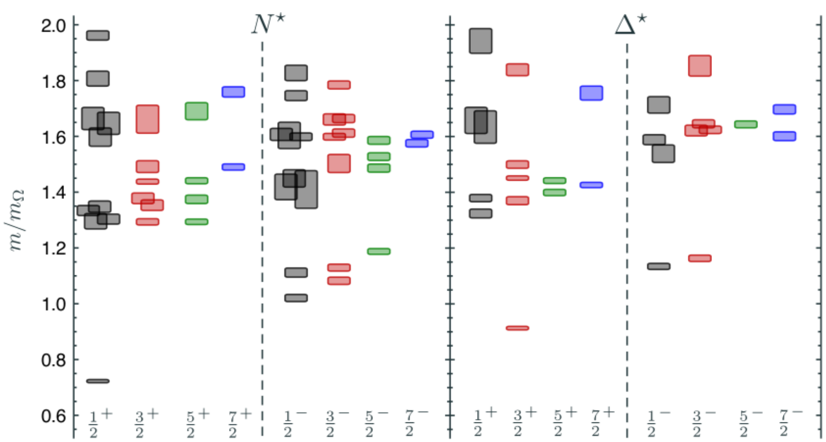
The Missing Resonance Problem

Overview: Int.J.Mod.Phys. E22 (2013) 1330015

- above 1.8 GeV much more states are predicted than observed,

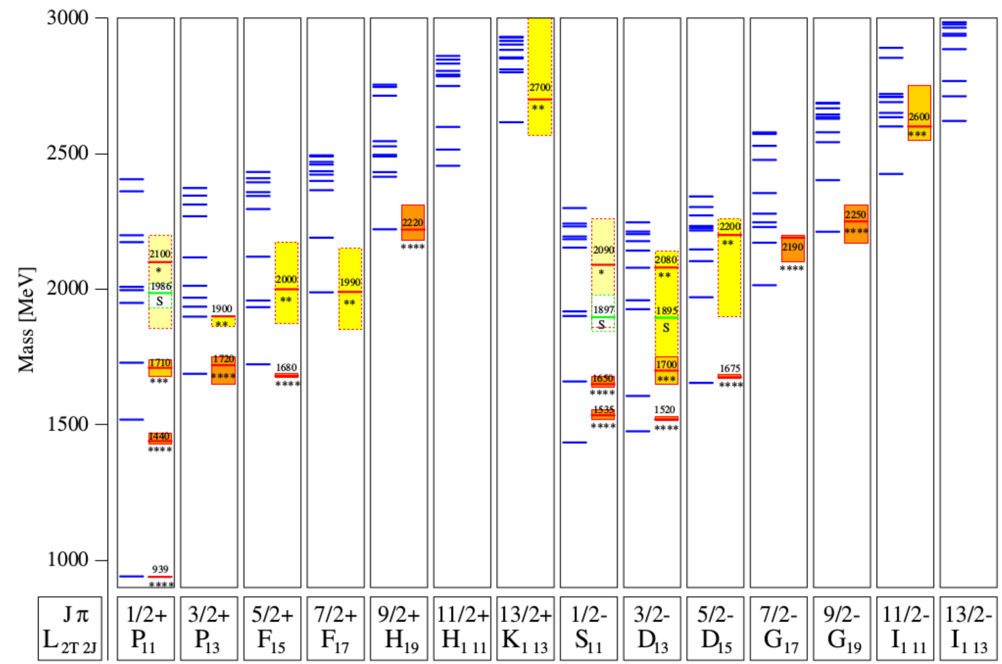
  "Missing resonance problem"

Lattice calculation (single hadron approximation):



[Edwards *et al.*, Phys.Rev. D84 (2011)]

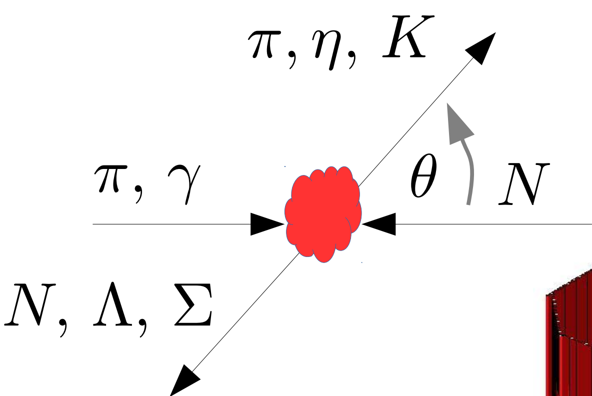
N^* spectrum in a relativistic quark model:

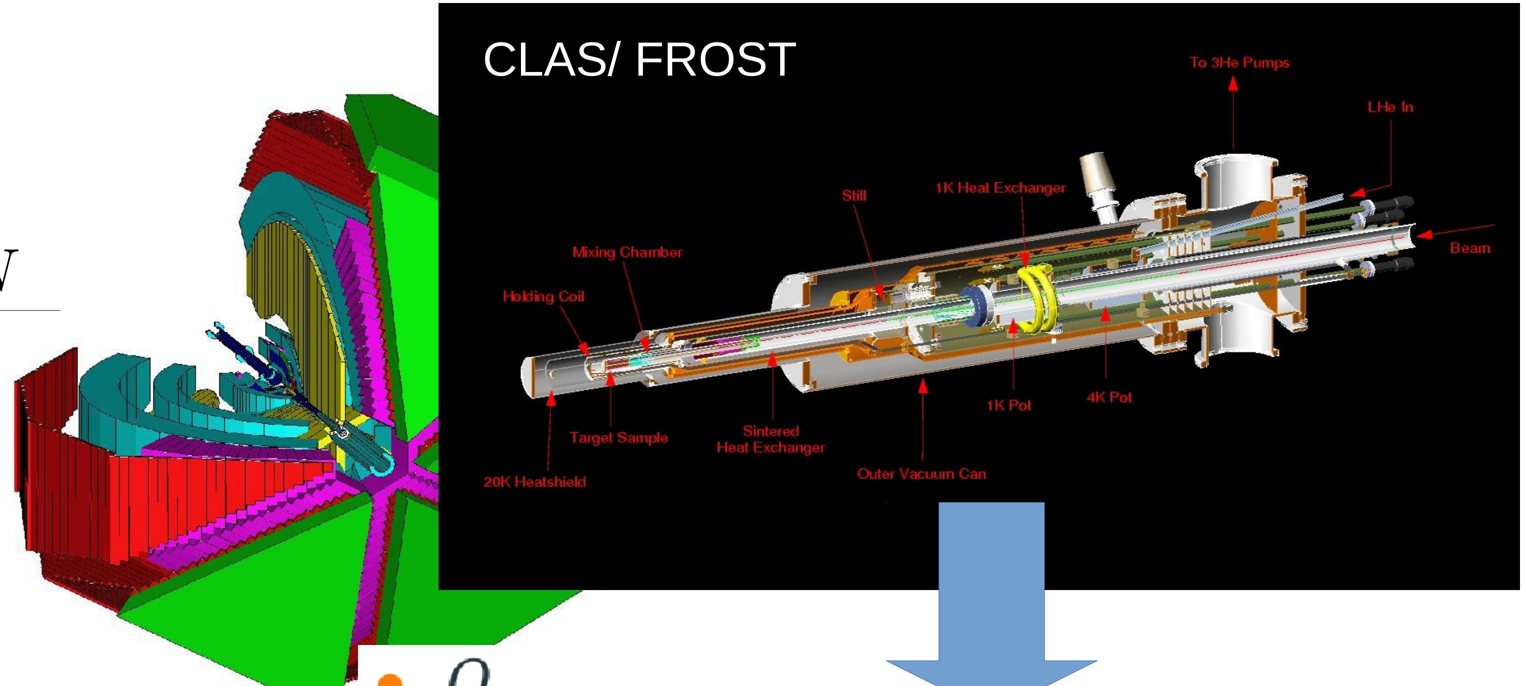


Löring *et al.* EPJ A 10, 395 (2001), experimental spectrum: PDG 2000

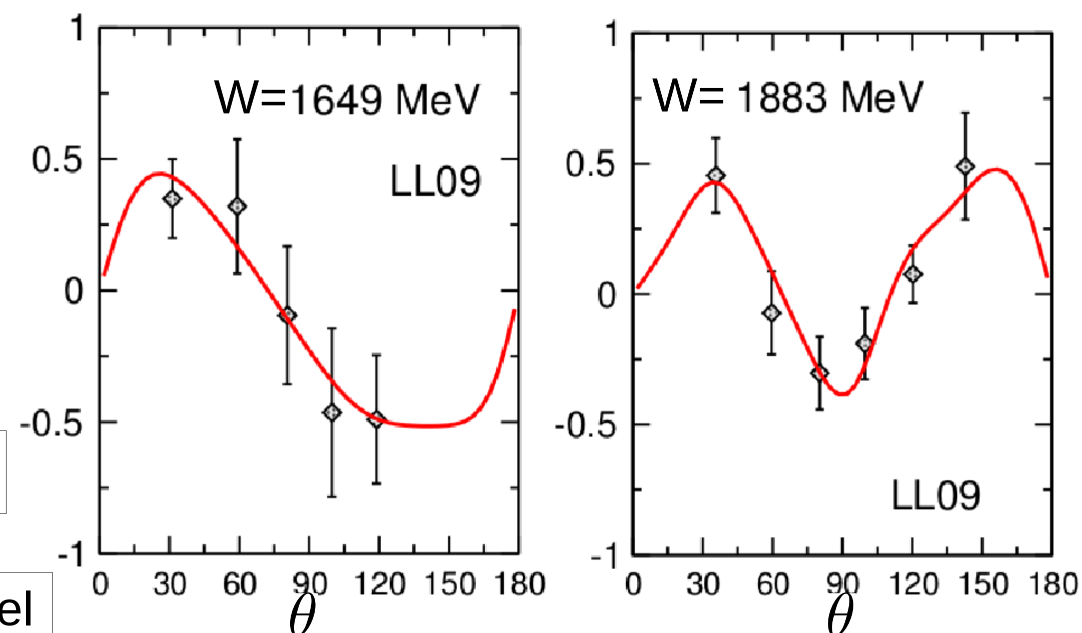
- only 15 established N^* states (PDG 2015)
- ~ 48% of the states have **** or *** status (PDG 1982: 58% with **** or ***)

Workflow (I)





- O_x



LL09: Lleres EPJA 39 (2009)

Independent, known variables:

W – Scattering energy (binned)

θ – Scattering angle (binned)

One collision, different produced states, all from same underlying amplitude: $\pi N,\ \pi\pi N,\ \eta N,\ K\Lambda,\ K\Sigma,\ \ldots$

Coupled-channels, 2- and 3-body unitarity,

50K data points → e.g. Jülich-Bonn model (Rönchen et al.)

Workflow (II): The data

Observable	σ	Σ	T	P	E	F	G	H	T_x	T_z	L_x	L_z	O_x	O_z	C_x	C_z
$p\pi^0$	✔	✔	✓		✓	✓	✓	✓								
$n\pi^+$	✔	✔	✓		✔	✓	✓	✓								
$p\eta$	✔	✓	✓		✔	✓	✓	✓								
$p\eta'$	✔	✓	✓		✓	✓	✓	✓								
$K^+\Lambda$	✔	✔	✔	✔	✓	✓	✓	✓	✓	✓	✓	✓	✔	✔	✔	✔
$K^+\Sigma^0$	✔	✔	✔	✔	✓	✓	✓	✓	✓	✓	✓	✓	✔	✔	✔	✔
$p\omega/\phi$	✔	✓	✓		✓	✓	✓	✓	✔ SDME							
$K^{+*}\Lambda$	✔			✔					SDME							
$K^{0*}\Sigma^+$	✔	✓									✓	✓	SDME			
$p\pi^-$	✔	✔			✔	✓	✓									
$p\rho^-$	✓	✓			✓	✓	✓									
$K^-\Sigma^+$	✔	✓			✓	✓	✓									
$K^0\Lambda$	✔	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
$K^0\Sigma^0$	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
$K^{0*}\Sigma^0$	✓	✓									✓	✓				

$\gamma p \to X$

$\gamma n \to X$

[Slide: V. Burkert]

Workflow (III): Partial-wave decomposition

$$\frac{d\sigma}{d\Omega} = (|g|^2 + |h|^2)\frac{k_f}{k_i} \quad \text{and} \quad P\frac{d\sigma}{d\Omega} = \frac{k_f}{k_i}(gh^* + g^*h)$$

All observables are bilinears of *amplitudes* g, h.

$$g \equiv g(W,\theta),\; h \equiv h(W,\theta) \in C$$

"Observables": different kind of data from same reaction:

$\frac{d\sigma}{d\Omega}(W,\theta)$ "differential cross section"

$P(W,\theta)$ "Recoil polarization"

…. many more

$$g_I = \sum_{J=1/2}^{J_{max}} \frac{(2J+1)}{2\sqrt{k_f k_i}} \left[d^J_{\frac{1}{2}\frac{1}{2}}(\theta)\cos\left(\frac{\theta}{2}\right)\left(\tau_I^{J-} + \tau_I^{J+}\right) + d^J_{-\frac{1}{2}\frac{1}{2}}(\theta)\sin\left(\frac{\theta}{2}\right)\left(\tau_I^{J-} - \tau_I^{J+}\right)\right],$$

$$h_I = -i\sum_{J=1/2}^{J_{max}} \frac{(2J+1)}{2\sqrt{k_f k_i}} \left[d^J_{\frac{1}{2}\frac{1}{2}}(\theta)\sin\left(\frac{\theta}{2}\right)\left(\tau_I^{J-} + \tau_I^{J+}\right) - d^J_{-\frac{1}{2}\frac{1}{2}}(\theta)\cos\left(\frac{\theta}{2}\right)\left(\tau_I^{J-} - \tau_I^{J+}\right)\right]$$

General decompostion into partial waves τ. Now, choose an energy-dependent parametrization:

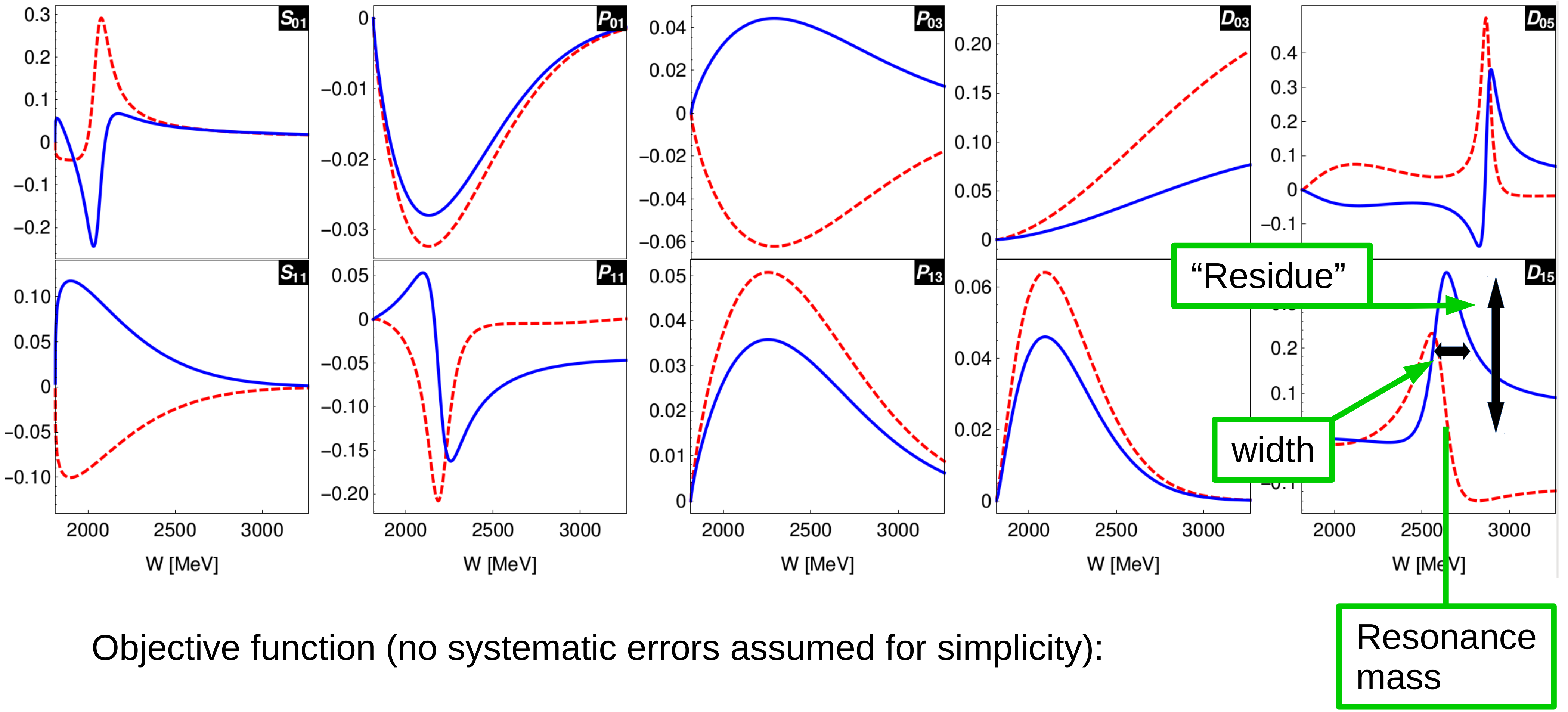
$$\tau(W) = e^{i\phi}\left(\frac{k_f(W)}{\Lambda}\right)^{L+1/2} \times \left(a\, e^{-\alpha^2\left(\frac{k_f(W)}{\Lambda}\right)^2} - x\, e^{i\Phi}\frac{\Gamma/2}{W - M + i\Gamma/2}\right)$$

Meromorphic function of W with 1st -order poles ("resonances") and smooth "background"

[plus branch points and cuts, not of relevance here]

How many resonances, with which mass M, width Γ and residue $xe^{i\Phi}$?

Real (solid) and imgaginary parts (dashed) of different partial waves $\mathcal{T}_i$



Objective function (no systematic errors assumed for simplicity):

$$\chi^2 = \chi^2_{\frac{d\sigma}{d\Omega}} + \chi^2_P + \ldots$$

$$\chi^2_j = \sum_{k_j=1}^{n_j} \frac{(O_{k_j}(\boldsymbol{\beta}) - y_{k_j})^2}{\sigma^2_{k_j}}$$

Motivation:

- Usually, add more and more resonances to the amplitude:
  - The data description improves.
  - Resonances are “seen”.
  - Their * rating in the PDG increases.
- No. Try to avoid resonances at all costs. Try to remove them from fit at all cost.
- Or make them irrelevant by pushing them to large width
  = indistinguishable from background.

→ Make missing resonance problem either worse or pointless

- If you **cannot**: Better signal for a resonance

→ Model selection

Least Absolute Shrinkage and Selection Operator (LASSO)

$$\sum_{i=1}^{n} \frac{(y_i - f(x_i, \beta_j))^2}{\sigma_i^2} + \lambda \sum_{j=1}^{m} |\beta_j|^n$$

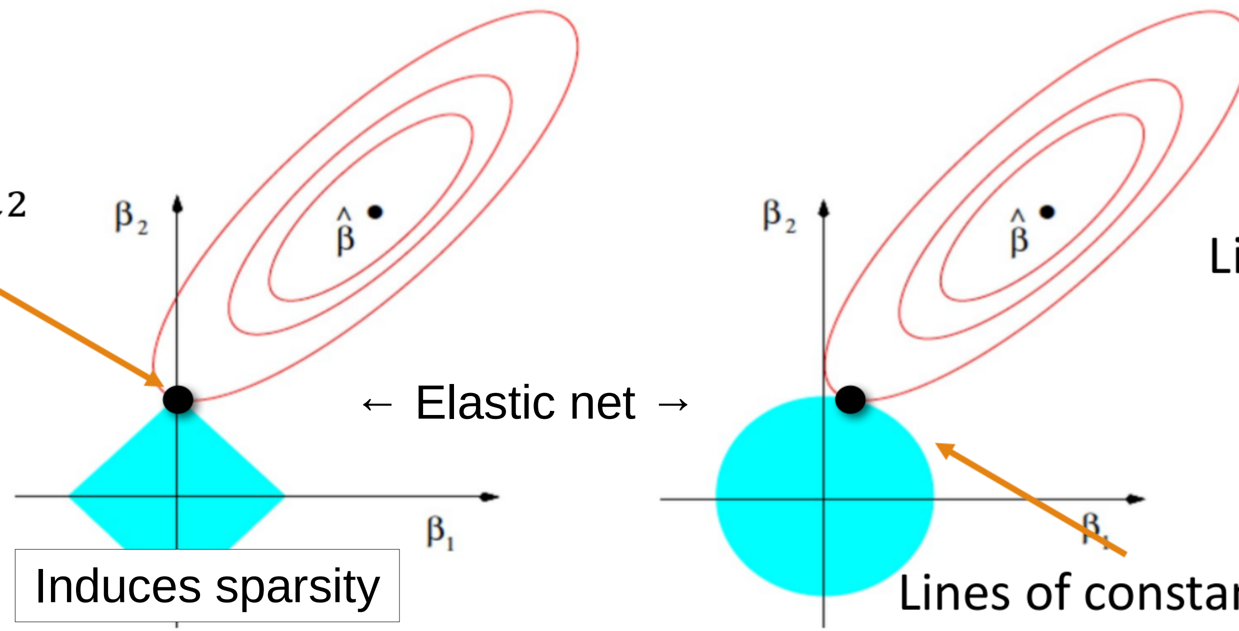
Normal X^2 — Penalty Term

$\hat{\beta}_i$: Best parameters without penalty
$\beta_i = 0$: Best parameters only penalty

LASSO $(n = 1)$

Ridge Regression $(n = 2)$

Simultaneous minimization of χ^2 and Penalty



Lines of constant χ^2

Lines of constant λ

See, e.g.: *The Elements of Statistical Learning*: Data Mining, Inference, and Prediction, T. Hastie, R. Tibshirani, J. Friedman, Springer 2009 second ed.

Regularizations as Bayesian Priors

Ridge regression $\ell_2 \text{norm}$

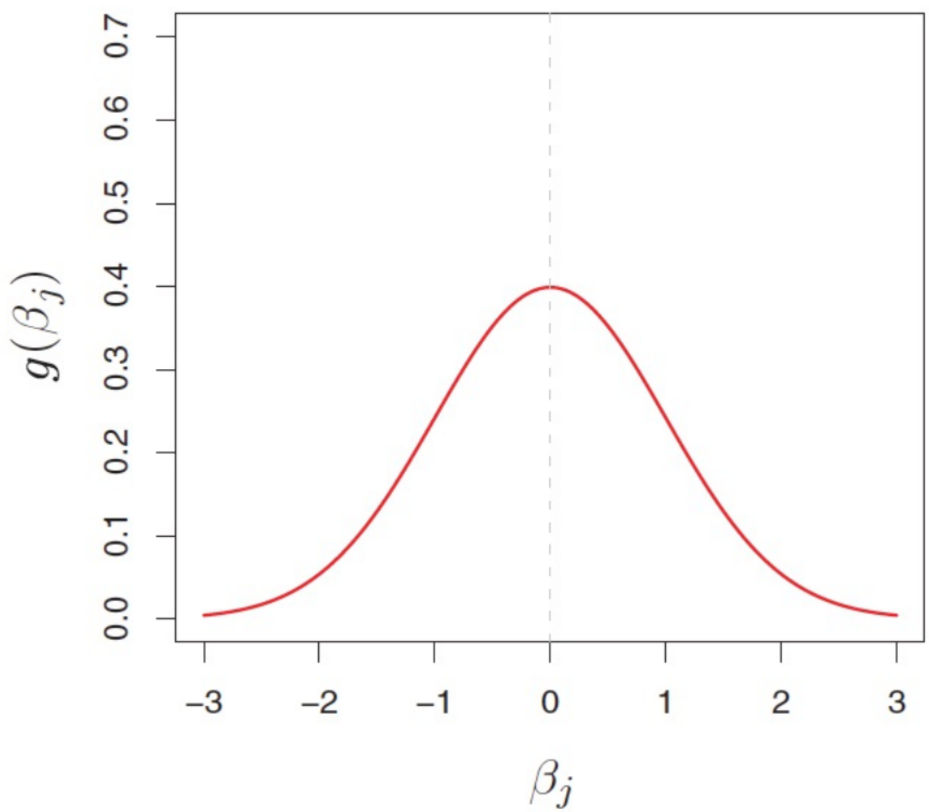
⇕

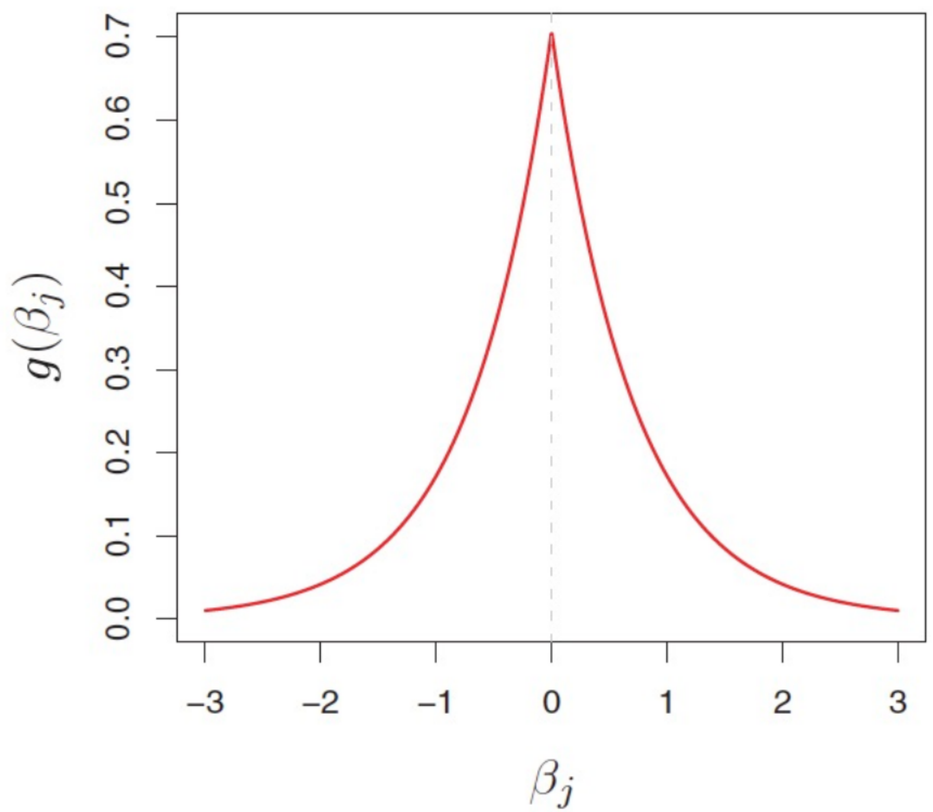
Gaussian prior

LASSO $\ell_1 \text{norm}$

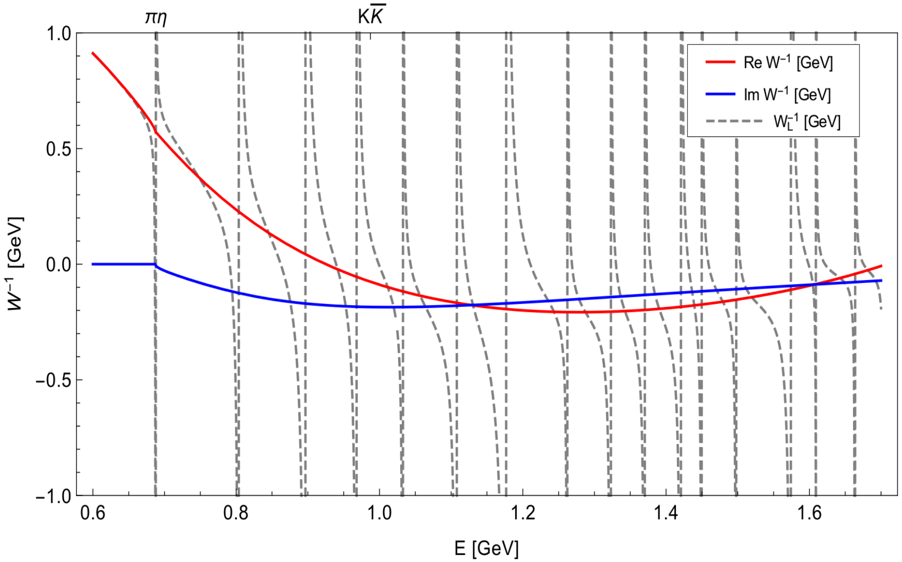
⇕

Laplacian prior





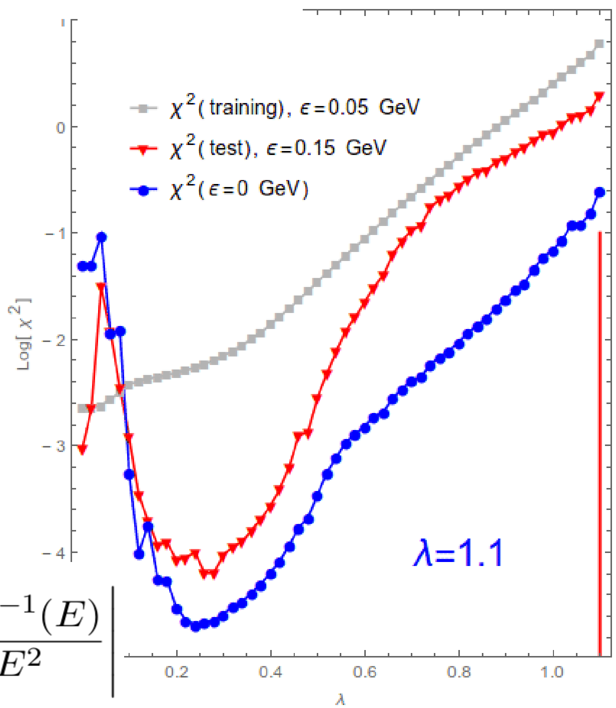
Figures: R. Molina

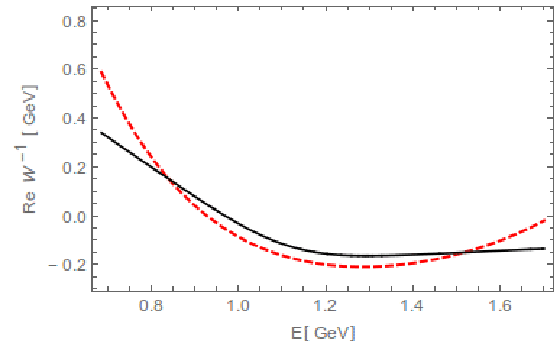


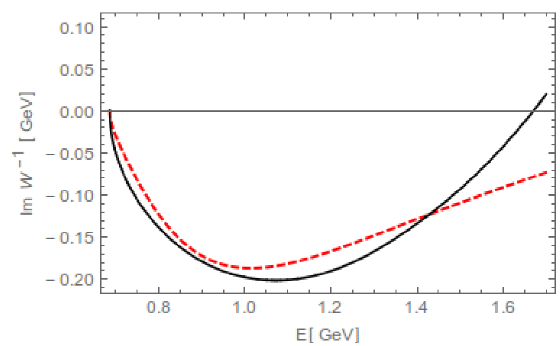
Application: Curve smoothing

$$W^{-1} = \lim_{\epsilon \to 0} \lim_{L \to \infty} W_L^{-1}$$

Physics background:
Infinite volume extrapolation of lattice QCD calculations
[Agadjanov, M.D., Mai, Meißner, Rusetsky, JHEP (2016)]



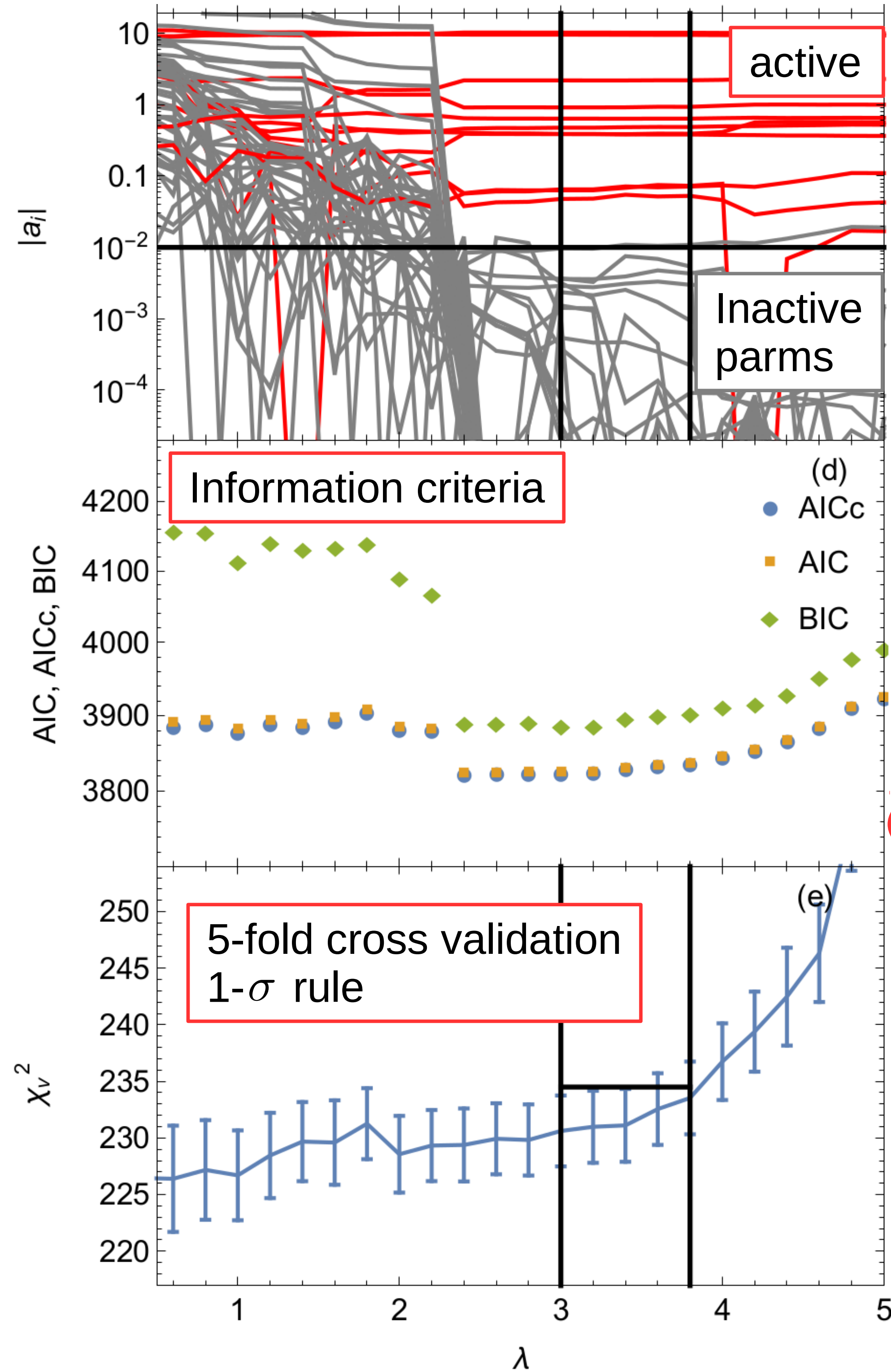


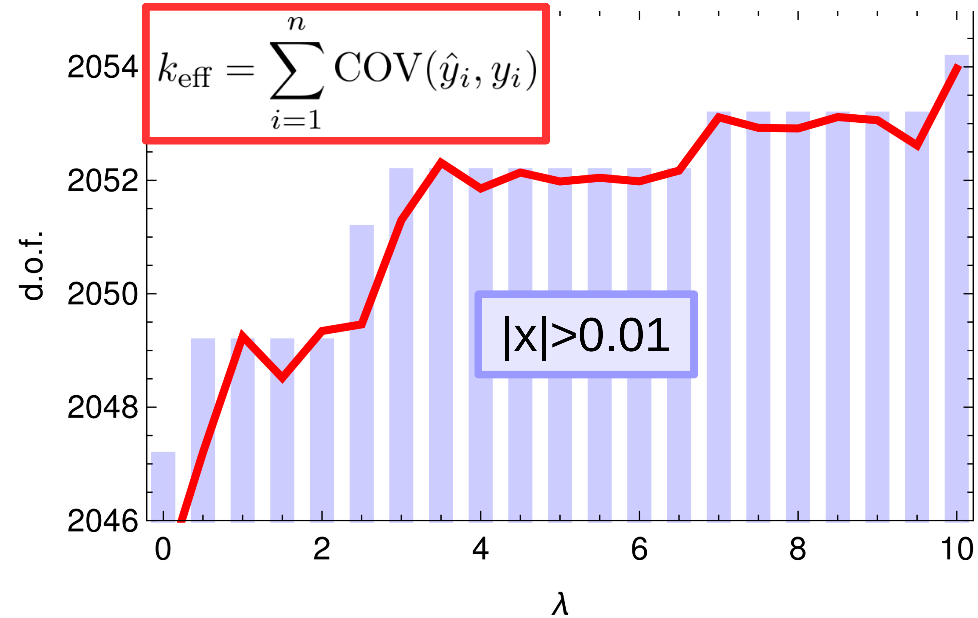


$$P_1(a_j,\, b_j) = \lambda^4 \int_{E_{\min}+i\varepsilon}^{E_{\max}+i\varepsilon} dE \left| \frac{\partial^2 \hat{W}^{-1}(E)}{\partial E^2} \right|$$

Correct Choice of penalization parameter λ through cross validation:

Fit at finite ϵ, validate at different ϵ' ($E \to E + i\epsilon$).





$k \to k_{\text{eff}}$ subsequently

$$AIC = -2 \max \log(L(\hat{\theta}|data)) + 2k = \chi^2 + 2k$$

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

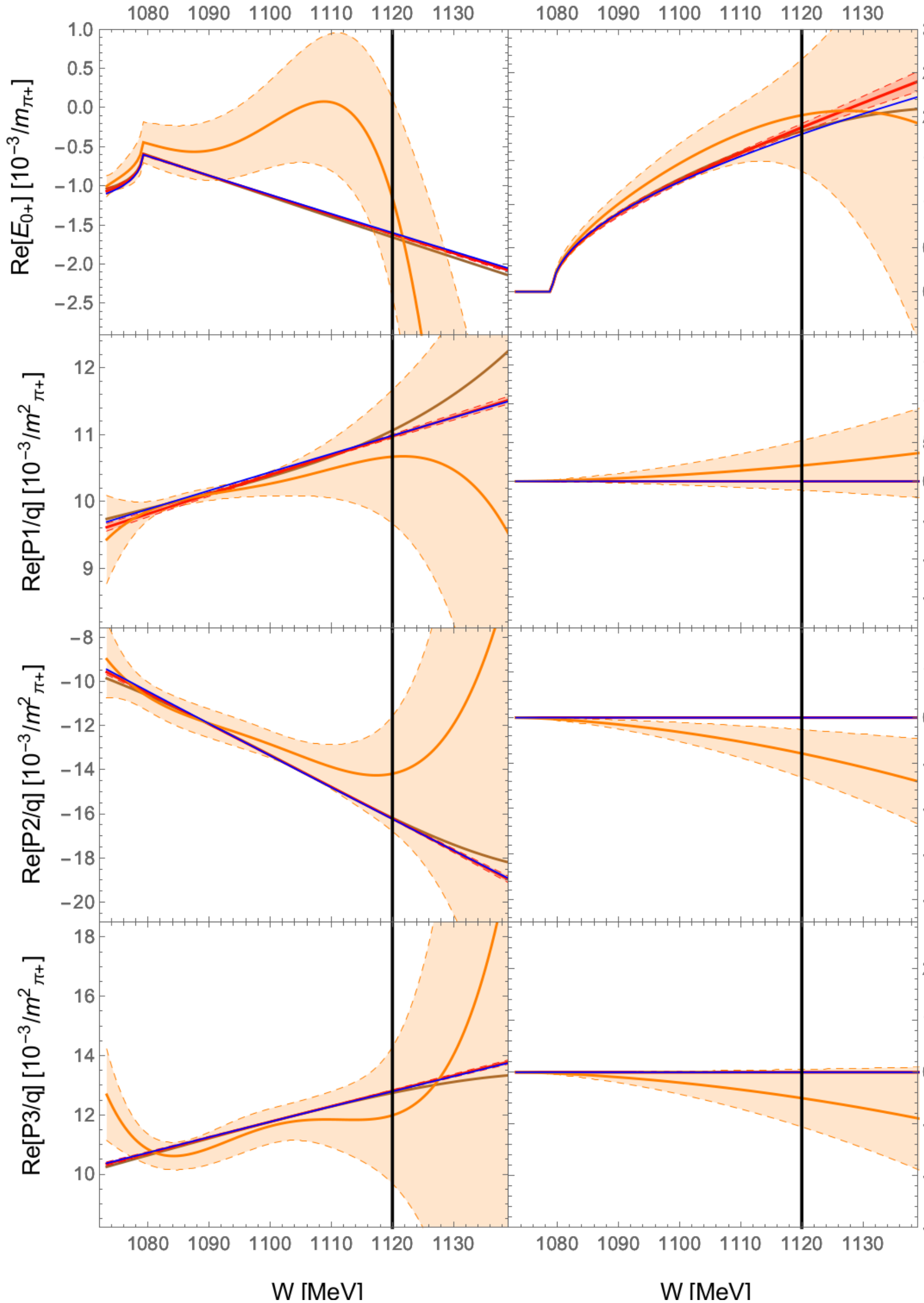
$$BIC = -2 \max \log(L(\hat{\theta}|data)) + 2log(n) = \chi^2 + k\, log(n)$$

Close relation to Bayesian model comparison (here: $n \gg k$)

See, e.g.: Andrew A. Neath, Joseph E. Cavanaugh, DOI: 10.1002/wics.199

Toy Model Results

[J. Landay, M.D., C. Fernandez, B. Hu. R. Molina, PRC 2017]

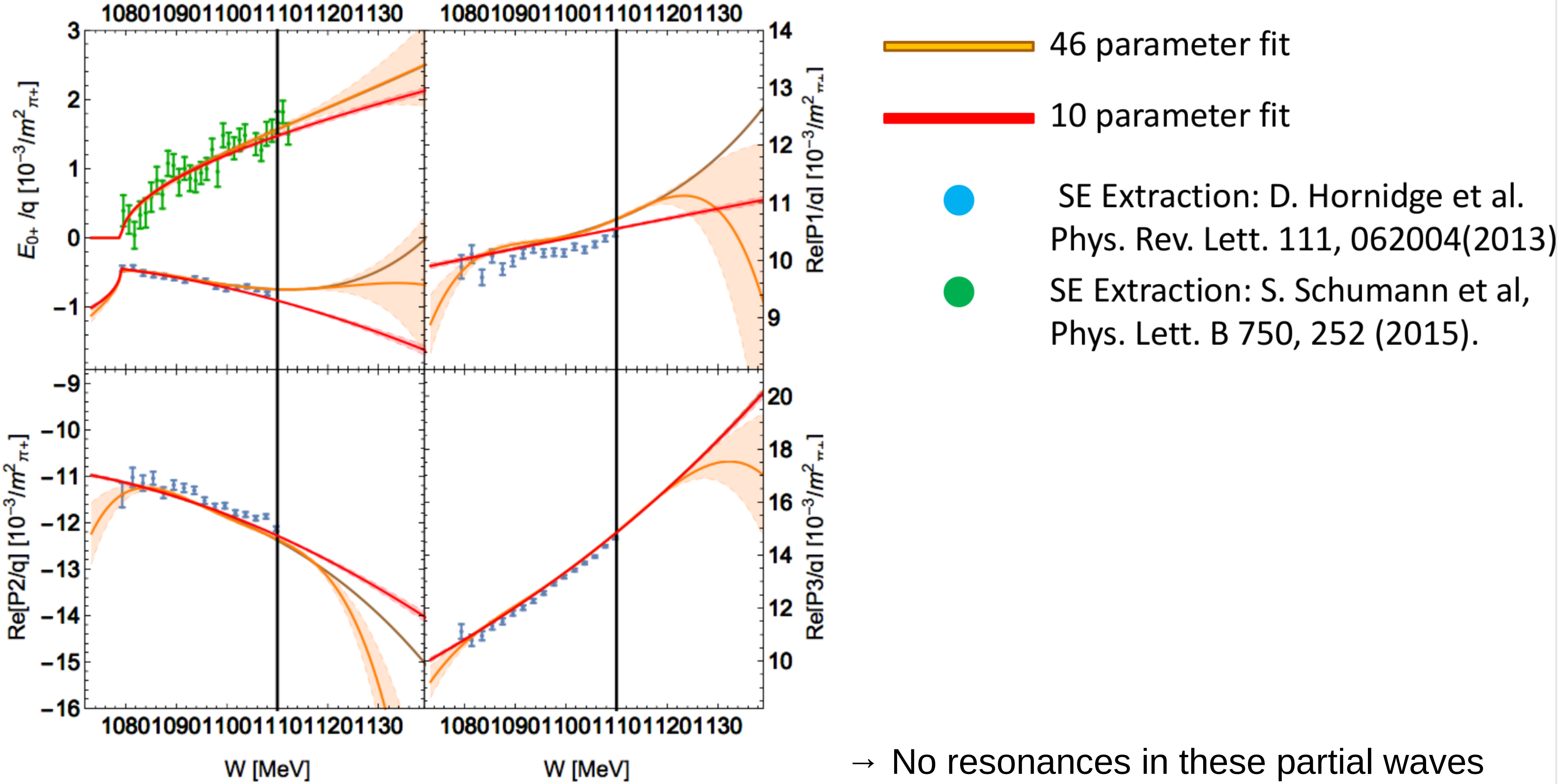


- Generate data from a toy model using a 9 parameter model (2 real S-waves, 1 imaginary S-wave, and 2 real $P_{1,2,3}$ –waves shown in blue
- LASSO (red) eliminates 36 parameters from a 46 parameter fit (orange) and reconstructs the true solution (blue) quite accurately
- LASSO sets all imaginary parts of P-waves and D- waves correctly to 0
- LASSO solution predicts true solution quite accurately beyond the fitted W_{max} =1120 MeV

See also: B. Guegan, J. Hardin, J. Stevens, and M. Williams, JINST 10 (2015)

Model selection with real data $(\gamma p \to \pi^0 p)$

[J. Landay, M.D., C. Fernandez, B. Hu. R. Molina, PRC 2017]



→ No resonances in these partial waves

→ What happens with resonances?

Selection of baryon spectrum

Example: $K^- p \to K\Xi$

[Landay, Mai, M.D., Haberzettl, Nakayama, arXiv: arXiv:1810.00075 [nucl-th]]

Synthetic data tests

- 10 partial waves
- 10 resonance candidates
- Synthetic data with 4 active resonances

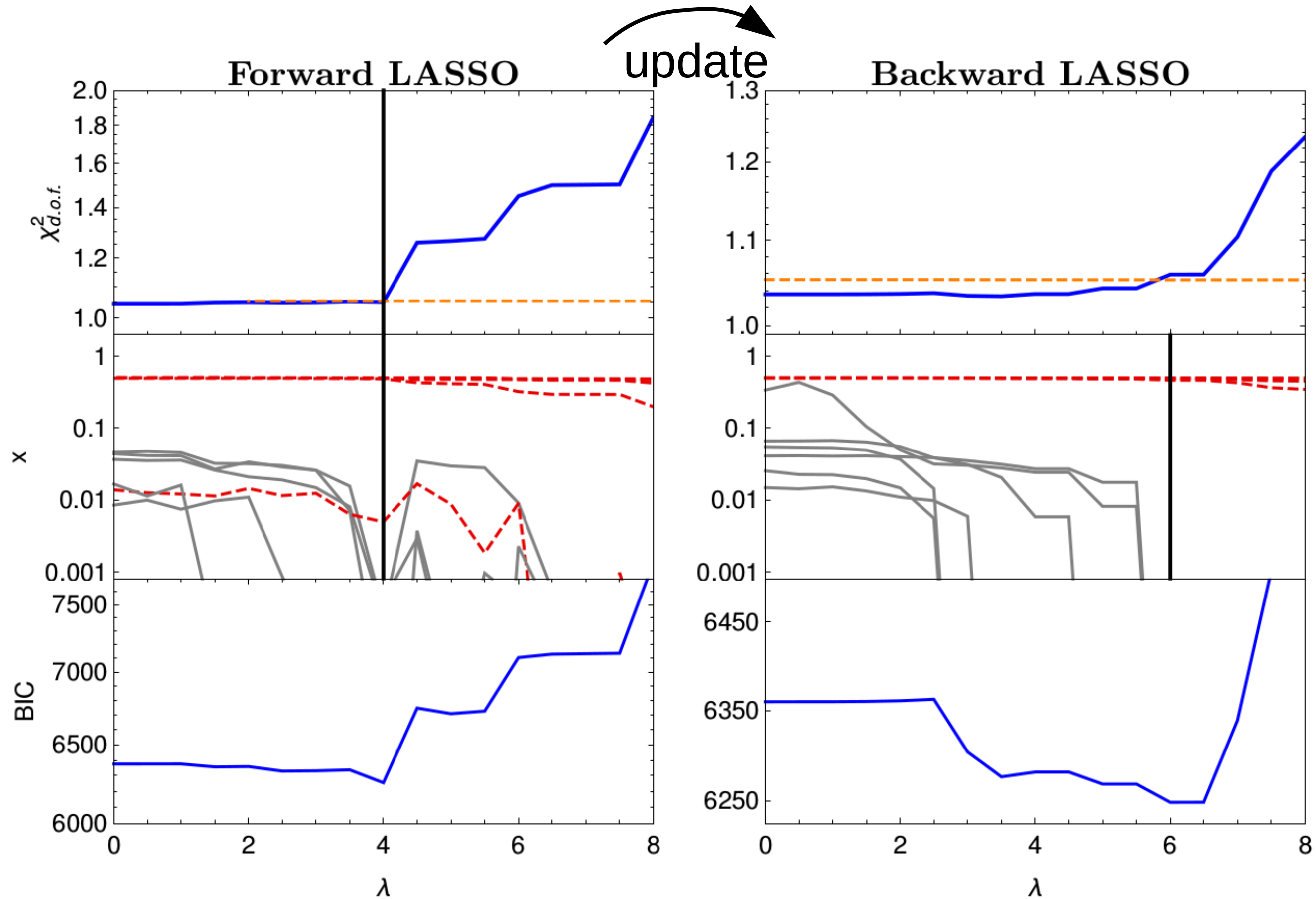
$$\tau(W) = e^{i\phi}\left(\frac{k_f(W)}{\Lambda}\right)^{L+1/2} \times \left(a\, e^{-\alpha^2\left(\frac{k_f(W)}{\Lambda}\right)^2} - \boxed{x}\, e^{i\Phi}\frac{\Gamma/2}{W - M + i\Gamma/2}\right)$$

- Penalty (group LASSO):

$$P_{gr}(\lambda) = \lambda^4 \sum_{i=1}^{i_{\max}} \sqrt{p_i}\,|x_i|$$

Large $\lambda \rightarrow$ small λ

Small $\lambda \rightarrow$ large λ



Selected:
4 active resonances
1 inactive resonance

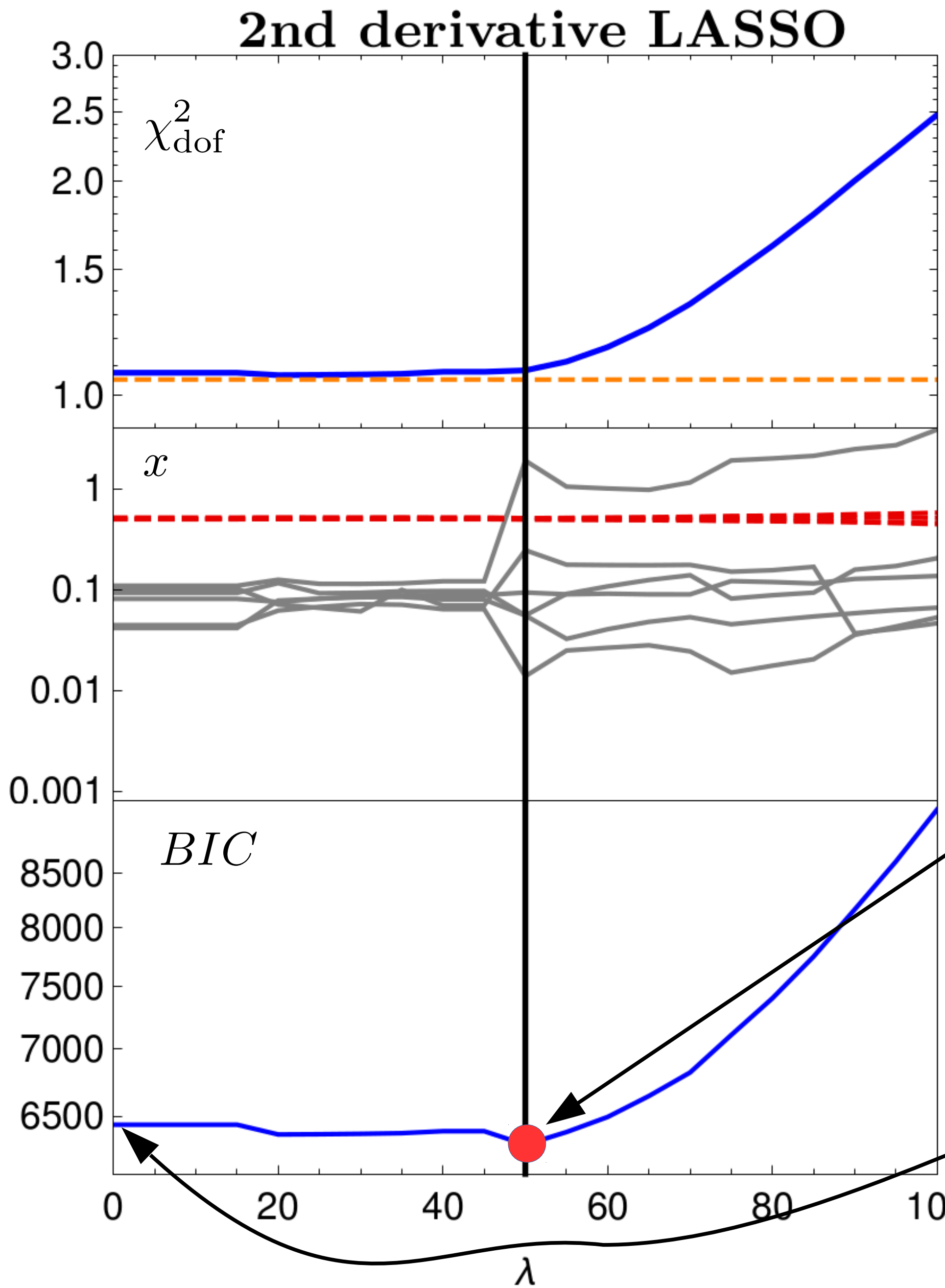
Finds good local minima!

Selected:
4 active resonances
0 inactive resonance

Greediness built in

Another penalty

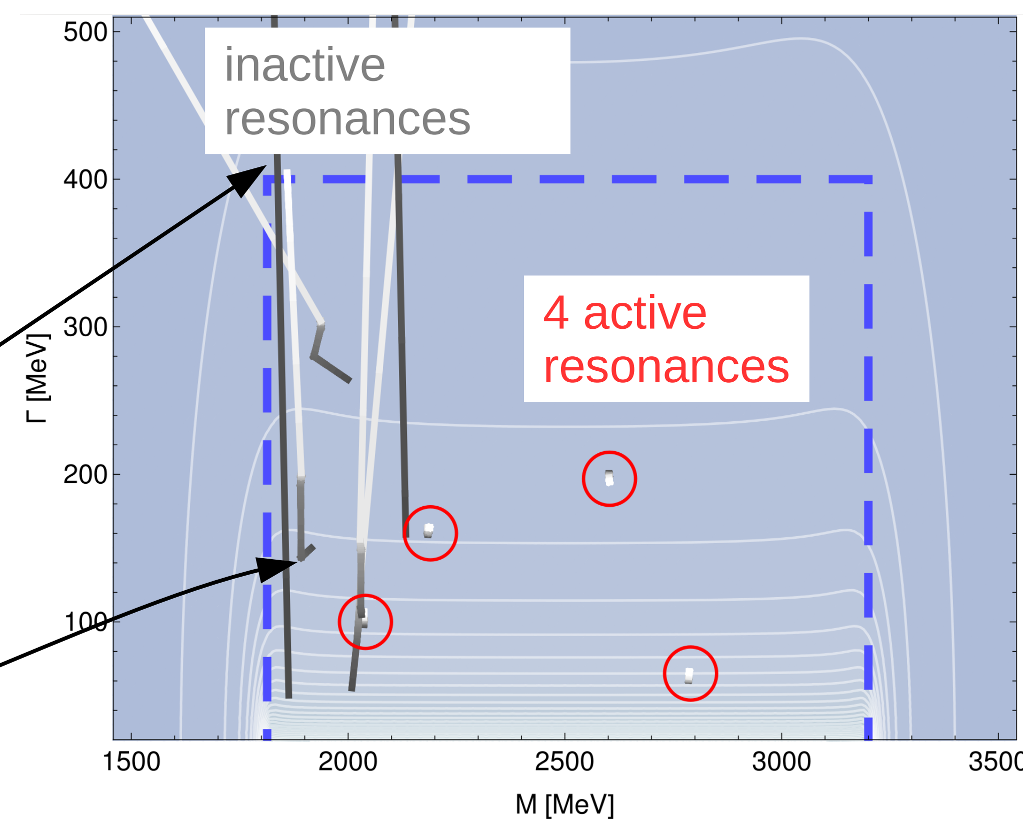
Render resonances pointless by driving out of physical region; less aggressive, less danger of trading variance for bias (type I error)



- Ten partial waves; 10 resonances in Ansatz
- Penalty: $\chi^2 = \chi^2_{\rm stat.} + P$

$$P(\lambda) = \lambda^5 \sum_{i=1}^{10} \frac{\int_{m_K+m_\Xi}^{W_{\rm max}} \left| \frac{\partial^2}{\partial W^2} \tau_i(W) \right|^2 dW}{\int_{m_K+m_\Xi}^{W_{\rm max}} |\tau_i(W)|^2 \, dW}$$

- LASSO picks the 4 correct ones:



Questions

- Do used penalties make sense? How to choose?
- How to deal with systematic uncertainties/outliers?
- Bayesian techniques (The real code is very high-dim & slow)?
- Bias-Variance tradeoff between background and resonances?

M. Williams, DOI: 10.1088/1748-0221/12/09/P09034

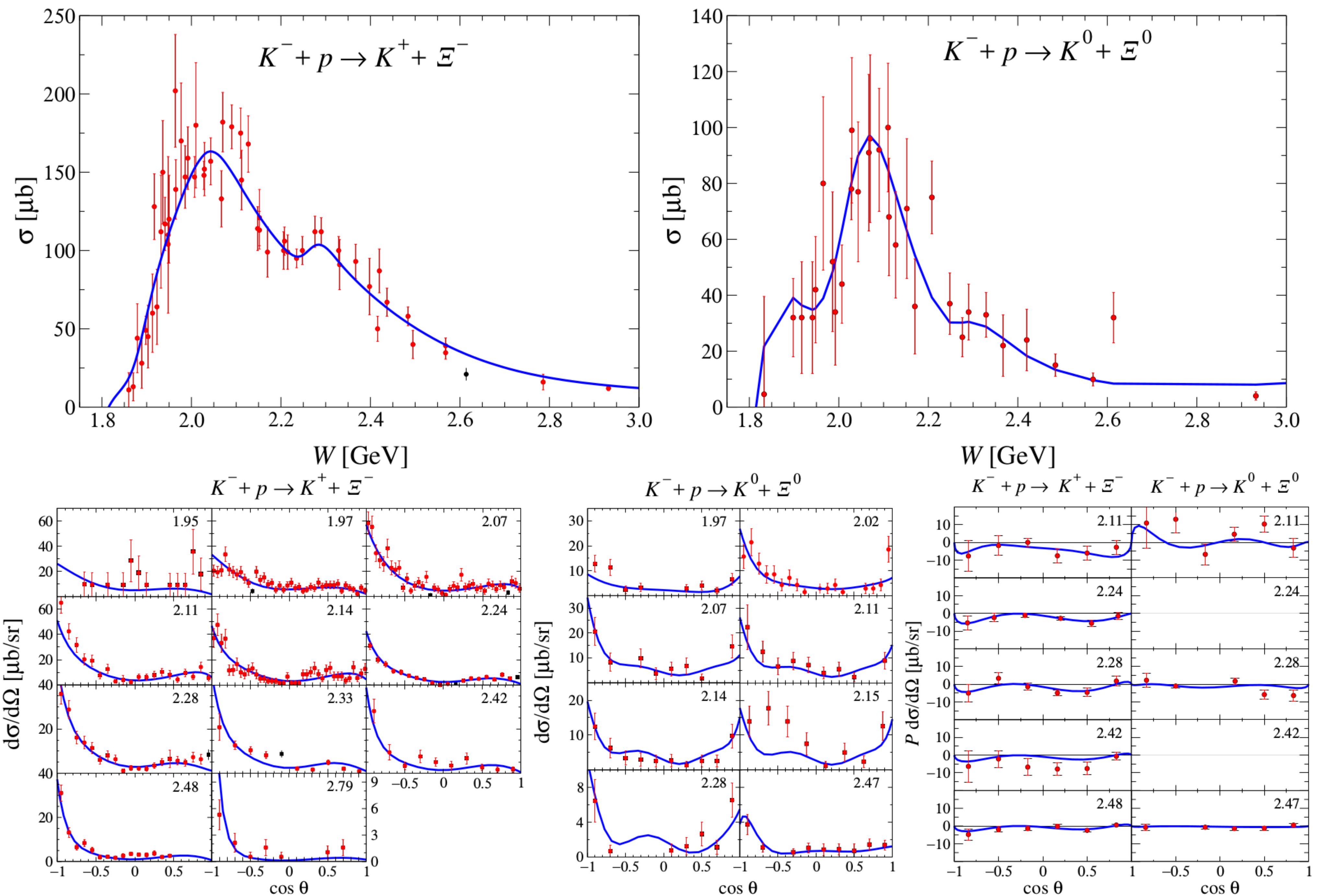
$$\tau(W) = e^{i\phi}\left(\frac{k_f(W)}{\Lambda}\right)^{L+1/2} \times \left(a\, e^{-\alpha^2\left(\frac{k_f(W)}{\Lambda}\right)^2} - x\, e^{i\Phi}\frac{\Gamma/2}{W - M + i\Gamma/2}\right)$$

- Stability selection & random LASSO?

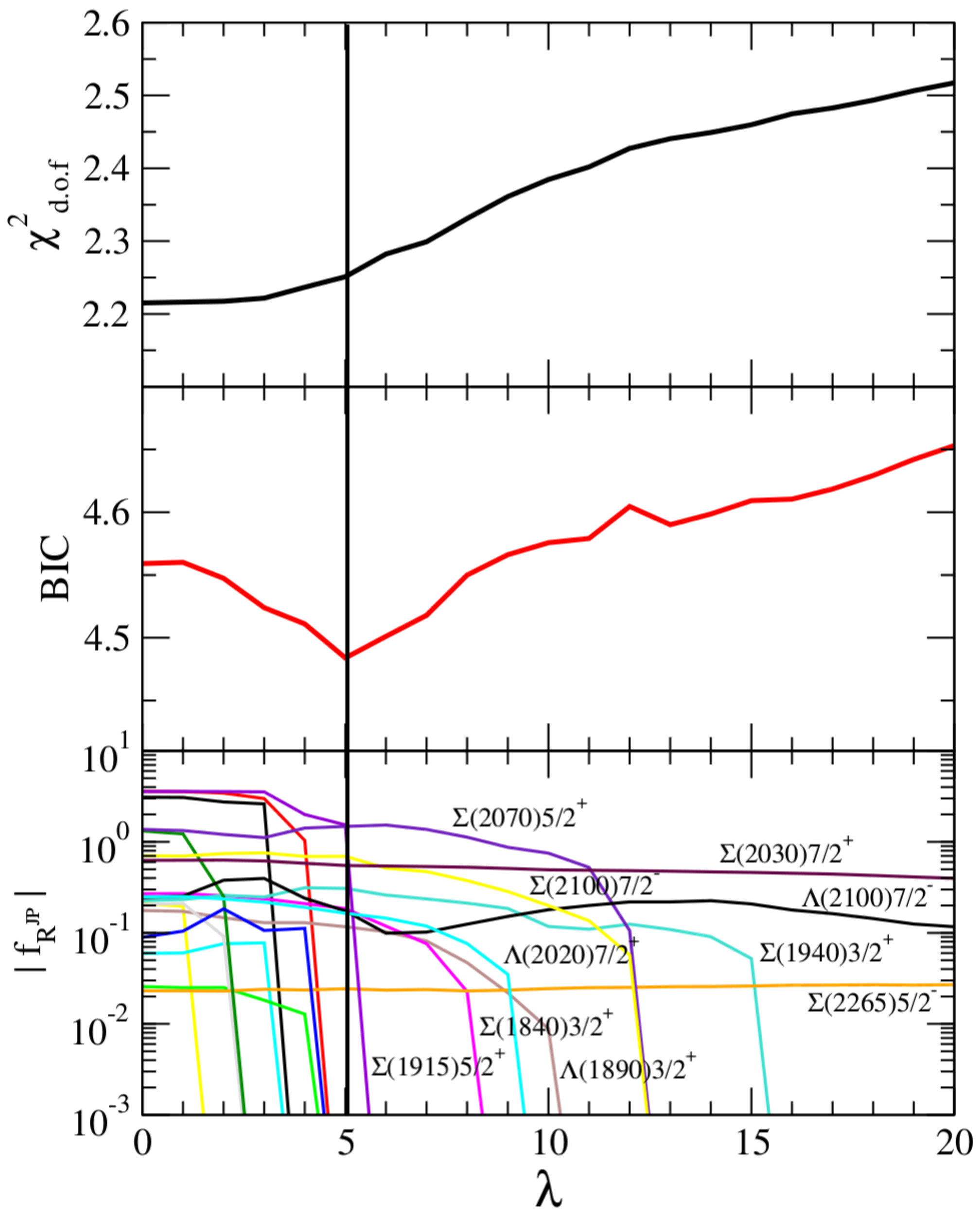
Meinshausen, Bühlmann, DOI: 10.1111/j.1467-9868.2010.00740.x

Spare slides

Real data for $K^- p \to K\Xi$



Real data (II)

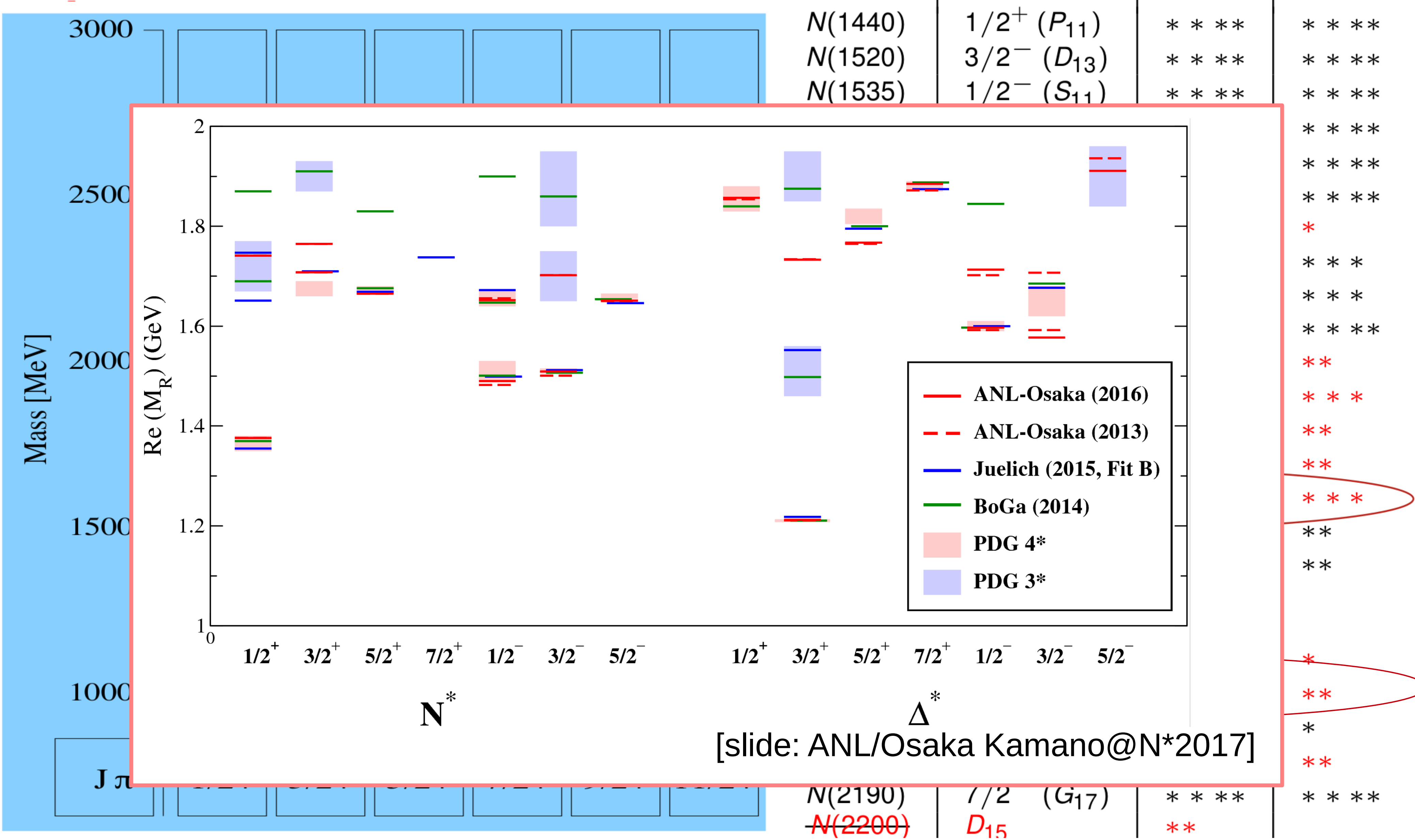


- Data pruning for outliers using smoothness as criterion (10 out of 448 removed)
- Backward selection
- Automatic shutoff LASSO (greedy)
- All 21 resonance candidates from PDG
- Masses and widths fixed to PDG values
- Only x fitted

  →

- 10 out of 21 resonances selected.

Spectrum of N* resonances



- Most new resonances by Bonn-Gatchina group;
- Many from kaon photoproduction

[Slide: V. Crede/Nstar 2017, slight modifications]

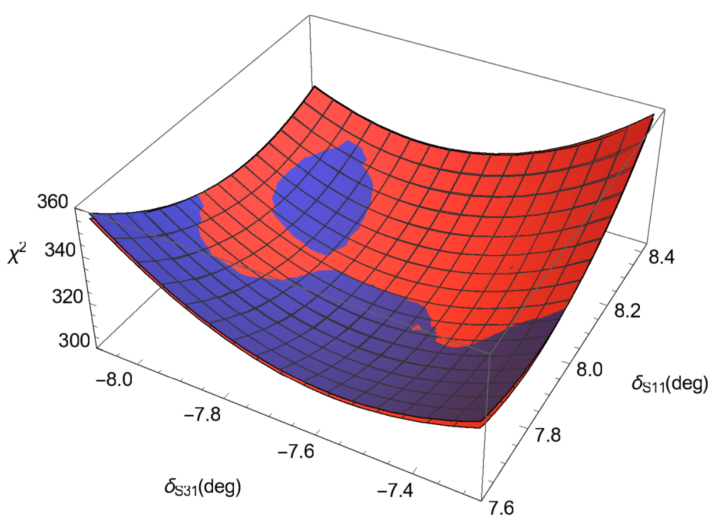
[See also: Crede, Roberts, Rep. Prog. Phys. 76 (2013)]

Toward Data-driven Analyses

[M.D., Revier, Rönchen, Workman, arXiv:1603.07265, PRC 2016]

- Multi-channel analyses to detect faint resonance signals
- All groups use GW/SAID partial waves for $\pi N \to \pi N$
  - The chi-square obtained in fits to single-energy solutions is not related to chi-square of a fit to data → Statistical interpretation of resonance signals difficult.
- Provide online covariance matrices etc. to allow other groups to perform *correlated chi-square* fits.

Slight adaptation of their code allows other groups to obtain a χ^2 (almost) as if they fitted to $\pi N \to \pi N$ directly.

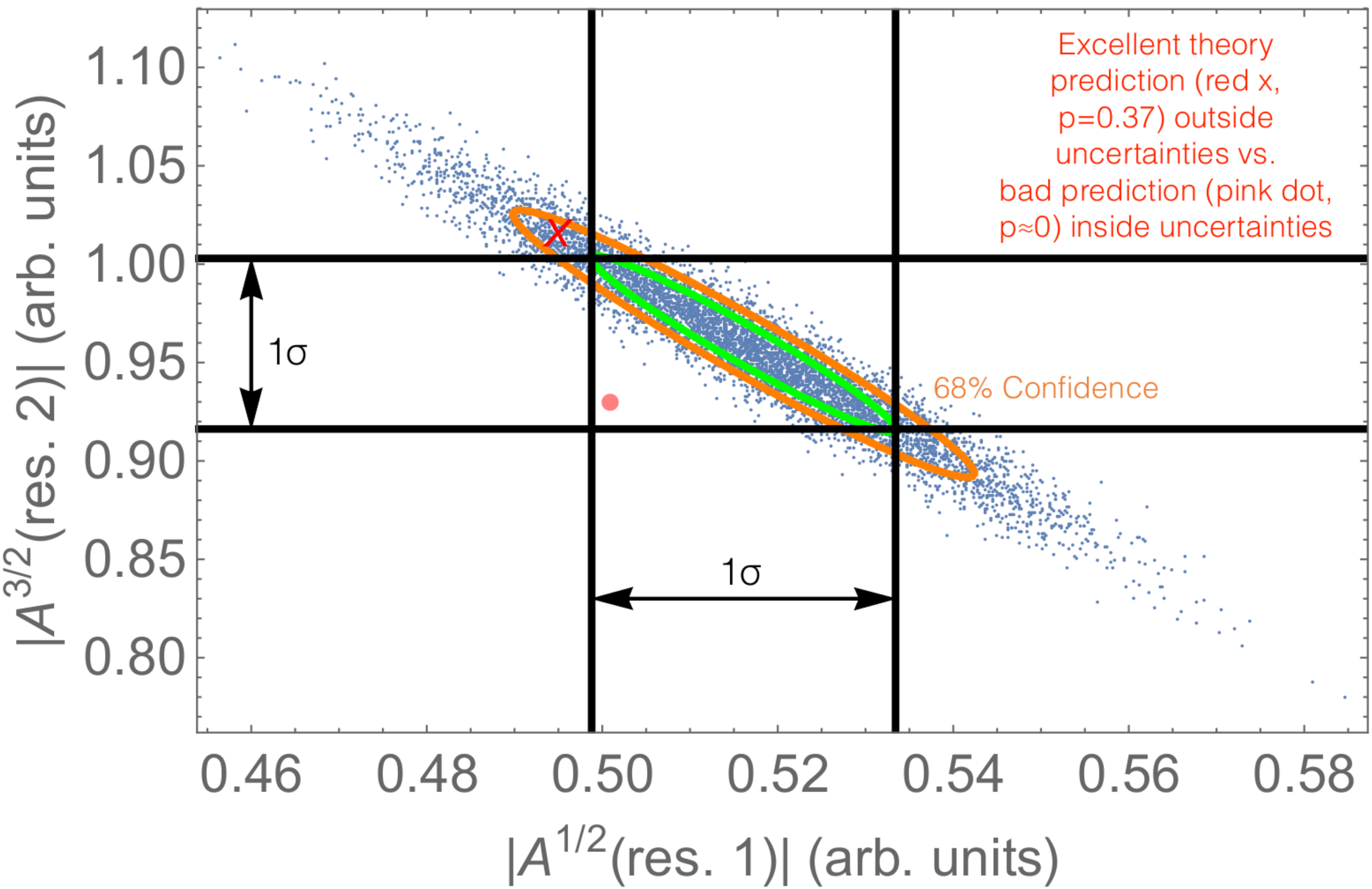


$$\chi^2(\mathbf{A}) = \chi^2(\hat{\mathbf{A}}) + (\mathbf{A} - \hat{\mathbf{A}})^T \hat{\Sigma}^{-1} (\mathbf{A} - \hat{\mathbf{A}}) + \mathcal{O}(\mathbf{A} - \hat{\mathbf{A}})^3$$

Covariance matrices etc. can be downloaded on the SAID and JPAC web pages.

How to quantify the impact of new measurements?

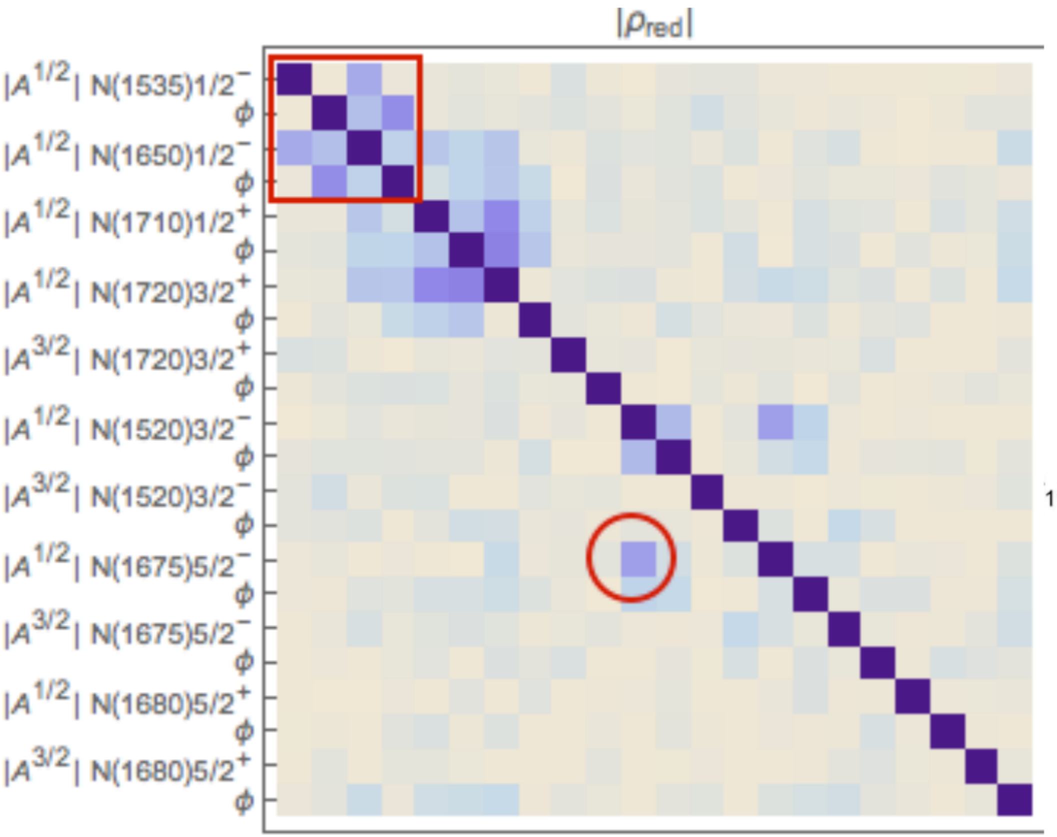
Consider correlations of helicity couplings extracted from experiment

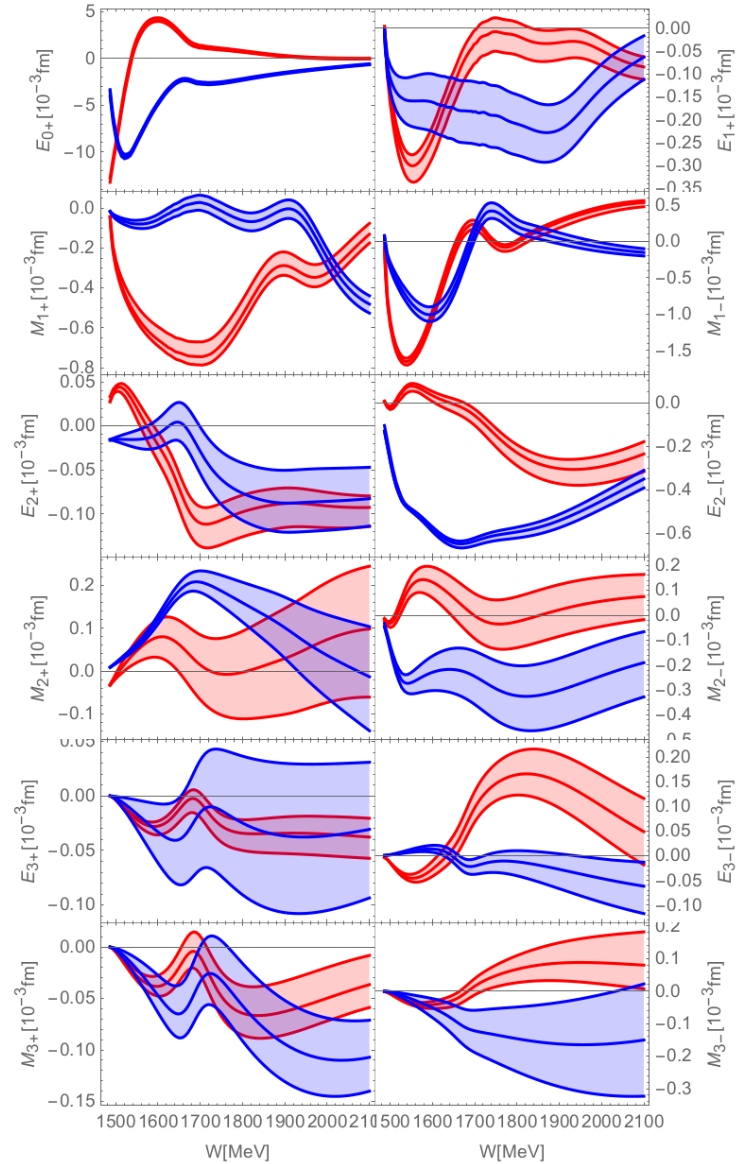


Results from analysis of world data of η photoproduction

[M.D., D. Sadasivan, in preparation]

Here $A = |A|e^{i\phi}$ defined at the resonance pole.





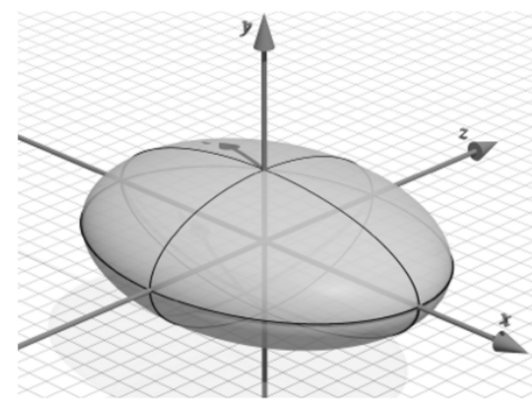
Bulk properties of uncertainties from different data sets

Helicity Coupling	All	No E	No F	No T	No Σ
Number of Data Points	6425	6369	6281	6281	6022
Generalized Variance	0.0494	0.0521	0.1288	0.1239	6.664
$\sqrt{\mathrm{Tr}\, C}$	10.4965	10.51	12.00	11.423	19.85
Multicollinearity	8.173	8.203	9.280	9.5323	10.371
Condition number	133.61	132.10	173.664	164.1	322.66

C=Covariance Matrix

Generalized Variance = Det[C] ~Volume of the Error Ellipsoid

Helicity Coupling	No artificial data	Cx	Cz	Cx and Cz
Number of Data Points	6425	6569	6569	6713
Generalized Variance	0.0494	0.03758	0.0362	0.0132
$\sqrt{\mathrm{Tr}\, C}$	10.4965	10.72	10.487	10.102
Multicollinearity	8.173	7.599	6.770	6.157
Condition number	133.61	112.47	109.69	107.683



- Allows to trace quantitatively the impact of data sets and observables
- Helpful in design of new measurements
- Correlations allow to assess quality of theory predictions