Designing Optimal Experiments: An Application to Proton Compton Scattering

J. A. Melendez* and R. J. Furnstahl†
Department of Physics, The Ohio State University, Columbus, OH 43210, USA

H. W. Grießhammer‡
Institute for Nuclear Studies, Department of Physics, The George Washington University, Washington, DC 20052, USA§
Department of Physics, Duke University, Box 90305, Durham NC 27708, USA and High Intensity Gamma-Ray Source, Triangle Universities Nuclear Laboratories, Box 90308, Durham NC 27708, USA

J. A. McGovern¶
School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

D. R. Phillips**
Department of Physics and Astronomy and Institute of Nuclear and Particle Physics, Ohio University, Athens, OH 45701, USA

M. T. Pratola††
Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

(Dated: March 3, 2021)

Interpreting measurements requires a physical theory, but the theory's accuracy may vary across the experimental domain. To optimize experimental design, and so to ensure that the substantial resources necessary for modern experiments are focused on acquiring the most valuable data, both the theory uncertainty and the expected pattern of experimental errors must be considered. We develop a Bayesian approach to this problem, and apply it to the example of proton Compton scattering. Chiral Effective Field Theory (χEFT) predicts the functional form of the scattering amplitude for this reaction, so that the electromagnetic polarizabilities of the nucleon can be inferred from data. With increasing photon energy, both experimental rates and sensitivities to polarizabilities increase, but the accuracy of χEFT decreases. Our physics-based model of χEFT truncation errors is combined with present knowledge of the polarizabilities and reasonable assumptions about experimental capabilities at HIγS and MAMI to assess the information gain from measuring specific observables at specific kinematics, *i.e.* to determine the relative amount by which new data are apt to shrink uncertainties. The strongest gains would likely come from new data on the spin observables Σ_{2x} and $\Sigma_{2x'}$ at $\omega \simeq 140$ to 200 MeV and 40° to 120°. These would tightly constrain $\gamma_{E1E1} - \gamma_{E1M2}$. New data on the differential cross section between 100 and 200 MeV and over a wide angle range will substantially improve constraints on $\alpha_{E1} - \beta_{M1}$, γ_π and $\gamma_{M1M1} - \gamma_{M1E2}$. Good signals also exist around 160 MeV for Σ_3 and $\Sigma_{2z'}$. Such data will be pivotal in the continuing quest to pin down the scalar polarizabilities and refine understanding of the spin polarizabilities.

I. INTRODUCTION

Six low-energy parameters known as polarizabilities characterize the response of the nucleon to low-frequency light: the electric and magnetic dipole polarizabilities α_{E1} and β_{M1}, and four spin polarizabilities γ_i; see, *e.g.*, refs. [1–3] for recent reviews. Despite their fundamental importance to understanding the proton and neutron, the value of only one combination is known with better than 2% accuracy, with current uncertainties for the rest varying from 10% to more than 100%. Most recent values and uncertainties are collected in ref. [4] and references therein, and summarized in table III below. Recent advances in Chiral Effective Field Theory (χEFT) have enabled precise quantitative predictions of Compton scattering that take the polarizabilities as inputs [1, 5–8].[1] This new ability to precisely trace the impact of these fundamental nucleon-structure constants on experimental observables is opportune. It comes at a time when photon facilities of unprecedented luminosity and sensitivity are now available. These concurrent developments have inspired new experimental campaigns to refine knowledge of nucleon polarizabilities; see, for example, recent overviews in refs. [9–13].

* melendez.27@osu.edu
† furnstahl.1@osu.edu
‡ hgrie@gwu.edu
§ Permanent address
¶ judith.mcgovern@manchester.ac.uk
** phillid1@ohio.edu
†† mpratola@stat.osu.edu

[1] In the χEFT we are using, the lowest-lying nucleonic resonance, the $\Delta(1232)$, is retained as an explicit degree of freedom.

But not all measurements are created equal, and beam time is not cheap. In Nuclear Physics, as in many other advanced disciplines, the costs of running an experiment include not only the workforce, time and money invested, but also the opportunity cost of measurements that could have been carried out with those same resources but were not. Thus, when planning an experiment, it is important to consider which data are most likely to provide the largest information gain.

How, then, does one assess the impact of measurements that have yet to be made? One possibility is to simulate various possible experimental scenarios and compute the extent to which each improves constraints on theory parameters. For example, ref. [14] recently investigated the ability of data on proton- and neutron elastic scattering from nuclei to constrain optical-model parameters. In the context of Compton scattering from the proton and neutron, ref. [6] assessed the sensitivity of many different observables to the polarizabilities, but provided only an initial theory perspective on those sensitivities. Neither of these papers used a single quantitative measure to choose between competing experimental designs. Such a measure should assess competing experimental designs in light of existing data and the feasibility of new data, and include a rigorous assessment of theory uncertainties.

Here, we argue for the framework of Bayesian statistics—Bayesian experimental design, in particular—as an enlightened way forward. Experimental design, in the context of statistics, provides insight into the allocation of scarce resources that have alternative uses. This is a highly practical field of study, with applications including engineering, biology, environmental processes, computer experiments, and psychology [15–25]. One begins by encoding as a utility function the goals of the experiment and the constraints inherent in carrying it out. Then, one considers the range of possible future experimental measurements, and computes the expected utility for each. The optimal design is then the one that maximizes the expected utility function. In this context, a *design* refers to the choice of observable, the kinematic points at which to measure it, and the relative allocation of beam time. With Bayesian experimental design, we incorporate *a priori* knowledge of the parameters we wish to constrain—in this case the nucleon polarizabilities—via Bayesian priors. The utility function must also account for the accuracy future experiments will reasonably achieve, and factor in the kinematic regimes where collecting data will likely be excessively difficult.

As we emphasize in this work, accounting for theory uncertainties in such an analysis is essential for a proper assessment of the optimal design. All models are wrong, but some are useful [26]. Still fewer are wrong in a way that is useful. Effective Field Theory calculations are carried out up to a particular order in a systematically improvable expansion. This predictable character of an EFT's uncertainty permits quantitative answers regarding this trade-off between increased experimental sensitivity and decreased theoretical accuracy. We address this trade-off using the physics-based Bayesian machine learning model proposed in ref. [27]. The known order-by-order EFT predictions allow the algorithm to learn the convergence pattern and how it is correlated in kinematic space, and then to use that pattern to formulate a statistical model of the truncation error, thus ensuring that high-energy data are not over-weighted.

χEFT incorporates all the physics of Compton scattering at photon energies between 0 and about 300 MeV. According to ref. [28], the χEFT Compton amplitude at a given order in the EFT expansion has a theory truncation error approximately proportional to $(m_\pi/\Lambda_b)^{\nu/2}$. Here, Λ_b is the breakdown scale of the theory, and ν depends on, but is not necessarily equal to, the order of the calculation (cf. sec. III below). A naïve application of experimental design that ignored the theory uncertainty might suggest running experiments at energies which are so high that χEFT is unreliable. That would lead to experiments with high precision, whose information content is, however, very small, wasting scarce human and financial resources.

Although we devote considerable space to the explanation of χEFT's theoretical errors, it is of course mandatory to also account for experimental realities. Our recommended experimental design should not involve kinematics at which measurements are notoriously difficult. Difficulties can arise in two different directions. The first is that beam time is limited. We account for that experimental constraint by considering three levels of possible precision of Compton data. These are intended to cover a range of plausible future experiments. Since these results correspond to different numbers of photons on target, they give us a sense of how knowledge of the different polarizabilities will scale with beam intensity and experimental run time. The second issue is that physical limitations can make it difficult to place detectors at particular locations, or to run a certain machine at specific energies. While these constraints are probably best assessed using a facility-specific factor in the utility function, in this first study we account for them crudely by precluding designs involving photon scattering angles where physical limitations will make it difficult to place detectors.

One of the key aspects of our experimental-design approach—or indeed of any Bayesian analysis—is that the underlying assumptions must be stated clearly and quantitatively. In our case, all results are conditional upon the use of χEFT to extract polarizabilities. We see this as a virtue of a Bayesian approach: one is forced to be explicit about one's assumptions, and acknowledge that any conclusions drawn are contingent upon them. In this spirit we also point out that a full experimental-design analysis would begin with an extraction of the polarizabilities from extant Compton data that uses the Bayesian framework adopted here—including its model of theory errors. This would establish the present knowledge of the polarizabilities, *i.e.* what is known about them before any new experiments are performed. We

do not perform such an analysis in this work. Instead we use previous determinations of polarizabilities in ChiEFT from Refs. [4, 5] to form the prior that encodes present knowledge of the polarizabilities. Our experimental-design conclusions are thus indicative, not definitive; our main goal is to show the power of Bayesian experimental design.

In the interest of accessibility to all readers, we try to ensure the sections listed below are self-contained. For example, an understanding of the EFT truncation model should not be essential to understand the experimental design, or the results of the analysis. We begin in sec. II by recounting the relevant facts of nucleon Compton scattering. Next, sec. III describes the important results of the Bayesian methodology employed in this work, including the model of EFT truncation errors and experimental design, adapted to χEFT for Compton scattering. These methods are applied to Compton scattering on a proton for various choices of experimental goals in sec. IV. Finally we conclude in sec. V. We provide details of our derivations in app. A and details of the truncation error model in app. B. Results for a different prior on the polarizabilities, for different levels of experimental precision and for the case of neutron Compton scattering are reserved for the Supplemental Material: app. S. We provide all data and codes needed to reproduce our results [29].

II. BASIC FACTS OF NUCLEON COMPTON SCATTERING

We start with an enumeration of those aspects of Compton scattering on the nucleon relevant for this presentation, to remind experts and introduce the minimal necessary vocabulary for non-experts. Motivations, context and details can be found in refs. [1, 4–6] and elsewhere.

The primary physical degrees of freedom probed in Compton scattering up to about 350 MeV are nucleons, pions and the $\Delta(1232)$ resonance. The relative importance of these changes with energy; at low energies, the process is dominated by the Born terms: a point-nucleon with anomalous magnetic moment, plus the π^0 t-channel coupling. As the energy increases, effects from the pion cloud around the nucleon become apparent, and finally the Delta resonance dominates over all other effects. In the context of an EFT, this means that the power counting changes with energy, something which is explained in more detail in sec. III B 2.

At all these energies, the polarizability contributions are well-described by six dipole polarizabilities which are labeled by the multipolarities of the incoming and outgoing electromagnetic field. In ref. [6], the following linear combinations were identified as most convenient for exploring sensitivities while exploiting the best available prior knowledge: the scalar (dipole) polarizabilities in the combinations

$$\alpha_{E1} \pm \beta_{M1}\,, \tag{1}$$

and the mutually orthogonal spin-polarizability combinations

$$\gamma_{0,\pi} \equiv -(\gamma_{E1E1} \pm \gamma_{M1M1} + \gamma_{E1M2} \pm \gamma_{M1E2}) \tag{2}$$

$$\gamma_{E-} \equiv \gamma_{E1E1} - \gamma_{E1M2} \tag{3}$$

$$\gamma_{M-} \equiv \gamma_{M1M1} - \gamma_{M1E2}\,. \tag{4}$$

These combinations map onto tight constraints on $\alpha_{E1} + \beta_{M1}$ and γ_0 from sum rules. For the proton, these have error bars which are better than those from direct Compton experiments.

Polarizabilities are fundamental hadron properties. The scalar polarizabilities are also important ingredients in, for example, the proton-neutron mass splitting [30–33], and the spin polarizabilities parametrize the response of the nucleon spin to electromagnetic fields (such as the nucleonic Faraday effect).

Thirteen independent observables per nucleon parametrize the process when at most two of the photon beam, nucleon target or recoil nucleon are polarized. The (unpolarized) differential cross section $\mathrm{d}\sigma$ (in nb/sr) is larger than zero but otherwise unbounded. The beam-target asymmetries Σ_3, Σ_y, Σ_{1x}, Σ_{1z}, Σ_{2x}, Σ_{2z}, Σ_{3y} and the polarization-transfer observables from a polarized beam to the recoil nucleon $\Sigma_{1x'}$, $\Sigma_{1z'}$, $\Sigma_{2x'}$, $\Sigma_{2z'}$, $\Sigma_{3y'}$ are ratios of differences over sums of rates and take values between -1 and 1. Below the pion-production threshold $\omega_\pi(\mathrm{lab}) \approx 150\,\mathrm{MeV}$, only six observables are non-zero: $\mathrm{d}\sigma$, Σ_3, Σ_{2x}, Σ_{2z}, $\Sigma_{2x'}$, $\Sigma_{2z'}$.

The following data on these proton observables is available: about 420 points of widely varying quality for the cross section (see extensive discussions in refs. [1, 5]), about 120 points for Σ_3 [10, 34–36], 9 for Σ_{2x} [37, 38], and 10 for Σ_{2z} [10, 39]; no direct neutron data exists.

III. BAYESIAN METHODOLOGY IN EFT

A. Problems and Solutions of Design Strategy

To estimate the best design strategy, our approach must incorporate two distinct sources of uncertainty: (1) the EFT truncation error and (2) the unknown measurements from future experiments, including their likely measurement uncertainties. Our Bayesian approach can handle both of these problems in one coherent framework while being candid about our uncertainties. The two problems and the solutions we propose are summarized as:

1. *Problem:* χEFT must be truncated at a finite order, leading to a truncation error that is correlated in kinematic space. That is, we trust our theory more in some kinematic regimes than we do in others, and the discrepancy itself is a smooth function.

This should be reflected when assessing how well the experimental data from these regimes constrain polarizabilities.

Solution: An estimate of the truncation error is found by summing over all plausible values for the higher-order terms in the EFT. This results in a covariance matrix for the theory error that weights experimental data from trusted regimes more heavily than data from less trusted regimes [40].

2. *Problem:* Given a choice of design, we still do not know the results of the yet-to-be-performed experiment. But such results are needed in order to estimate how well they would constrain the polarizabilities.

   Solution: Bayesian experimental design considers all data that could plausibly be measured. For each of these we compute corresponding polarizability posteriors. The *expected* utility, or worth, of such an experiment can then be judged by sampling a utility function over all the data possibilities that have been evaluated.

Sampling is often computationally quite expensive; however, in our case, a controlled approximation allows it to be done analytically, leading to a simple and intuitive formula for the expected utility of an experiment. In the following subsections we describe in detail our approach to the problem of truncation errors and experimental design.

B. EFT Truncation Errors

A Bayesian model of EFT truncation errors has been proposed and discussed thoroughly in refs. [27, 41, 42]. Here we recapitulate the main results of the convergence model that are relevant to Compton scattering, and discuss how it must be modified to account for the rearrangement of the power counting in the regime of the Delta resonance. For a much more thorough introduction to this model of EFT truncation errors, we refer to ref. [27].

Suppose we are interested in the prediction of an observable $y(x)$ at some kinematic point x. Here, $x \equiv \{\omega, \theta\}$ is the incident-photon energy ω and scattering angle θ in the lab frame. EFTs provide a hierarchy of predictions $\{y_n(x;\vec{a})\}$, with each order n more precise than the last. These predictions depend on low-energy constants—the polarizabilities[2]—which we denote collectively as a vector $\vec{a}$. Let k be the highest order at which the complete EFT process has been calculated to date. Then there is a theory truncation error δy_k associated with all higher order terms left out of the state-of-the-art EFT prediction. Furthermore, if we are to compare our predictions to experimental measurements $y_{\rm exp}$, there is the problem of experimental noise, $\delta y_{\rm exp}$, to contend with. In ref. [27], the authors assume the following relationship, where theory and experimental uncertainties are independent:

$$y_{\rm exp}(x) = y_k(x;\vec{a}) + \delta y_k(x) + \delta y_{\rm exp}(x)\,. \tag{5}$$

Because $\delta y_k(x)$ and $\delta y_{\rm exp}(x)$ are unknown, they are treated as random variables. Given statistical models for δy_k and $\delta y_{\rm exp}$, we can use eq. (5) to tell us the kinematics x that will result in the most stringent constraints on the polarizabilities $\vec{a}$.

We extend refs. [27, 41, 42] by writing the observable expansion as

$$y_k(x;\vec{a}) = y_{\rm ref}(x)\sum_{n=0}^{k} c_n(x;\vec{a})Q^{\nu_n(\omega)}(x)\,, \tag{6}$$

where Q is the dimensionless expansion parameter of the EFT and $y_{\rm ref}$ is a reference scale for the observable y_k. For the EFT power counting to hold, the (dimensionless) observable coefficients c_n should be approximately of order unity. (Note: we leave implicit the dependence of c_n on $\vec{a}$ in subsequent equations.) equation (6) differs from refs. [27, 41, 42] by the inclusion of $\nu_n(\omega)$ rather than a simple n as the exponent of the expansion parameter. The use of $\nu_n(\omega)$ reflects the fact that the power counting changes as one moves from $\omega \simeq m_\pi$ to $\omega \simeq \Delta$, resulting in a re-ordering of contributions to the amplitude [28].

Because the EFT amplitudes are squared to compute observables (and in the case of spin observables, further divided by the differential cross section), the c_n are not linearly related to the polarizabilities; instead they appear naturally sized and randomly distributed. We will exploit this fact in the convergence model.

For given choices of $y_{\rm ref}$, Q and $\nu_n(\omega)$, the coefficients c_n are in 1-to-1 correspondence with the results y_n for orders $n \le k$. We will begin by discussing these choices and how they lead to a physically motivated distribution for δy_k. The choice of $\nu_n(\omega)$ is more technical, and is described in sec. III B 2.

1. The Distribution of δy_k

The reference scale $y_{\rm ref}$ should capture the overall size of y_k in the appropriate units. The spin observables Σ_i are dimensionless and bounded in $[-1, 1]$, hence the natural choice is $y_{\rm ref} = 1$. In contradistinction, the cross section varies over orders of magnitude, and can contain cusp-like behavior near the pion-production threshold. We capture its overall trend, without cusps, by using a $y_{\rm ref}$ comprised of the basic Born, pion pole, and Delta-pole parts of the proton (and neutron) cross section. With this choice, the proton cross section still shows

[2] In Compton scattering, these include the six nucleon polarizabilities in the combinations defined in sec. II: $\alpha_{E1} \pm \beta_{M1}$ and γ_0, γ_π, γ_{E-}, γ_{M-}.

some growth of the c_n near ω_π and at forward angles, so we multiply this reference by a shifted 2-dimensional Lorentzian

$$\left[\left(\frac{\omega_{\rm lab}-\omega_\pi}{50\,{\rm MeV}}\right)^2+\left(\frac{\theta_{\rm lab}}{150^\circ}\right)^2+\frac{1}{3}\right]^{-1}+1\,. \qquad (7)$$

There is no particular physics in this function. It serves only to produce c_n that look similar across kinematic space.

EFTs exploit a separation of scales, from which one can construct a small expansion parameter Q. Here we choose the expansion parameter

$$Q(x)=\sqrt{\frac{\omega_{\rm cm}+m_\pi}{2\Lambda_b}}\,, \qquad (8)$$

where the low-momentum scale is the average of m_π and $\omega_{\rm cm}$, the photon momentum in the center-of-momentum frame. This extends the expansion parameter $Q=\sqrt{m_\pi/\Lambda_b}$ proposed for $\omega\lesssim m_\pi$ in ref. [28]. The Q of eq. (8) explicitly builds in our expectation that the EFT's accuracy degrades at large ω. The breakdown scale of this χEFT in the single-nucleon sector was identified in ref. [28] as $\Lambda_b=650\,{\rm MeV}$. We also ran the analysis using the expansion parameter $Q\to\sqrt{m_\pi/\Lambda_b}$ and found the results were essentially unchanged.

The crux of the EFT truncation error model is induction on the c_n: the coefficients for $n>k$, which we have not yet seen, are assumed to have approximately the same size and dependence on (ω,θ) as the lower-order c_n that we already have from our EFT calculation. To formalize this inductive step, we model the c_n as independent and identically distributed (i.i.d.) curves and assign them a Gaußian process (GP) prior. This is notated as:

$$c_n(x)\,|\,\bar{c}^2,\ell_\omega,\ell_\theta\overset{\text{i.i.d.}}{\sim}\mathcal{GP}[0,\bar{c}^2 r(x,x';\ell_\omega,\ell_\theta)]\,. \qquad (9)$$

Here and below $z\,|\,g$ is read as "z given g", while $z\sim\ldots$ is statistical shorthand for "z is distributed as $\ldots$." In this case the c_n's are distributed as $\mathcal{GP}[m(x),\kappa(x,x';\boldsymbol{\theta})]$, which denotes a GP with mean function $m(x)$ and covariance function $\kappa(x,x';\boldsymbol{\theta})$. GPs are popular machine-learning algorithms that have been employed in a wide variety of disciplines to perform nonparametric regression and classification [43–45]. The samples from a GP are *functions*, as opposed to numbers or vectors. A brief introduction to GPs in this context is given in ref. [27]; see also refs. [45–47] for more in-depth discussions. We adopt a mean function of 0 since, *a priori*, the χEFT corrections c_n are just as likely to be positive as they are to be negative.

The values of our GP hyperparameters $\bar{c}^2$ and ℓ_i, whose meaning are discussed below, are tuned to the known c_n with $n\le k$ (at the best known values of the polarizabilities $\vec{a}$ for each EFT order). An example of how these hyperparameters, combined with symmetry constraints on the observables, lead to a distribution for higher-order

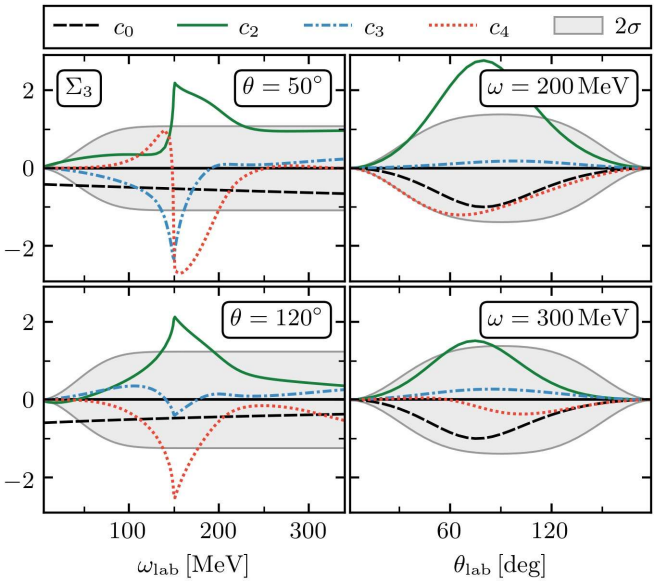

FIG. 1. (Colour online) Observable coefficients for Σ_3. The gray 2σ bands indicate the expected 95% credible interval for all higher order coefficients. Note that, although $c_0\neq 0$ at $\omega=0$, all of the corrections c_n, along with their derivatives, do vanish. A similar situation occurs at forwards and backwards angles. These constraints, and the corresponding ones for other observables, are built into the EFT truncation error model.

c_n is shown in fig. 1. We observe that there is no obvious systematic increase or decrease of the coefficients with n. This supports our adoption of a breakdown scale $\Lambda_b=650$ MeV. But we also observe that cusps in the observables around the pion-production threshold, $\omega\approx\omega_\pi$, can grow rather large. These are expected, and are not a problem in and of themselves (as discussed when tuning ℓ_ω below). But, for spin observables Σ_i, their large size is uncharacteristic compared to the c_n away from ω_π; hence we choose to exclude $125<\omega_{\rm lab}<200\,{\rm MeV}$ when training $\bar{c}^2$ and ℓ_i.[3] Despite providing the most rigorous accounting of uncertainties to date, we are thus less confident in the estimate of the χEFT uncertainty δy_k for Σ_i very close to ω_π. [The convergence pattern of $d\sigma$, in contrast, remains regular at ω_π, giving us confidence in our design results there.] The details of the fitting procedure, the symmetry constraints on the observables, and figures for the remaining observables, are reserved for app. B. The results of the fits are shown in table I. By the inductive step [eq. (9)], these hyperparameters tell us about the unknown higher-order c_n.

The marginal variance $\bar{c}^2$ in eq. (9) controls the size of the c_n. If the c_n are naturally sized, then $\bar{c}$ should be of

[3] Interestingly, the spikes disappear when considering the c_n for rate-differences $d\sigma\times\Sigma_i$, but a Gaußian uncertainty in these and in $d\sigma$ does not lead to a simple Gaußian uncertainty in Σ_i. Further study of the convergence patterns of Σ_i near ω_π is needed.

TABLE I. The optimized truncation error hyperparameters for each observable listed in ref. [6], which also discusses the definition and properties of the observables. The length scale ℓ_ω is given in units of MeV, whereas ℓ_θ is in units of degrees. The c_n are dimensionless quantities, so $\bar{c}$ is dimensionless as well.

	Proton			Neutron		
	$\bar{c}$	ℓ_ω	ℓ_θ	$\bar{c}$	ℓ_ω	ℓ_θ
$\mathrm{d}\sigma$	0.9	56	64	3.2	40	78
Σ_{1x}	0.76	35	46	0.68	58	49
Σ_{1z}	0.48	35	56	0.4	56	42
Σ_{2x}	0.59	42	37	0.7	50	39
Σ_{2z}	1.5	50	45	2.1	46	52
Σ_3	0.7	49	35	0.5	70	44
Σ_y	0.63	41	52	0.57	61	44
Σ_{3y}	0.83	36	45	0.66	60	45
$\Sigma_{3y'}$	0.64	41	46	0.66	49	46
$\Sigma_{1x'}$	0.66	37	47	0.49	55	43
$\Sigma_{1z'}$	0.28	33	44	0.29	55	43
$\Sigma_{2x'}$	1.1	32	58	1.3	40	53
$\Sigma_{2z'}$	0.75	34	52	1.8	48	57

order unity. We place an inverse chi-squared prior on $\bar{c}^2$:

$$\bar{c}^2 \sim \chi^{-2}(\nu_0, \tau_0^2)\,, \tag{10}$$

where ν_0 and τ_0 are the prior degrees of freedom and scale parameters, respectively. This is a conjugate prior and allows the posterior for $\bar{c}^2$ to be found analytically; see ref. [27] for details. We choose $\nu_0 = 1$ and $\tau_0 = 1$, which is weakly informative [48]. We take the posterior mean as an estimate for $\bar{c}^2$ in eq. (9).

The smoothness of the c_n is dictated by the correlation function r. We take the correlation of c_n between two kinematic points $x = (\omega, \theta)$ and $x' = (\omega', \theta')$ to be given by a radial basis function (RBF)

$$r(x, x'; \ell_\omega, \ell_\theta) = \exp\left\{-\frac{(\omega-\omega')^2}{2\ell_\omega^2} - \frac{(\theta-\theta')^2}{2\ell_\theta^2}\right\}, \tag{11}$$

where the correlation lengths ℓ_i control how quickly the c_n vary as a function of ω and θ. There is no conjugate prior for ℓ_i; rather, we use a uniform prior and find the best fits by maximizing the log likelihood. Choosing the RBF as a correlation function for the c_n implies that they are quite smooth. This assumption is validated empirically, except, as already noted, at the pion-production threshold, which occurs at photon energy $\omega_\pi \approx 150\,\mathrm{MeV}$ in the lab frame. Tuning ℓ_ω to data with cusps will bias it towards very small values. To fix this bias, we set the correlations between $c_n(x)$ and $c_n(x')$ to zero if they are on opposite sides of the pion-production threshold, but still use the same correlation lengths ℓ_ω and ℓ_θ below and above the cusps. This procedure is consistent with the behavior of the coefficients excluding the threshold region and with tests using GP toy models. The details for ℓ_ω are not critical for the present analysis as when we look at multiple design points they are all at the same energy: while correlations in angle matter for the design assessment, correlations in energy do not.

Assuming we have estimates of $\bar{c}^2$ and ℓ_i, we can construct the distribution for δy_k. It follows from extending eq. (6) that

$$\delta y_k(x) = y_{\mathrm{ref}}(x) \sum_{n=0}^{\infty} c_{n+k+1}(x) Q^{\nu_{\delta k}(\omega)+n}(x)\,, \tag{12}$$

where $\nu_{\delta k}(\omega)$ captures the first incomplete order of the EFT, and we assert for simplicity that powers of Q increment in integer steps afterwards. [For an EFT with a single power counting, one might expect $\nu_{\delta k}(\omega) = k+1$.] equation (12) is a geometric sum of Gaußian random variables, from which it follows that

$$\delta y_k(x)\,|\,\bar{c}^2, \ell_\omega, \ell_\theta \sim \mathcal{GP}[0, \bar{c}^2 R_{\delta k}(x, x'; \ell_\omega, \ell_\theta)]\,, \tag{13}$$

where

$$R_{\delta k}(x, x'; \ell_\omega, \ell_\theta) \equiv y_{\mathrm{ref}}(x) y_{\mathrm{ref}}(x') \frac{Q^{\nu_{\delta k}(\omega)}(x) Q^{\nu_{\delta k}(\omega')}(x')}{1 - Q(x)Q(x')} \times r(x, x'; \ell_\omega, \ell_\theta)\,. \tag{14}$$

Given choices of y_{ref}, Q, $\nu_{\delta k}$, and $r(x, x')$, along with estimates of $\bar{c}^2$ and ℓ_i, the above equations completely define a physics-based uncertainty due to truncation.

2. *The Power-Counting and Its Rearrangement*

An EFT begins with an infinite set of operators that one orders via a power counting in a ratio of a small to a large scale, denoted Q. There is then a finite number of parameters, and a finite set of diagrams, that contribute to the process of interest at any given order in the EFT expansion.

Here, we briefly describe the power counting employed in our calculations of Compton scattering; details and amplitudes are given in refs. [4–6]; see also references therein. Because of a different hierarchy of physical mechanisms, the power counting is different at different energies, and so is the first order at which the EFT amplitude is incomplete. We thus also detail the model used to interpolate between the two regimes in which the EFT power counting is well understood: $\omega \sim m_\pi$ and $\omega \sim M_\Delta$. The EFT expansion breaks down entirely as $\omega \to \Lambda_b \approx 650\,\mathrm{MeV}$.

The first regime concerns low photon energies up to around the pion mass,[4] $\omega \lesssim m_\pi \approx 140\,\mathrm{MeV}$. Here, the

[4] In this regime, the relation between the expansion used here and

power-counting mirrors that of standard heavy-baryon χPT for graphs involving nucleons and pions; all pion and nucleon Born graphs are counted as LO, or $\mathcal{O}(Q^0)$, where $Q \sim \sqrt{(\omega, m_\pi)/\Lambda_b}$. There is no NLO [$\mathcal{O}(Q^1)$] correction. The first corrections come at N^{2}LO [$\mathcal{O}(Q^2)$] from the pion cloud around the nucleon. It is at this order that the polarizabilities enter first. At N^{3}LO [$\mathcal{O}(Q^3)$], effects from the lowest-lying nucleonic resonance, the $\Delta(1232)$, and its pion cloud are added; the counting exploits the numerical fact that $M_\Delta/\Lambda_b \sim Q$ to avoid handling multiple expansion parameters. At N^{4}LO [$\mathcal{O}(Q^4)$], corrections to pion-cloud effects are accounted for, and no new effects from the Delta enter. Our implementation of χEFT for Compton scattering from the nucleon contains all contributions up to and including N^{4}LO for photon energies $\omega \simeq m_\pi$ as well as some terms that are N^{5}LO there. Therefore, the theory error δy_k in this regime follows eq. (12) with $\nu_{\delta k} = k + 1$, and proceeds indeed in integer steps.

However, the contributions are re-ordered in the regime where the photon energy approaches the excitation energy of the Delta resonance, $\omega \approx 300\,\text{MeV}$. (Hereafter, we use the symbol Δ for this energy, $\Delta = M_\Delta - M_N$.) As this resonance is very strong, it is physically intuitive that the most dramatic reordering involves Delta-pole diagrams. On the formal level, this change stems from the propagator in the Delta pole graph which becomes large near the resonance. The onset of the Delta-dominated region can be estimated from the Delta resonance width to occur at $\omega \sim 230\,\text{MeV}$.

In order to describe the relative order of contributions in the Delta-resonance region, and the transition, we first define $\nu_n(\Delta)$ as the lowest order at which those diagrams which are of order n at $\omega \lesssim m_\pi$ contribute when $\omega \simeq \Delta$. The leading Delta pole diagram is promoted from $\mathcal{O}(Q^3)$ (N^{3}LO) for $\omega \simeq m_\pi$, to $\mathcal{O}(Q^{-1})$ (LO) for $\omega \simeq \Delta$. Hence $\nu_3(\Delta) = -1$. This is now the leading contribution in the resonance region. (The Delta-π loops are less strongly promoted, to N^{2}LO, so it is the pole graph that determines $\nu_3(\Delta)$.) Subleading Delta pole diagrams (with dressed vertices) transition from $\mathcal{O}(Q^5)$ (N^{5}LO) to $\mathcal{O}(Q^0)$ (NLO). (Indeed, we include them in our amplitudes even though they are not a complete set of low-energy N^{5}LO diagrams, because we want to catch all NLO contributions in the resonance region.) Born effects and corrections to resonance parameters also enter at NLO (now $\mathcal{O}(Q^0)$), so $\nu_0(\Delta) = 0$. This makes the NLO contributions be complete in the resonance region. The contributions from pion loops round the nucleon change from $\mathcal{O}(Q^2)$ (N^{2}LO) to $\mathcal{O}(Q^1)$ so $\nu_2(\Delta) = 1$, and from $\mathcal{O}(Q^4)$ (N^{4}LO) to $\mathcal{O}(Q^2)$ so $\nu_4(\Delta) = 2$. These therefore enter in the resonance region at N^{2}LO and N^{3}LO, respectively. However, in the resonance region our amplitude is not complete at N^{2}LO and N^{3}LO: there are further contributions at these orders that we have not accounted for. For $\omega \sim \Delta$ the amplitude is complete only up to and including NLO.

Summarizing the previous paragraph: the relevant reorderings turn out to follow the rule that, for diagrams of order n at $\omega \simeq m_\pi$, $\nu_n(\Delta) = n/2$ for even orders but $\nu_n(\Delta) = (n-5)/2$ for odd orders. The first omitted order for $\omega \sim \Delta$, $\nu_{\delta k}(\Delta)$, is then given by the smaller of $\nu_{k+1}(\Delta)$ or $\nu_{k+2}(\Delta)$.

Now, we wish to be able to handle data in the transition region lying somewhere in between $\omega \simeq m_\pi$ and $\omega \simeq \Delta$, by defining an interpolator $\nu_n(\omega)$ that is a function of ω. If we define a suitable monotonic function $f(\omega)$ satisfying

$$f(\omega \approx m_\pi) \approx 0\,, \quad f(\omega \approx \Delta) \approx 1\,, \tag{15}$$

the reordering is smoothly captured by

$$\nu_n(\omega) = \begin{cases} [1 - f(\omega)/2]n\,, & n \text{ even} \\ [1 - f(\omega)/2]n - 5f(\omega)/2\,, & n \text{ odd} \end{cases} \tag{16}$$

which is tabulated in table II. For definiteness we use a logistic form inspired by the Fermi function

$$f(\omega) = \left[1 + \exp\left(-4\ln 3 \cdot \frac{\omega - \omega_m}{\omega_2 - \omega_1}\right)\right]^{-1}, \tag{17}$$

where $\omega_1 = 180\,\text{MeV}$ and $\omega_2 = 240\,\text{MeV}$ are the locations where $f(\omega_1) = 1/10$ and $f(\omega_2) = 9/10$, and $\omega_m = (\omega_1 + \omega_2)/2 \approx 210\,\text{MeV}$ is the midpoint $f(\omega_m) = 1/2$. The same form was already used in the plots of ref. [6] to parametrize the "gray mist" at high energies, but our framework puts this "mist" on a quantitative footing via the EFT-inspired theory error, eqs. (12)–(14). These choices are consistent with the estimate above that the Delta resonance region starts around 230 MeV. This form for $f(\omega)$ is only one of several possibilities. Other sensible models for $f(\omega)$ lead to results which are compatible with those presented below.

In fig. 2, we translate the first column of table II and the logistic function (17) to a graphical representation of the re-ordering of contributions. It is straightforward to read off the dominant theory uncertainty that an amplitude which is complete up to $\mathcal{O}(Q^n)$ in the $\omega \simeq m_\pi$ regime has in the $\omega \simeq \Delta$ regime. This defines $\nu_{\delta n}(\Delta)$. The resulting form of the leading truncation error $\nu_{\delta n}(\omega)$ as function of ω is given in the third column of table II.

We observe that starting with the full amplitude up to and including $\mathcal{O}(Q^4)$ (N^{4}LO) for $\omega \simeq m_\pi$ only yields an amplitude that is complete at $\mathcal{O}(Q^{-1})$ (LO) for $\omega \simeq \Delta$. However, there are only a small number of diagrams that are missing at $\mathcal{O}(Q^0)$ for $\omega \simeq \Delta$. These were identified, computed and added to the amplitude in both regimes

the notation of refs. [4–6, 28] is $Q^n = e^2\delta^n$; the simpler symbol leads to more compact formulae later. The LO defined in this presentation corresponds to performing the χEFT power counting on the structure part of the nucleon Compton amplitude, *i.e.*, what remains after the (relativistic) nucleon and pion Born terms are subtracted, and thus differs in that detail from the power counting described in refs. [4–6].

TABLE II. The rearrangement of orders of diagrams between the m_π regime and the regime of the Delta resonance, and the approximate power of the first omitted term in the expansion. The most general transition function $f(\omega)$ is defined in eq. (15), with a particular choice (logistic function) given in eq. (17).

Order at $\omega \simeq m_\pi$	Transition with ω	Leading Truncation Error
0 (LO)	$\nu_0(\omega) = 0$	$\nu_{\delta 0}(\omega) = 2 - 3f(\omega)$
2 ($\mathrm{N^2LO}$)	$\nu_2(\omega) = 2 - f(\omega)$	$\nu_{\delta 2}(\omega) = 3 - 4f(\omega)$
3 ($\mathrm{N^3LO}$)	$\nu_3(\omega) = 3 - 4f(\omega)$	$\nu_{\delta 3}(\omega) = 4 - 4f(\omega)$
4 ($\mathrm{N^4LO}$)	$\nu_4(\omega) = 4 - 2f(\omega)$	$\nu_{\delta 4}(\omega) = 5 - 5f(\omega)$
5 ($\mathrm{N^5LO}$)	$\nu_5(\omega) = 5 - 5f(\omega)$	$\nu_{\delta 5}(\omega) = 6 - 4f(\omega)$

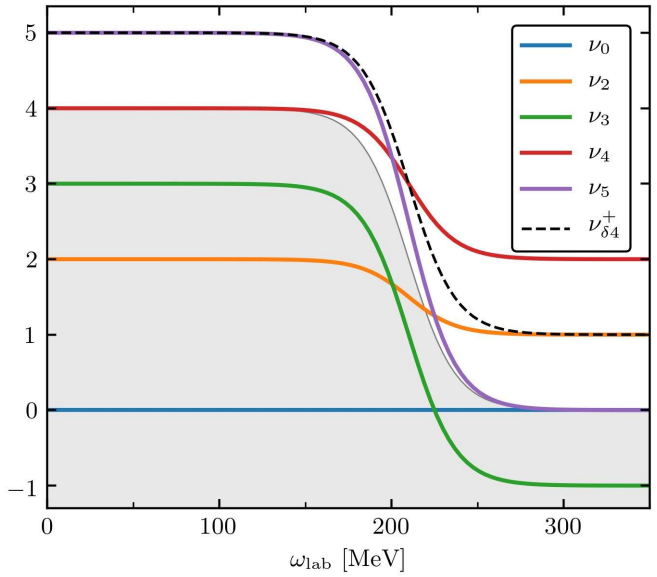

FIG. 2. (Colour online) The power counting transitions from the m_π regime to the regime around the Delta resonance. Solid lines corresponding to ν_n capture the most relevant reordering of diagrams, as described in the text. The shaded region is the approximate order up to which the $\mathrm{N^4LO^+}$ EFT is complete. The dashed line $\nu_{\delta 4}$ is one unit above the shaded boundary and represents the first order to be included in the EFT truncation error.

in ref. [5]. This produces an amplitude that is complete up to $\mathcal{O}(Q^0)$ for $\omega \simeq \Delta$ and is "$\mathrm{N^4LO^+}$", *i.e.*, more than $\mathrm{N^4LO}$ but not fully $\mathrm{N^5LO}$, for $\omega \simeq m_\pi$. Since the truncation error must include all orders that do not contain a complete set of diagrams, we therefore identify

$$\nu^+_{\delta 4}(\omega) = 5 - 4f(\omega) \tag{18}$$

as the first omitted power of our $\mathrm{N^4LO^+}$ EFT.

C. Experimental Design

The process of designing an experiment must begin with defining a goal. For example, this goal could be to make an accurate prediction of some future measurement, to discriminate between competing models, or to precisely constrain parameters of the theory. The goal could even be designed with a compromise between several different experimental aims in mind. It could also incorporate time and cost constraints. But in this work, we simply take constraining the nucleon polarizabilities as the goal of the experiments we are designing—although time and cost constraints will be assessed indirectly when we define different scenarios for the experimental accuracy.

The next step is to encode as mathematical objects the experimental goal and all uncertainties. Once encoded, our goal is known as a utility function, or design criterion, $U(\mathbf{d}, \vec{a}, \mathbf{y})$, that depends on the design points[5] $\mathbf{d}$ in the design space D from which experimental data $\mathbf{y}$ is then measured, and the theory parameters $\vec{a}$. But, of course, $\mathbf{y}$ will not be known until the experiment is conducted, and $\vec{a}$ is exactly the quantity we have constructed our experiment to find. Hence the optimal design $\mathbf{d}^\star$ is that which maximizes the *expected* utility $U(\mathbf{d}) = \mathbb{E}[U(\mathbf{d}, \vec{a}, \mathbf{y})]$. That is,

$$\begin{aligned}
\mathbf{d}^\star &= \arg\max_{\mathbf{d}\in D} U(\mathbf{d}) \\
&= \arg\max_{\mathbf{d}\in D} \int U(\mathbf{d}, \vec{a}, \mathbf{y})\, \mathrm{pr}(\vec{a}, \mathbf{y} \,|\, \mathbf{d})\, \mathrm{d}\vec{a}\, \mathrm{d}\mathbf{y} \\
&= \arg\max_{\mathbf{d}\in D} \int \Big\{ U(\mathbf{d}, \vec{a}, \mathbf{y})\, \mathrm{pr}(\vec{a} \,|\, \mathbf{y}, \mathbf{d})\, \mathrm{d}\vec{a} \Big\}\, \mathrm{pr}(\mathbf{y} \,|\, \mathbf{d})\, \mathrm{d}\mathbf{y} \,.
\end{aligned} \tag{19}$$

These integrals are usually intractable for nonlinear theories such as the observable predictions from χEFT, but we show in figs. 3 and 4 that linearizing χEFT predictions around the best known $\vec{a}$ is a very good approximation, and we employ it from here on.

Equation (19) says that the process of experimental design requires a theory $y(x; \vec{a})$ and a probabilistic model relating data to theory parameters, $\mathrm{pr}(\vec{a}, \mathbf{y} \,|\, \mathbf{d})$. To calculate that pdf we employ a different (but equivalent) form of the product rule to that used below eq. (19) and write $\mathrm{pr}(\vec{a}, \mathbf{y} \,|\, \mathbf{d}) = \mathrm{pr}(\mathbf{y} \,|\, \vec{a}, \mathbf{d})\, \mathrm{pr}(\vec{a})$. Our truncation error model from eq. (5) now comes into play. If a Gaußian prior is placed on the polarizabilities,[6]

$$\vec{a} \sim \mathcal{N}(\vec{\mu}_0, V_0)\,, \tag{20}$$

[5] A single *design* $\mathbf{d}$ in this work is specified by an observable and a set of kinematic points at which to measure it, and possibly the experimental noise levels. The space D is the set of all considered experiments over which the utility is optimized, *e.g.*, all possible 5-angle measurements at a given energy.

[6] The notation $\mathcal{N}(\vec{\mu}_0, V_0)$ denotes a Gaußian with mean $\vec{\mu}_0$ and covariance V_0. See also discussion of notation below eq. (9).

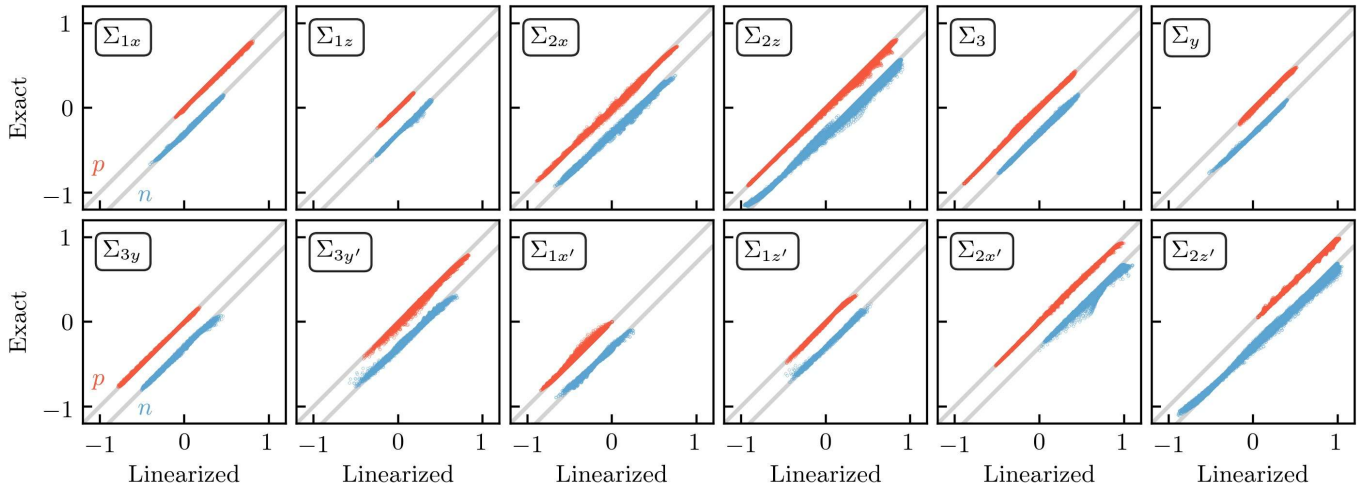

FIG. 3. (Colour online) Exact vs linearized predictions of proton ('p', red) and neutron ('n', blue) observables from the N^{4}LO$^+$ EFT. The neutron predictions have been vertically offset for clarity. The points shown use all predictions at $\omega = 60, 70, \ldots, 340\,\text{MeV}$ and $\theta = 40, 50, \ldots, 140°$, with 1000 sets of polarizability values sampled from the prior. The markers are so tightly clustered around the gray diagonals that they are difficult to distinguish. Here we have used the linear combinations of polarizabilities listed in table III. This significantly improves the linear approximation.

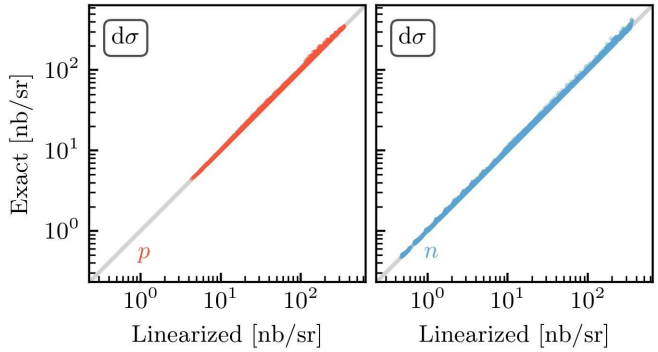

FIG. 4. (Colour online) Exact vs linearized predictions of the differential cross section $d\sigma$ for proton ('p', red) and neutron ('n', blue) from the N^{4}LO$^+$ EFT, with points as in fig. 3.

then under the assumption that $y(x;\vec{a})$ is linear in $\vec{a}$, one can show that the posterior is given by

$$\vec{a}\,|\,\mathbf{y},\mathbf{d} \sim \mathcal{N}(\vec{\mu}, V)\,, \tag{21}$$

where $\vec{\mu}(\mathbf{y},\mathbf{d})$ and $V(\mathbf{d})$ take into account both the truncation error and the experimental errors, and both depend on the values of the GP hyperparameters that have already been tuned to the EFT convergence pattern (see sec. III B 1 and app. A). The linearization point is chosen to be $\vec{\mu}_0$ and the prior for each nucleon is given in table III. We will discuss these priors momentarily.

Our goal is to constrain polarizabilities, so the optimal design is that which is likely to provide the most information about $\vec{a}$. It is reasonable then to choose the utility to be the gain in Shannon information for $\vec{a}$ based on the experiment $(\mathbf{d},\mathbf{y})$. This is equivalent to the Kullback-Leibler (KL) divergence between the prior and posterior for $\vec{a}$, followed by marginalizing over $\mathbf{y}$:

$$U_{\text{KL}}(\mathbf{d}) = \int\left\{\ln\left[\frac{\text{pr}(\vec{a}\,|\,\mathbf{y},\mathbf{d})}{\text{pr}(\vec{a})}\right]\text{pr}(\vec{a}\,|\,\mathbf{y},\mathbf{d})\,d\vec{a}\right\}\text{pr}(\mathbf{y}\,|\,\mathbf{d})\,d\mathbf{y}. \tag{22}$$

The assumptions of eqs. (20) and (21) allow (22) to be computed exactly (see app. A) with the result

$$U_{\text{KL}}(\mathbf{d}) = \frac{1}{2}\ln\frac{|V_0|}{|V(\mathbf{d})|} \equiv \ln\mathcal{S}(\mathbf{d}) \geq 0\,, \tag{23}$$

where we have defined the posterior shrinkage factor $\mathcal{S} \geq 1$. Consider the hyperellipsoids defined by given confidence levels for the $\vec{a}$ prior and posterior, (20) and (21). Then $\mathcal{S}$ is the factor by which the volume of the prior ellipsoid shrinks as it is updated to the posterior, with larger values of $\mathcal{S}$ (or U_{KL}) being more informative than smaller values. An experiment yielding $\mathcal{S} = 1$ (or $U_{KL} = 0$) is then completely uninformative. The utility of an experiment designed to constrain any subset of $\vec{a}$, without regard to the others, can be assessed by simply computing eq. (23) with the corresponding submatrices of V_0 and V.

Although the posterior shrinkage has the benefit of being strictly non-negative and increasing with increasing information, it is unbounded, making it difficult to compare plots on different scales. Thus, we choose to show the percent decrease in uncertainty (sometimes referred to as "information gain" below)

$$\begin{aligned}\%\text{ Decrease} &= \frac{|V_0|^{\frac{1}{2}} - |V|^{\frac{1}{2}}}{|V_0|^{\frac{1}{2}}} \times 100\% \\ &= \left(1 - \frac{1}{\mathcal{S}}\right) \times 100\%\,.\end{aligned} \tag{24}$$

TABLE III. Priors for the linear combinations of polarizabilities defined in eqs. (12) and (13) of ref. [6]. The prior covariance matrix V_0 for the neutron is constructed as a diagonal matrix with $[V_0]_{ii} = [\vec{\sigma}_0]_i^2$. This formula also defines the diagonal elements of the proton covariance matrix. But for $\alpha_{E1} - \beta_{M1}$ and γ_{M-} there are correlations in the prior, as described in the text. Unless otherwise referenced, values are from table 1 in ref. [4].

	Proton			Neutron		
$\vec{a}$	$\vec{\mu}_0$	$\vec{\sigma}_0$	ref.	$\vec{\mu}_0$	$\vec{\sigma}_0$	ref.
$\alpha_{E1} + \beta_{M1}$	14.0	0.2	[49]	15.2	0.4	[50]
$\alpha_{E1} - \beta_{M1}$	7.5	0.9	[5]	7.9	3.0	[51, 52]
γ_0	−0.90	0.14	[53]	0.4	2.2	
γ_π	7.7	1.8	[54]	7.8	2.2	
γ_{E-}	−0.7	2.0		−3.9	2.0	
γ_{M-}	0.3	0.9		−1.1	0.9	

This shares the beneficial aspects of $\mathcal{S}$, but is bounded in the range of 0–100%.

Our assumptions lead to a form of the expected utility that is analytic, easy to understand, and quick to compute. This makes eq. (23) very attractive. It allows quick assessment of both:

- Optimal designs for various assumptions, such as experimental noise levels and truncation error forms.
- Which polarizability subsets will have their constraints improved by a particular experiment—and by how much.

Constraints from previous experiments are built in naturally via the prior on the polarizabilities. For example, a large utility in a previously well-measured observable or region of kinematic space means that there is still valuable constraining information to be gained there. One might be concerned that this choice of prior has undue influence on the final results, so here we note two ways in which it does *not* influence them. First, eq. (23) is invariant under any linear transformation of $\vec{a}$, meaning, *e.g.*, that the choice of units for $\vec{a}$ is irrelevant, and that this analysis would be consistent if we had instead used $\vec{a} = \{\alpha_{E1}, \beta_{M1}, \gamma_{E1E1}, \gamma_{M1M1}, \gamma_{E1M2}, \gamma_{M1E2}\}$, so long as V_0 were transformed accordingly (see sec. II). Second, the final covariance matrix V is independent of the data y: it is determined by the accuracy of the experiment, the sensitivity of observables to the polarizabilities at the design points, and the prior covariance matrix V_0. Because we have linearized the problem, the central values shown in table III only enter our experimental-design assessment insofar as they affect the value of the vector B that encodes the sensitivity of the observable y to the six polarizabilities (see eq. A5). The *width* of the prior—the choice of V_0—does affect the design analysis though.

D. Choice of Priors

The priors summarized in table III are the uncertainties to which the polarizabilities are known at present. As we base our design on the results of the χEFT variant of refs. [1, 4–6], it is natural to resort to table 1 of ref. [4] for the central values and uncertainties for all polarizabilities which are not well-determined by other means. Uncertainties (theory and, as applicable, statistical) were combined in quadrature in table 1 of ref. [4]. Theory uncertainties were derived as advocated in ref. [41]: based on the order-by-order convergence of the series, one uses a Bayesian analysis to predict the size of the first omitted order in the EFT expansion.

Four values in table III are not taken from ref. [4]. The best known is $\alpha_{E1} + \beta_{M1}$ which is most accurately determined not from Compton scattering experiments, but from evaluations of the Baldin Sum Rule for the proton [49, 55] and neutron [50]. This recasts it as an energy-weighted integral over photoproduction cross sections. Likewise, the GDH Sum Rule provides a highly precise value for the proton's γ_0 [53].

Meanwhile, for the spin-polarizability γ_π of the proton and neutron, there is information available from backscattering Compton experiments. For the proton we take the result $\gamma_\pi = 7.7 \pm 1.8$ from ref. [54]. For the neutron the spin-polarizability situation is less clear, since "experimental" results that do exist are obtained from experiments on a deuteron target after nuclear effects have been removed. Therefore in the neutron case we employ the predictions inferred from the χEFT results for the spin polarizabilities in table 1 of ref. [4]. We also employ the χEFT result from that table for the proton spin polarizability, γ_{E-}. The last proton spin polarizability, γ_{M-}, is derived from the the χEFT result for γ_{M1E2} and a fit value for γ_{M1M1} [5], see below for further comments.

Alternative values for the spin polarizabilities with overall similar uncertainty estimates are available [49, 55–59], as well as some from recent data analyses [35, 37–39]. Other recent extractions of polarizabilities from unpolarized data should also be mentioned [60–62].

The values of $\alpha_{E1} - \beta_{M1}$ in table III were determined in the χEFT variant we employ here from Compton scattering data on the proton [5] and, for the neutron values, on the deuteron [51, 52]. The fit of proton Compton cross-section data in ref. [5] revealed a correlation between $\alpha_{E1} - \beta_{M1}$ and γ_{M-}. Since our knowledge of these two parameters comes from these data, that correlation should be included in the prior. Translating the correlation matrix from fig. 12 of ref. [5] into the polarizability basis used in this paper and inflating the errors to account for theory uncertainty in the extraction per the

procedure of ref. [4] we find a covariance matrix of:

$$\begin{pmatrix} \sigma^2_{\alpha_{E1}-\beta_{M1}} & \sigma^2_{\alpha_{E1}-\beta_{M1}\ \gamma_{M-}} \\ \sigma^2_{\alpha_{E1}-\beta_{M1}\ \gamma_{M-}} & \sigma^2_{\gamma_{M-}} \end{pmatrix} = \begin{pmatrix} 0.82 & 0.52 \\ 0.52 & 0.80 \end{pmatrix}. \tag{25}$$

Ultimately the prior represents how extant experimental data and χEFT analysis constrain the polarizabilities. Therefore it should be obtained from a χEFT analysis of Compton data that uses the statistical model for theoretical uncertainties in observables that we formulated here. While the analysis of ref. [5] included theoretical uncertainties in its final results, it did so in a less sophisticated manner. Redoing the extraction of polarizabilities in a way that is completely consistent with the experimental-design analysis of this work is an important project for the future. Until that is done though, we use a prior based on ref. [5] supplemented by other information, namely the prior we have described in this section. We reiterate that the conclusions of the design analysis are insensitive to the central values adopted for the polarizabilites in the prior, μ_0. Changes in the correlation structure could affect them. But adopting different, reasonable correlations in the prior does not produce marked changes in expected utilities or posterior shrinkages—as the results we now present will demonstrate.

IV. RESULTS

It is useful to start the presentation of results with the customary word of caution in mathematical statistics. The predictions which form the output of this formalism should be understood as likely outcomes, not as guarantees. They carry "errors on the errors". Details depend on our input choices (priors) and model assumptions, and it is an advantage of the Bayesian approach that these must be discussed explicitly. We found our results to be robust against other reasonable choices, though reasonable people can make other reasonable choices, which then leads to scientific progress by discussion. Overall, the choices we explored led to different details, but not to substantially different outcomes.

We would therefore not label one design's superiority as significant if its decrease in uncertainty [see eq. (24)] is within a few percentage points of others. But we are confident that a difference of, say, ten percent indicates a clear preference of one design over others.

The guidance we provide for observables and kinematic locations is documented in a publicly available Jupyter notebook [29]. We hope this will facilitate improvements on this analysis, which is meant to be the first, not the last, word in the ongoing conversation regarding the best way to improve the constraints on the nucleon polarizabilities.

TABLE IV. The experimental noise levels for the differential cross section dσ and spin observables Σ_i. The standard deviation for dσ is given in a percentage of its predicted value (using the prior mean for the polarizabilities), whereas the standard deviation for the spin observables is on an absolute scale.

Level	dσ (%)	Σ_i (abs.)
Standard	$\pm$5.00	$\pm$0.10
Doable	$\pm$4.00	$\pm$0.06
Aspirational	$\pm$3.00	$\pm$0.03

A. Precision Levels and Constraints of Compton Experiments

We attempt to choose estimates of experimental input which are realistic for modern accelerators and detectors, but also realize that the specifics depend on experimental details. For a first take, we focus on a scenario which is not optimized to a *particular* facility but should be at least of some use for planning and design at *any* facility. We also do not assess the impact of common-mode errors on the accuracy with which polarisabilities can be extracted, but instead consider only point-to-point ("statistical") errors.

We consider three levels of point-to-point detector precision, see table IV. We believe these provide a range of plausibly achievable experimental uncertainties for measurements on the proton. The "standard" scenario assumes uncertainties in the cross section of $\pm$5% (point-to-point systematic and statistical uncertainties combined in quadrature), and an absolute uncertainty in spin observables of $\pm$0.10. This is state-of-the-art for proton Compton experiments for the cross section and those spin observables that have already been measured [9, 10, 12, 35, 37, 39, 63]. A second scenario lists experimental error bars Δd$\sigma \approx \pm$4.0% and an absolute Σ_i error of $\pm$0.06 which are deemed "doable" nowadays without excessive improvements. The "aspirational" scenario assumes considerable but realistic new resources and possibly new equipment. Our choices were informed by discussions with our experimental colleagues who work on Compton scattering at MAMI and HIγS, for whose input we are very grateful [63]. Unless otherwise stated, all results in figures assume the "doable" level of experimental precision, with the remaining levels reserved for the Supplemental Material.

It should be noted that achieving even "standard" errors of $\pm$0.10 for some of the hitherto-unmeasured spin observables is not simple. Especially for the spin-polarization transfer observables, the experimental challenges of detecting recoil spin polarizations are considerable. In that case, the estimate can serve as benchmark, with the "standard" scenario already an "aspirational goal." Due to the absence of quasi-stable free-neutron

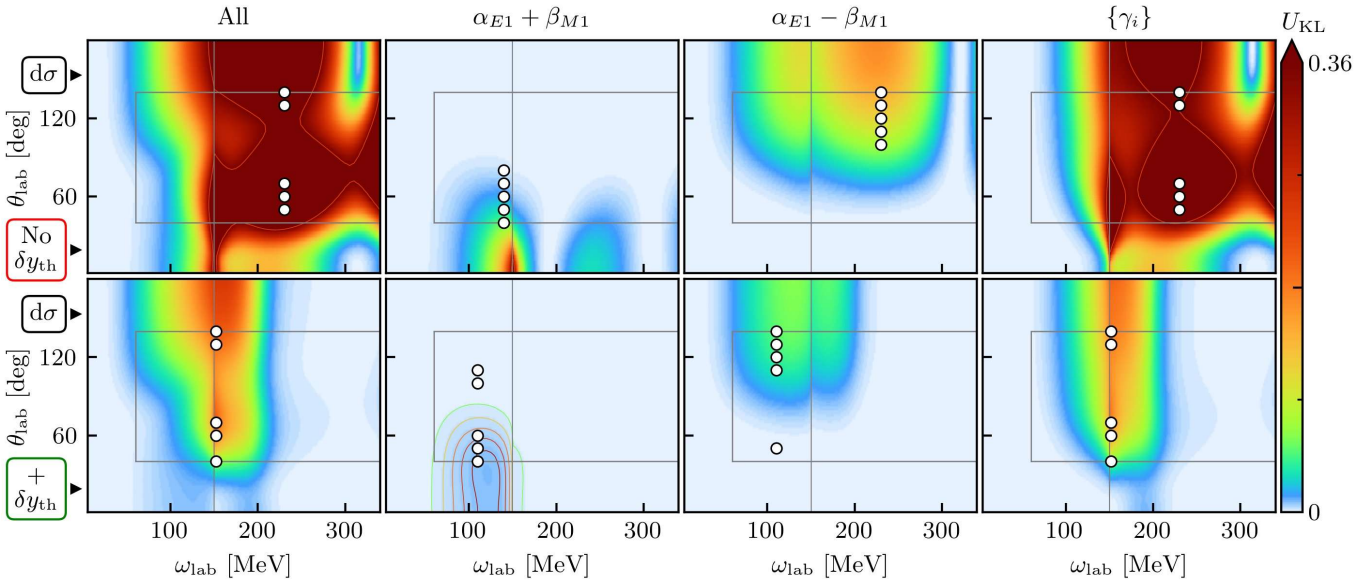

FIG. 5. (Colour online) The expected utility eq. (23) of proton differential cross section (dσ) measurements. Colors indicate the utility of one measurement conducted at each kinematic point $(\omega_{\rm lab}, \theta_{\rm lab})$, with the point of largest utility U_{KL} being by definition the optimal 1-point design. (The color bar is on a linear scale, though the hue varies *much* more quickly for small U_{KL}.) The top row (with the red, "No $\delta y_{\rm th}$" box) does not include EFT truncation estimates, whereas the bottom row does include the EFT uncertainty. Each column shows the utility one could expect to achieve for a subset of the proton polarizabilities: the first column considers all polarizabilities together, while the second and third show the gain for $\alpha_{E1} \pm \beta_{M1}$ individually, and the final column reflects the collective information to be learned about the four spin polarizabilities $\{\gamma_i\}$. The interior box that excludes $\omega < 60\,$MeV along with forward and backward angles marks the experimentally accessible regime (see text for further discussion). The vertical line marks the cusp at the pion-production threshold. To visualize the full range of variation in the subplots, the color ranges from zero to a saturation point calculated by averaging the optimal utilities across all subplots. Contour lines are added for utilities above the saturation point, with the first contour at saturation and subsequent contour lines at intervals of 50% of the color range, see the top left subplot. To help visualize the utilities from less-constraining measurements, additional contours are added to any subplot whose maximum utility is less than 10% of the color range, see, *e.g.*, fig. 6. Unless otherwise stated, the color ranges are common to all subplots within a figure, but different between figures. The white circles with black borders show the optimal five-point design kinematics, as described in the text. The effect of including EFT truncation errors is striking: it shifts the region of optimal utility to lower energies and moderates the expected utility.

targets, a "standard" uncertainty for neutron Compton scattering is of course well beyond "aspirational." We nonetheless chose to use the same error bars for the neutron, to ease comparison.

We search for the optimal one-point design and the optimal five-point design, *i.e.*, a search over all accessible combinations of five unique angles at a given $\omega_{\rm lab}$ ("5-point design"). In line with experimentalists' constraints on the placement of bulky detectors, we require that the angles be at least 10° apart. The focus on one photon energy and multiple angles mirrors the capabilities of "monochromatic-beam" facilities like HIγS which measure several angles at one energy simultaneously, but many other choices could be made. The assessment can easily be extended to "bremsstrahlung facilities," where a number of both angles and energies can be measured simultaneously. In that case, a typical spacing between the central energy of each energy bin of about 10 to 20 MeV appears realistic, given that a sufficient number of events must be collected in each "bin" for meaningful statistics [9, 10, 12, 35, 37, 39, 63]. Hence, our results attempt to be as realistic as possible given the choices above, and are a proof of principle for further, more specific research.

While the plots show a full range of energies and angles, we also indicate on them regions defined by $\omega \leq 60\,$MeV or $\theta \leq 40°$ or $\theta \geq 150°$ in which experiments are unlikely to be conducted, because forward and backward angles are physically hard to access, or because sensitivity to polarizabilities at very low energies is negligible. Therefore, we do not elaborate on designs that involve these kinematic regions.

Only the cross section and Σ_3 are non-vanishing as $\omega \to 0$, the physics of both being governed by the Thomson limit, with polarizability corrections very small. Our LO result provides the correct Thomson limit for each observable automatically, and we constrained the unknown higher-order corrections so that they do not change this; see app. B. Indeed, we find a typical energy correlation length of $\ell_\omega \approx 50\,$MeV for the proton; see table I, so that such a constraint becomes less important around

and above 60 MeV.

In addition, due to the coordinate singularity at $\theta = 0°$ and 180°, observables or their derivatives with respect to θ must be zero there. We implemented these constraints as described in app. B. As the angular correlation lengths from table I are all smaller than $\ell_\theta \approx 60°$, this is not a strong constraint on observables at intermediate angles where experiments are most feasible.

Our design model does not include the constraint that spin observables Σ_i can only have values between -1 and 1. This is a reasonable omission because the mean value of most Σ_i and their χEFT uncertainties are mostly well contained within these bounds (except maybe at the largest ω), see app. B for details.

Finally, we reiterate that around the pion-production threshold ($\omega_\pi \pm 20\,\text{MeV}$ or so, with $\omega_\pi \approx 150\,\text{MeV}$ marked by a vertical line in plots), the χEFT truncation error estimates for the spin observables Σ_i are less understood, and experimental conditions are difficult as well; see sec. III B 1 and app. B.

B. First Discussion and Impact of Accounting for Theory Uncertainties: The Cross Section

The unpolarized differential cross section is the most extensively studied nuclear Compton scattering observable. Therefore we begin by showing the expected utility of further measurements, with the goal of constraining various subsets of the polarizabilities. Our results focus on proton observables unless otherwise stated, due to the difficulty of performing experiments on neutrons; see the Supplemental Material for the corresponding neutron design results. We start by considering the following subsets: all polarizabilities simultaneously, only $\alpha_{E1} + \beta_{M1}$, only $\alpha_{E1} - \beta_{M1}$, and only the spin polarizabilities $\{\gamma_i\} \equiv \{\gamma_0, \gamma_\pi, \gamma_{E-}, \gamma_{M-}\}$, or each of them separately.

Figure 5 shows the expected utility of future proton experiments, with and without an estimate of the truncation error. Without truncation, the utility of an experiment to measure $\alpha_{E1} \pm \beta_{M1}$ mirrors the sensitivity analysis of fig. 8 in ref. [6]. There, the derivative of the observable with respect to a particular polarizability was plotted, and the truncation error of the EFT was only accounted for indirectly by casting a "gray mist" over the plot which thickens into the Delta resonance region, starting at $\omega \gtrsim 210\,\text{MeV}$, *i.e.*, where our transition region starts.

In fact, the constraining power of truncation-free 1-point measurements (as measured by U_{KL}) on each individual polarizability follow exactly the patterns of the local sensitivities for all observables and polarizabilities. This can be verified by comparing the remainder of our zero-truncation-error results in the Supplemental Material with the appropriate subplots of figs. 9–20 in ref. [6].

However, when truncation-error estimates are included, the optimal designs are pushed to lower ω_{lab}, with a particularly dramatic shift for $\alpha_{E1} - \beta_{M1}$. This is expected because the χEFT uncertainty δy_k increases with energy. Still, the optimal locations for constraining the spin polarizabilities often remain at or above the pion-production threshold.

One of the benefits of our Bayesian analysis over a purely derivative-based approach—like that of ref. [6]—is that we can examine the collective gain in information for multiple polarizabilities. The second, third, and fourth panels in the lower row of fig. 5 provide the optimal kinematics at which to constrain either of the scalar polarizabilities, or the spin-polarizabilities collectively. In these second and third panels (fourth panel) the design choice is computed unconditional on the other five (two) polarizabilities, *i.e.* we marginalize over the polarizabilities we are not interested in for that panel. Looking across them reveals that the collective utility in the first panel is approximately the sum of the utilities of each subset. In this study, we found that the correlations between these linear combinations of polarizabilities that are induced by fitting are rather small; see the extended discussion of fig. 8. Equation (23) then says that, to the extent that the covariance matrices V_0 and V are diagonal, the total utility is the sum of the individual utilities. The feature seen here is thus generic in the absence of correlations: the amount of benefit derived from collectively constraining $\vec{a}$ is related to how much the utilities for individual components of $\vec{a}$ overlap in kinematic space.

C. All Observables: Discussion

We now extend our analysis of the differential cross section to the spin observables Σ_i. For the remainder of this work, we include truncation error estimates, because otherwise the constraining power of any measurement would be overstated; see the Supplemental Material for corresponding results without truncation errors included. Figures 6 and 7 show heat maps of the expected utility of all proton observables, with truncation-error estimates included. Note that the scale has changed dramatically, as can be seen by comparing the results in the bottom row of fig. 5 to the same results repeated in the top row of fig. 6.

Figures 6 and 7 contain a wealth of information about the relative utility between observables at various kinematics, but for more readily interpretable statements about potential constraining power, we turn first to fig. 8. In what follows, we will discuss the observables in turn based on the promise shown in that figure, at the same time looking at the relevant row of Figures 6 and 7 for more detailed information.

From fig. 8 we can see the largest percent decrease in uncertainty or "information gain" [eq. (24)] of all optimal 5-point designs for each observable, with utilities split up based on the polarizabilities one might be interested in measuring. Thus, given a decision about which polarizability is of most interest—a decision which we

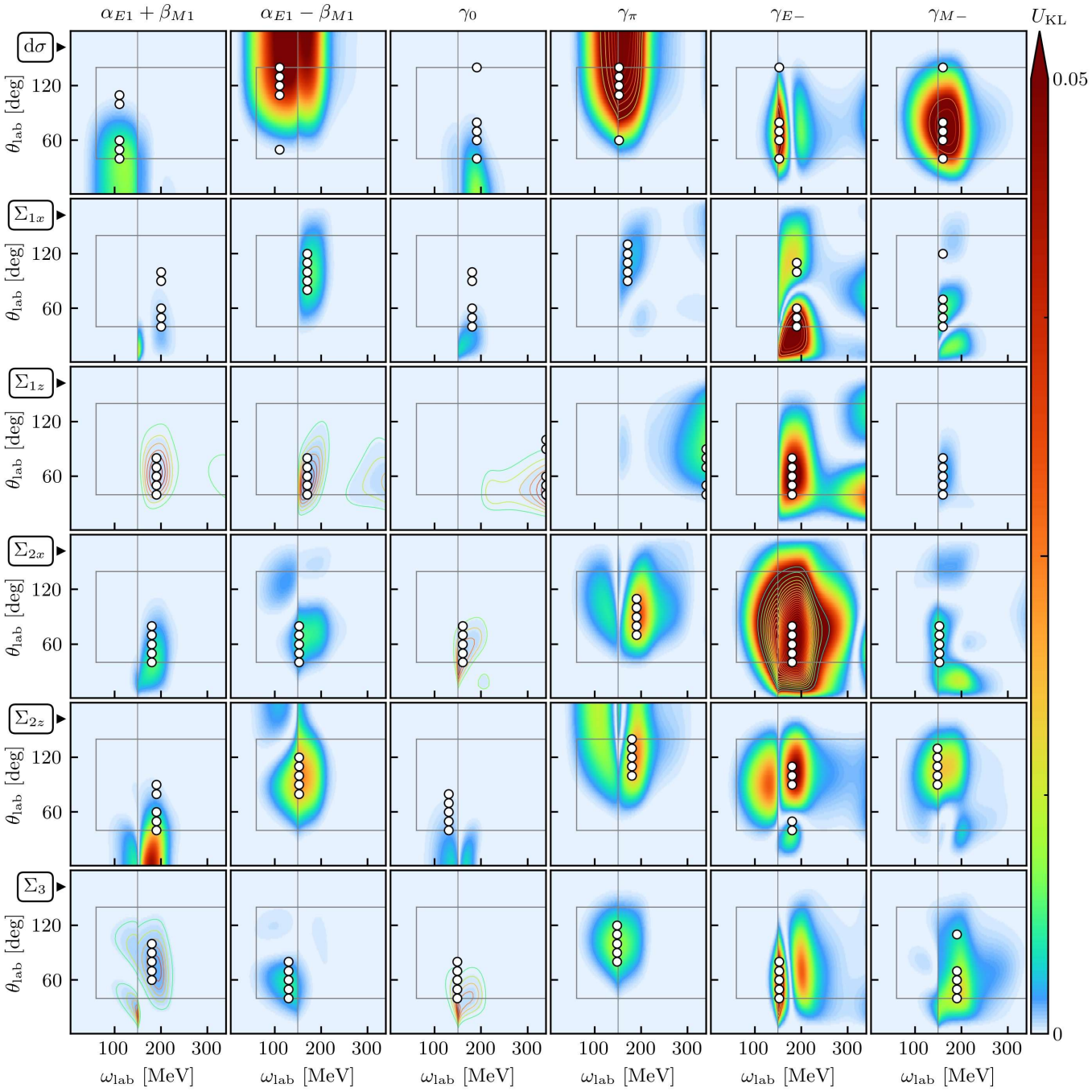

FIG. 6. (Colour online) Expected utilities of conducting various experiments, with the goal of constraining individual proton polarizabilities. Truncation errors are included throughout. See fig. 5 for a detailed description of the figure notation. Again, the color scale is common among all subplots so that both the location and relative magnitudes of utilities can be uncovered. To facilitate comparisons between observables, the color range is identical to that in fig. 7. However it must be noted that the color scale is very, very different from fig. 5, and that dark red in the current plot reflects only a modest increase in information.

do not encode mathematically—our approach provides a quantitative method for evaluating the worth of future experiments.

Any *set* of utilities, such as "All" or "$\{\gamma_i\}$" are guaranteed to be greater than or equal to the optimal utility of any individual polarizability that contributes to it. How much more is learned by considering multiple polarizabilities, depends on how much their optimal designs overlap in kinematic space—because we have found that in this case only small correlations are induced in the covariance matrix V by fitting.

For the proton, the combinations $\alpha_{E1}+\beta_{M1}$ and γ_0

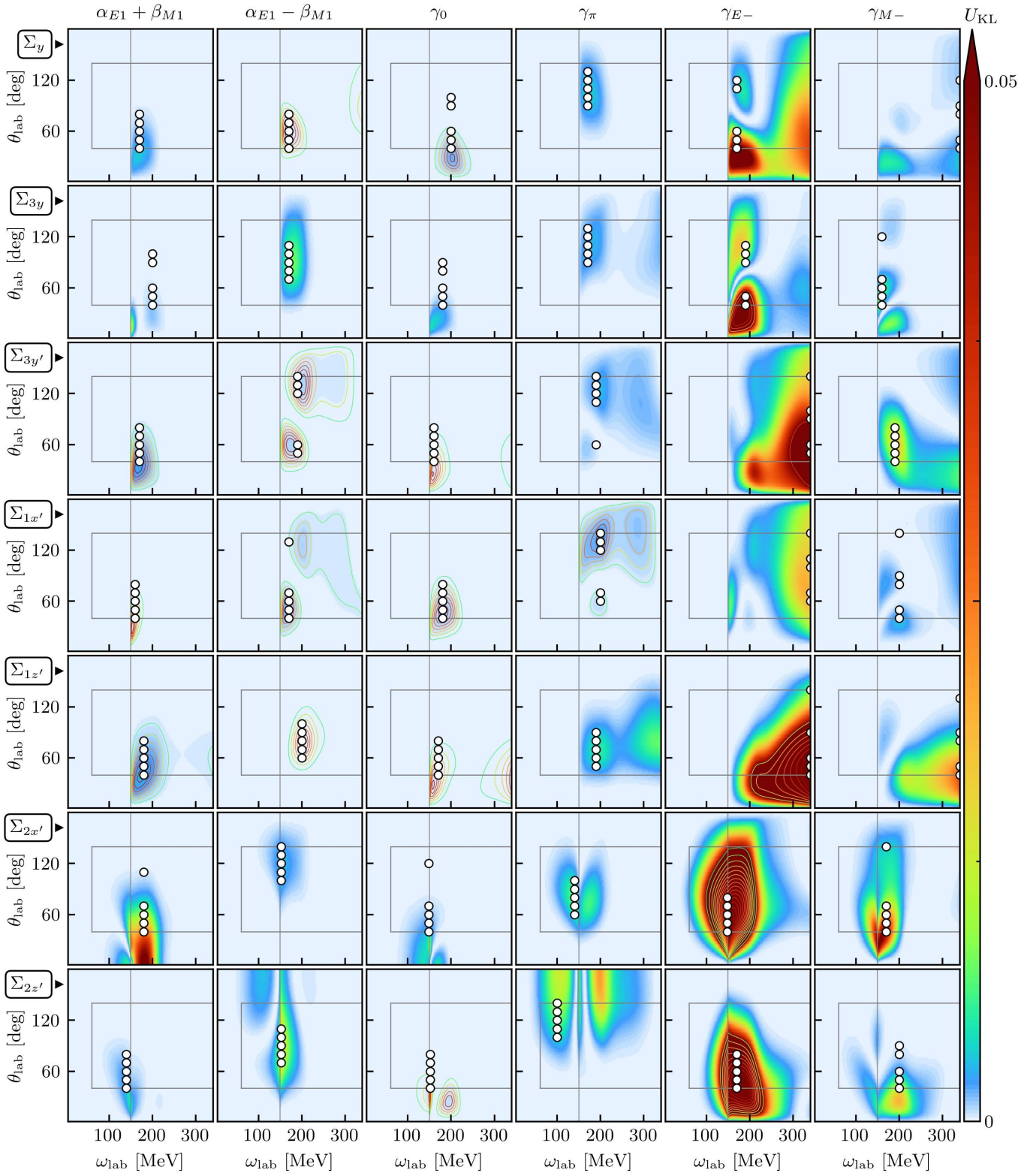

FIG. 7. (Colour online) The remaining observables, see fig. 6. Even with a saturated color scale that greatly exaggerates small values, the utilities from different measurements cover a vast range; see comments in fig. 6.

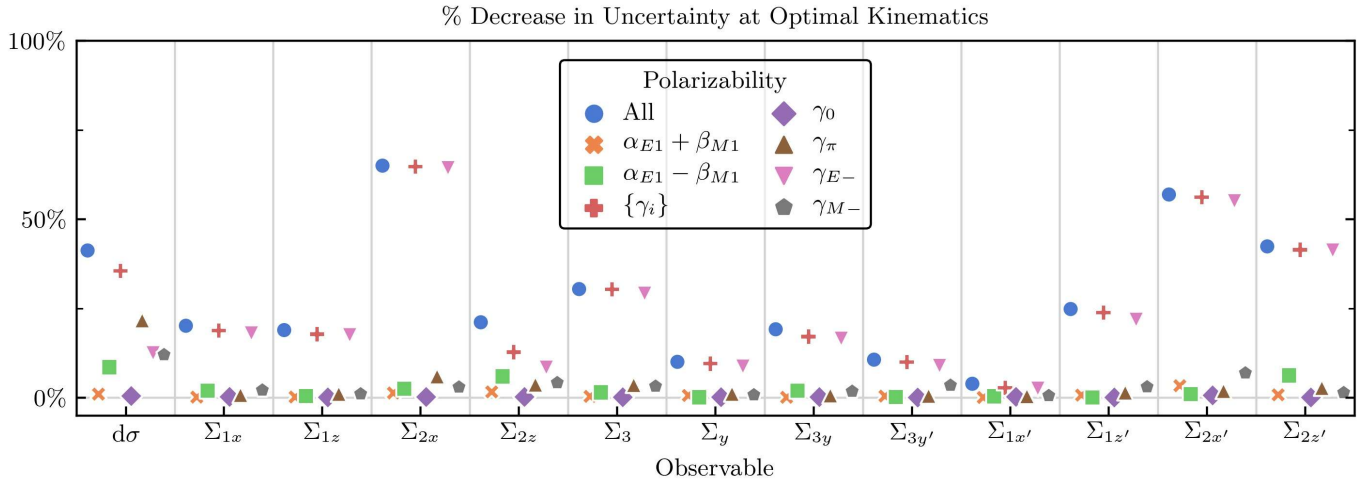

FIG. 8. (Colour online) The percent decrease in uncertainty or "information gain" [eq. (24)] computed while including truncation errors for the optimal 5-point proton Compton experiments. The maximum for each polarizability and observable is likely to be obtained at its *own* unique kinematic points, see figs. 5, 6, and 7. The spin observables Σ_{2x} and $\Sigma_{2x'}$ show powerful constraining power of γ_{E-}, with the potential to shrink its uncertainty by 50%. In contrast, measurements of the differential cross section $d\sigma$ can reduce the uncertainty in γ_π, γ_{E-} and γ_{M-} by about 30% (collectively), and also can shrink the error on $\alpha_{E1}-\beta_{M1}$ by about 25%. Other observables show slight constraining power. Note that the (primed) spin polarization transfer observables are notoriously difficult to measure.

are well-constrained by sum rules [49, 55]; see the small error bars in table III. In Compton scattering, these are the only two linear combinations of polarizabilities which enter the cross section as $\theta \to 0$. Figure 8 reveals that indeed little information on them can be gained from direct Compton experiments. The other four combinations, $\alpha_{E1}-\beta_{M1}$, γ_π, γ_{E-} and γ_{M-}, will therefore dominate our discussion.

One observable that stands out in fig. 8 is Σ_{2x} (circularly polarized photons on a transversely polarized target). Its overall utility of over 60% stems near-exclusively from the gain in γ_{E-}. Combining this with fig. 6, we see that the gain occurs in a quite robust and large region which extends from about ω_π to above 200 MeV, at angles between $\theta \approx 30°$ and 90°. The other polarizabilities are optimally constrained in a similar kinematic region, but according to fig. 8, their contribution to the overall utility from such an experiment is negligible ($\lesssim 5\%$) compared to what would be learned about γ_{E-}. This means that measurements of Σ_{2x} in that region allow for an extraction of γ_{E-} which is highly insensitive to the particular values of α_{E1}, β_{M1} and the other spin-polarizabilities used. This observable was already explored in a pioneering experiment at MAMI for $\omega \approx 290\ldots330$ MeV [37], where unfortunately the information content of an EFT interpretation is not very high. (Remember that the color scale in fig. 6 changes *rapidly* with decreasing utility.) This analysis implies that much more information can be gained from an experiment at $\omega \lesssim 200$ MeV.

A similarly large constraint on the polarizabilities comes from the analogous polarization-transfer observable $\Sigma_{2x'}$ (incident circularly polarized photon on unpolarized target, transverse spin polarization of recoil proton detected). In this observable γ_{E-} is a little less constrained than it is in Σ_{2x}, but the information gain for $\Sigma_{2x'}$ is still more than 50%. Meanwhile, the shrinkage for γ_{M-} increases slightly compared to Σ_{2x}: it is now a bit less than 10%. Some limited information about γ_{M-} can perhaps be gained here, but such a small information gain should not be over-interpreted. To achieve these shrinkages a $\Sigma_{2x'}$ experiment would need to be made near—or a few dozen MeV above—ω_π and towards forward angles.

The polarization transfer $\Sigma_{2z'}$ (incident circularly polarized photon on unpolarized target with detection of longitudinal recoil polarization) provides a gain of about 40% overall, at similar energies but slightly smaller angles. Most of the gain is again in γ_{E-}, followed by a small gain in $\alpha_{E1}-\beta_{M1}$.

Decent information gain on γ_{E-} (about 30%) can also be found from measuring the beam asymmetry Σ_3 (linearly polarized beam on unpolarized target) at intermediate angles in two narrow corridors, namely close to the pion-production threshold and slightly higher, $\omega \approx 200$ MeV. Some data are actually available there [10, 34–36] but these have not yet been analyzed in EFT and thus do not enter the prior in table III. These results for Σ_3 suggest that such an analysis could be valuable.

Measurements of Σ_3 at lower ω have been used in attempts to constrain the scalar polarizability β_{M1} [35]. But we see that even in the most sensitive kinematics the impact of this observable on β_{M1} amounts to just a few percent.

Instead, the combination $\alpha_{E1}-\beta_{M1}$ can best be measured from the cross section in a region somewhat above 100 MeV at back-angles. The overall information gain from a five-point measurement there is $\approx 10\%$. Qualita-

tively, this angle regime is not surprising since it is well known that this particular linear combination enters the cross section as $\theta \to 180°$, as does γ_π. Nevertheless, the conclusion may at first be surprising since there is already an extensive data set from ref. [64] in this energy region, and contemporary evaluations of $\alpha_{E1} - \beta_{M1}$, encoded in our prior, lean heavily on this data set. In fact, the possibility seen in fig. 8 to improve knowledge of $\alpha_{E1} - \beta_{M1}$ by measuring the cross section is supported by simple estimates, as follows. There are 20 Olmos de León *et al.* data points corresponding to lab photon energy > 135 MeV. How then, can five additional data points result in significant information gain for $\alpha_{E1} - \beta_{M1}$? In fact, these Olmos de León *et al.* points have precision of 7–15%. When combined with the point-to-point systematic of 5% their statistical power is slightly less than that of five additional points with 5% precision.

Interestingly, the next-largest information gain for $\alpha_{E1} - \beta_{M1}$ appears to be found in Σ_{2z} (circularly polarized beam on longitudinally polarized target) and the corresponding polarization transfer, $\Sigma_{2z'}$, but these amount to only 5–10%. In both these cases the region of greatest sensitivity lies right at the pion-production threshold, where experiments are particularly challenging and where our χEFT uncertainties may be less accurate (see sec. III B 1).

According to fig. 8, γ_{M-} is quite elusive. Only the differential cross section shows appreciable information gain (about 15%), while the next-largest gain, in $\Sigma_{2x'}$, is $\approx 10\%$. In all three observables, the region of largest sensitivity to γ_{M-} is right at the ω_π cusp, where the dσ convergence pattern is well behaved (see fig. 12). This makes us more confident in our design predictions for dσ than for the spin observables, which fluctuate more strongly. Taking all this into account, we find that a measurement of the cross section in a broad band around ω_π and at intermediate angles is the best chance to constrain γ_{M-}. As a bonus, such a measurement would concurrently constrain other polarizabilities "for free."

For γ_π, any information gain can only be found in the cross section and is about 20%; several other polarizabilities contribute similar amounts; see discussion above. A dedicated $\theta = 180°$ experiment, like that of ref. [65], may be able to resolve γ_π and $\alpha_{E1} - \beta_{M1}$, but needs a special design. No other polarizability combination enters at that angle.

Indeed, optimal 5-point measurements of the differential cross section dσ can decrease the collective uncertainty of all polarizability combinations by about 40%, but the information gain is spread out amongst individual polarizabilities: about 15–20% for γ_π, γ_{E-}, and γ_{M-}, 10% for $\alpha_{E1} - \beta_{M1}$, and no perceptible information gain for γ_0 and $\alpha_{E1} + \beta_{M1}$. In part, different kinematic regions are sensitive to individual combinations, so measurements across a wide array of energies and angles can be used to disentangle individual contributions.

The correlation structure implemented in the prior can affect these experimental-design conclusions. To check this, we also ran an "uncorrelated" design analysis in which the off-diagonal elements of the correlation matrix for $\alpha_{E1} - \beta_{M1}$ and γ_{M-} given in eq. (25) were set to zero. Results for this alternative prior choice are provided in the Supplemental Material, see figs. S.7–S.9. The results for these two priors are very similar. The most notable difference is that the information gain expected in $\alpha_{E1} - \beta_{M1}$ from differential cross section measurements is a little larger if an uncorrelated prior is employed, cf. fig. 8 and S.9. The amount that we expect to be learnt from Σ_{2z} and $\Sigma_{2x'}$ also increases slightly with use of an uncorrelated prior. This, as well as the other alterations in shrinkage and utilities that result from changing the prior's correlation structure, are well within the impact of different "reasonable prior choices" that we anticipated at the beginning of this section.

We now highlight an important point regarding the fact that our experimental design analysis is actually "aware" of the experimental information that is presently available. The utility function does not explicitly trace the kinematics and quality of available data. But the significant amount of proton Compton cross section data enters via the priors on α_{E1}, β_{M1} and γ_{M-} in table III, via the experimental (statistical plus systematic) uncertainties in the fits of those quantities. The fact that additional information can be gained from more high-quality data in specific kinematic regions implies that the quantity and, most importantly, the quality of future Compton data in that region can provide important information gains on the polarizabilities, even if that region appears at first glance to be already saturated. On the other hand, the available data for both Σ_{2x} [37, 38] and Σ_{2z} [10, 39] did not enter in the priors of table III. That is because those data were taken in the Delta resonance region, where the sensitivity of these two observables to any polarizability is minuscule, according to fig. 6. Adding their information to the priors will therefore not change our conclusions or improve polarizability error bars.

The relatively steep differences in information-gain reflect to a large extent the fact that the uncertainties for the spin polarizabilities and for $\alpha_{E1} - \beta_{M1}$ are substantially larger than for the sum-rule constrained combinations $\alpha_{E1} + \beta_{M1}$ and γ_0. In particular, the very small error bar on γ_0 makes it very hard to gain information via Compton scattering; fig. 8 shows that the expected shrinkage is indeed close to zero for all observables.

We caution, however, that the prior size of a polarizability's error bar is not by itself a reliable predictor of possible information gain. The error bar of γ_{M-} is about half of that of γ_π or γ_{E-}, so one might expect the information gain in measuring it to be about half of that for γ_π or γ_{E-}. Instead, a measurement with substantial information gain is much more elusive than that, as explained above.

The biggest sensitivity of $\Sigma_{3y'}$ to γ_{E-} and of $\Sigma_{1z'}$ to both γ_{E-} and γ_{M-} is pushed to the maximum considered energies $\omega \approx 300$ MeV. On the one hand, such behavior might be interpreted as in apparent tension with the fact

that χEFT is significantly less reliable in the Delta resonance region than at lower energies. On the other hand, the uncertainty of χEFT is accounted for in our experimental design; see discussion in sec. IV B. Possibly, the 5-point design in that region probes a sensitivity of the correlated angular dependence at high energies, rather than on individual values/rates at a particular angle. If so, and if χEFT predicts these correlations more robustly than overall sizes of an observable, then the phenomenon would be explained and measurements of the functional dependence of observables on angle at such high energies could provide determinations of γ_{E-} and γ_{M-}. However, optimal 1-point designs would not be sensitive to correlations and still appear sometimes at very high energies in these same observables. Apparently, the sensitivity is so strong at such kinematics as to win over the decreased theory uncertainties. As we did not find an intuitively obvious resolution, this merits further study.

Equally as notable as these substantial information gains on polarizabilities are those observables that seem to provide almost no information about the polarizabilities at this level of experimental and theoretical precision. The most prominent such example is $\Sigma_{1x'}$ (total gain $< 5\%$). Given our truncation error estimates for this quantity, there is little information on the polarizabilities to be gained from any 5-point experiment. Measuring it, or indeed any observable, in a region where the information gain for polarizabilities is negligible, can still be useful though. It provides information about how accurately χEFT describes the Compton process, independent of the polarizabilities. This is an important cross-check of χEFT, even though it is not part of the utility used in this work.

Such an analysis raises a further question: if experimental resources are limited, does it make sense to measure 1 point very precisely or many points less precisely? Our framework can supply answers to this and many other such questions. By comparing the optimal designs of 1- and 5-point experiments at both the "doable" and "aspirational" level of experimental precision (see table IV) we get an idea of how to design the most effective experiment. The results are given in fig. 9. Again, it is clear that the details depend on the observable, which proves the usefulness of our approach: one need not rely on heuristics when a quantitative scheme is readily available.

For example, the differential cross section does not appear to benefit as much from an increase in precision (red circle) as it would from more data across θ (blue cross). In almost all other cases, such as Σ_{2x}, $\Sigma_{2x'}$ or $\Sigma_{2z'}$, the gain in information an "aspirational" 1-point experiment is about the same as 5 measurements from a "doable" experiment. Surprisingly, some observables, such as $\Sigma_{1z'}$, benefit very little from either increased precision or an increased number of data points: one "doable" measurement in the right spot already realizes most of the information gain to be had from them.

For completeness, we show the utility of performing neutron Compton scattering experiments, with more plots for the neutron in the Supplemental Material. Figure 10 shows the 1-point profile of the differential cross section with truncation error, which is similar to the corresponding utility in fig. 5. Such measurements are notoriously difficult, so the interpretation of these results should proceed with caution. More realistically, our analysis should be applied to χEFT predictions of light incident on the deuteron, ^{3}He or other few-nucleon targets for which calculations of Compton scattering are available in the same χEFT formulation [1, 7, 8, 66].

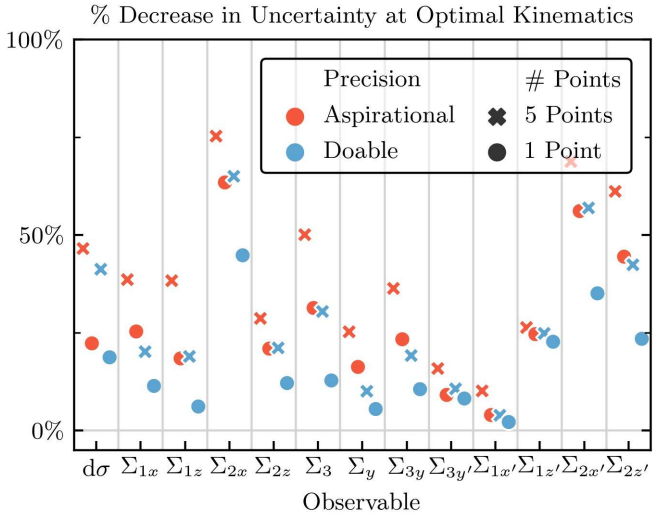

FIG. 9. (Colour online) The percent decrease in $\vec{a}$ uncertainties, as in fig. 8, applied to decide on the trade-off between different allocations of experimental resources (exploration vs. exploitation). Larger values imply that the measurement is more informative. The 1-point (5-point) optimal design is denoted by a circle (cross), and the experimental precision levels are given in table IV. The decision to increase precision or measure at more kinematic points (or neither) can vary significantly by observable.

V. SUMMARY

We have proposed a powerful and versatile framework to help plan experiments which rely on EFT to extract or check parameters. Using the example of Compton experiments in order to constrain nucleon polarizabilities, this method quantifies the expected gain in information from an experiment: it maximizes shrinkage of the posterior. The framework solves the problems of theoretical errors conflicting with experimental considerations, and finds a compromise between the two. Under reasonable assumptions, we obtain an algorithm that is analytic, easy to understand, and quick to compute.

Furthermore, we employed a Bayesian machine learning algorithm for estimating EFT truncation errors whose power counting varies across the domain. This is a novel extension of the model introduced in ref. [27].

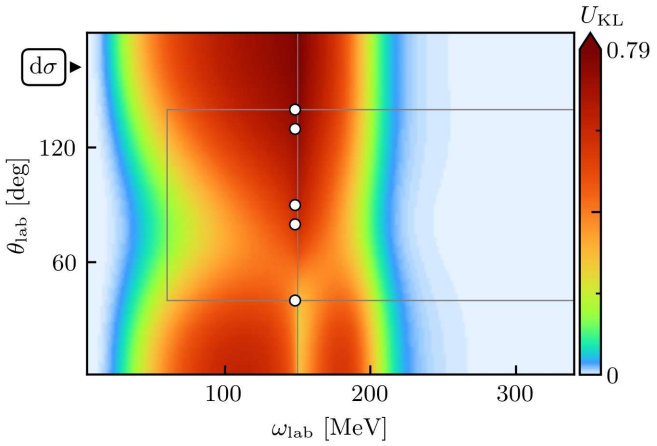

FIG. 10. (Colour online) The expected utility from measuring the differential cross section for the neutron while including truncation error. All neutron polarizabilities are included in this analysis. Its profile resembles that of the analogous proton observable, see fig. 5, though the color scales differ.

Gaußian processes efficiently and accurately account for correlations in the EFT truncation error, and impose the symmetry constraints on observables and their derivatives that must vanish, *e.g.*, at $\theta = 0°$ or $180°$. This physically motivated model is crucial to the study of experimental design with EFTs, as otherwise errors will be underestimated.

To facilitate reproduction and extension of our results, we provide all of the codes and data that generated our results [29].

Our Bayesian experimental design framework has the following benefits:

1. It can incorporate the effects of both experimental and theoretical uncertainties.

2. Its output contains both the optimal design and an estimate of the gain in information for that design which can be understood quite easily.

3. It can include the effects from measuring multiple kinematic points and can assess the interaction of multiple polarizabilities at once.

4. It permits a quantitative analysis of competing choices, *e.g.*, one can answer the question: should an experiment measure one point with high precision or many points with less precision?

5. Bayesian statistics mandates us to clearly specify our assumptions. Those who disagree with any assumption (size of error bars, priors on GP hyperparameters, the power counting in the transition region, design constraints, etc.) can readily modify our calculations, provided at [29], to their own specifications, thereby facilitating an ongoing dialog regarding the robustness of our experimental design results.

We also make the obvious point that while we have focused on Compton scattering experiments here, our EFT-based Bayesian approach to experimental design is easily adapted to other experiments informed by EFT calculations.

We tried to make a realistic assessment of experimental capabilities in this work, but realize that experiments can differ greatly. We reiterate that our analysis was meant to open the discussion on experimental design for Compton scattering, not provide the last word. In the future, our framework could be applied to a specific experiment at, *e.g.*, MAMI or HIγS, with additional details accounted for. For example, the analysis carried out here did not consider common-mode (correlated) errors, which can significantly affect the accuracy of an absolute measurement of the differential cross section. In fact, the correlation matrix of systematic errors across the space of data points could in general be quite complex. Including a more sophisticated treatment of such issues in the assessment of experimental designs would be one place where collaboration with experimental colleagues could be very fruitful. We emphasize that since we provide a Jupyter notebook for the experimental design problem, colleagues who wish to make different assumptions about the nature of correlations in errors between data taken in different kinematics can rerun the analysis with a statistical model improved thusly.

Bayesian experimental design can then determine the expected amount or quality of data required to reach a given level of polarizability precision within a particular experimental setup. However, we caution that such guidance is *always* in the light of particular prior information. Because experiments are assessed by the information gain they are expected to provide relative to a particular prior, that implies that a different choice of prior can produce different conclusions about which experiments will be most enlightening.

In this context, we point out that a fully consistent design analysis would begin with an extraction of scalar and spin polarizabilities from the present proton Compton scattering database that uses the error model (5) + (6), *i.e.* the same error model as used to assess the information gain of future experiments. Obtaining this six-dimensional probability distribution is an important topic for future work, as it would accurately represent the current state of knowledge of polarizabilities in χEFT. The prior we have used here only approximately represents that state. But the results of the design analysis are insensitive to the central values of the polarizabilities in that prior. And adopting different, reasonable correlations in the prior does not produce marked changes.

This framework can be extended to sequential designs, where experimental campaigns are split into a sequence of parts and the design of future experiments depends on the results of the initial experiments [17]. This "sequential experimental design" analysis can be computationally intensive and we have not attempted it here. Although we have found the assumption of

linearity to be good in this case, one could perform a full Bayesian experimental design if this assumption no longer holds [17, 67]. Our theoretical truncation estimates are the most comprehensive to date, but further study of chiral EFT convergence for Compton observables should be performed. These are all tasks for future work.

ACKNOWLEDGMENTS

We thank Ian Vernon for useful discussions, and M. Ahmed, E. Downie, G. Feldman, P. P. Martel, as well as the MAMI-A2/CB and Compton@HIγS teams for their patience in discussing experimental constraints. We gratefully acknowledge the stimulating atmosphere created by organizers and participants of the workshops Uncertainty Quantification at the Extremes (ISNET-6) at T.U. Darmstadt (Germany) and Bayesian Inference in Subatomic Physics - A Wallenberg Symposium (ISNET-7) at Chalmers U. (Göteborg, Sweden), which triggered and expanded these investigations. H.W.G. gratefully acknowledges the warm hospitality and financial support of the A2/Crystall Ball Collaboration Meeting 2020 at MAMI (U. Mainz, Germany), of both Ohio University and the Ohio State University, and of the University of Manchester, where part of this work was conducted. The work of R.J.F. and J.A.M. was supported in part by the National Science Foundation under Grant Nos. PHY–1614460 and PHY–1913069 and the NUCLEI SciDAC Collaboration under US Department of Energy MSU subcontract RC107839-OSU. The work of D.R.P. was supported by the US Department of Energy under contract DE-FG02-93ER-40756 and by the ExtreMe Matter Institute EMMI at the GSI Helmholtzzentrum für Schwerionenphysik, Darmstadt, Germany. The work of H.W.G. was supported in part by the US Department of Energy under contract DE-SC0015393, by the High Intensity Gamma-Ray Source HIγS of the Triangle Universities Nuclear Laboratory TUNL in concert with the Department of Physics of Duke University, and by The George Washington University: by the Dean's Research Chair programme and an Enhanced Faculty Travel Award of the Columbian College of Arts and Sciences; and by the Office of the Vice President for Research and the Dean of the Columbian College of Arts and Sciences. His work was conducted in part at GW's Campus in the Closet. The work of J.McG. was supported by the UK Science and Technology Facilities Council grant ST/P004423/1. The work of M.T.P. was supported in part by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-2018-CRG7-3800.3.

Appendix A: Experimental Design Details

Suppose that our theoretical model $y_k(x;\vec{a})$ is related to measurements $y_{\rm exp}(x)$ via additive theoretical and experimental noise, as in eq. (5). We can linearize $y_k(x;\vec{a})$ about some point $\vec{a}_\star$ by keeping only the first order terms in its Taylor expansion, *i.e.*,

$$\begin{aligned} y_k(x;\vec{a}) &\approx y_k(x;\vec{a}_\star) + \sum_i b_i(x)[\vec{a}_i - \vec{a}_\star] \\ &= c(x;\vec{a}_\star) + \vec{b}(x)\cdot\vec{a}\,, \end{aligned} \tag{A1}$$

where $\vec{b}(x) \equiv \partial y_k(x;\vec{a})/\partial\vec{a}$ evaluated at $\vec{a}_\star$ are our basis functions and $c(x;\vec{a}_\star) \equiv y_k(x;\vec{a}_\star) - \vec{b}\cdot\vec{a}_\star$ is constant with respect to the polarizabilities $\vec{a}$ but depends on the kinematic point x. Thus, the vector of N measurements $\mathbf{y}$ is related to the polarizabilities via the likelihood

$$\mathbf{y}\,|\,\vec{a} \sim \mathcal{N}[B\vec{a} + \mathbf{c}, \Sigma] \tag{A2}$$

where $B \equiv \vec{b}(\mathbf{x})$ is an $N\times 6$ matrix, $\mathbf{c} \equiv c(\mathbf{x};\vec{a}_\star)$ is a length N vector, and Σ is the $N\times N$ covariance matrix due to theoretical and experimental error. That is, given some experimental covariance $\Sigma_{\rm exp}$ and a theoretical covariance $\bar{c}^2 R_{\delta k}$ from eqs. (13) and (14), then

$$\Sigma = \bar{c}^2 R_{\delta k} + \Sigma_{\rm exp}\,. \tag{A3}$$

Note that $R_{\delta k}$ depends on the values of the tuned ℓ_ω and ℓ_θ, whose estimates from the order-by-order convergence pattern are given in table I.

The linear model of eq. (A2) is well known in the statistics literature [68, 69], so here we will simply state the relevant results. If a Gaußian prior with mean $\vec{\mu}_0$ and covariance V_0 is placed on $\vec{a}$ as in eq. (20), then the resulting posterior is also Gaußian, with mean and covariance given by

$$\vec{\mu} = V\left[V_0^{-1}\vec{\mu}_0 + B^\intercal\Sigma^{-1}(\mathbf{y}-\mathbf{c})\right]\,, \tag{A4}$$

$$V = (V_0^{-1} + B^\intercal\Sigma^{-1}B)^{-1}\,. \tag{A5}$$

Importantly to our study of experimental design, the posterior covariance V depends on the kinematic points $\mathbf{x}$ where the experiment is performed, and on the specifics of the observable through Σ, but not on the exact results of the experiment $\mathbf{y}$.

Given that we choose to maximize the expected information gain in the polarizabilities, then the integrals of eq. (22) must still be performed. The integral over $\vec{a}$ splits into the difference of two terms: the differential entropy of the prior $\mathrm{pr}(\vec{a})$ and of the posterior $\mathrm{pr}(\vec{a}\,|\,\mathbf{y},\mathbf{d})$. The differential entropy of a Gaußian $\mathcal{N}(\mu,\Sigma)$ is well

known to be $\frac{1}{2}\ln|2\pi e\Sigma|$. Therefore

$$\begin{aligned} U_{\mathrm{KL}}(\mathbf{d}) &= -\int \ln[\mathrm{pr}(\vec{a})]\,\mathrm{pr}(\vec{a})\,\mathrm{d}\vec{a} \\ &\quad + \int \ln[\mathrm{pr}(\vec{a}\,|\,\mathbf{y},\mathbf{d})]\,\mathrm{pr}(\vec{a}\,|\,\mathbf{y},\mathbf{d})\,\mathrm{d}\vec{a}\,\mathrm{pr}(\mathbf{y}\,|\,\mathbf{d})\,\mathrm{d}\mathbf{y} \\ &= \frac{1}{2}\ln|2\pi e V_0| - \frac{1}{2}\ln|2\pi e V| \int \mathrm{pr}(\mathbf{y}\,|\,\mathbf{d})\,\mathrm{d}\mathbf{y} \\ &= \frac{1}{2}\ln\frac{|V_0|}{|V|}\,, \end{aligned} \tag{A6}$$

where we used the fact that V does not depend on $\mathbf{y}$ and then performed the trivial integration over all possible measurements $\mathbf{y}$.

Appendix B: Observable Constraints and EFT Truncation Model Details

Constraints on Compton observables are discussed in detail in ref. [6]. Some of this is reproduced here, with particular attention paid to nth-order chiral *corrections* to observables Δy_n rather than the value of the observable y itself. The Δy_n impact the distribution for the χEFT uncertainty δy_k, but because we restrict the "experimentally accessible regime" in this study from small ω, and forwards/backwards angles, these constraints are not as important as they otherwise would be. These constraints are summarized for particular ω_{lab} and θ_{lab} values in table V.

All observables that are nonzero below ω_π approach the Thomson limit as $\omega \to 0$ [1]. Thus, higher-order corrections must vanish there , and approach $\omega = 0$ as at least ω^2. Therefore, at least the first derivative of all corrections must vanish there as well.

The remaining observables must vanish for $\omega \leq \omega_\pi$, but there is no constraint on the the derivative of corrections at $\omega = \omega_\pi$. We have found that the corrections approach 0 *very* quickly, so that imposing the constraint $\Delta y_n(\omega_\pi, \theta) = 0$ for all higher order terms is actually a worse approximation than not imposing the constraint at all; see, *e.g.*, fig. S.4 in the Supplemental Material. This comes back to the large cusps in the spin-observable c_n found near ω_π, discussed in sec. III B 1, which remain an unresolved aspect of this model.

Due to the coordinate singularity at $\theta = 0°$ and 180°, observables or their derivative with respect to θ must vanish there [6]. But this does not preclude *both* the value and their derivatives from vanishing there. These constraints can be deduced by symmetry arguments, and are summarized in table V.

The hyperparameters $\bar{c}^2$ and ℓ_i, shown in table I, are tuned to coefficients c_n at the best known $\vec{a}$ (see table III) for $\Lambda_b = 650\,\mathrm{MeV}$. The training data are on a grid with $\theta_{\mathrm{lab}} = \{30°, 50°, 70°, 90°, 110°, 130°\}$ and $\omega_{\mathrm{lab}} = \{200, 225, 250\}\,\mathrm{MeV}$ for observables which are zero below ω_π. For observables that are non-zero below ω_π, the additional training points $\omega_{\mathrm{lab}} = \{50, 75, 100, 125\}\,\mathrm{MeV}$ are included, and common ℓ_ω and ℓ_θ are used between the two regions. The training region is well outside the kinematic endpoints where additional constraints arise on observables or their derivatives, and excludes the pion-production threshold region.

Because the first nonzero order often behaves differently than the corrections, we do not use it for induction on the c_n; that is we only train the hyperparameters on *corrections*. Hence, we train on c_2–c_4 for $\mathrm{d}\sigma$ and Σ_3, but otherwise we train on c_3 and c_4.

The coefficients for various observable slices are shown in figs. 12 for the cross section, and figs. S.1–S.6 of the Supplemental Material for the spin observables. These plots also include uncertainty bands for higher order coefficients, with the symmetry constraints given in table V included. These constraints on both the coefficient functions and their derivatives propagate directly to the truncation error δy_k by replacing $r(x, x'; \ell_\omega, \ell_\theta)$ in eq. (14) by its *conditional* form $\tilde{r}(x, x'; \ell_\omega, \ell_\theta)$, see refs. [27, 45]. For example, if the value of c_n is known at the set of points $\mathbf{x}$, then one can compute its conditional GP, with covariance kernel given by

$$\tilde{r}(x, x') = r(x, x') - r(x, \mathbf{x}) r(\mathbf{x}, \mathbf{x})^{-1} r(\mathbf{x}, x')\,. \tag{B1}$$

See refs [70–72] for details about adding derivative observations to GPs. Because the RBF kernel [eq. (11)] is separable in ω and θ, these constraints can simply be applied to each one-dimensional kernel separately, and multiplied to yield the total constrained kernel. We employ the `gptools` python package for easily implementing derivative constraints [73].

For completeness, we also provide the profile for the truncation error standard deviation (up to factors of $\bar{c}$, which vary by observable); see fig. 11. It assumes the form of Q provided in eq. (8) along with the first omitted χEFT order given in eq. (18).

This allows us to return to the discussion of the omitted constraints $\Sigma_i \in [-1, 1]$ on the spin observables in sec. IV A. Over the physically interesting kinematic range, the actual value of most spin observables lies in the much more narrow interval $[-0.7, 0.7]$; see fig. 5 in ref. [6]. So, then the question becomes: are the mean prediction *and* its theory uncertainty contained in $[-1, 1]$ with a high degree of probability? From eq. (12), one can see that the 1σ interval for the truncation error δy_k is $y_{\mathrm{ref}}\bar{c}$ times another factor $Q^{\nu_{\delta k}(\omega)}/\sqrt{1 - Q^2(\omega)}$. Here $y_{\mathrm{ref}} = 1$ and $\bar{c} \lesssim 0.7$ for most spin observables (see table I). The third factor is plotted in fig. 11 and does not exceed ≈ 0.3 at $\omega \lesssim 230\,\mathrm{MeV}$, where our analysis shows the biggest sensitivities. Therefore, even for the spin observables with large magnitudes, the 1σ upper range of a GP will only give values about 0.2 larger than the established maximum of 0.7, namely about 0.9 in total. This is close but still below $|\Sigma_i| = 1$. Therefore, a majority of our test functions in the GP will not probe, let alone exceed, the strict bounds on those spin observables. Furthermore, if observables and their truncation errors vanish at $\theta = 0°$ or 180°, this will make the constraint even more

TABLE V. The constraints on *corrections* to observables Δy and their derivatives $\Delta y'$ at particular θ and ω. The LO amplitude as well as all calculated higher orders fulfill them automatically, so these must only be enforced in the GP. The observables marked by a dagger † are zero below the pion-production threshold, but we impose no constraint on them at $\omega = \omega_\pi$, as discussed in the text.

	θ [deg]		ω [MeV]	
	$\Delta y = 0$	$\Delta y' = 0$	$\Delta y = 0$	$\Delta y' = 0$
$\mathrm{d}\sigma$	—	0, 180	0	0
Σ_{1x}	0, 180	—	†	†
Σ_{1z}	0, 180	0, 180	†	†
Σ_{2x}	0, 180	—	0	0
Σ_{2z}	—	0, 180	0	0
Σ_3	0, 180	0, 180	0	0
Σ_y	0, 180	—	†	†
Σ_{3y}	0, 180	—	†	†
$\Sigma_{3y'}$	0, 180	—	†	†
$\Sigma_{1x'}$	0, 180	0	†	†
$\Sigma_{1z'}$	0, 180	180	†	†
$\Sigma_{2x'}$	180	0	0	0
$\Sigma_{2z'}$	0	180	0	0

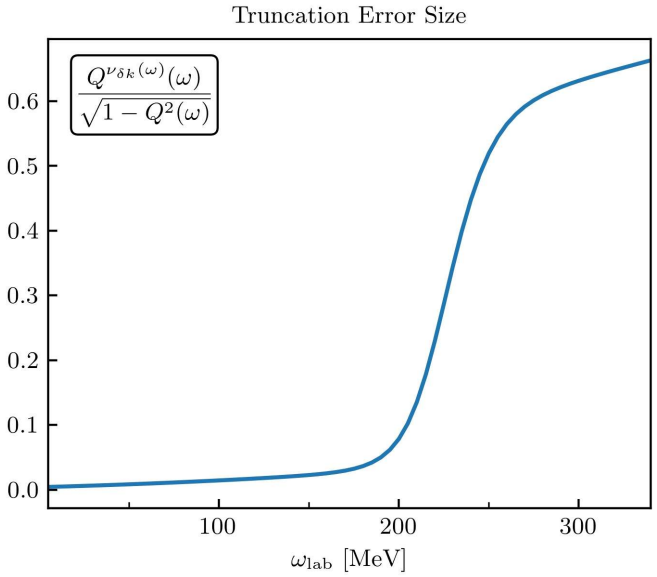

FIG. 11. (Colour online) A component of the standard deviation due to χEFT uncertainty at $\mathrm{N^4LO^+}$, see eq. (14). The factor of $\bar{c}$ is unique to each observable, and is not included. See table I.

trivially satisfied near these regions. We are therefore confident that implementing the constraint $\Sigma_i \in [-1, 1]$ would not impact our results for $\omega \lesssim 220\,\mathrm{MeV}$, and cautiously optimistic that the impact would be small even at higher energies.

Though we likewise do not constrain the cross section to be non-negative, we are confident that within our constrained angle range, corrections are highly unlikely to be large enough for this to be a worry. According to fig. 4 in ref. [6], the $\mathrm{N^4LO}$ cross section is small ($< 10\,\mathrm{nb/sr}$) in a narrow region at forward angles around ω_π. Figure 11 shows that the expansion parameter is small, and fig. 12 shows that the coefficients c_i are natural-sized. Therefore, the GP corrections are highly unlikely to exceed the size of the predicted cross section and create negative (unphysical) values.

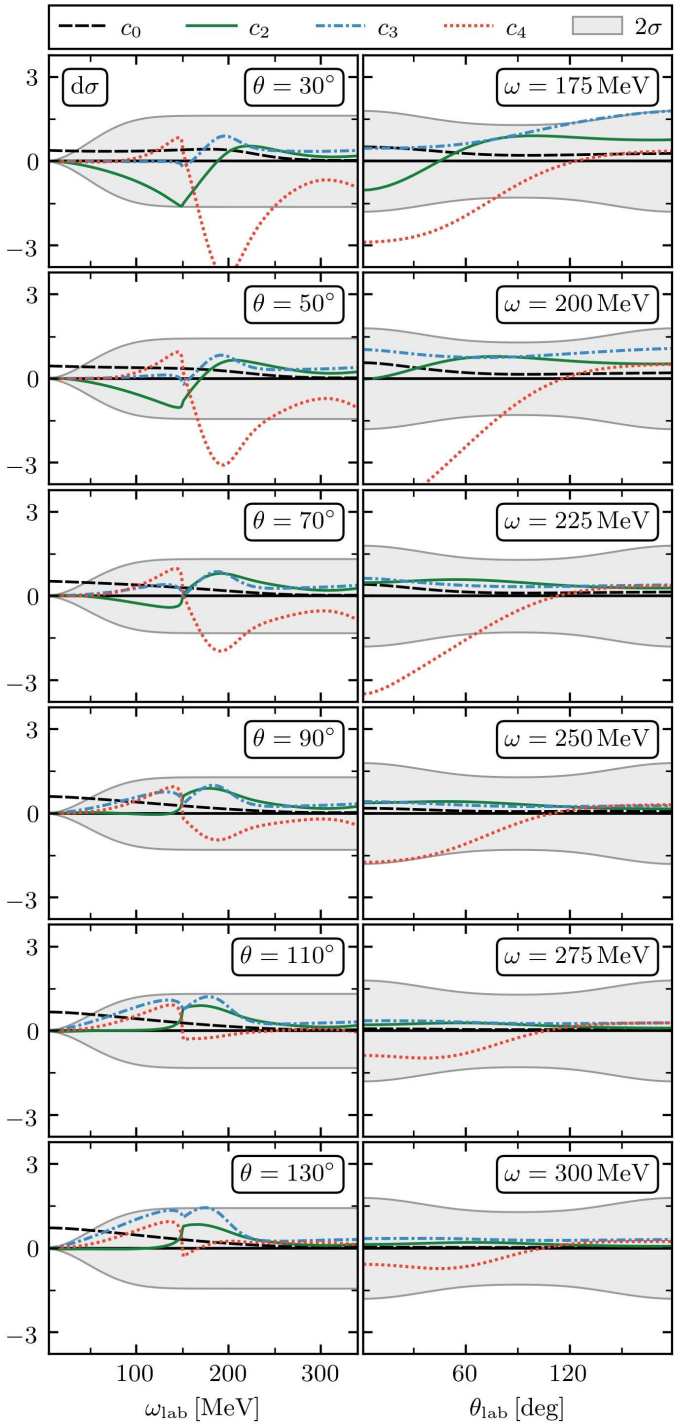

FIG. 12. (Colour online) Coefficients for the differential cross section $\mathrm{d}\sigma$.

Appendix S: Supplemental Material

1. Coefficients and Uncertainties for the Proton Spin Observables

In this section, we present plots of the sizes of the coefficients c_i as function of ω and slices of θ, and vice versa, as figs. S.1 to S.6. The coefficients for the cross section are shown in fig. 12 of App. B.

2. Proton Observables for an Uncorrelated Prior

In this section, we present results for the analysis of proton observables where we neglect the correlations between $\alpha_{E1} - \beta_{M1}$ and γ_{M-}, *i.e.* we neglect the off-diagonal elements in eq. (25).

Figures S.7, S.8, and S.9 are the analogs of figs. 6, 7, and 8 but for this uncorrelated prior.

3. Proton Observables with Other Experimental Precision

Figure S.10 compares the maximal information gain in each observable for "standard", "doable" and "aspirational" experiments. Not surprisingly, data with aspirational experimental error bars are far superior to those with only standard ones. If theory errors were absent, one would naively assume the information gain of the scenarios to scale roughly like $1/\sqrt{\Delta\Sigma_i}$. This appears to be largely fulfilled, except for $\Sigma_{2x'}$ and, less noticeably, $\Sigma_{3y'}$.

Figures S.11 and S.12 are the analogs of figs. 6 and 7 in the main text, but this time without any χEFT truncation error included. Figures S.13, S.14, and S.17 are the analogs of figs. 6, 7, and 8 but for the "standard" level of precision, rather than the "doable" one employed for results in the main text. Figures S.15, S.16, and S.18 show the corresponding results for the "aspirational" precision level.

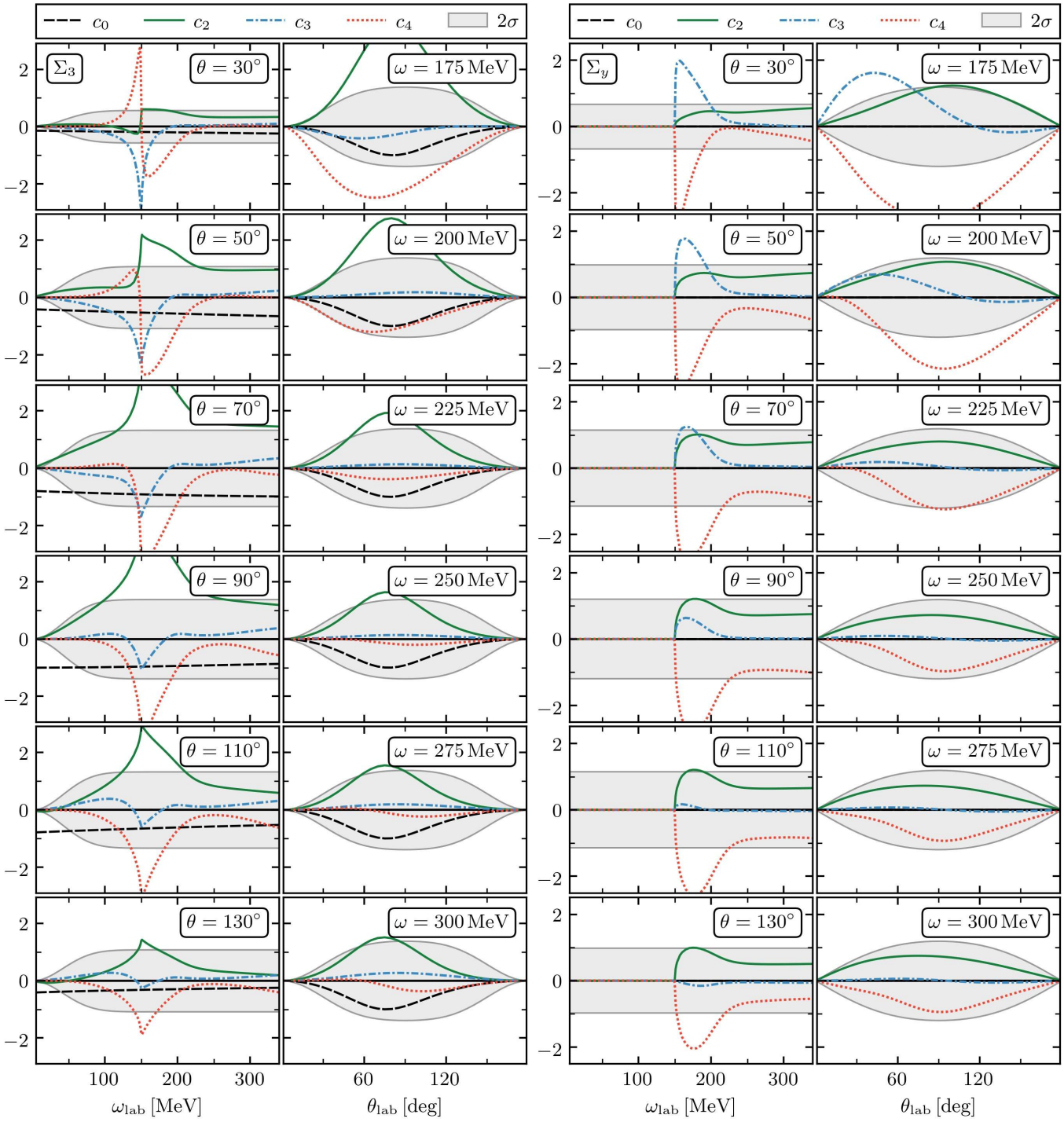

FIG. S.1. (Colour online) Coefficients for Σ_3 and Σ_y.

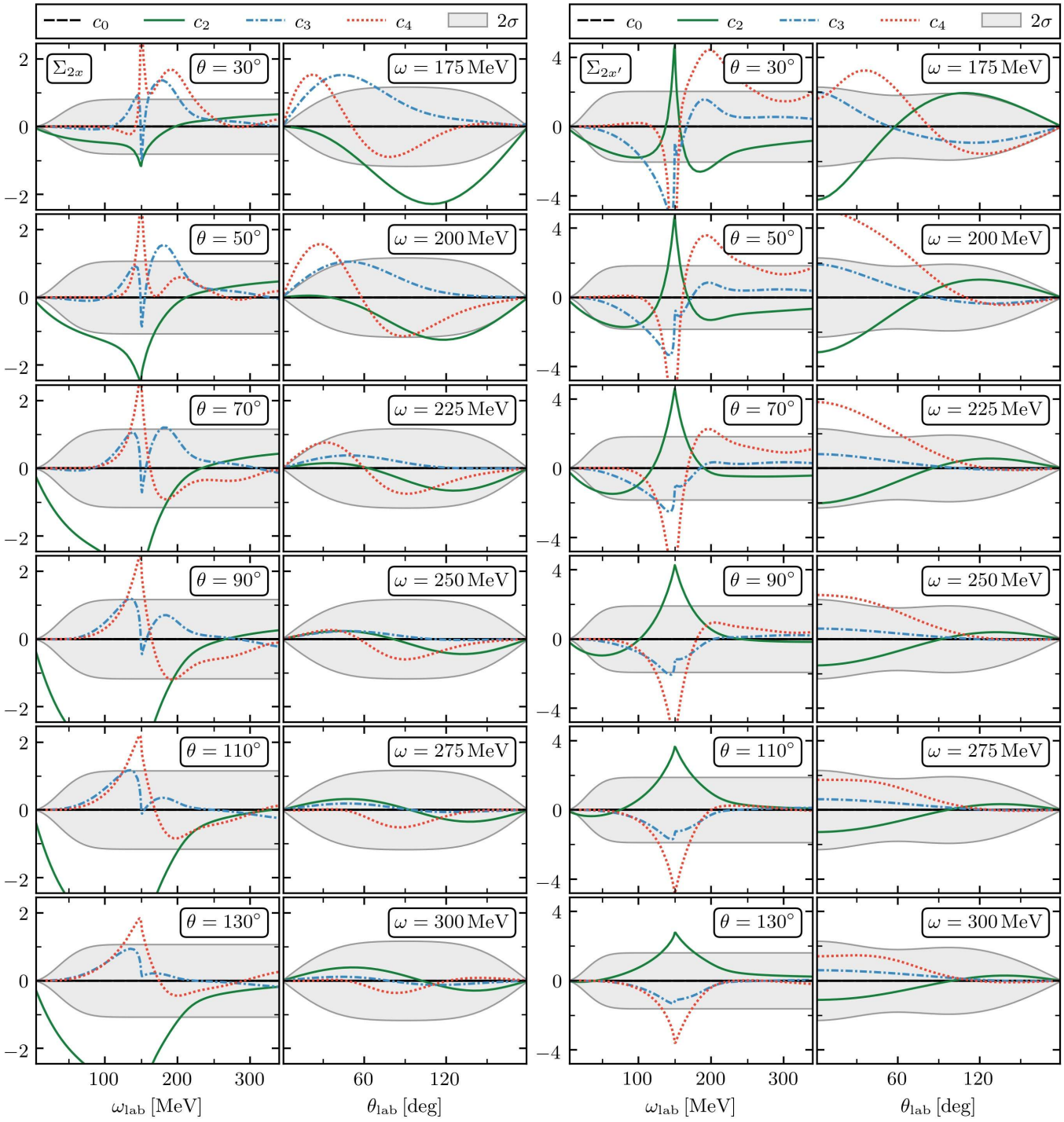

FIG. S.2. (Colour online) Coefficients for Σ_{2x} and $\Sigma_{2x'}$.

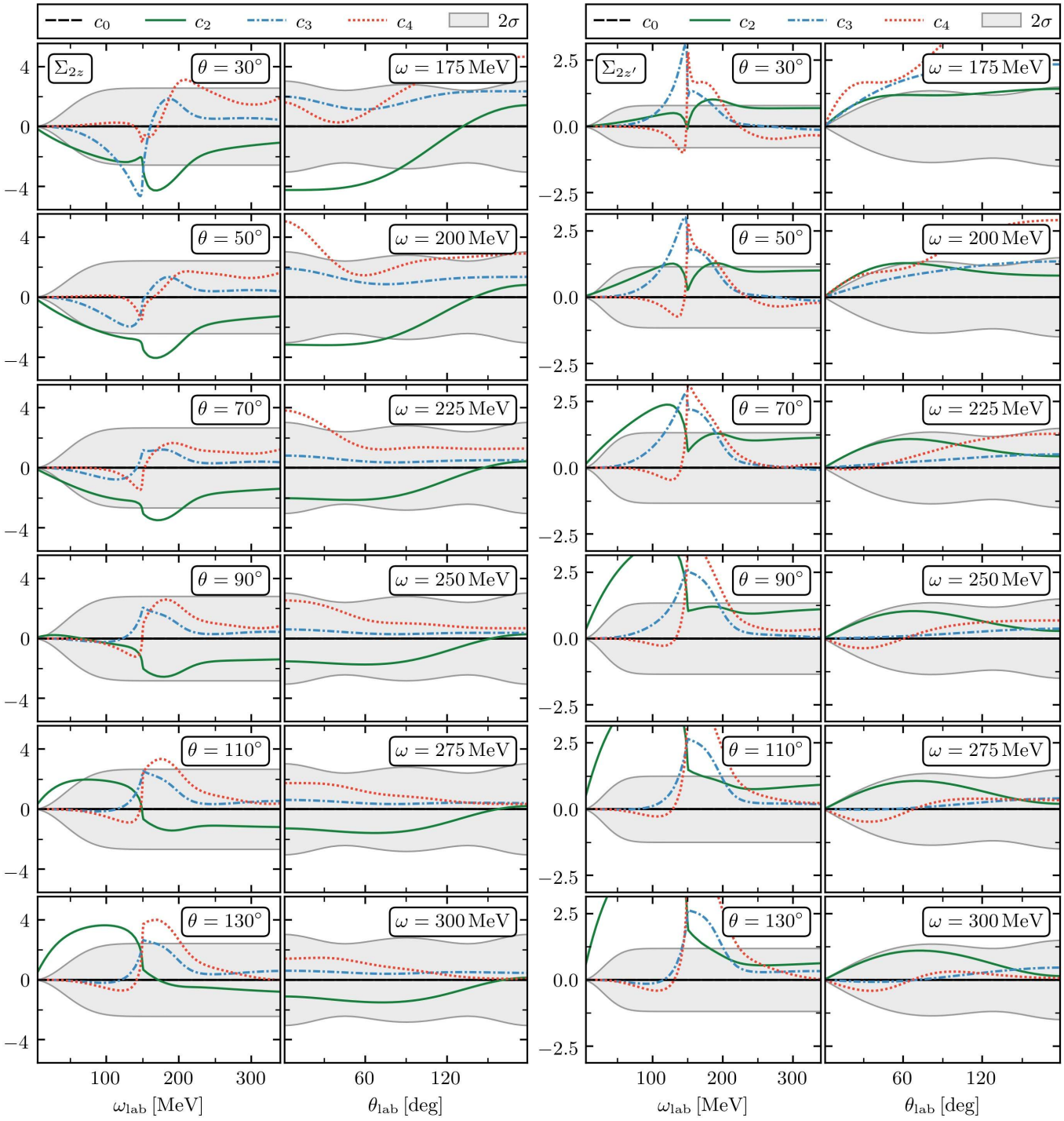

FIG. S.3. (Colour online) Coefficients for Σ_{2z} and $\Sigma_{2z'}$.

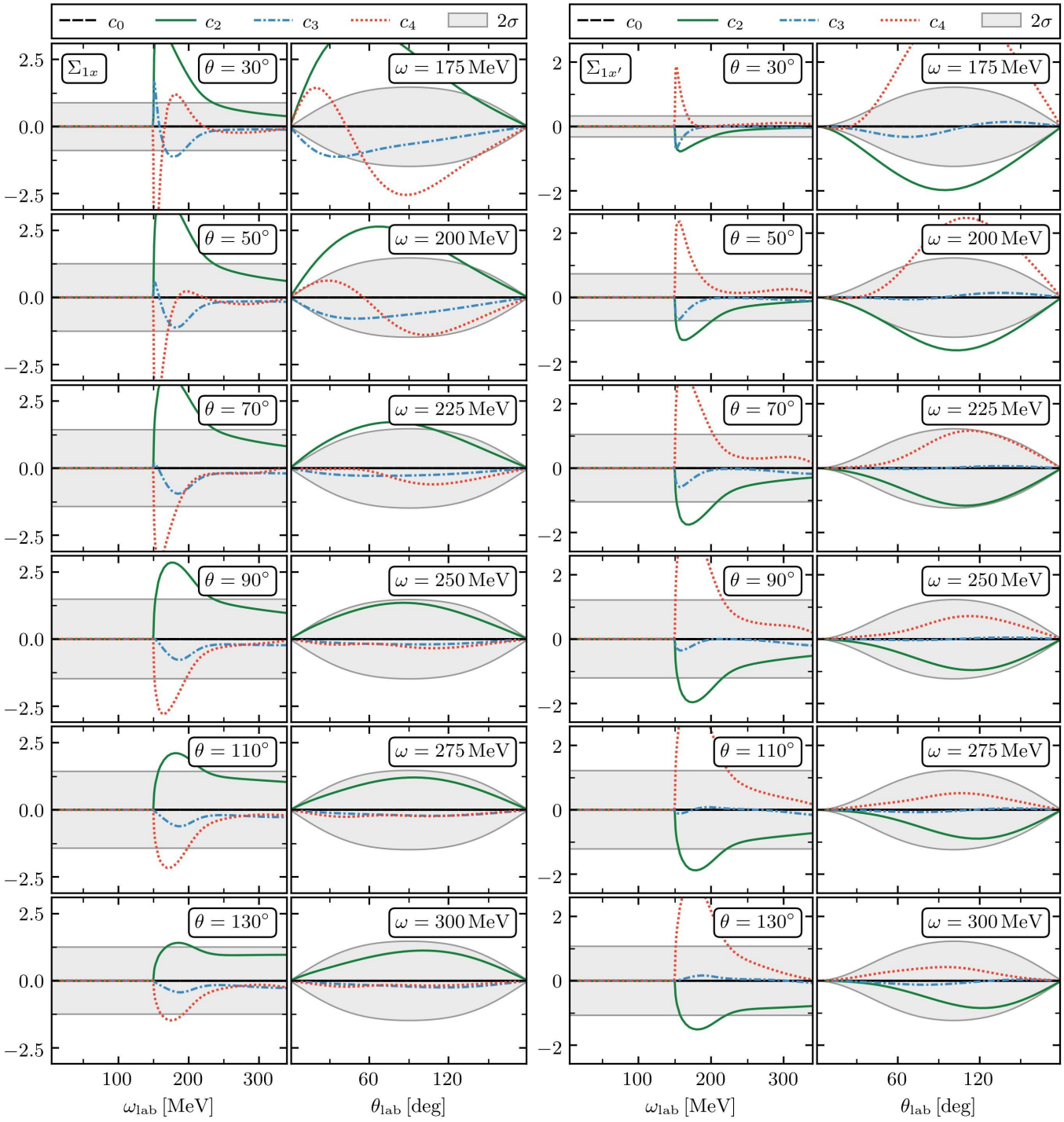

FIG. S.4. (Colour online) Coefficients for Σ_{1x} and $\Sigma_{1x'}$.

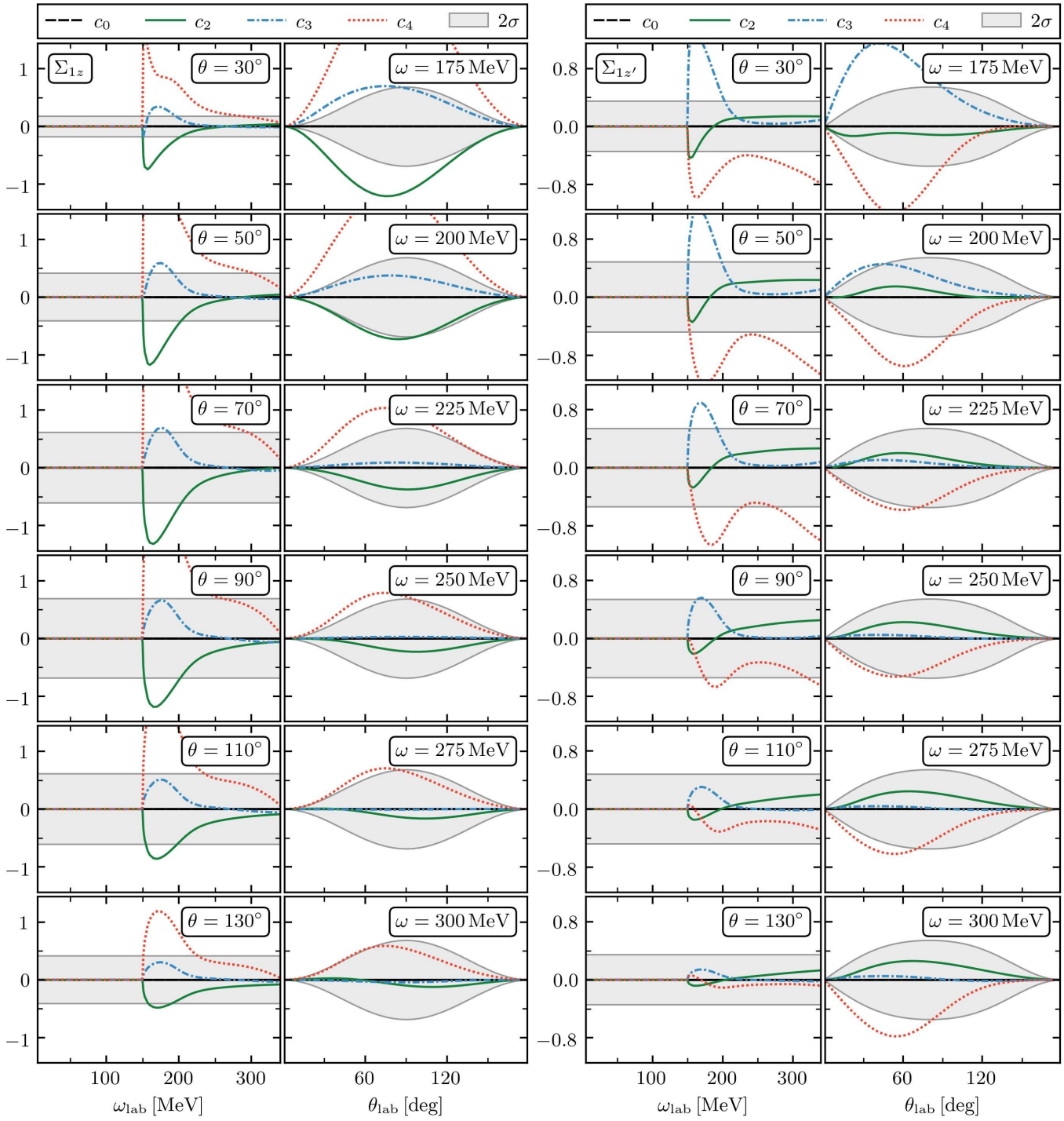

FIG. S.5. (Colour online) Coefficients for Σ_{1z} and $\Sigma_{1z'}$.

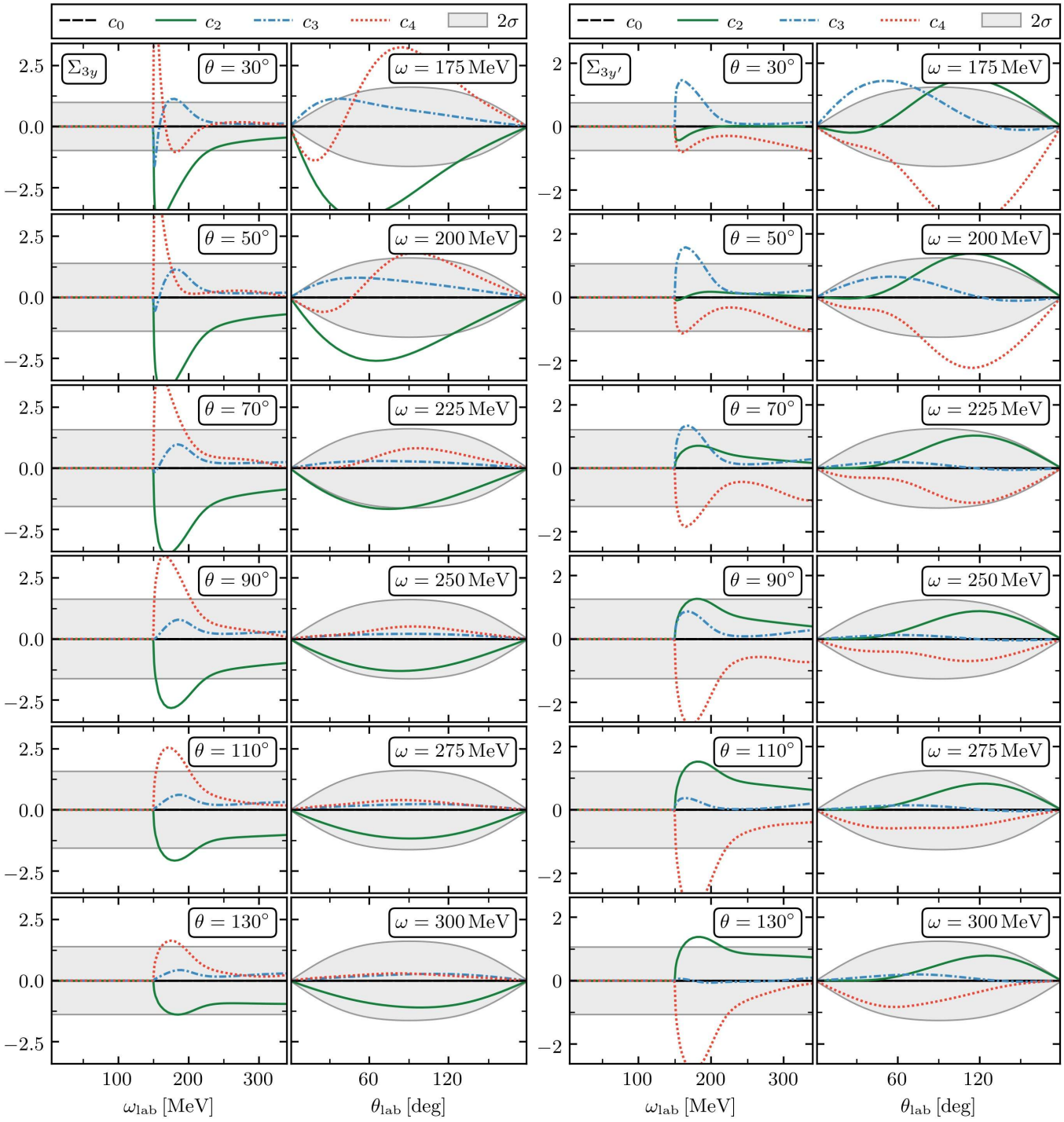

FIG. S.6. (Colour online) Coefficients for Σ_{3y} and $\Sigma_{3y'}$.

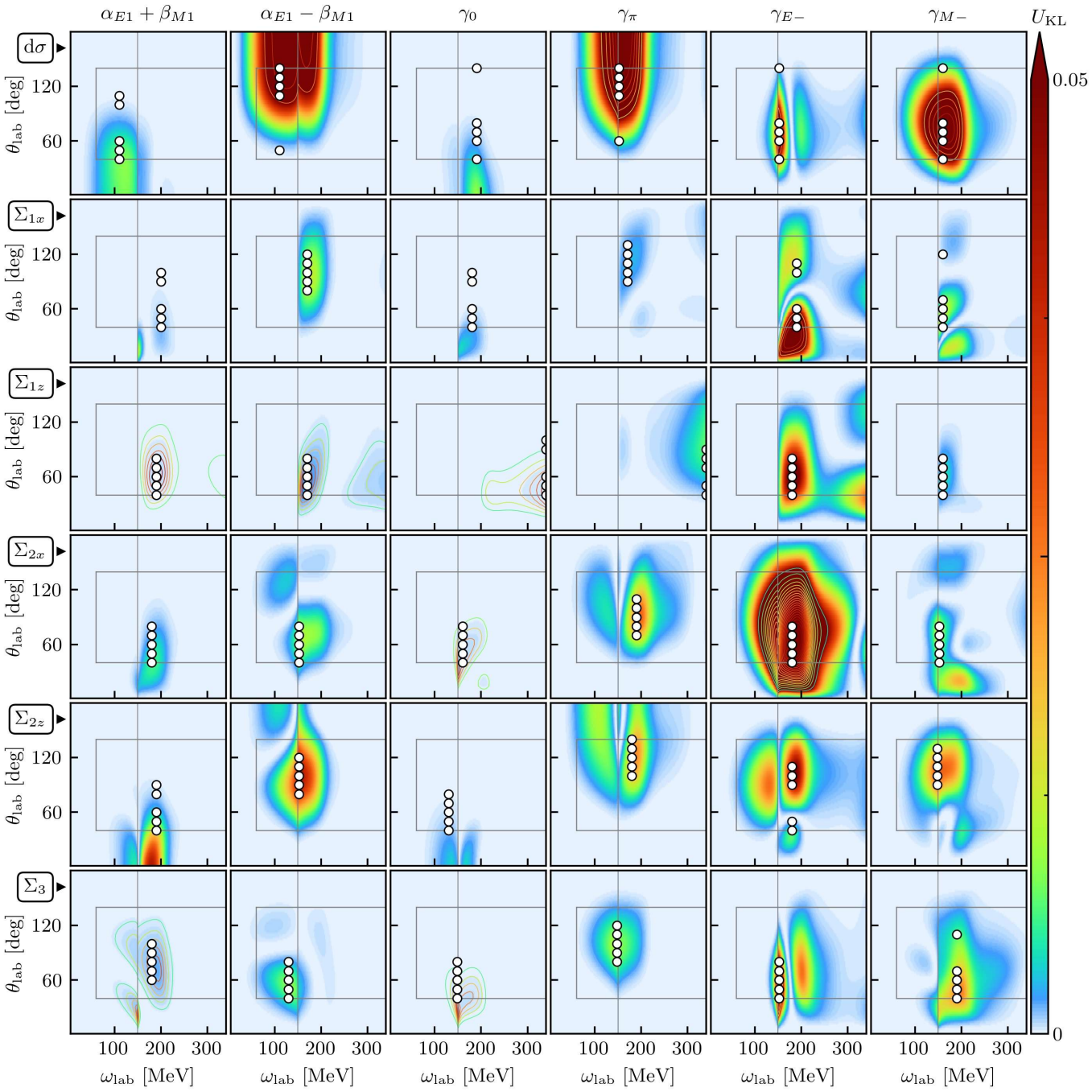

FIG. S.7. (Colour online) As in fig. 6, with "doable" experimental precision, but for the case of an uncorrelated prior.

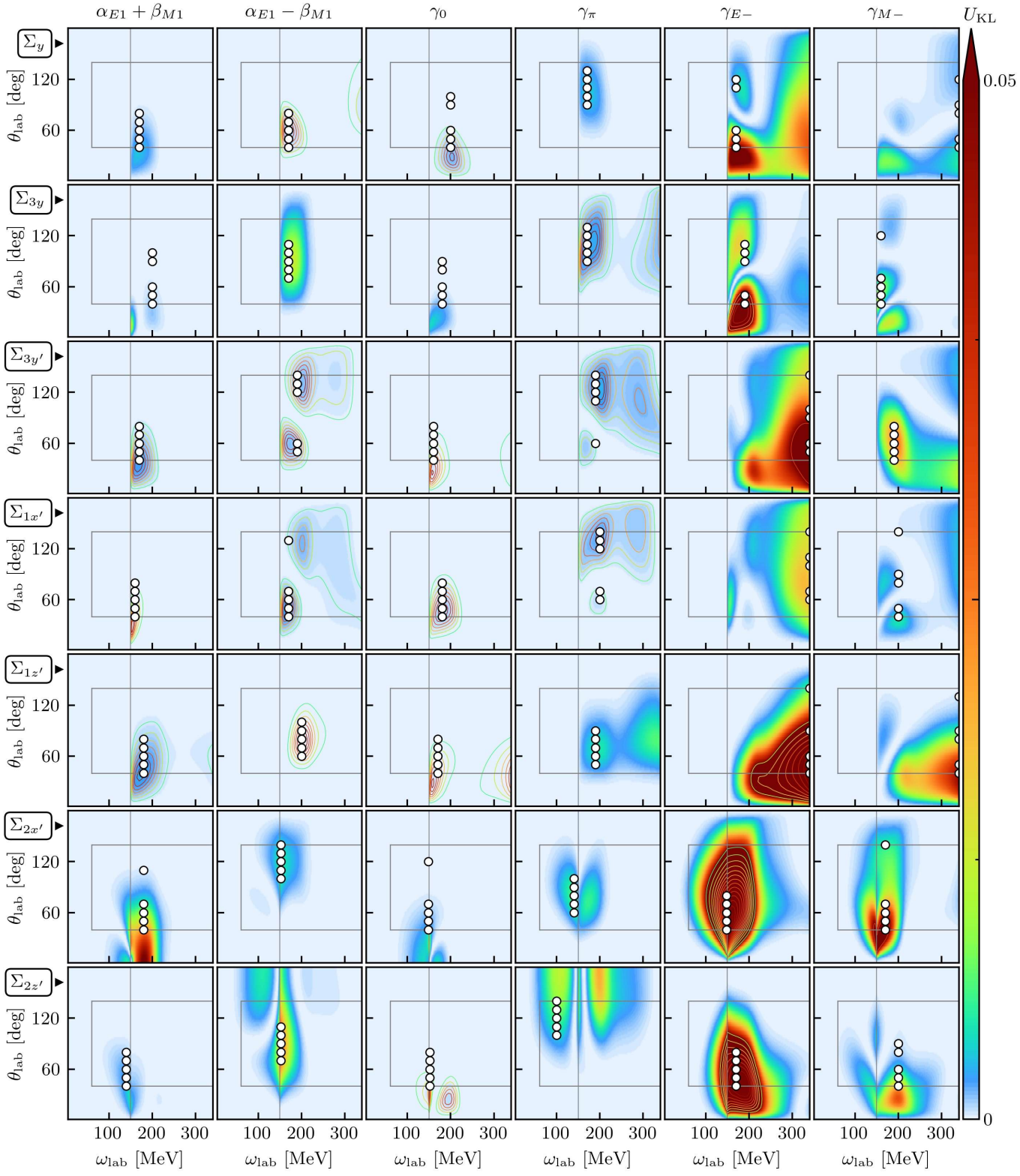

FIG. S.8. (Colour online) As in fig. 7, with "doable" experimental precision, but for the case of an uncorrelated prior.

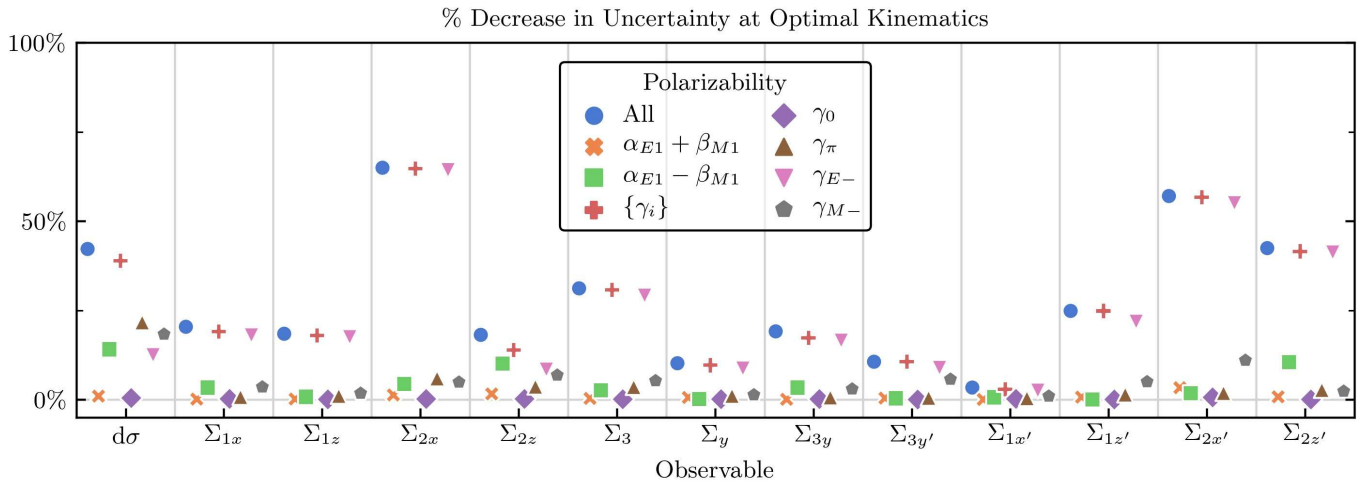

FIG. S.9. (Colour online) As in fig. 8, but for the case of an uncorrelated prior.

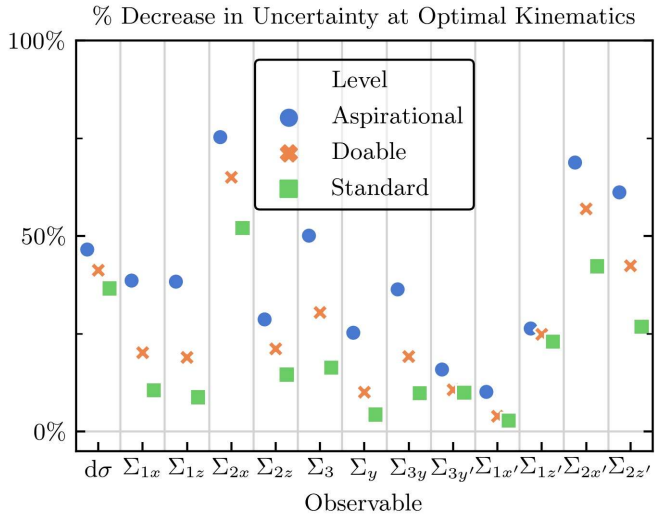

FIG. S.10. (Colour online) Comparison of the shrinkage power of the optimal designs for each precision level. Note that the optimal design likely differs between precision levels.

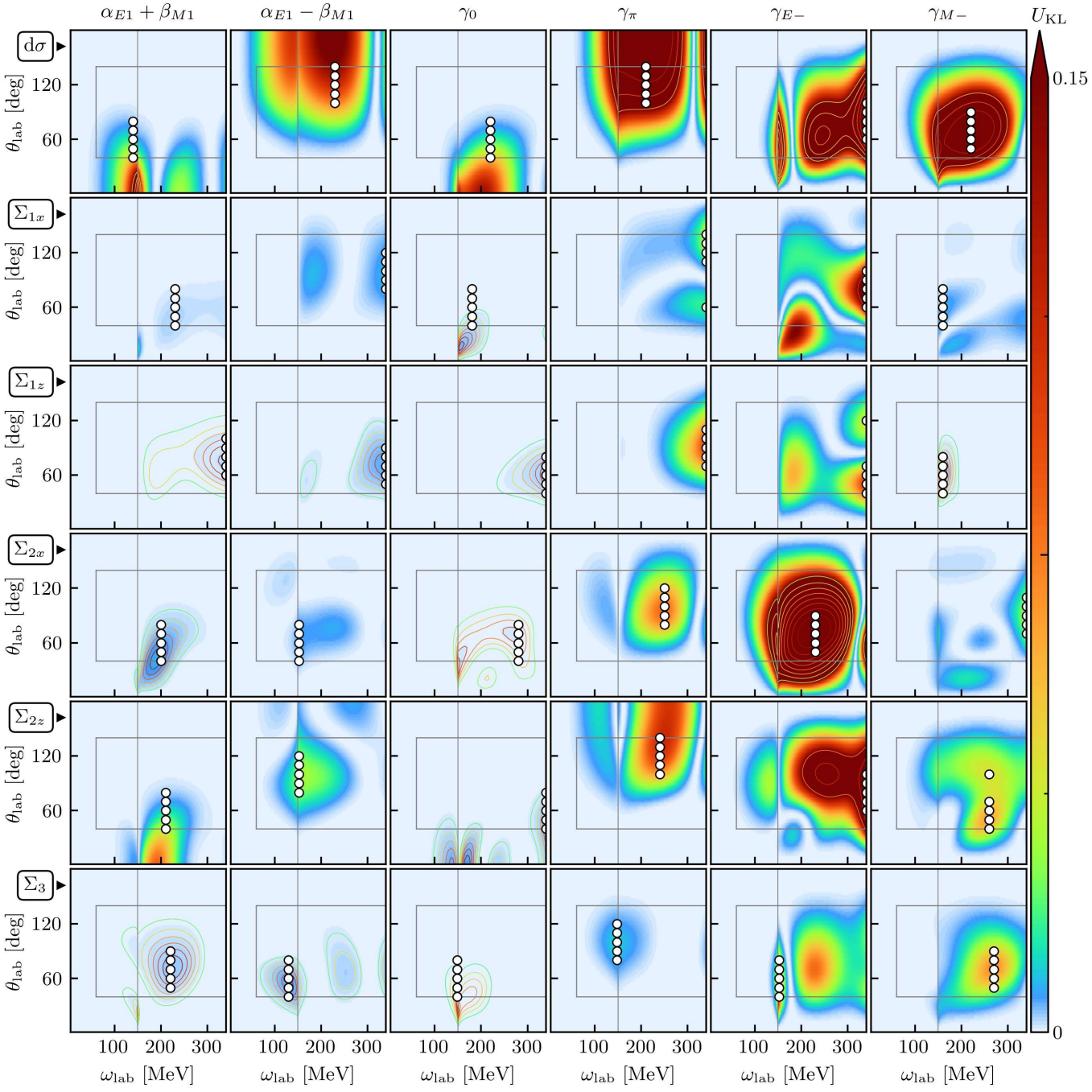

FIG. S.11. (Colour online) As in fig. 6, with "doable" experimental precision, but without including an estimate of truncation error.

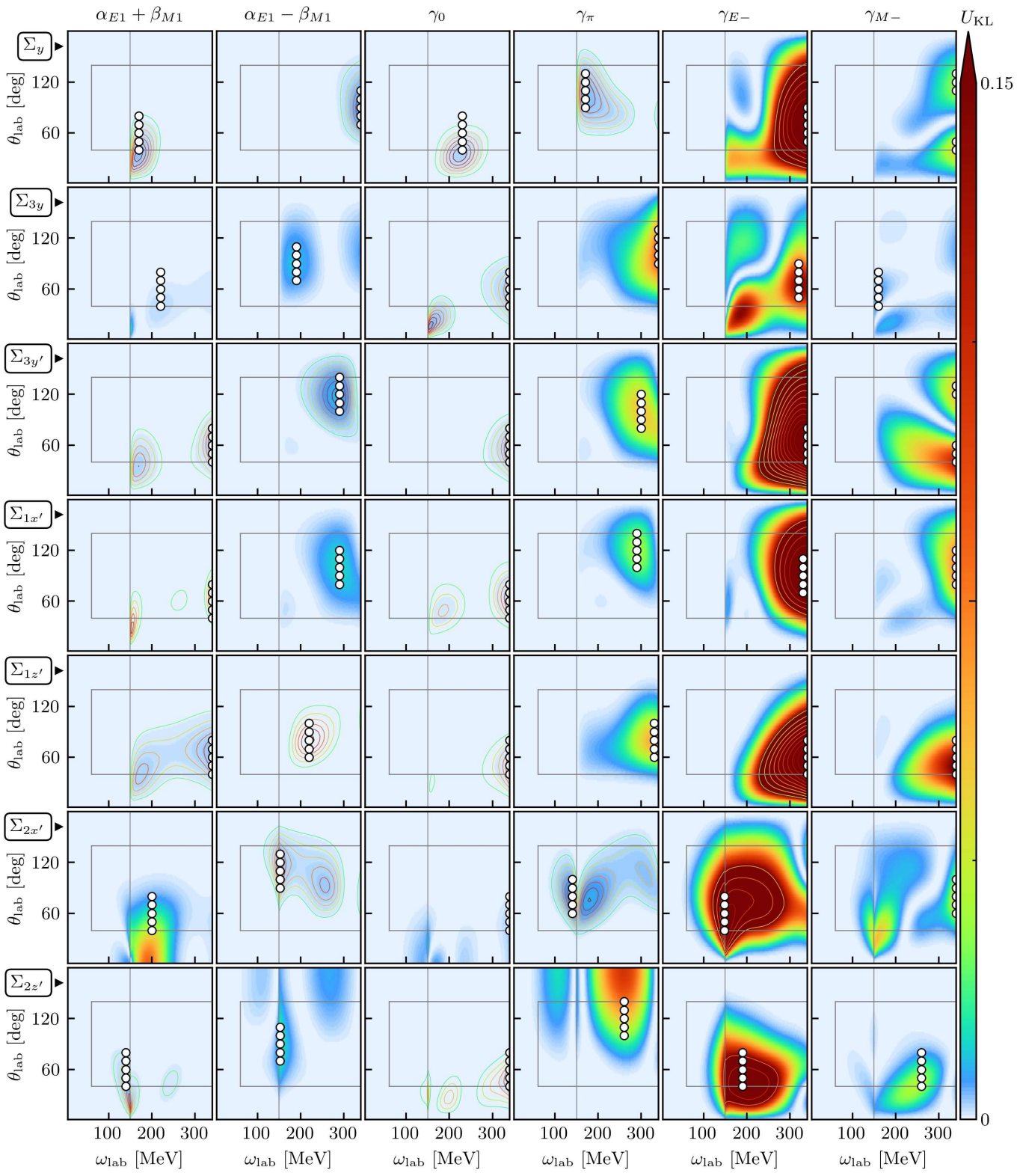

FIG. S.12. (Colour online) As in fig. 7, with "doable" experimental precision, but without including an estimate of truncation error.

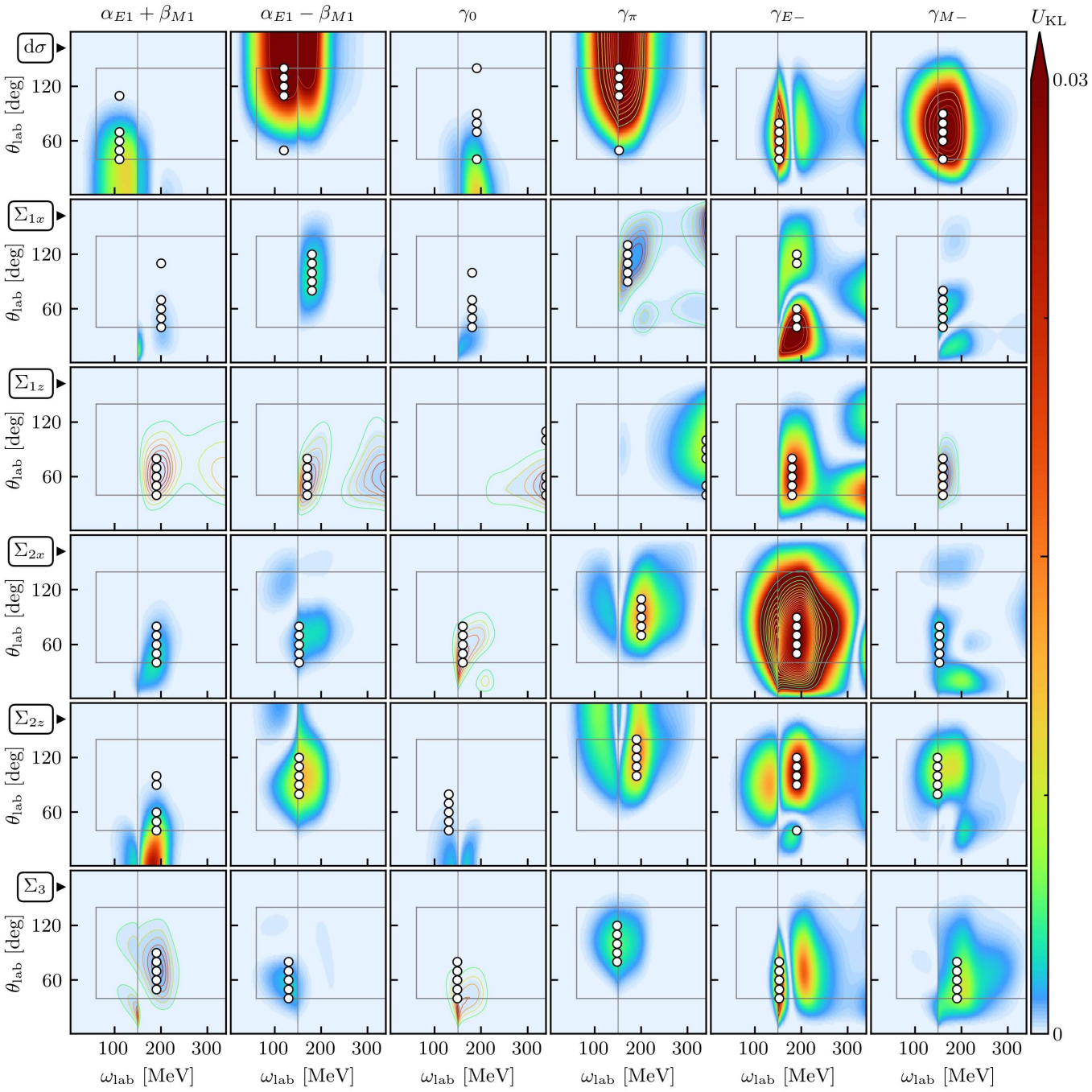

FIG. S.13. (Colour online) As in fig. 6, but with the "standard" level of experimental precision; see table IV.

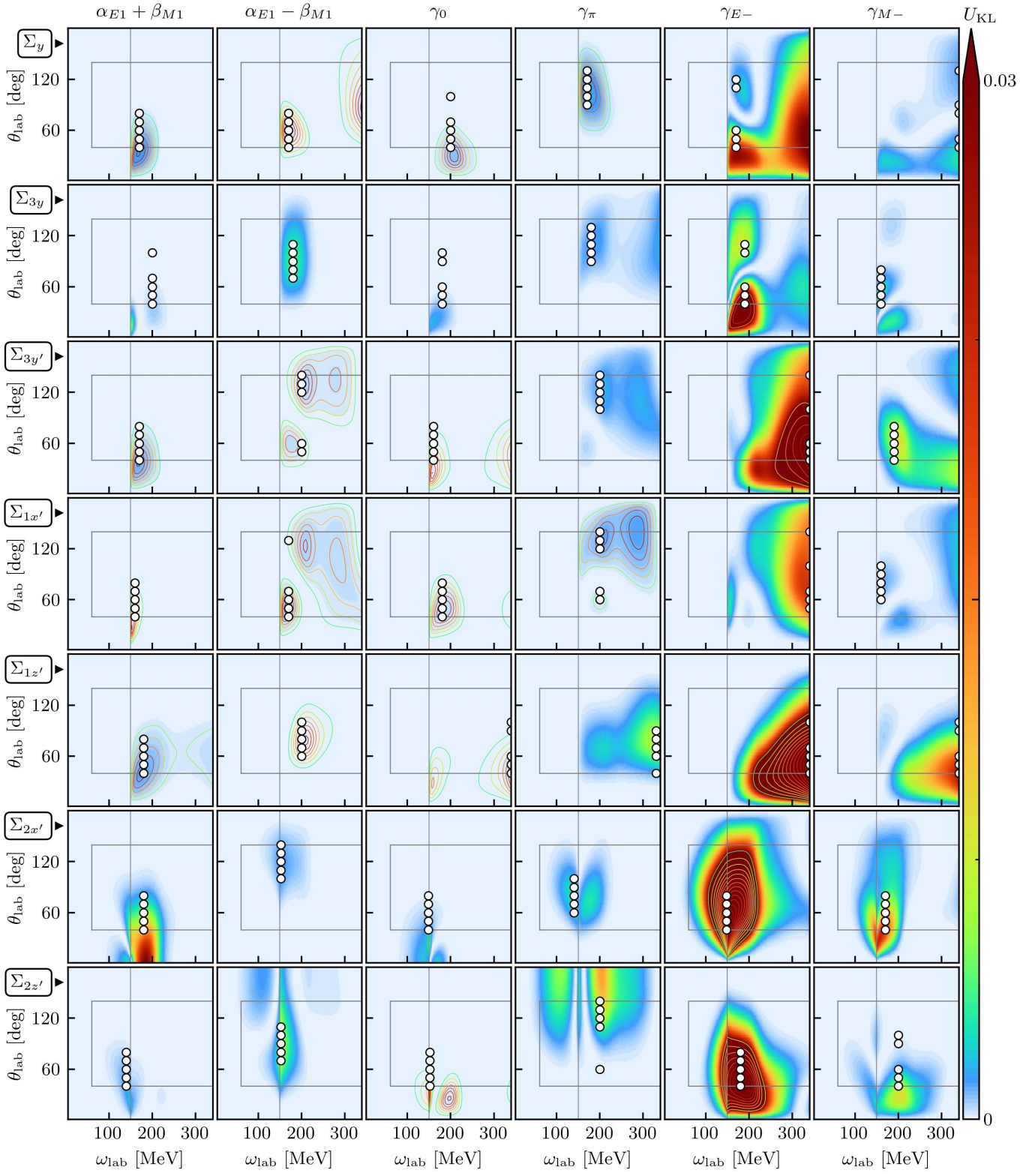

FIG. S.14. (Colour online) As in fig. 7, but with the "standard" level of experimental precision; see table IV.

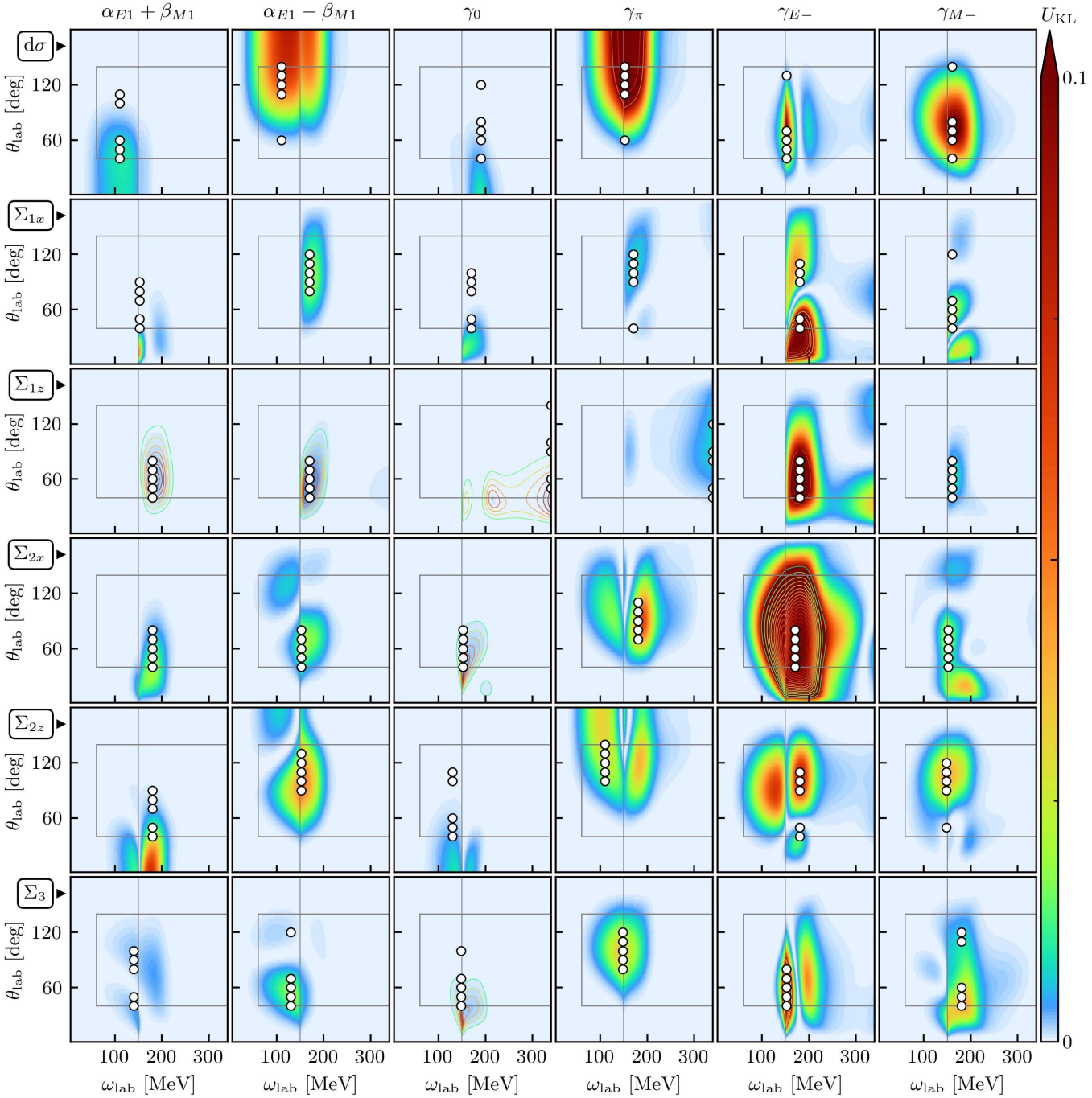

FIG. S.15. (Colour online) As in fig. 6, but with the "aspirational" level of experimental precision; see table IV.

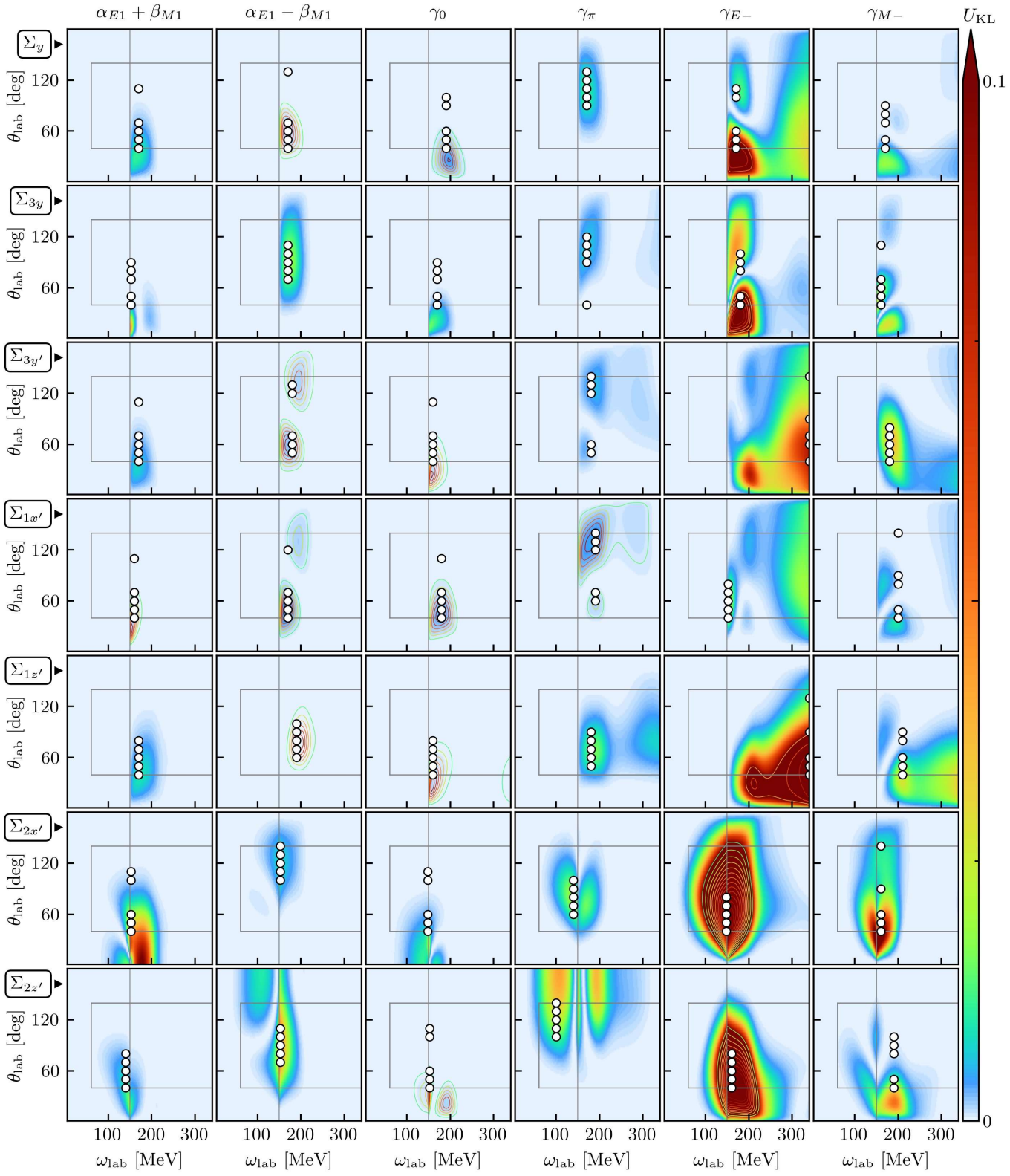

FIG. S.16. (Colour online) As in fig. 7, but with the "aspirational" level of experimental precision; see table IV.

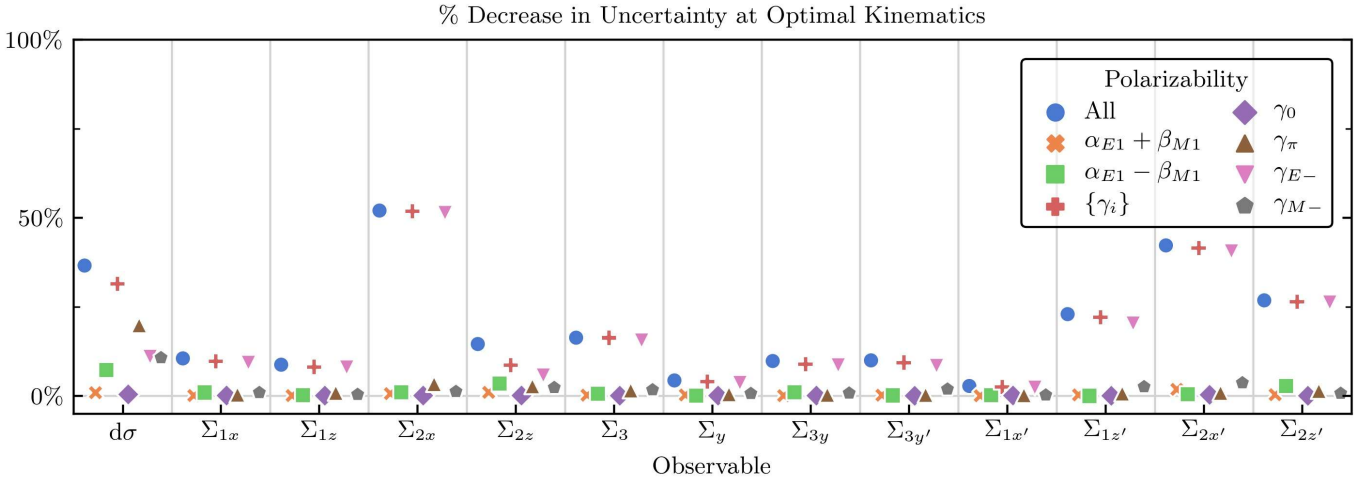

FIG. S.17. (Colour online) As in fig. 8, but with the "standard" level of experimental precision; see table IV.

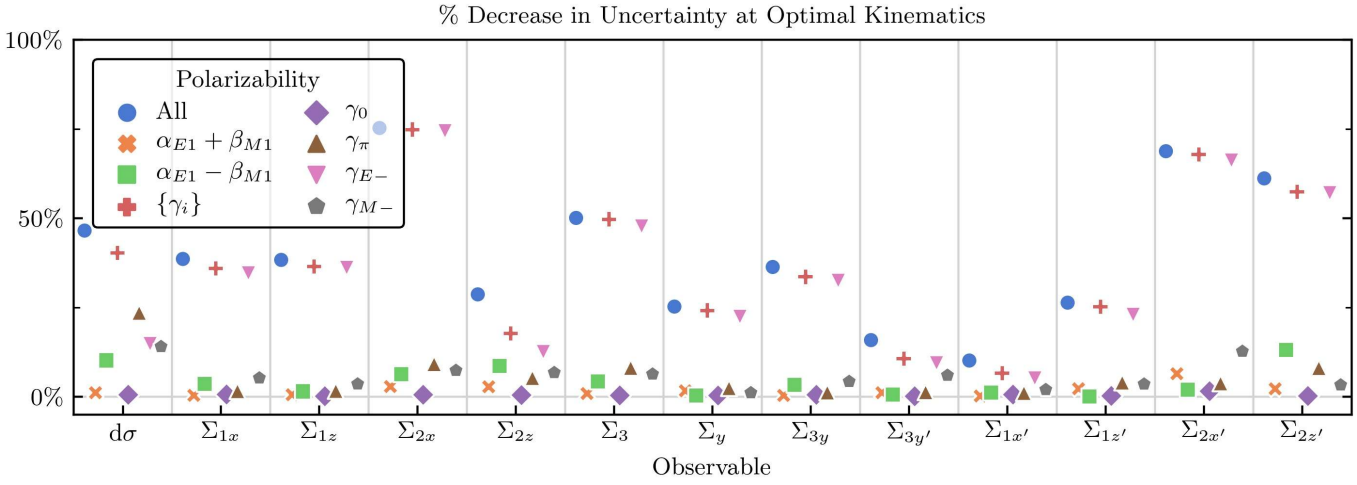

FIG. S.18. (Colour online) As in fig. 8, but with the "aspirational" level of experimental precision; see table IV.

4. Neutron Observables

Here we show the corresponding results for the neutron observables. Because such experiments are difficult, only the "standard" level of precision is used (see table IV). Even this is likely optimistic for such measurements, as discussed in the main text.

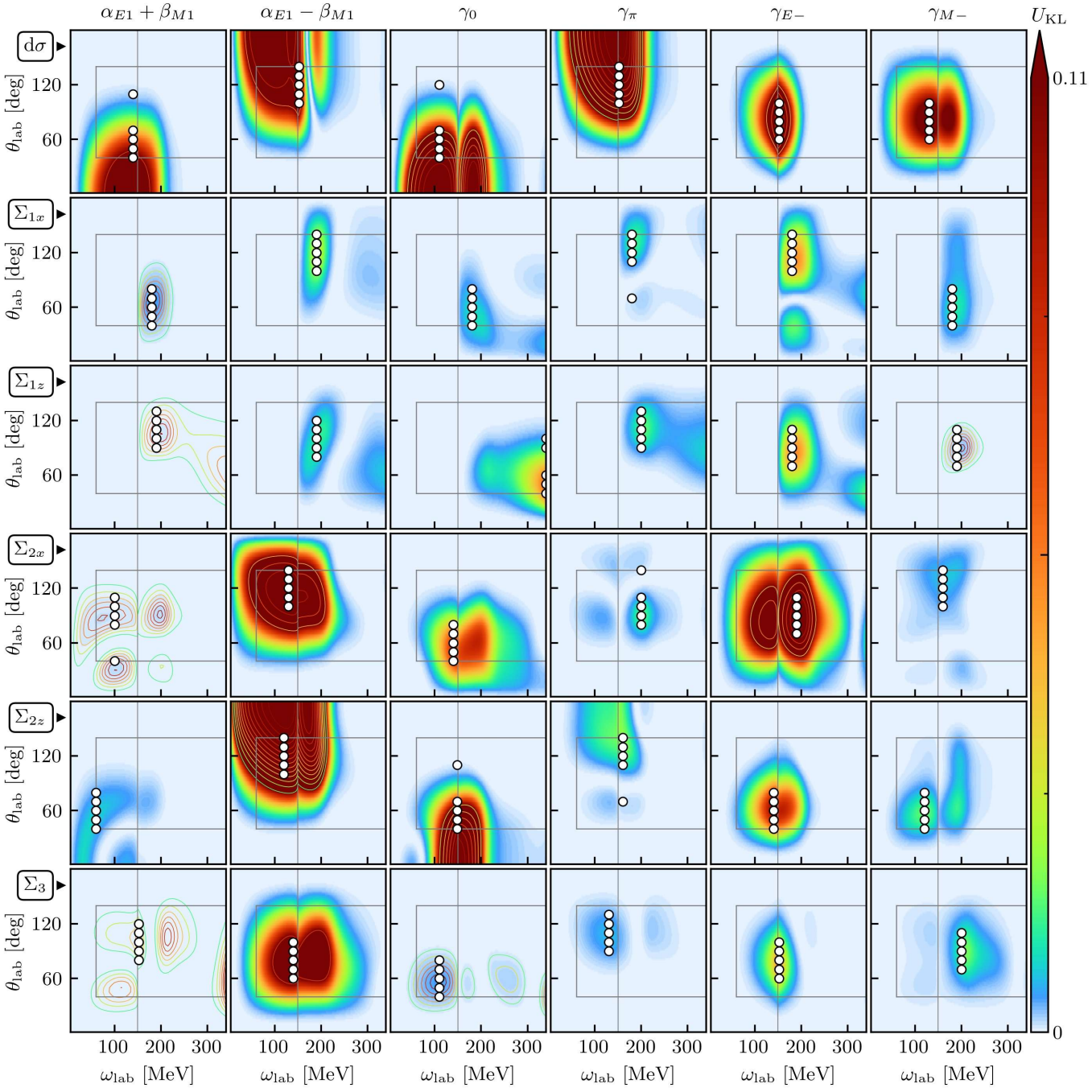

FIG. S.19. (Colour online) As in fig. 6, but for neutron observables with the "standard" level of experimental precision; see table IV.

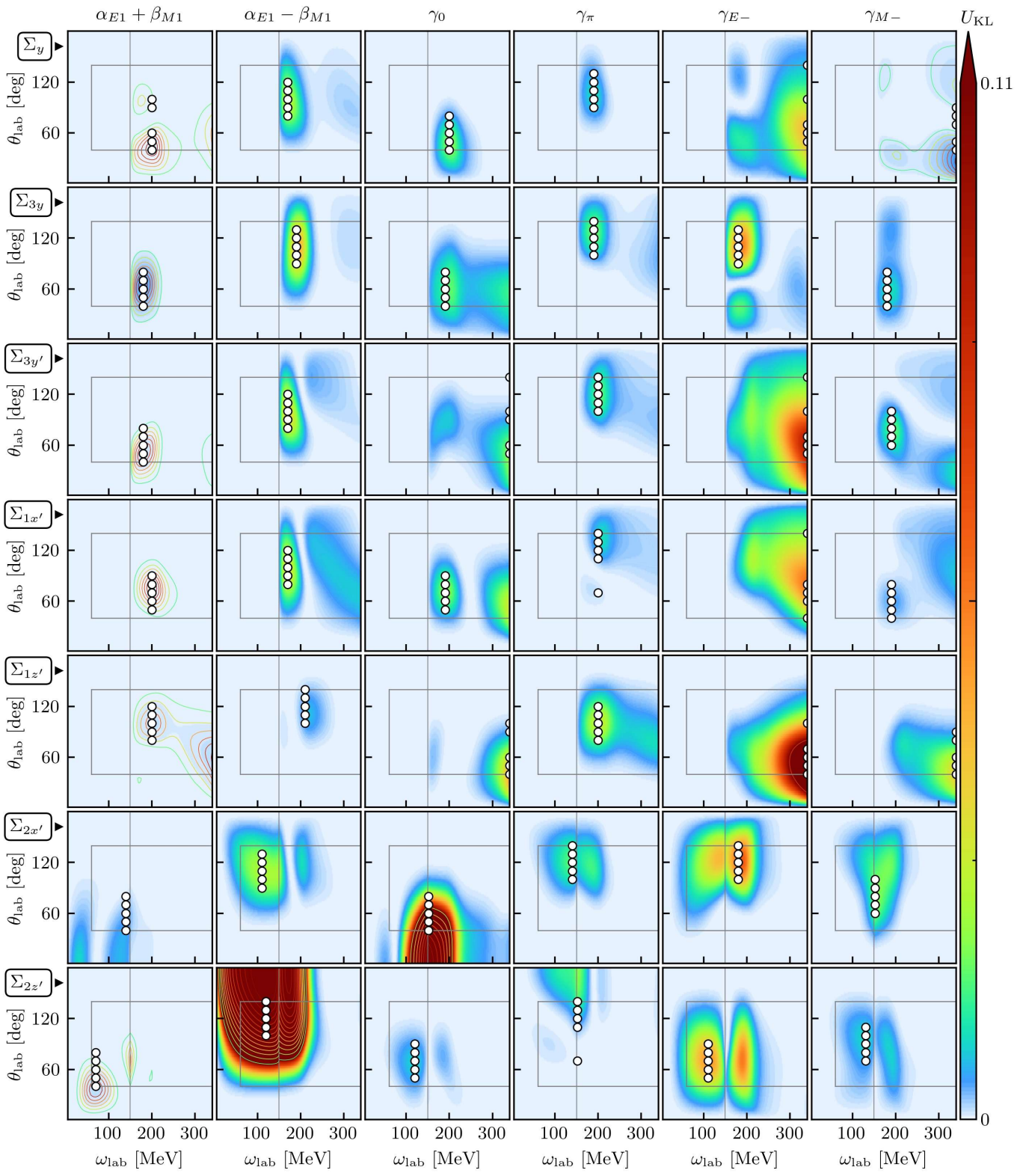

FIG. S.20. (Colour online) As in fig. 7, but for neutron observables with the "standard" level of experimental precision; see table IV.

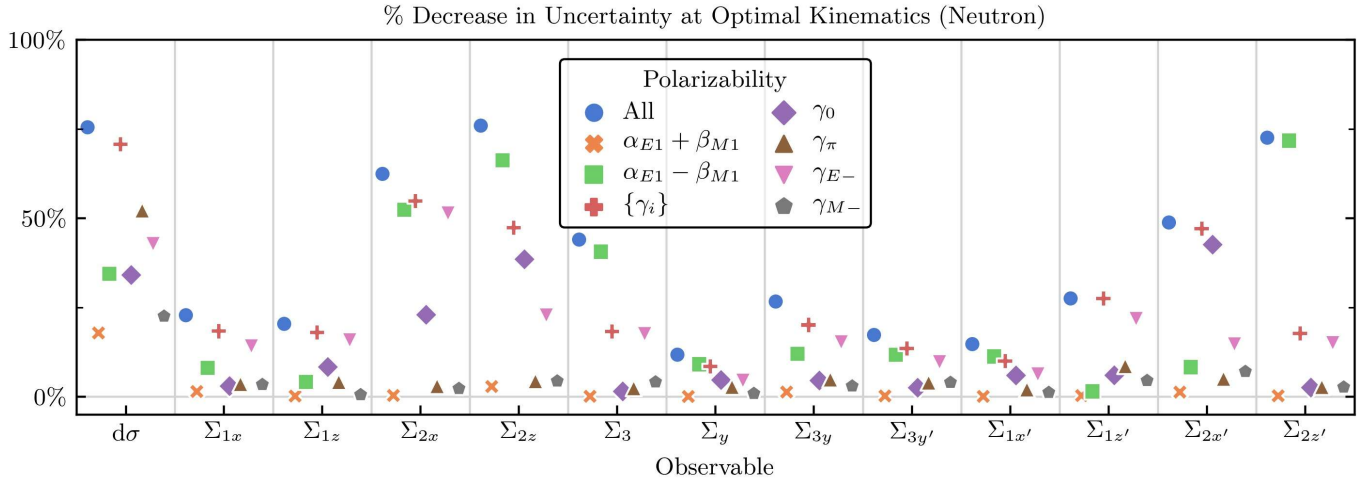

FIG. S.21. (Colour online) As in fig. 8, but for neutron observables with the "standard" level of experimental precision; see table IV.

[1] H. W. Griesshammer, J. A. McGovern, D. R. Phillips, and G. Feldman, Prog. Part. Nucl. Phys. **67**, 841 (2012), arXiv:1203.6834.

[2] B. R. Holstein and S. Scherer, Ann. Rev. Nucl. Part. Sci. **64**, 51 (2014), arXiv:1401.0140 [hep-ph].

[3] F. Hagelstein, R. Miskimen, and V. Pascalutsa, Prog. Part. Nucl. Phys. **88**, 29 (2016), arXiv:1512.03765 [nucl-th].

[4] H. W. Griesshammer, J. A. McGovern, and D. R. Phillips, Eur. Phys. J. A **52**, 139 (2016), arXiv:1511.01952 [nucl-th].

[5] J. A. McGovern, D. R. Phillips, and H. W. Griesshammer, Eur. Phys. J. A **49**, 12 (2013), arXiv:1210.4104.

[6] H. W. Griesshammer, J. A. McGovern, and D. R. Phillips, Eur. Phys. J. **A54**, 37 (2018), arXiv:1711.11546.

[7] A. Margaryan, B. Strandberg, H. W. Grießhammer, J. A. Mcgovern, D. R. Phillips, and D. Shukla, Eur. Phys. J. **A54**, 125 (2018), arXiv:1804.00956 [nucl-th].

[8] H. W. Grießhammer, Eur. Phys. J. **A49**, 100 (2013), [Erratum: Eur. Phys. J.A53,no.5,113(2017); Erratum: Eur. Phys. J.A54,no.4,57(2018)], arXiv:1304.6594 [nucl-th].

[9] P. Martel (A2), *Proceedings, 9th International Workshop on Chiral Dynamics (CD18): Durham (NC), United States*, PoS **CD2018**, 038 (2019).

[10] P. Martel, M. Biroth, C. Collicott, D. Paudyal, and A. Rajabi (A2), *Proceedings, International Symposium Advances in Dark Matter and Particle Physics (ADMPP16): Messina, Italy, October 24-27, 2016*, EPJ Web Conf. **142**, 01021 (2017).

[11] G. M. Huber and C. Collicott, in *Proceedings, 12th Conference on the Intersections of Particle and Nuclear Physics (CIPANP 2015): Vail, Colorado, USA, May 19-24, 2015* (2015) arXiv:1508.07919 [nucl-ex].

[12] M. W. Ahmed, *Proceedings, 9th International Workshop on Chiral Dynamics (CD18): Durham (NC), United States*, PoS **CD2018**, 001 (2020).

[13] H. R. Weller, M. W. Ahmed, H. Gao, W. Tornow, Y. K. Wu, M. Gai, and R. Miskimen, Prog. Part. Nucl. Phys. **62**, 257 (2009).

[14] M. Catacora-Rios, G. King, A. Lovell, and F. Nunes, Phys. Rev. C **100**, 064615 (2019).

[15] K. Chaloner and I. Verdinelli, Statistical Science **10**, 273 (1995).

[16] J. Liepe, S. Filippi, M. Komorowski, and M. P. H. Stumpf, PLoS Computational Biology **9**, e1002888 (2013).

[17] E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt, International Statistical Review **84**, 128 (2016), https://onlinelibrary.wiley.com/doi/pdf/10.1111/insr.12107.

[18] J. I. Myung, M. Pitt, Y. Tang, and D. R. Cavagnaro (2009).

[19] I. Vernon, M. Goldstein, R. G. Bower, *et al.*, Bayesian analysis **5**, 619 (2010).

[20] L. M. Berliner and Y. Kim, Journal of Climate **21**, 1891 (2008).

[21] J. A. Cumming and M. Goldstein, Technometrics **51**, 377 (2009).

[22] M. Farrow and M. Goldstein, Journal of Statistical Planning and Inference **136**, 498 (2006).

[23] T. J. Santner, B. J. Williams, and W. I. Notz, *The design and analysis of computer experiments*, Vol. 1 (Springer, 2003).

[24] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker, *A Bayesian approach to the design and analysis of computer experiments*, Tech. Rep. (Oak Ridge National Lab., TN (USA), 1988).

[25] M. Jones, M. Goldstein, P. Jonathan, and D. Randell, Journal of Statistical Planning and Inference **171**, 115 (2016).

[26] G. E. P. Box, in *Robustness in Statistics*, edited by R. L. Launer and G. N. Wilkinson (Academic Press, 1979) p. 201.

[27] J. A. Melendez, R. J. Furnstahl, D. R. Phillips, M. T. Pratola, and S. Wesolowski, Phys. Rev. C **100**, 044001 (2019), arXiv:1904.10581.

[28] V. Pascalutsa and D. R. Phillips, Phys. Rev. **C67**, 055202 (2003), arXiv:nucl-th/0212024.

[29] BUQEYE collaboration (2020) `https://buqeye.github.io/software/`.

[30] A. Walker-Loud, C. E. Carlson, and G. A. Miller, Phys. Rev. Lett. **108**, 232301 (2012), arXiv:1203.0254 [nucl-th].

[31] A. Walker-Loud, *Proceedings, 9th International Workshop on Chiral Dynamics (CD18): Durham (NC), United States*, PoS **CD2018**, 045 (2019), arXiv:1907.05459 [nucl-th].

[32] J. Gasser, M. Hoferichter, H. Leutwyler, and A. Rusetsky, Eur. Phys. J. **C75**, 375 (2015), arXiv:1506.06747 [hep-ph].

[33] J. Gasser, H. Leutwyler, and A. Rusetsky, Phys. Lett. **B814**, 136087 (2021), arXiv:2003.13612 [hep-ph].

[34] G. Blanpied *et al.*, Phys. Rev. **C64**, 025203 (2001).

[35] V. Sokhoyan *et al.* (A2), Eur. Phys. J. **A53**, 14 (2017), arXiv:1611.03769 [nucl-ex].

[36] C. Collicott, Ph.D. thesis, Dalhousie University (2015), https://wwwa2.kph.uni-mainz.de/images/publications/phd/thesis_Collicott-Cristina-2015.pdf.

[37] P. P. Martel *et al.* (A2), Phys. Rev. Lett. **114**, 112501 (2015), arXiv:1408.1576 [nucl-ex].

[38] P. P. Martel, Ph.D. thesis, University of Massachusetts, Amherst (2012), https://wwwa2.kph.uni-mainz.de/images/publications/phd/pmartel_dis.pdf.

[39] D. Paudyal *et al.* (A2), Phys. Rev. C **102**, 035205 (2020), arXiv:1909.02032 [nucl-ex].

[40] S. Wesolowski, R. J. Furnstahl, J. A. Melendez, and D. R. Phillips, J. Phys. G **46**, 045102 (2019), arXiv:1808.08211.

[41] R. J. Furnstahl, N. Klco, D. R. Phillips, and S. Wesolowski, Phys. Rev. C **92**, 024005 (2015), arXiv:1506.01343.

[42] J. A. Melendez, S. Wesolowski, and R. J. Furnstahl, Phys. Rev. C **96**, 024003 (2017), arXiv:1704.03308.

[43] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, Statistical science **4**, 409 (1989).

[44] N. Cressie, Terra Nova **4**, 613 (1992).

[45] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, Adaptative computation and machine learning series (University Press Group Limited, 2006).

[46] D. J. C. MacKay, in *Neural Networks and Machine Learning*, NATO ASI Series, Vol. 168, edited by C. M. Bishop (Springer, 1998) pp. 133–166.

[47] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, UK ; New York, 2003).

[48] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis* (CRC press, 2013).

[49] O. Gryniuk, F. Hagelstein, and V. Pascalutsa, Phys. Rev. **D92**, 074031 (2015), arXiv:1508.07952 [nucl-th].

[50] M. I. Levchuk and A. I. L'vov, Nucl. Phys. **A674**, 449 (2000), arXiv:nucl-th/9909066 [nucl-th].

[51] L. S. Myers *et al.*, Phys. Rev. **C92**, 025203 (2015), arXiv:1503.08094 [nucl-ex].

[52] L. S. Myers *et al.* (COMPTON@MAX-lab), Phys. Rev. Lett. **113**, 262506 (2014), arXiv:1409.3705 [nucl-ex].

[53] B. Pasquini, P. Pedroni, and D. Drechsel, Phys. Lett. B **687**, 160 (2010), arXiv:1001.4230 [hep-ph].

[54] M. Schumacher, Prog. Part. Nucl. Phys. **55**, 567 (2005), arXiv:hep-ph/0501167 [hep-ph].

[55] O. Gryniuk, F. Hagelstein, and V. Pascalutsa, Phys. Rev. **D94**, 034043 (2016), arXiv:1604.00789 [nucl-th].

[56] V. Lensky, J. McGovern, and V. Pascalutsa, Eur. Phys. J. **C75**, 604 (2015), arXiv:1510.02794 [hep-ph].

[57] D. Babusci, G. Giordano, A. I. L'vov, G. Matone, and A. M. Nathan, Phys. Rev. **C58**, 1013 (1998), arXiv:hep-ph/9803347 [hep-ph].

[58] R. P. Hildebrandt, H. W. Grießhammer, T. R. Hemmert, and B. Pasquini, Eur. Phys. J. **A20**, 293 (2004), arXiv:nucl-th/0307070 [nucl-th].

[59] B. R. Holstein, D. Drechsel, B. Pasquini, and M. Vanderhaeghen, Phys. Rev. **C61**, 034316 (2000), arXiv:hep-ph/9910427 [hep-ph].

[60] B. Pasquini, P. Pedroni, and S. Sconfietti, Phys. Rev. **C98**, 015204 (2018), arXiv:1711.07401 [hep-ph].

[61] N. Krupina, V. Lensky, and V. Pascalutsa, Phys. Lett. **B782**, 34 (2018), arXiv:1712.05349 [nucl-th].

[62] B. Pasquini, P. Pedroni, and S. Sconfietti, J. Phys. G **46**, 104001 (2019), arXiv:1903.07952 [hep-ph].

[63] P. P. Martel, E. Downie, M. Ahmed, and G. Feldman, Private Communication (2020).

[64] V. O. de León *et al.*, Eur. Phys. J. **A10**, 207 (2001).

[65] A. Zieger, R. Van de Vyver, D. Christmann, A. De Graeve, C. Van den Abeele, and B. Ziegler, Phys. Lett. **B278**, 34 (1992).

[66] R. P. Hildebrandt, H. W. Grießhammer, and T. R. Hemmert, Eur. Phys. J. **A46**, 111 (2010), arXiv:nucl-th/0512063 [nucl-th].

[67] S. Jackson, *Design of Physical System Experiments Using Bayes Linear Emulation and History Matching Methodology with Application to Arabidopsis Thaliana*, Doctoral, Durham University (2018).

[68] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science (Taylor & Francis, 2013).

[69] A. O'Hagan, *Bayesian Inference*, Kendall's Advanced Theory of Statistics (Edward Arnold, London, 1994).

[70] C. Rasmussen, Bayesian Statistics 7 , 651 (2003).

[71] E. Solak, R. Murray-smith, W. E. Leithead, D. J. Leith, and C. E. Rasmussen, in *Advances in Neural Information Processing Systems 15*, edited by S. Becker, S. Thrun, and K. Obermayer (MIT Press, 2003) pp. 1057–1064.

[72] D. Eriksson, K. Dong, E. Lee, D. Bindel, and A. G. Wilson, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) pp. 6867–6877, arXiv:1810.12283.

[73] M. Chilenski, M. Greenwald, Y. Marzouk, N. Howard, A. White, J. Rice, and J. Walk, Nucl. Fusion **55**, 023012 (2015).