University of Glasgow | Experimental Particle Physics

Applications of Linearised Oscillation Models on the T2K Experiment

Paul Morrison

Supervised by Dr Phillip Litchfield

School of Physics and Astronomy
University of Glasgow

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Abstract

T2K is a long-baseline neutrino oscillation experiment measuring ν_μ disappearance and ν_e appearance with a baseline of 295 km and a narrow-band beam peaking in energy at around 0.6 GeV. This thesis presents a linearised oscillation probability model for the phase angle $\delta_{\mathcal{CP}}$, designed to complement standard analyses used in this experiment. We demonstrate that the framework provides a universal method for performing T2K sensitivity studies, and characterisation of systematics, that avoids almost all of the problems with commonly used metrics, most notably the dependence on the choice of trial point. Furthermore, we show how it restores the validity of well-known asymptotic approximations, allowing the likelihood to be described by a simple Gaussian form which could reduce or eliminate the need for computationally expensive tools such as MCMC or Frequentist toys sampling when quoting results at high significance.

Declaration

The research presented in this thesis is the result of my own work, carried out within the Experimental Particle Physics Group of the School of Physics & Astronomy at the University of Glasgow. It has not been submitted for any other qualification at this or any other institution.

The research was conducted while I was a member of the T2K Collaboration, by whom the data used in this thesis were generated. The software and fitting tools used in this work rely heavily on existing code developed and maintained by the T2K P-Theta OA group.

Paul Morrison

Acknowledgements

I firmly believe that science is a collective endeavour, which must be particularly true of experimental particle physics. Therefore, any insights contained in this thesis are the product of countless exchanges with fellow experimentalists. Nevertheless, there are several individuals to whom I owe particular gratitude.

First and foremost, I thank my supervisor Phill Litchfield, for his mentorship, for giving and trusting me with such a unique and interesting project, and, despite my best efforts, for getting me to the finish line. I also thank Paul Soler, my second supervisor, for the many useful conversations and encouragement over the years, and even before my PhD; it was the literature review I completed during my undergraduate degree under his guidance that sparked my interest in this field. I must also thank the P-Theta T2K group, and particularly Lukas Berns, for your advice and for being an endless source of ideas.

It wasn't all work, though, and I'm also grateful for all the relief and distraction provided by my friends and colleagues. This includes the PPE group at Glasgow and all the people I met during my seven-month attachment in Tokai, who made the experience truly memorable. I also want to thank my friends outside particle physics, including my very close friend Kerr and the entire Beagan Mafia.

I want to thank my partner, Liam, for all the support and endless laughs along the way (the cat got the cream, but who got that cat?). Your visit to Japan is one of my most cherished memories, although I still twitch a little whenever Osaka is mentioned, and I very much maintain that bag was your responsibility!

This next part might seem a little odd, but I want to acknowledge my late childhood cat, Bobo, who has been a near-constant presence and companion for… well, nearly my whole life. I'll carry you with me always, and you will be deeply missed.

Of course, none of this would have been possible without my family: my sister Hazel, my parents Paul and Lorraine, and my amazing new nephew Nathan (whose

favourite colour, as of writing, is definitely purple). Your unwavering support and belief throughout nine years of university have meant everything to me. I share this accomplishment with you.

Preface

Neutrino oscillations represent the first, and as of writing this, the only generally accepted experimental evidence of physics beyond the Standard Model. Their discovery has opened the possibility of a new and potentially significant source of $\mathcal{CP}$-violation, as well as the existence of a new, and yet-to-be-understood, mass scale (< 1 eV).

In line with quark mixing, the four degrees of freedom of the neutrino mixing matrix are conventionally parameterised in terms of three mixing angles and one complex phase, $\theta_{23}, \theta_{13}, \theta_{12}, \delta_{\mathcal{CP}}$, using the ordering $R(\theta_{23})\Gamma_\delta R(\theta_{13})\Gamma_\delta^\dagger R(\theta_{12})$. This parameterisation is well established within the neutrino community and is generally regarded as the most effective framework for describing mixing. However, from a statistical perspective it poses challenges, as at current and foreseeable statistics it violates the standard asymptotic approximations upon which many useful statistical tools rely. As a result, reported likelihoods are highly non-Gaussian and strongly data dependent. This complicates the interpretation of sensitivity studies and significantly slows down T2K analyses and updates.

In this thesis we present and implement an oscillation probability model in which the neutrino event rates are expressed as linear functions of new parameters $X_S = \sin\delta_{\mathcal{CP}}$ and $X_C = \cos\delta_{\mathcal{CP}}$. We show how this framework greatly simplifies T2K sensitivity studies, restores Gaussian-like behaviour and enables results that are both easier to interpret and reveal insights usually obscured by the conventional parametrisation.

Chapter 1 provides a brief history of neutrinos and neutrino oscillations. It also introduces the Standard Model, with the purpose of contextualising neutrino physics within the broader HEP community and helping to motivate the physics aims of current and future experiments. Chapter 2 focuses on the T2K experiment itself, describing the detectors, relevant technologies, and their role in achieving T2K's physics goals. Chapter 3 introduces the basics of asymptotic statistics and demonstrates how the conventional PMNS parameterisation violates its main assumptions. Chapter 4 discusses how this undermines and complicates many of T2K's analysis strategies,

such as the Asimov Fit. It also explains the linearisation concepts and presents a scheme for implementing this for $\delta_{\mathcal{CP}}$ (the X_S–X_C framework). Chapter 5 uses the X_S–X_C parameterisation to demonstrate how much of the complex $\delta_{\mathcal{CP}}$ behaviour is easier to interpret in extended space, and that sensitivity studies in this space are much simpler. Chapter 6 validates the method with real T2K data by reproducing the OA21 analysis in the X_S–X_C parameterisation.

Finally, the appendices show proof-of-concept for further directions: generalisation of the concept to the parameter θ_{23} and the mass ordering, and an application to quickly generate Frequentist intervals with correct coverage.

Contents

Chapter 1.

Introduction

1.1. Neutrino history

1.1.1. Early history

The neutrino has its origins in early studies of radioactivity phenomena. In 1911, studies on the beta decay of radium unexpectedly observed a continuous energy spectrum (depicted in Fig. 1.1). This was in contrast to the discrete energy spectrum seen for alpha and gamma radiation [1, 2]. At the time the neutron had not yet been discovered, so the decay was presumed to take the form,

$$A \rightarrow B + e^{-}. \tag{1.1}$$

Here, A is the parent nucleus, B is the daughter nucleus, and e^{-} the emitted beta particle [3]. Given that the energy released in a true two-body decay is well defined, momentum conservation constrains the beta particle's kinetic energy to a unique value, meaning its energy spectrum should be a discrete, mono-energetic peak. The discovery that it instead followed a broad spectrum posed a serious and uncomfortable problem from a theoretical point of view. Some prominent physicists even suggested that energy conservation held only statistically and could be violated for single events. The puzzling anomaly formed a key topic of the 1930 nuclear physics conference in Tübingen, Germany, where Wolfgang Pauli, in an open letter, proposed a solution [4]: he suggested that if another undetected neutral fermion were emitted along with the electron, it could carry away varying amounts of energy and thus allow the kinetic energy of the beta particle to vary. One of the original supporters of

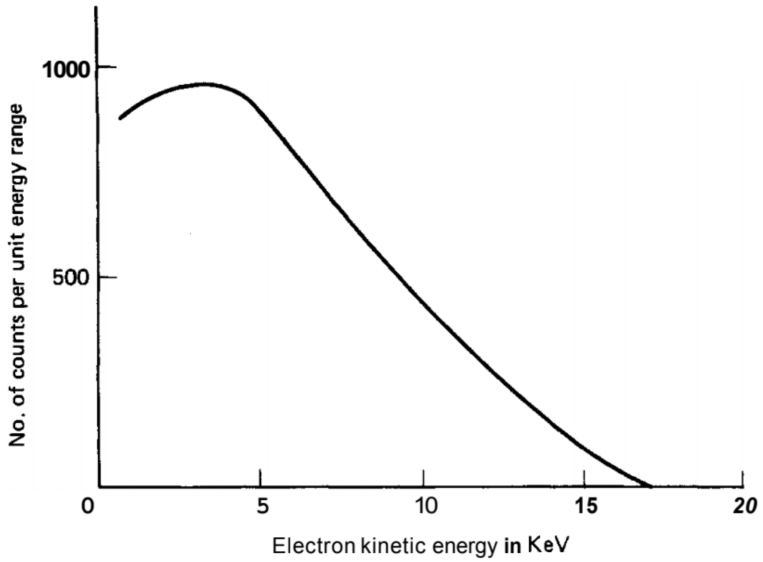


Figure 1.1.: The observed continuous energy spectrum of beta particles. Reproduced from Ref. [3].

Pauli's hypothesis, Enrico Fermi, proposed that the theorised particle be called the neutrino (*"little neutral one"* in Italian) [5]. Shortly after, Fermi developed and published his theory of beta decay – essentially a low-energy effective theory for electroweak interactions. Using this theory, the neutrino cross section was estimated to be of the order $\sigma \leq 1 \times 10^{-44}\,\mathrm{cm}^2$ [6]. This astonishingly small cross-section led them to conclude that direct detection of neutrinos would be impossible. Luckily around the same time, it was realised that nuclear fission could provide a copious source of neutrinos. The decisive breakthrough came with the Cowan–Reines neutrino experiment, which detected neutrinos using a large tank of cadmium chloride ($CdCl_2$) placed near a nuclear reactor [7]. The strategy involved first detecting a prompt signal of two back-to-back 511 keV photons from the annihilation of positrons produced via $\bar{\nu}_e + p \to n + e^+$. The second element was the delayed detection, approximately 5μs later, of a de-excitation photon from neutron capture on cadmium ($n + {}^{113}\mathrm{Cd} \to {}^{114}\mathrm{Cd} + \gamma$). This time-correlated signal provided a distinctive signature of neutrino interactions and offered conclusive proof of the existence of the neutrino.

1.1.2. Early hints and observations of neutrino oscillation

The discovery of neutrino oscillation emerged from two major experimental anomalies that challenged the Standard Model's treatment of neutrinos as massless particles with fixed flavour.

The solar and atmospheric neutrino problem

The first indication that there was a problem with our understanding of neutrinos came in the late 1960s when the Homestake experiment detected a significant deficit in the observed solar neutrino flux (ν_e's , produced in fusion reactions in the Sun's core) reaching Earth. Using a 600 kL tank of tetrachloroethylene (C_2Cl_4) located 1.5 km underground, at the Homestake Mine in South Dakota, it measured the rate of ^{37}Ar production through the reaction $\nu_e + {}^{37}\mathrm{Cl} \rightarrow {}^{37}\mathrm{Ar} + e^-$. The experiment found the rate of ^{37}Ar production was only one third of that predicted by the Standard Solar Model [8–11].

Around two decades later, a similar anomaly relating to neutrinos produced in Earth's atmosphere (atmospheric neutrinos) emerged. When cosmic rays, predominantly high energy protons, interact with nucleons in the atmosphere, they produce hadronic showers whose constituents decay into neutrinos. The dominant decay process for this is via pion decay, followed by muon decay:

$$\begin{aligned} \pi^{\pm} &\rightarrow \mu^{\pm} + \nu_{\mu}\,(\bar{\nu}_{\mu}), \\ \mu^{\pm} &\rightarrow e^{\pm} + \nu_e\,(\bar{\nu}_e) + \bar{\nu}_{\mu}\,(\nu_{\mu}). \end{aligned} \tag{1.2}$$

This produces two μ-like neutrinos for every e-like neutrino, resulting in the expected ratio $R = (\nu_\mu + \bar{\nu}_\mu)/(\nu_e + \bar{\nu}_e) = 2$. However, the large Japanese water-Čerenkov detector Super-Kamiokande observed a substantially smaller ratio, of $R = 0.61 \pm 0.03_{\text{stat.}} \pm 0.05_{\text{sys.}}$ [12, 13].

The solution

The solution to these anomalies was provided by the Sudbury Neutrino Observatory (SNO) and Super-Kamiokande (SK) collaborations, which conclusively demonstrated that the observed deficits were due to neutrino oscillations [14, 15].

SNO achieved this using a heavy-water (D_2O) detector, which enabled sensitivity to three channels with distinct interaction mechanisms: the elastic scattering (ES) channel, $\nu_x + e^- \rightarrow \nu_x + e^-$; the charged-current (CC) channel, $\nu_e + d \rightarrow e^- + p + p$; and, crucially, the flavour-agnostic neutral-current (NC) channel, $\nu_x + d \rightarrow \nu_x + p + n$. Measurements of the NC channel showed that the total neutrino flux was consistent with the Standard Solar Model. By combining measurements from all three channels, SNO was able to constrain the ν_e to $\nu_{\mu,\tau}$ flux ratio to $\phi_e : \phi_{\mu\tau} \approx 1 : 2$ (see Fig. 1.2). This result explained the two-thirds deficit in the ν_e flux observed by the Homestake experiment by confirming that it was due to two-thirds of the original ν_e flux oscillating into ν_μ or ν_τ.

SK resolved the atmospheric neutrino anomaly by demonstrating a clear angular dependence in the observed deficit. Event rates for downward-going neutrinos, which traveled only short distances through the atmosphere, were consistent with expectations from cosmic ray interactions. In contrast, the event rate for upward-going neutrinos, which traveled much larger distances through the Earth before detection, deviated substantially from predictions. This revealed a propagation length dependence that is characteristic of oscillation phenomena (see Fig. 1.3).

In 2015, Takaaki Kajita and Arthur B. McDonald, representing the SK and SNO collaborations respectively, were jointly awarded the Nobel Prize in Physics for this achievement [16]. Since then, neutrino oscillations have been observed in a plethora of different contexts and from various sources, including artificial ones like reactor neutrinos, and those produced in particle accelerators.

1.2. The Standard Model

The Standard Model (SM) constitutes one of the most mathematically complete and successful theories in fundamental physics. It is formulated as a quantum field theory characterised by a set of symmetry properties. Namely, it is invariant under Poincaré transformations and respects a specific gauge symmetry.

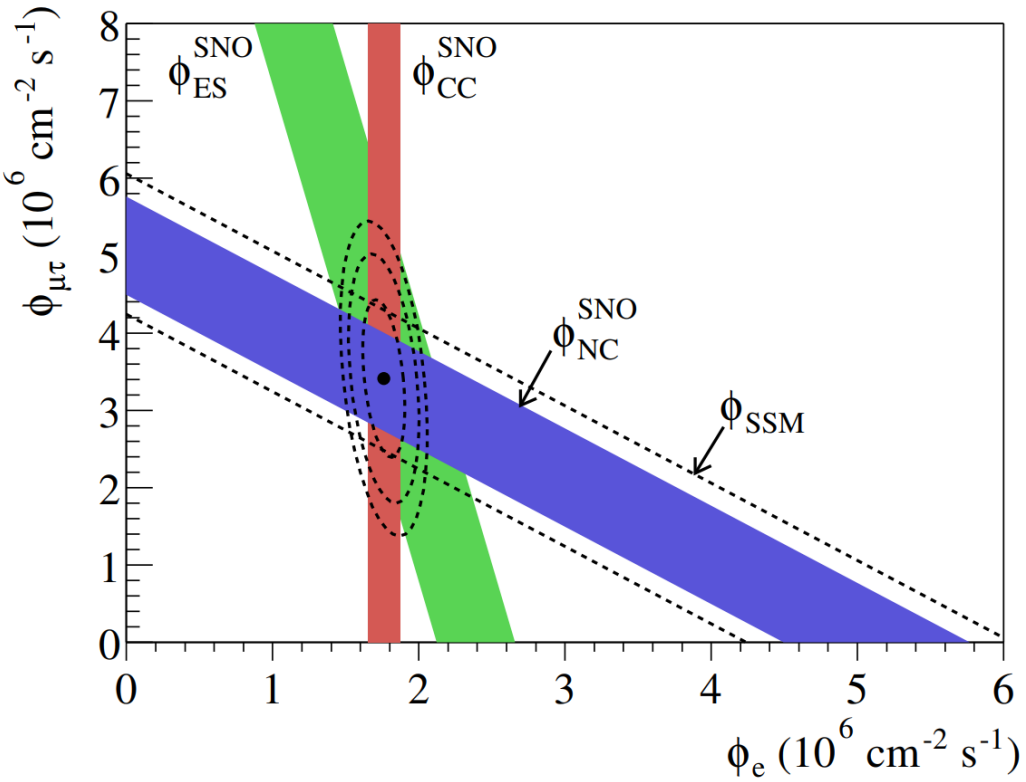
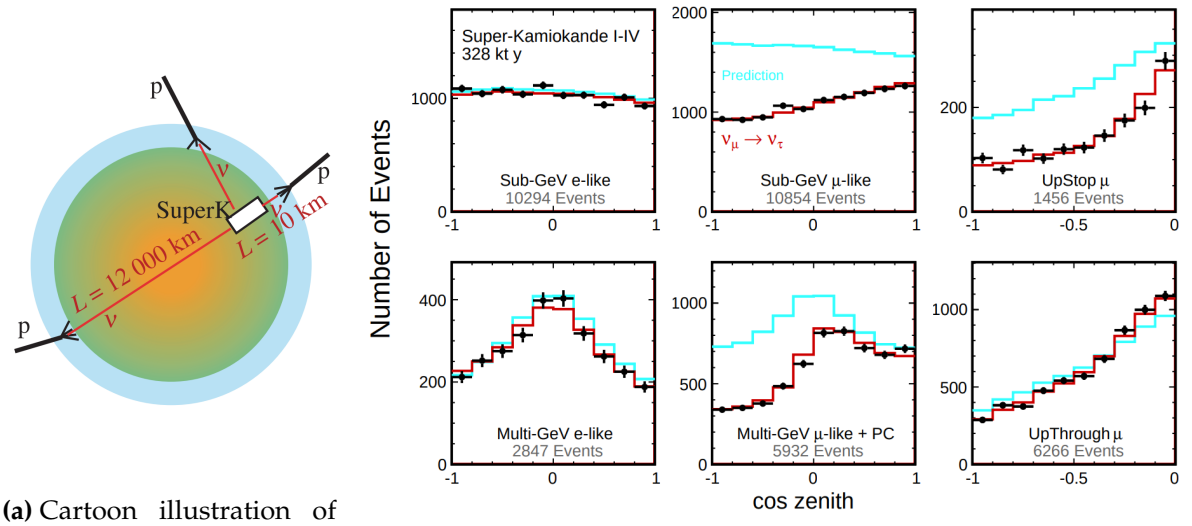


Figure 1.2.: The 1σ confidence-level bands for the solar neutrino flux components $\phi(\nu_e)$ and $\phi(\nu_{\mu,\tau})$ measured by SNO using CC, ES, and NC interactions. Each coloured band corresponds to a single interaction channel, with the overlap indicating consistency between the measurements and agreement of the total flux with the Standard Solar Model. Reproduced from Ref. [15].



(a) Cartoon illustration of how neutrinos incident at different angles have different propagation lengths. Reproduced from Ref. [17].

(b) Zenith angle distributions of atmospheric neutrinos observed at SK. The blue-edged histogram shows the Monte Carlo prediction for assuming no oscillation. The red edged histogram is the best-fit oscillation prediction, which is in significantly better agreement with data. Reproduced from Ref. [18].

Figure 1.3.: Super-Kamiokande Zenith-Angle diagram and oscillation data.

1.2.1. Poincaré symmetry

Poincaré symmetry is the generalisation of Lorentz Symmetry that includes spacetime translations. It is described by the group $SO(1,3) \rtimes \mathbb{R}^4$ and is required to impose locality and keep the theory consistent with relativity. $SO(1,3)$ is the Lorentz group, which is the part that preserves Minkowski spacetime and does most of the heavy lifting. The particle content of the SM is built and classified by those fields that have good transformation properties under this group:

- **Spin-0 scalar fields:** These transform trivially under Lorentz transformations. Within the Standard Model, the Higgs field is the only fundamental field of this type.
- **Spin-1 vector fields:** These transform as four-vectors and correspond to the gauge fields of the theory. The associated quanta are collectively referred to as gauge bosons.
- **Spin-$\frac{1}{2}$ spinor fields:** These fields describe the fermionic content of the Standard Model. Their transformation properties are more subtle than those of scalars or vectors, as spinor representations form a double cover of the Lorentz group. A Dirac spinor can be decomposed into two two-component Weyl spinors of

definite chirality: left-handed (LH) and right-handed (RH). The operation that interchanges these chiralities is parity. Because the Standard Model is a chiral theory, this distinction is physically meaningful: the weak interaction maximally violates parity by coupling only to LH spinor fields.

Additionally, objects built from these fields that relate to physically observable quantities, such as masses or conserved currents like electric current, must be Lorentz invariant. This restricts how these fields can appear in the action. For instance, terms involving Dirac-spinors appear as bilinear contractions of the form $\bar{\psi}\Gamma\psi$ (where Γ is some matrix representation respecting the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\mathbb{1}$).

1.2.2. Gauge symmetry

Gauge transformations are "local transformations", meaning they have an explicit space-time dependence. They are associated with Lie groups, which are the groups of infinitesimal transformations continuously connected to the identity. Finite elements of these groups are expressed by exponentiating its generators, which satisfy the associated Lie-algebra $[T^A, T^B] = if^{ABC}T^C$. Every Lie-algebra (and by extension Lie-group) is uniquely defined by its structure constants f^{ABC}. A local gauge transformation assigns to each space-time point an element of the Lie group, such that the matter-field, $\psi(x)$, transforms as,

$$\psi'(\boldsymbol{x}) \to e^{-i\alpha^a(\boldsymbol{x})T^a}\psi(\boldsymbol{x}), \tag{1.3}$$

where the T^a's are the generators of the Lie group and the $\alpha^a(x)$'s parametrise the transformation at space-time location $\boldsymbol{x}$. The simplest example, which is useful to quickly examine, is the Abelian group $U(1)$, which has the single trivial generator $T^1 = 1$ (meaning of course, $f^{111} = 0$): any QFT Lagrangian will include a kinetic term defining the dynamics of the theory. Such kinetic terms will necessarily contain time-derivatives, which generally are not covariant under gauge transformations. Take for instance, the kinetic term of the Dirac-Lagrangian under a $U(1)$ gauge transformation,

$$\bar{\psi}i\partial\!\!\!/\psi \to \bar{\psi}i\partial\!\!\!/\psi - \bar{\psi}\gamma^\mu\psi\,\partial_\mu\alpha(x). \tag{1.4}$$

The derivative acting on the gauge-shifted ψ generates an extra $\bar{\psi}\gamma^\mu\psi\,\partial_\mu\alpha(x)$ term, which breaks gauge invariance. To compensate, the differential operator ∂ is replaced

by the covariant derivative:

$$\partial_\mu \to D_\mu \equiv \partial_\mu + igA_\mu, \tag{1.5}$$

where we have introduced a vector field (called a gauge field in this context) and also prescribed that this gauge field transforms as,

$$A_\mu \to A_\mu - \frac{1}{g}\partial_\mu \alpha(x), \tag{1.6}$$

in keeping with the fact that Lie groups are related to transformations infinitesimally close to the identity. The transformation now takes the form:

$$\begin{aligned}
\mathcal{L}_\psi \to \mathcal{L}'_\psi &= \bar{\psi}' \left[\left(i \not{\partial} - g \not{A}' \right) \right] \psi' \\
&= \bar{\psi} e^{i\alpha} \left[i \not{\partial} - g \left(\not{A} + \frac{1}{g} \not{\partial}\alpha \right) \right] e^{-i\alpha} \psi \\
&= \bar{\psi}\, e^{i\alpha} \left[i \not{\partial} - g\not{A} - \not{\partial}\alpha \right] e^{-i\alpha} \psi \\
&= \bar{\psi} \left[i \not{\partial}(e^{-i\alpha}\psi) - \not{\partial}\alpha\, e^{-i\alpha}\psi - g\not{A}\, e^{-i\alpha}\psi \right] e^{i\alpha} \\
&= \bar{\psi} \left[i(-i \not{\partial}\alpha\, e^{-i\alpha}\psi + e^{-i\alpha} \not{\partial}\psi) - \not{\partial}\alpha\, e^{-i\alpha}\psi - g\not{A}\, e^{-i\alpha}\psi \right] e^{i\alpha} \\
&= \bar{\psi}\, e^{i\alpha} \left[i \not{\partial}\psi - g\not{A}\psi \right] e^{-i\alpha} = \bar{\psi} i \partial^\mu \gamma_\mu \psi - g \bar{\psi} A^\mu \gamma_\mu \psi \\
&= \bar{\psi} i \left[\not{\partial} - ig\not{A} \right] \psi = \bar{\psi} i \not{D} \psi = \mathcal{L}_\psi,
\end{aligned} \tag{1.7}$$

where $\not{A} = \gamma^\mu A_\mu$. This procedure, called the minimal coupling scheme, ensures that the kinetic part of the Lagrangian transforms covariantly ($\bar{\psi} i \not{D} \psi \mapsto \bar{\psi} i \not{D} \psi$). Thus the necessary modification to preserve gauge invariance introduces an additional interaction term and new vector field. The second-to-last line of Eq. (1.7), highlighted in blue, shows the additional term that arises from this procedure. This extra term suggests that these gauge fields should be interpreted as those that couple to fermions and mediate interactions (e.g. in QED ψ would correspond to the electron field, A^μ the photon field, and g the electron charge). Since these gauge fields are interpreted as physical, they must also possess their own dynamics, and so require a kinetic term in the SM Lagrangian. The canonical choice is

$$\mathcal{L}_A = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.8}$$

with the most general form of the field strength tensor being,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu]. \tag{1.9}$$

The gauge group of the SM is the twelve dimensional group $G_{SM} = \mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. There are therefore twelve generators corresponding to twelve gauge-fields with field strength tensors $B_{\mu\nu}$, $W^i_{\mu\nu}$ $(i \in \{1,2,3\})$, and $G^j_{\mu\nu}$ $(j \in \{1,\ldots,8\})$ defined by,

$$B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$$

for $\mathrm{U}(1)_Y$,

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^i_{jk} W^j_\mu W^k_\nu$$

for $\mathrm{SU}(2)_L$ (where ϵ^i_{jk} is the Levi-Civita antisymmetric tensor), and

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^a_{bc} G^b_\mu G^c_\nu$$

for $\mathrm{SU}(3)_C$.

Note that it is because $\mathrm{U}(1)_Y$ is Abelian, with a trivial structure constant, that its field strength tensor contains no non-linear terms, and therefore the photon carries no electric charge (thus has no self-interactions). Additionally, the covariant derivative generalises to

$$D_\mu = \partial_\mu + i\sum_A g_A A^A_\mu T^A, \tag{1.10}$$

where the sum runs over the generators of the gauge group, and the explicit form depends on the gauge representation of the fermion field. Consequently, fermion gauge interactions are fully determined by their transformation properties under G_{SM}, requiring each fermion to be assigned a specific representation of the Standard Model gauge group. $\mathrm{U}(1)$ is simple as there is only one non-trivial irreducible representation, characterised by the particle's charge. For the non-Abelian components, $\mathrm{SU}(N)$, fermions either transform trivially, where the field maps to itself, or in the fundamental representation, where the field is an N-vector acted on by an $N \times N$ matrix. Table 1.1 shows their representations under each individual subgroup of G_{SM}. For instance, the left-handed quark doublet, being charged under all gauge interactions, has the

	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
Q_L	3	2	1/6
L_L	1	2	−1/2
u_R	3	1	2/3
d_R	3	1	−1/3
e_R	1	1	−1

Table 1.1.: Gauge representations of the SM fermion fields.

covariant derivative,

$$\not{D}_\mu = \partial_\mu + ig_s T^a G_{a\mu} + ig\tau^a W_{a\mu} + ig' Y B_\mu \tag{1.11}$$

where Y is the weak hypercharge, τ^a are the $\mathrm{SU}(2)_L$ generators represented by the Pauli matrices, T^a are the $\mathrm{SU}(3)_C$ generators represented by the Gell-Mann matrices, and g_s, g, g' are the coupling constants for the strong, weak, and hypercharge interactions respectively. Note that the quarks are explicitly written as doublets in the Lagrangian because up-type and down-type quarks after spontaneous symmetry breaking acquire distinct masses and electrical charges, and are thus treated as individual particles (unlike $\mathrm{SU}(3)$ colour symmetry, which is unbroken and so interpreted as different internal states of the same quark). The LH-quark (Q_L), and leptons (L_L) are written as,

$$Q_L \equiv \begin{pmatrix} u_{L_i} \\ d_{L_i} \end{pmatrix} \equiv \left(\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \right); \quad L_L \equiv \begin{pmatrix} \nu_{L_\ell} \\ \ell_L \end{pmatrix} \equiv \left(\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \right) \tag{1.12}$$

Conversely, because the weak interaction is chiral, right-handed quarks and leptons do not participate in weak isospin. They therefore transform as singlets under $\mathrm{SU}(2)_L$:

$$U_R \equiv (u_R, c_R, t_R); \quad D_R \equiv (d_R, s_R, b_R); \quad E_R \equiv (e_R, \mu_R, \tau_R). \tag{1.13}$$

1.2.3. Electroweak symmetry breaking and mass

The chiral nature of the weak interaction means that the LH and RH components transform differently under $SU(2)_L$. As a result, Dirac masses like $-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$, which couple the LH and RH components together, violate gauge invariance. Similarly, gauge mass terms like $\frac{1}{2}m^2A_\mu A^\mu$ also violate this symmetry. Therefore, the SM forbids explicit mass terms, and instead relies on an alternative dynamic mass-generating mechanism, called the Higgs-mechanism. The idea is that a complex doublet scalar field $\Phi = \begin{pmatrix}\Phi^+\\ \Phi^0\end{pmatrix}$ is introduced with a potential that is symmetric and vanishes at high energies, but acquires a non-zero vacuum expectation value (VEV) at lower energies (this is the famous "Mexican hat shape"). This allows the theory to manifestly preserve the required symmetry, but contain a vacuum that does not. The Lagrangian for this scalar field (which is of course known as the Higgs field) is,

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(|\Phi|) \tag{1.14}$$

with the particular Mexican-hat potential,

$$V(|\Phi|) = -\mu^2|\Phi|^2 + \lambda|\Phi|^4 \tag{1.15}$$

and with $\mu^2 < 0$. When the Higgs field condenses and acquires its non-zero VEV, it breaks the electroweak component of the SM gauge group, $G_{\text{SM}} \to \text{SU}(3)_\text{C} \times \text{U}(1)_\text{EM}$, in a process called spontaneous symmetry breaking (SSB) [19–21]. It is through this process that the $W^\pm$ and Z gauge bosons obtain their masses. The photon, whose gauge field remains invariant under the new broken symmetry of $\text{U}(1)_\text{EM}$ is massless. Similarly, the fermions also acquire mass via this process through Yukawa interactions with the Higgs field. To see this explicitly, consider the most general form of the Yukawa-interaction Lagrangian:

$$\mathcal{L}^{\text{Quark}}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}}\lambda_d^{ij}\,\bar{Q}_{L_i}\Phi D_{R_j} - \frac{1}{\sqrt{2}}\lambda_u^{ij}\,\bar{Q}_{L_i}\tilde{\Phi}U_{R_j} + \text{h.c.} \tag{1.16}$$

where λ_{ij} are the 3×3 Yukawa coupling matrices with (i,j) indexing the fermion generations. After SSB, the Higgs field takes its vacuum expectation value $\Phi \to \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}$, and Eq. (1.16) becomes,

$$\mathcal{L}^{\text{Quark}}_{\text{Yukawa}} = -\left(\lambda_d^{ij}v\right)\bar{d}_{L_i}d_{R_j} - \left(\lambda_u^{ij}v\right)\bar{u}_{L_i}u_{R_j} + \text{h.c.} \tag{1.17}$$

Comparing this to $-m\bar{\psi}\psi$, the terms highlighted in blue are interpreted as Dirac masses. Diagonalising the Yukawa matrices,

$$\lambda_u = U_L^u \Lambda_u U_R^{(u)\dagger}, \quad \lambda_d = U_L^d \Lambda_d U_R^{(d)\dagger} \tag{1.18}$$

the quark masses are given by, $\Lambda_u = \text{diag}(m_u, m_c, m_t), \quad \Lambda_d = \text{diag}(m_d, m_s, m_b)$. The transformed fields that leave the form of this Yukawa Lagrangian invariant correspond to the mass states,

$$u_L \to U_L^u u_L, \quad d_L \to U_L^d d_L, \quad u_R \to U_R^u u_R, \quad d_R \to U_R^d d_R. \tag{1.19}$$

Of course this also happens for leptons, although in that case there is only one Yukawa term, $U_L^u \Lambda_u U_R^{(u)\dagger}$. This diagonalisation has important implications for flavour mixing, as it is the origin of mixing between generations.

1.2.4. The Standard Model Lagrangian

The full Standard Model Lagrangian consists of the simplest renormalisable terms that respect the symmetries described in the previous sections. It is conventionally written as the sum:

$$\begin{aligned}
\mathcal{L}_{\text{SM}} = & \underbrace{-\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}}_{\text{Gauge kinetic term}} \\
& + \underbrace{\bar{Q}_L i \not{D} Q_L + \bar{U}_R i \not{D} U_R + \bar{D}_R i \not{D} D_R + \bar{L}_L i \not{D} L_L + \bar{E}_R i \not{D} E_R}_{\text{Fermion kinetic terms}} \\
& + \underbrace{(D^\mu \Phi)^\dagger (D_\mu \Phi)}_{\text{Higgs kinetic term}} - \underbrace{\left(\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2\right)}_{\text{Higgs potential}} \\
& - \underbrace{\left(\lambda_d \bar{Q}_L \Phi D_R + \lambda_u \bar{Q}_L \tilde{\Phi} U_R + \lambda_e \bar{L}_L \Phi E_R + \text{h.c.}\right)}_{\text{Yukawa interactions}}
\end{aligned} \tag{1.20}$$

With each line describing:

1. Gauge kinetic term: Describes the dynamics of the gauge fields and their self-couplings (or lack thereof for the Non-Abelian field).

2. Fermion kinetic terms: Describes the dynamics of the fermion fields and their couplings to the gauge fields via covariant derivatives.

3. Higgs-field terms: Describes the dynamics of the Higgs field and the specific potential responsible for SSB.

4. Yukawa interactions: Describes the Higgs-fermion field interactions, which generate mass after SSB.

1.2.5. The Standard Model neutrino

The Standard Model contains no right-handed neutrino. There is no experimental evidence for exotic or non-standard neutrino interactions, implying that neutrino interactions are well described by electroweak theory. As such, measurements of the total and partial decay widths of Z-bosons produced in e^+e^- experiments have been used to constrain the number of active light neutrino species (defined by $m_\nu \leq m_Z/2$) to

$$N_\nu = \frac{\Gamma_{\text{inv}}}{\Gamma_\ell}\left(\frac{\Gamma_\ell}{\Gamma_\nu}\right)_{\text{SM}} = 2.9840 \pm 0.0082, \tag{1.21}$$

where the invisible partial width, Γ_{inv}, is defined as the difference between the total decay width and the sum of the partial widths into quarks and charged leptons [22,23].

These active neutrinos only appear as members of the left-handed lepton doublet and carry weak isospin $T_3 = \frac{1}{2}$, and hypercharge $Y = -\frac{1}{2}$. Therefore, unlike all other fermions in the Standard Model, neutrinos do not participate in either strong or electromagnetic interactions, making them exceptionally elusive and experimentally challenging to detect.

A relevant aside here concerns the definition of "flavour generations"in the context of neutrinos. The structure of Table 1.1 is repeated twice over. These additional particle species are the generation of fermions, with the only distinction being their masses; that is they possess the same gauge interactions but have distinct Yukawa interactions. The lack of RH-neutrinos means that no neutrino Yukawa-Higgs coupling can be constructed, and thus all SM neutrinos are massless. This definition of flavour therefore breaks down for neutrinos. Instead neutrino flavours are defined in terms of their corresponding $\text{SU}(2)_L$-partner (e.g. the neutrino of the first generation is called the electron neutrino (ν_e), the second, the muon neutrino (ν_μ), and the third, the tau neutrino (ν_τ)). This leaves a degree of freedom in defining the neutrino mass states that we now know exist. We may use the structure of the mass spectrum (schematically

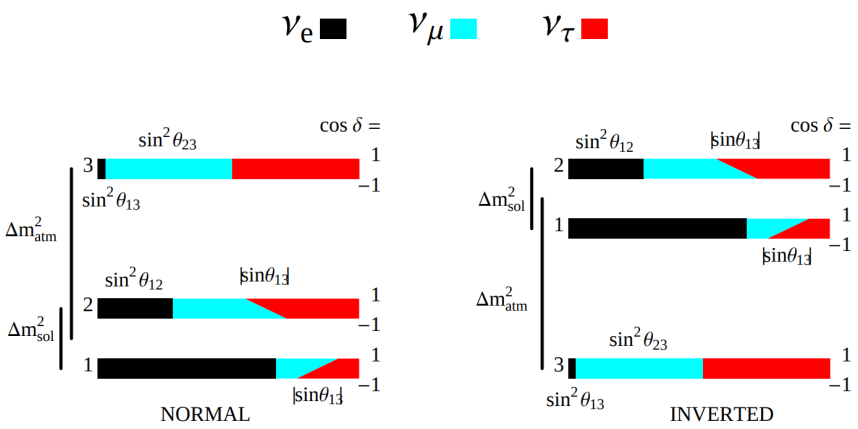


Figure 1.4.: The three-neutrino mass spectrum for normal and inverted mass ordering. The mass states are defined using the convention prescribed earlier in Sec. 1.2.5. The flavour content of each mass eigenstate is indicated by colour (ν_e black, ν_μ cyan, ν_τ red). Reproduced from Ref. [24].

depicted in Fig. 1.4) to define the mass states relative to each other: the lightest (heaviest) of the two nearly degenerate mass states is ν_1 (ν_2), while the distant other mass state is ν_3. However, we also need a scheme for relating them to the matrix elements. Because $\left(|U_{e1}|^2 \sim 0.66\right) > \left(|U_{e2}|^2 \sim 0.33\right) > \left(|U_{e3}|^2 = 0.02\right)$ one of most common schemes is to define them in terms of decreasing ν_e content, where ν_1 has the most and ν_3 least.

1.2.6. Flavour and generational mixing

Expanding out the gauge-fermion couplings of the fermion kinetic terms in Eq. (1.20) will show how the gluon, photon, and Z fields couple fermions exclusively to their conjugate states, and are therefore flavour diagonal. In contrast, interactions involving W-exchange couple up-type and down-type flavours:

$$\begin{aligned}\mathcal{L}_{\text{CC}}^{\text{Quark}} = g\,\overline{Q_L}\gamma^\mu(\tau_1 W_\mu^1 + \tau_2 W_\mu^2)Q_L &= \frac{g}{\sqrt{2}}\overline{Q_L}\gamma^\mu\tau^+ W_\mu^+ Q_L + \text{h.c.} \\ &= \frac{g}{\sqrt{2}}\overline{(u\ c\ t)}_L\gamma^\mu \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L W_\mu^+ + \text{h.c.}\end{aligned} \tag{1.22}$$

where $W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right)$ are the physical bosons, arising after SSB as a linear combination of the original gauge fields of the unbroken symmetry, and $\tau^\pm = \frac{1}{2}\left(\tau_1 \pm i\tau_2\right)$ are weak isospin ladder operators expressed in terms of two of the generators of $\mathrm{SU}(2)_L$. As a result, the charged-current (CC) weak interaction is the only flavour-changing process in the SM. Furthermore, since the up-quark and down-quark Yukawa matrices are not diagonal in the same basis, the CC-interaction also couples quarks across different generations. This is because the field redefinitions performed in Eq. (1.18) to rotate into the mass basis do not cancel in the W-couplings of Eq. (1.22). Consequently, CC-weak interactions couple quarks of different generations.

$$\begin{aligned}\mathcal{L}_{\mathrm{CC}}^{\mathrm{Quark}} &= \frac{g}{\sqrt{2}} W_\mu^+ \left(U_L^{(d)} U_L^{(u)\dagger}\right)_{ij} \overline{u_L^i} \gamma^\mu d_L^j + \mathrm{h.c.} \\ &= \frac{g}{\sqrt{2}} W_\mu^+ U_{ij}^{\mathrm{CKM}} \overline{u_L^i} \gamma^\mu d_L^j + \mathrm{h.c.}. \end{aligned} \tag{1.23}$$

where U^{CKM} is the quark mixing matrix known as the CKM matrix [25, 26]. In the SM, there is only a single lepton–Higgs interaction term. This allows for the simultaneous diagonalisation of both the gauge and Yukawa interactions, meaning that neutrinos of a given flavour couple exclusively to charged leptons of the same flavour. However, the observation of neutrino oscillations clearly contradicts this. Since this is the only SM mechanism that involves inter-generational mixing, an analogous situation must exist for neutrinos and so neutrinos must be massive.

There are at least two ways to introduce a neutrino mass term. The simplest is to add a RH-neutrino that generates an orthodox Dirac mass by participating in Yukawa couplings with the Higgs field. However, as neutrinos are uncharged, it is also possible that they are their own antiparticles (Majorana fermions), meaning the distinction between LH-neutrinos and RH-antineutrinos is not fundamental, as they can transform into each other ($\nu_R = \nu_L^c$)[1] [27]. This would permit the construction of a Majorana mass term, $\frac{1}{2} m_\psi \left(\bar{\psi}_L^c \psi_L + \bar{\psi}_L \psi_L^c\right)$, which involves only one chirality of neutrino (the factor $\frac{1}{2}$ compensates for double counting $\bar{\nu}_L^c \nu_L = \bar{\nu}_L \nu_L^c$). Generating such a Majorana mass term via the Higgs mechanism requires the introduction of the dimension-five Weinberg operator, which is the lowest-dimensional extension that can achieve this. However, its inclusion renders the theory non-renormalisable, thereby reducing the SM to an

[1] Because of helicity suppression, the rate of this transformation only becomes appreciable well below the energies that neutrinos are experimentally observed, $(m_\nu/p_\nu)^2 \sim (0.1/10^6)^2 = 10^{-14}$. As a result, this flipping is not a viable way of determining the mass nature of neutrinos.

effective field theory valid only below some higher energy scale [28]. Importantly, this operator explicitly violates lepton number by two units.

Lepton number is one of the "accidental symmetries" of the SM, in that it is not imposed by construction, but rather emerges because none of the SM interactions happen to violate it. If neutrinos are Majorana, then there should exist examples of lepton-number violating processes in nature, which would conclusively show that neutrinos are Majorana particles. One of the most well-known of these is neutrinoless double beta decay, which is currently being searched for in many experiments (e.g. CUORE [29] and CUPID [30] to name a few).

The specific origins of neutrino masses and their nature are of little relevance to neutrino oscillations though; from an oscillation perspective, all that is required is the existence of a nontrivial neutrino mass matrix whose eigenstates differ from those of the charged leptons. Writing the rotated lepton fields as $\ell_\alpha = V^{\ell}_{\alpha j}\ell_j$ and $\bar{\nu}_\alpha = V^{(\nu)\dagger}_{\alpha j}\bar{\nu}_j$, we obtain,

$$\mathcal{L}^{\text{Lepton}}_{\text{CC}} = -\frac{g}{\sqrt{2}} \sum_{\substack{\alpha=e,\mu,\tau \\ j=1,2,3}} W^+_\mu (V^{(\nu)\dagger}V^{\ell})_{\alpha j}\bar{\nu}^j_L \gamma^\mu \ell^j_L + h.c. \tag{1.24}$$

where $\mathbf{U}_{\text{PMNS}} = (V^{(\nu)\dagger}V^{\ell})_{\alpha j}$ is the leptonic analogue of the CKM matrix, named in honour of Pontecorvo, Maki, Nakagawa and Sakata [31–34]. The literal physical interpretation of these matrix elements is that they define the vertex amplitude, $\frac{g}{\sqrt{2}}U_{\alpha j}$ (from here on U refers to $\mathbf{U}_{\text{PMNS}}$), for a neutrino ν_i to produce a charged lepton ℓ_α by emission or absorption of a W-boson. Technically speaking, $\mathbf{U}_{\text{PMNS}}$ need only be non-diagonal for mixing to occur; that is, the relative rotation between the charged leptons and neutrinos is what generates the mixing. For the CKM-matrix, the convention is to define the up-type quark mass and interaction states as aligned. Similarly, we define the neutrino flavour states to coincide with their mass eigenstates so that all of the rotation is done by the neutrinos ($V^{\ell} = \mathbb{1}, V^{(\nu)} = U$). For leptons, however, this choice has a more practical motivation; the much larger charged-lepton masses would decohere over extremely short distances, which would make it difficult to probe this physics through interference phenomena. Thus, the $\mathbf{U}_{\text{PMNS}}$ is interpreted as the matrix

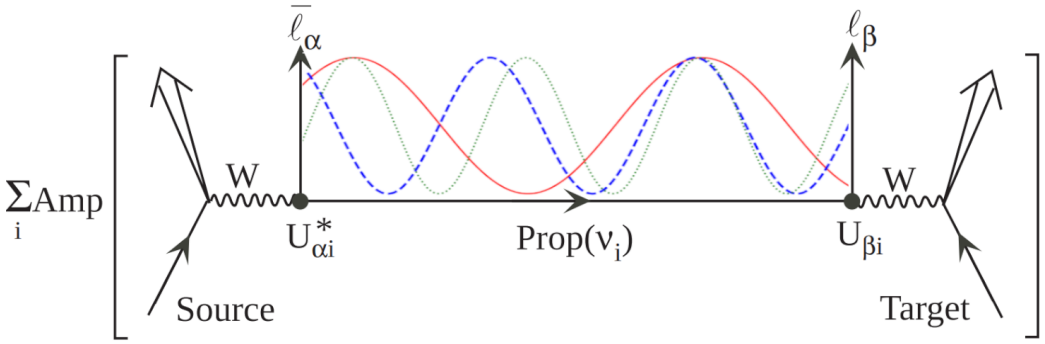


Figure 1.5.: Pictorial representation of neutrino oscillation with the different colours denoting the states of definite mass. Altered figure from Ref. [35].

relating the neutrino flavour and mass states:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{1.25}$$

This is consistent with the usual way neutrino oscillations are considered: as purely a quantum phenomenon in which the neutrino interferes with itself. If a neutrino, initially prepared as a pure flavour state, evolves out of that state over time, then the flavour states are not eigenstates of the Hamiltonian. As they span Hilbert space, they can be written as a superposition of the states of definite mass. Since each mass state accumulates phase at a different rate, the superposition changes, and the probability of detecting a neutrino produced in $W \to \ell_\alpha + \nu_\alpha$ to interact later and produce a lepton of flavour different than α, can deviate from 1. The oscillation probability for this is determined by considering the usual quantum amplitude yielded by $|\langle \nu_\alpha | \nu(t) \rangle|$, where $|\nu(t)\rangle$ is the coherent sum over all mass states (this is pictorially depicted in Fig. 1.5).

1.3. Oscillation probability

1.3.1. General oscillation probability

By making a few basic assumptions, it is fairly easy to derive a general expression for the probability of flavour transitions. In the context of single-particle systems we can

write,

$$|\nu_\alpha\rangle = \sum_i U^*_{\alpha i}|\nu_i\rangle, \tag{1.26}$$

where $|\nu_\alpha\rangle$ is created by acting the conjugate neutrino field on the vacuum, $|0\rangle$ [5]. This state only contains information on the particle properties (e.g. the internal degrees of freedom). We write the full state-vector, complete with energy-momentum dependence as $|\Psi_\alpha(\mathbf{x},t)\rangle$, and therefore the expression for a propagating weak eigenstate as,

$$\begin{aligned} |\Psi_\alpha(\mathbf{x},t)\rangle &= \sum_{i=1,2,3} U^*_{\alpha i}|\Psi_i(\mathbf{x},t)\rangle \\ \implies |\Psi_\alpha(\mathbf{x},t)\rangle &= \sum_{i=1,2,3} U^*_{\alpha i}\psi_i(\mathbf{x},t)|\nu_i\rangle. \end{aligned} \tag{1.27}$$

where $\psi_i(\mathbf{x},t)$ is the operator that evolves the mass states in space and time. If a neutrino of flavour α is produced at $x = 0$, $t = 0$, the transition amplitude at a later arbitrary time, t, will be the projection

$$A(\alpha \to \beta, t) = \langle \Psi_\beta(\mathbf{x},t)|\Psi_\alpha(0,0)\rangle. \tag{1.28}$$

Squaring, we find the corresponding probability is

$$\begin{aligned} P(\alpha \to \beta, t) &= |\langle \Psi_\beta(\mathbf{x},t)|\Psi_\alpha(0,0)\rangle|^2 \\ &= \left| \left(\sum_{j=1,2,3} U_{\beta j}\psi^*_j(\mathbf{x},t)\langle\nu_j| \right) \left(\sum_{i=1,2,3} U^*_{\alpha i}\psi_i(0,0)|\nu_i\rangle \right) \right| \\ &= \left| \sum_{i=1,2,3}\sum_{j=1,2,3} U^*_{\alpha i}U_{\beta j}\psi^*_j(\mathbf{x},t)\psi_i(0,0)\langle\nu_j|\nu_i\rangle \right|. \end{aligned} \tag{1.29}$$

Given that $\langle\nu_j|$ and $|\nu_i\rangle$ are by definition orthogonal,

$$\begin{aligned} P(\alpha \to \beta, t) &= \left| \sum_{i=1,2,3}\sum_{j=1,2,3} U^*_{\alpha i}U_{\beta j}\psi^*_j(\mathbf{x},t)\psi_i(0,0)\delta_{ij} \right|^2 \\ &= \left| \sum_{i=1,2,3} U^*_{\alpha i}U_{\beta i}\psi^*_i(\mathbf{x},t)\psi_i(0,0) \right|^2. \end{aligned} \tag{1.30}$$

The rest of this derivation (which roughly follows that produced in Ref. [5]) is based on the following simplifying assumptions:

1. The wave function of a propagating mass eigenstate can be described as a plane wave with definite momenta. This implies that,

$$\int \frac{d\mathbf{p}}{(2\pi)^{3/2}} \psi_i(\mathbf{p}) = 1, \tag{1.31}$$

as the Fourier space will consist of a single delta function centred on $\mathbf{p_i}$. Thus the space-time evolution operator reduces to

$$\psi_i(\mathbf{x}, t) = e^{i(\mathbf{p_i}\cdot\mathbf{x} - E_i t)}. \tag{1.32}$$

While this assumption is unrealistic, since a particle with definite momentum is completely delocalised, a rigorous consideration, by modelling neutrinos as propagating wave-packets, demonstrates this is justified, provided the mass states maintain coherence from production to detection [36].

2. That all flavour states have the same definite momentum.

3. Neutrinos travel at ultrarelativistic velocities. This means that the particle's energy is almost entirely described by the momentum ($|\mathbf{p_i}| \gg m_i$) and thus the relativistic dispersion relation reduces approximately to,

$$E_i = \sqrt{|\mathbf{p}|^2 + m_i^2} \approx |\mathbf{p}| + \frac{m_i^2}{2}|\mathbf{p}|, \tag{1.33}$$

where we have expanded in terms of a Taylor series and justified our truncation with ($|\mathbf{p_i}| \gg m_i$). This also allows us to approximate ultrarelativistic time to distance ($x \approx t$).

Applying these assumptions and letting the direction of propagation be along the x-axis, Eq. (1.31) yields,

$$\psi_i(x, t) = e^{i(|\mathbf{p}|x - E_i t)} \approx e^{i\left(|\mathbf{p}|x - (|\mathbf{p}| + \frac{m_i^2}{2|\mathbf{p}|})t\right)} = e^{i\left(|\mathbf{p}|(x-t) - \frac{m_i^2}{2|\mathbf{p}|}t\right)}. \tag{1.34}$$

Substituting this result into Eq. (1.29), we obtain:

$$\begin{aligned} P(\alpha \to \beta, t) &= \left| \sum_i U^*_{\alpha i} U_{\beta i} e^{i\left(|\mathbf{p}|(x-t) - \frac{m_i^2}{2|\mathbf{p}|}t\right)} \right|^2 \\ &= \left(\sum_i U^*_{\alpha i} U_{\beta i} e^{i\left(|\mathbf{p}|(x-t) - \frac{m_i^2}{2|\mathbf{p}|}t\right)} \right) \left(\sum_j U_{\alpha j} U^*_{\beta j} e^{-i\left(|\mathbf{p}|(x-t) - \frac{m_i^2}{2|\mathbf{p}|}t\right)} \right) \\ &= \sum_i |U^*_{\alpha i}|^2 |U_{\beta i}|^2 + \sum_{i \neq j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{-i\left(\frac{\Delta m^2_{ij}}{2|\mathbf{p}|}t\right)}. \end{aligned} \tag{1.35}$$

Using the identity,

$$\sum_{i \neq j} X_{ij} = \sum_{i>j} (X_{ij} + X_{ji}) = 2\,\mathrm{Re}\left[\sum_{i>j} X_{ij}\right],$$

and applying the ultrarelativistic approximations, $|\mathbf{p}| \approx E$ and $x \approx t$ (it is more common to write the propagation distance, x, as L), we obtain the general transition probability:

$$P(\alpha \to \beta, L) = \sum_i |U^*_{\alpha i}|^2 |U_{\beta i}|^2 + 2\Re\mathfrak{e}\left[\sum_{i>j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{-i\left(\frac{\Delta m^2_{ij}}{2E}L\right)}\right]. \tag{1.36}$$

If neutrinos are produced in an incoherent way (without interference), or they decohere via wave packet separation, the oscillatory term of Eq. (1.36) vanishes and the oscillation probability takes the form of a constant classical transition probability that is independent of L/E.

This transition probability may be expressed in several equivalent forms. A popular one is as a sum of the real and imaginary components of these so called quartets (rephasing-invariant products of the form $V^*_{\alpha i} V_{\beta i} V_{\alpha j} V^*_{\beta j}$). This is particularly useful because it explicitly separates the real and imaginary components of the oscillation probability. This is achieved using the unitarity relation of the PMNS matrix, which allows the non-oscillatory contribution to be written as

$$\sum_i |U^*_{\alpha i}|^2 |U_{\beta i}|^2 = \delta_{\alpha\beta} - 2\sum_{i>j} \Re\mathfrak{e}\left[U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j}\right]. \tag{1.37}$$

Substituting this in to Eq. (1.36) and using the identity $\mathrm{Re}(ab) = \mathrm{Re}(a)\mathrm{Re}(b) - \mathrm{Im}(a)\mathrm{Im}(b)$ yields,

$$P(\alpha \to \beta) = \delta_{\alpha\beta} - 4\sum_{i>j} \Re\mathfrak{e}\left[U^*_{\alpha i}U_{\beta i}U_{\alpha j}U^*_{\beta j}\right] \sin^2\left(\frac{\Delta m^2_{ij}L}{4E}\right) + 2\sum_{i>j} \Im\mathfrak{m}\left[U^*_{\alpha i}U_{\beta i}U_{\alpha j}U^*_{\beta j}\right] \sin\left(\frac{\Delta m^2_{ij}L}{2E}\right). \quad (1.38)$$

Experimentally, this probability is the fundamental theory through which neutrino oscillation parameters are determined. The mixing parameters of $\mathbf{U}_{\mathrm{PMNS}}$ govern the oscillation amplitude, while the squared mass differences determine the oscillation frequency. In experimental language, when the source emits ν_α, the channel $P(\alpha \to \beta)$ is referred to as ν_β appearance, while $P(\alpha \to \alpha)$ is called ν_α disappearance.

From Eq. (1.38), it is clear that if more than two neutrinos are massless then, $\Delta m_{ij} = 0\ \forall\{i,j\}$. Additionally, the three neutrino masses must be distinct, otherwise two independent mass-difference terms would not manifest.

1.3.2. Standard parametrisation

A complex $N \times N$ matrix is fully specified by $2N^2$ real parameters. Given that the mixing matrix is unitary, this imposes $N(N-1)+N$ constraints and reduces the number of independent real parameters to N^2: N subtracted from the normality condition, and $2N(N-1)/2$ from the orthogonality conditions. Additionally, not all of the complex phases are physical since N phases may be be factorised into the lepton fields by field redefinition, $\ell_\alpha \to \ell_\alpha e^{-i\phi_i}$. If neutrinos have Dirac masses, we may similarly remove $N-1$ additional phases by redefining the neutrino field. If they are Majorana then we are not able to do this as it would result in a complex mass term. However, the oscillation-dependent part of Eq. (1.36) is composed of quartets of mixing matrix elements. Since Majorana phases appear as global phases, they cancel out in these products. As a result, oscillation experiments are insensitive to them [37].

We can factorise the full mixing matrix (F) as the product of a matrix containing rotations and Dirac phases, and a matrix containing the Majorana phases (P),

$$F = \mathbf{U}_{\mathrm{PMNS}}P. \quad (1.39)$$

For N generations, $\mathbf{U}_{\text{PMNS}}$ therefore has $(N-1)(N-2)/2$ physically significant phases, and $N(N-1)/2$ angles. Including the neutrino mass values, lepton mixing for three neutrinos therefore introduces 7–9 fundamental parameters to the SM (depending on the nature of neutrino masses). Oscillations are sensitive to six of these parameters, as only relative mass differences contribute to phase differences and therefore generate interference. The matrix $\mathbf{U}_{\text{PMNS}}$ is conventionally parametrised by three Tait–Bryan rotations and a single Dirac phase,

$$\mathbf{U}_{\text{PMNS}} = R(\theta_{23})\Gamma_\delta R(\theta_{13})\Gamma_\delta^\dagger R(\theta_{12}), \tag{1.40}$$

where $\Gamma_\delta = \text{diag}(1, 1, e^{i\delta})$. For historical reasons, the complex phase, δ, is associated with the θ_{13} rotation, and $\Gamma_\delta R(\theta_{13})\Gamma_\delta^\dagger$ is interpreted as a single complex rotation $R(\theta_{13}, \delta)$. Interestingly, the order of the rotations can impact the statements on their parameter constraints, although Eq. (1.40) has been found to have the most desirable properties [38]. Expanding $\mathbf{U}_{\text{PMNS}}$ explicitly we obtain:

$$\begin{aligned}\mathbf{U}_{\text{PMNS}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}\end{aligned} \tag{1.41}$$

with $c_{ij} = \cos\theta_{ij}, s_{ij} = \sin\theta_{ij}$. In terms of the $\mathbf{U}_{\text{PMNS}}$ elements, these are formally defined as,

$$\sin^2\theta_{23} = \frac{|U_{\mu 3}|^2}{1-|U_{e3}|^2} \quad \sin^2\theta_{13} = |U_{e3}|^2 \quad \sin^2\theta_{12} = \frac{|U_{e2}|^2}{1-|U_{e3}|^2}, \quad \delta_{\mathcal{CP}} = \arg(U_{e3}) \tag{1.42}$$

though this is typically visualised as the angles defining how the planes of two embedded coordinate systems are rotated relative to each other (e.g. θ_{13} quantifies the rotation in the (ν_e, ν_3)-plane). Fig. 1.6 shows the evolution of experimental constraints on these parameters over the past two and a half decades.

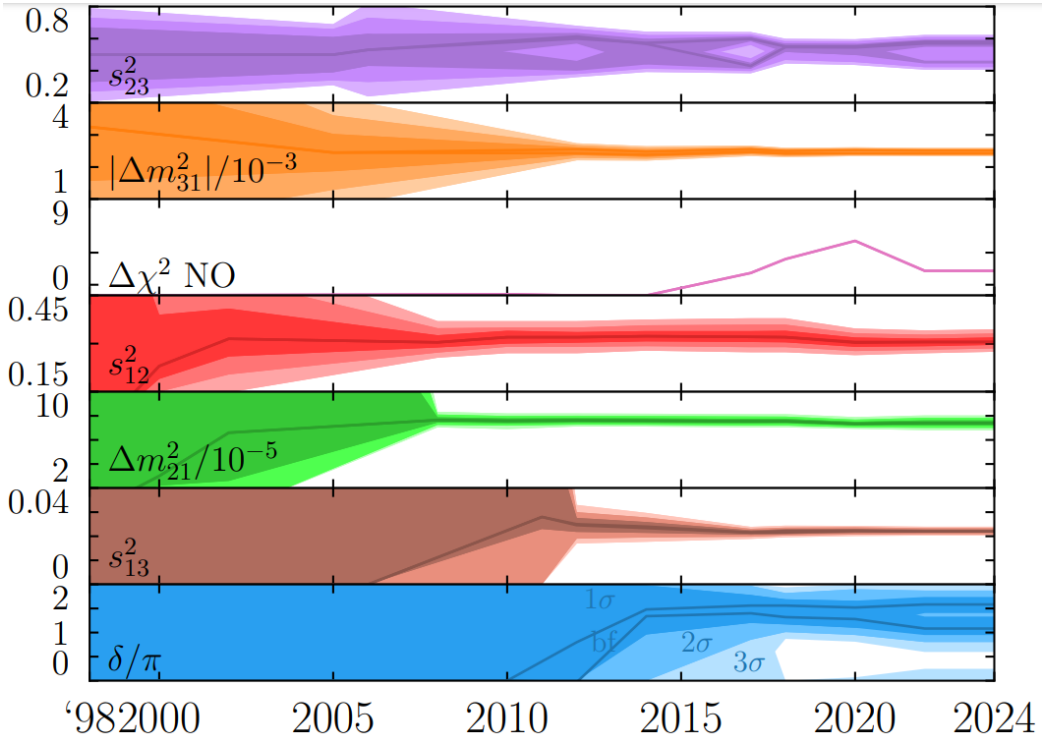


Figure 1.6.: Evolution of oscillation parameter constraints since the discovery of neutrino oscillations. The solid lines indicate the best-fit values, while the shaded regions show the 1σ, 2σ, and 3σ intervals. Reproduced from Ref. [27].

In this parametrisation, Fig. 1.7 can be used to construct the full vacuum probability for any channel from the general relation of Eq. (1.38): the middle column gives the real quartet terms in the sum, and the last column is the sign of the Jarlskog invariant $\Im\mathfrak{m}[U^*_{\alpha k}U_{\beta k}U_{\alpha j}U^*_{\beta j}] = s^{kj}_{\alpha\beta}J_{\mathcal{CP}}$ [39]:

$$s^{kj}_{\alpha\beta}J_{\mathcal{CP}} = \begin{cases} +J_{\mathcal{CP}} & \text{for } (\alpha,\beta) = (e,\mu),\ (\mu,\tau),\ (\tau,e) \\ +J_{\mathcal{CP}} & \text{for } (i,j) = (1,2),\ (2,3),\ (3,1) \\ -J_{\mathcal{CP}} & \text{otherwise} \end{cases} \tag{1.43}$$

1.3.3. $\mathcal{CP}$-violation in oscillation phenomena

$\mathcal{CP}$ transformations interchange neutrino processes with their corresponding antineutrino processes. To determine under which conditions antineutrino oscillations differ from neutrino oscillations, we must derive the corresponding transition probabilities

$\alpha\beta$	kj	$\Re\mathfrak{e}\left[U^*_{\alpha k}\,U_{\beta k}\,U_{\alpha j}\,U^*_{\beta j}\right]$	$s_{\alpha\beta;kj}$
$e\mu$	21	$-\frac{1}{4}\sin 2\vartheta_{12}\,c^2_{13}\left[\sin 2\vartheta_{12}\left(c^2_{23}-s^2_{23}s^2_{13}\right)+\cos 2\vartheta_{12}\,\sin 2\vartheta_{23}\,s_{13}\cos\delta_{13}\right]$	$+1$
	32	$-s_{12}s_{23}c^2_{13}s_{13}\left(s_{12}s_{23}s_{13}-c_{12}c_{23}\cos\delta_{13}\right)$	$+1$
	31	$-c_{12}s_{23}c^2_{13}s_{13}\left(c_{12}s_{23}s_{13}+s_{12}c_{23}\cos\delta_{13}\right)$	-1
$\mu\tau$	21	$\frac{1}{16}\sin^2 2\vartheta_{12}\,\sin^2 2\vartheta_{23}\left(1+s^2_{13}\right)^2-\frac{1}{4}\left(\sin^2 2\vartheta_{12}+\sin^2 2\vartheta_{23}\right)s^2_{13}$ $-\frac{1}{16}\sin 4\vartheta_{12}\,\sin 4\vartheta_{23}\left(1+s^2_{13}\right)s_{13}\cos\delta_{13}$ $+\frac{1}{4}\sin^2 2\vartheta_{12}\,\sin^2 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}$	$+1$
	32	$-\frac{1}{4}\sin 2\vartheta_{23}\,c^2_{13}\left[\sin 2\vartheta_{23}\left(c^2_{12}-s^2_{12}s^2_{13}\right)+\sin 2\vartheta_{12}\,\cos 2\vartheta_{23}\,s_{13}\cos\delta_{13}\right]$	$+1$
	31	$\frac{1}{4}\sin 2\vartheta_{23}\,c^2_{13}\left[\sin 2\vartheta_{23}\left(c^2_{12}s^2_{13}-s^2_{12}\right)+\sin 2\vartheta_{12}\,\cos 2\vartheta_{23}\,s_{13}\cos\delta_{13}\right]$	-1
τe	21	$\frac{1}{4}\sin 2\vartheta_{12}\,c^2_{13}\left[\sin 2\vartheta_{12}\left(c^2_{23}s^2_{13}-s^2_{23}\right)+\cos 2\vartheta_{12}\,\sin 2\vartheta_{23}\,s_{13}\cos\delta_{13}\right]$	$+1$
	32	$-s_{12}c_{23}c^2_{13}s_{13}\left(s_{12}c_{23}s_{13}+c_{12}s_{23}\cos\delta_{13}\right)$	$+1$
	31	$-c_{12}c_{23}c^2_{13}s_{13}\left(c_{12}c_{23}s_{13}-s_{12}s_{23}\cos\delta_{13}\right)$	-1
ee	21	$\frac{1}{4}\sin^2 2\vartheta_{12}\,c^4_{13}$	0
	32	$\frac{1}{4}s^2_{12}\,\sin^2 2\vartheta_{13}$	0
	31	$\frac{1}{4}c^2_{12}\,\sin^2 2\vartheta_{13}$	0
$\mu\mu$	21	$\frac{1}{4}\sin^2 2\vartheta_{12}\left(c^4_{23}+s^4_{23}s^2_{13}\right)+\frac{1}{4}\left(1-\frac{1}{2}\sin^2 2\vartheta_{12}\right)\sin^2 2\vartheta_{23}\,s^2_{13}$ $+\frac{1}{4}\sin 4\vartheta_{12}\,\sin 2\vartheta_{23}\left(c^2_{23}-s^2_{23}s^2_{13}\right)s_{13}\cos\delta_{13}$ $-\frac{1}{4}\sin^2 2\vartheta_{12}\,\sin^2 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}$	0
	32	$s^2_{23}c^2_{13}\left(c^2_{12}c^2_{23}+s^2_{12}s^2_{23}s^2_{13}-\frac{1}{2}\sin 2\vartheta_{12}\,\sin 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}\right)$	0
	31	$s^2_{23}c^2_{13}\left(s^2_{12}c^2_{23}+c^2_{12}s^2_{23}s^2_{13}+\frac{1}{2}\sin 2\vartheta_{12}\,\sin 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}\right)$	0
$\tau\tau$	21	$\frac{1}{4}\sin^2 2\vartheta_{12}\left(s^4_{23}+c^4_{23}s^2_{13}\right)+\frac{1}{4}\left(1-\frac{1}{2}\sin^2 2\vartheta_{12}\right)\sin^2 2\vartheta_{23}\,s^2_{13}$ $+\frac{1}{4}\sin 4\vartheta_{12}\,\sin 2\vartheta_{23}\left(s^2_{23}-c^2_{23}s^2_{13}\right)s_{13}\cos\delta_{13}$ $-\frac{1}{4}\sin^2 2\vartheta_{12}\,\sin^2 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}$	0
	32	$c^2_{23}c^2_{13}\left(c^2_{12}s^2_{23}+s^2_{12}c^2_{23}s^2_{13}+\frac{1}{2}\sin 2\vartheta_{12}\,\sin 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}\right)$	0
	31	$c^2_{23}c^2_{13}\left(s^2_{12}s^2_{23}+c^2_{12}c^2_{23}s^2_{13}-\frac{1}{2}\sin 2\vartheta_{12}\,\sin 2\vartheta_{23}\,s^2_{13}\cos^2\delta_{13}\right)$	0

Figure 1.7.: Real quartet terms. Reproduced from Ref. [36]

for antineutrinos. To do that we make use of the $\mathcal{CPT}$ invariance of Eq. (1.36):

$$\begin{aligned} & P(\alpha \to \beta, t) \overset{CPT}{=} P(\bar{\beta} \to \bar{\alpha}, t) \\ \implies & P(\bar{\beta} \to \bar{\alpha}, t) = \sum_i |U^*_{\alpha i}|^2 |U_{\beta i}|^2 + 2\Re\mathfrak{e}\left[\sum_{i \neq j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{\left(-\frac{i\Delta m^2_{ij}}{2E} L\right)}\right]. \end{aligned} \tag{1.44}$$

The probability $P(\bar{\alpha} \to \bar{\beta}, t)$ is found by interchanging α and β, or equivalently swapping the mixing matrix with its complex conjugates:

$$P(\bar{\alpha} \to \bar{\beta}, t) = \sum_i |U^*_{\beta i}|^2 |U_{\alpha i}|^2 + 2\Re\mathfrak{e}\left[\sum_{i \neq j} U^*_{\beta i} U_{\alpha i} U_{\beta j} U^*_{\alpha j} e^{(-\frac{i\Delta m^2_{ij}}{2E} L)}\right]. \tag{1.45}$$

That is, a necessary condition for the probabilities $\alpha \to \beta$ and $\bar{\alpha} \to \bar{\beta}$ to be the same is that the mixing matrix be real ($U^*_{\alpha i} = U_{\alpha i}, U^*_{\alpha j} = U_{\alpha j}, U^*_{\beta i} = U^*_{\beta i}, U^*_{\beta j} = U^*_{\beta j}$) [35]. Therefore, complex phases in the mixing matrix that cannot be eliminated by means of field re-definitions, and whose values are non-integer multiple of π, represent a source of $\mathcal{CP}$-violation ($\mathcal{CP}$v). Additionally, as there are no physically significant Dirac-phases for fewer than three-generations, $\mathcal{CP}$v from flavour mixing is inherently a three generation phenomenon[2]. Indeed, the observation of $\mathcal{CP}$v in neutral-kaon experiments was presented as evidence of the existence of a third generation [40]. It is curious that the only known source of $\mathcal{CP}$-violation in the SM, and therefore the fact that physics can appear different backwards in time (via the $\mathcal{CPT}$ theorem), arises from the seemingly benign question of whether the mixing matrix (for both quarks and leptons) is real or complex. An important note is that because the $\mathcal{CP}$-probability was obtained by swapping α and β, $\alpha \neq \beta$ is required for this change to be non-trivial. Consequently, $\mathcal{CP}$v terms manifest only in appearance probabilities. Interestingly though, terms proportional to $\mathrm{sgn}(\sin\delta)$, namely $\cos\delta_{\mathcal{CP}}$, are still present in the disappearance probabilities and are therefore informative about the size of $\mathcal{CP}$v, if it exists.

Because it is independent of parameterisation, it is also common to see $\mathcal{CP}$v reported using the Jarlskog invariant measure of $\mathcal{CP}$v. In the PDG convention, it takes the form:

$$J_{\mathcal{CP}} = \frac{1}{8}\cos(\theta_{13})\sin(2\theta_{12})\sin(2\theta_{23})\sin(2\theta_{13})\sin(\delta_{\mathcal{CP}}) \tag{1.46}$$

[2] Technically if neutrinos are Majorana particles there is a possibility of having $\mathcal{CP}$v with only two generations

with $J_{\mathcal{CP}} \neq 0$ signifying $\mathcal{CP}$v.

1.4. Neutrino propagation in matter

1.4.1. The matter effect

Neutrinos propagating through matter will of course interact with the particles of that medium. While the cross section for a neutrino interaction with a given particle is inherently small (i.e. for neutrino-proton scattering $\sigma \sim \frac{G_F^2 s}{\pi} \sim 10^{-43}\,\mathrm{cm}^2 \left(\frac{E}{\mathrm{MeV}}\right)^2$), if these interactions occur in such a way that the particles of the medium remain unchanged, so that individual scattering events cannot be identified, the corresponding amplitudes add coherently, enhancing the overall effect. In this regime, neutrinos interact with the medium as a whole through a process known as coherent forward scattering (CFS). This effect manifests as an effective background potential proportional to G_F, rather than as a sequence of individual scattering events with probabilities proportional to G_F^2. Therefore, CFS alters the Hamiltonian via the addition of a non-negligible effective potential [41]. Since matter is, leptonically speaking, predominantly composed of electrons, and since coherent forward scattering via W-exchange with electrons can only happen for ν_e, this extra potential affects the Hamiltonian asymmetrically.

Another approach of obtaining oscillation probabilities is by solving the Schrödinger-like equation:

$$i\frac{\partial}{\partial t}|\nu(t)\rangle = \mathcal{H}|\nu(t)\rangle \tag{1.47}$$

This formalism is useful as it provides a convenient way to include perturbations to the Hamiltonian, such as the matter induced effective potential:

$$i\frac{\partial}{\partial t}|\nu(t)\rangle = \mathcal{H}_{\text{Free}}|\nu(t)\rangle + \mathcal{H}_{\text{Interacting}}|\nu(t)\rangle. \tag{1.48}$$

To illustrate the effects of matter interactions, it is useful to consider a simple two-generation approximation. Since these additional potentials arise through weak inter-

actions, it is necessary to consider the Hamiltonian initially in the weak basis:

$$H_f = \begin{pmatrix} \langle \nu_\alpha | H_m | \nu_\alpha \rangle & \langle \nu_\alpha | H_m | \nu_\beta \rangle \\ \langle \nu_\beta | H_m | \nu_\beta \rangle & \langle \nu_\beta | H_m | \nu_\beta \rangle \end{pmatrix} \tag{1.49}$$

where H_m is the vacuum Hamiltonian in the mass-basis. The matrix elements of Eq. (1.49) are given by,

$$\begin{aligned} \langle \nu_\alpha | H_m | \nu_\beta \rangle &= \langle \sum_i U^*_{\alpha i} \nu_i | H | \sum_j U_{\beta j} \nu_j \rangle \\ &= \sum_i \sum_j U^*_{\alpha i} U_{\beta j} \langle \nu_i | H | \nu_j \rangle = \sum_j U^*_{\alpha j} U_{\beta j} \left(E + \frac{m_j^2}{2E} \right), \end{aligned} \tag{1.50}$$

where again the usual ultra-relativistic and equal momenta approximations, $\sqrt{p_i + m_i} = p + \frac{m_i^2}{2p} = E + \frac{m_i^2}{2E}$, have been used. Doing this, Eq. (1.49) becomes,

$$\begin{aligned} H_f &= \begin{pmatrix} [\frac{-\Delta m^2}{2E}] \cos^2\theta + (E + \frac{m_2^2}{2E}) & [\frac{\Delta m^2}{2E}] \cos\theta \sin\theta \\ [\frac{\Delta m^2}{2E}] \cos\theta \sin\theta & [\frac{\Delta m^2}{2E}] \cos^2\theta (E + \frac{m_1^2}{2E}) \end{pmatrix} \\ &= (E + \frac{m_1 + m_2}{2E}) \mathbb{I} + \frac{\Delta m^2}{2E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} \end{aligned} \tag{1.51}$$

where in the second line, $\frac{1}{2}\mathrm{Tr}(H_f)$ has been factored out. With the Hamiltonian now in the flavour basis, the effective potentials arising from NC-interactions, V_Z, and CC-interactions, $V_\mathrm{W} = \sqrt{2} G_\mathrm{F} N_e$ (G_F is the Fermi constant and N_e is electron density), can be introduced:

$$H_f = \underbrace{\left(E + \frac{m_1 + m_2}{4E} \right) \mathbb{I} + \frac{\Delta m^2}{4E} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}}_{H_\mathrm{Free}} + \underbrace{\begin{pmatrix} V_\mathrm{Z} & 0 \\ 0 & V_\mathrm{Z} \end{pmatrix} + \begin{pmatrix} V_\mathrm{W} & 0 \\ 0 & 0 \end{pmatrix}}_{H_\mathrm{Interacting}}. \tag{1.52}$$

Rotating back into the mass basis, the effective Hamiltonian is,

$$\begin{aligned} H = U^\dagger H_f U &= (E + V_Z)\,\mathbb{I} + \frac{1}{4E}\left[(m_1^2 + m_2^2)\mathbb{I} + \Delta m^2 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\right] + U^\dagger \begin{pmatrix} V_W & 0 \\ 0 & 0 \end{pmatrix} U \\ &= (E + V_Z)\mathbb{I} + \frac{1}{4E}\begin{pmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{pmatrix} + V_W \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}. \end{aligned} \tag{1.53}$$

Defining $A = 2EV_W = 2\sqrt{2}G_F N_e E$ and omitting the terms proportional to unity, $(E + V_Z)I$, as only relative phases are observable in oscillation phenomena, Eq. (1.53) simplifies to:

$$H = \frac{1}{4E}\begin{pmatrix} A\cos^2\theta + m_1^2 & A\cos\theta\sin\theta \\ A\cos\theta\sin\theta & A\sin^2\theta + m_2^2 \end{pmatrix}. \tag{1.54}$$

This Hamiltonian is not diagonal in the original mass basis, and so the eigenstates of the vacuum Hamiltonian are distinct from those of the effective Hamiltonian. The impact of the matter effect is therefore to introduce an effective neutrino mass. The effective masses $\mu_\pm$ are obtained by solving the characteristic equation,

$$\begin{aligned} &\begin{vmatrix} A\cos^2\theta + m_1^2 - \epsilon & A\cos\theta\sin\theta \\ A\cos\theta\sin\theta & A\sin^2\theta + m_2^2 - \epsilon \end{vmatrix} = 0 \\ \implies &\epsilon^2 - (m_1^2 + m_2^2 + A)\epsilon + m_1^2 m_2^2 + A(m_1^2\sin^2\theta + m_2^2\cos^2\theta) = 0 \\ \implies &\mu_\pm = \frac{1}{2}\left[m_1^2 + m_2^2 + A \pm \sqrt{(\Delta m^2\cos 2\theta - A)^2 + (\Delta m^2 \sin 2\theta)^2}\right] \end{aligned} \tag{1.55}$$

$$\begin{aligned} \Delta\mu^2 = \mu_2^2 - \mu_1^2 = \epsilon_+ - \epsilon_- &= \Delta m^2 \sqrt{\left(\cos 2\theta - \frac{A}{\Delta m^2_{21}}\right)^2 + \sin^2 2\theta} \\ &= \gamma\,\Delta m. \end{aligned} \tag{1.56}$$

The $\mathbf{U}_{\text{PMNS}}$ defines how the mass and flavour basis relate. The above illustrates how in the presence of matter, the mass basis is modified, meaning to compensate, the $\mathbf{U}_{\text{PMNS}}$ must also change. That is, matter-effects not only alter the mass terms but also the

mixing parameters. To understand how the mixing angles relate, we can compare the matrix elements of:

$$\left(\frac{\mu_1^2+\mu_2^2}{4E}\right)\mathbb{I}+\frac{\Delta\mu}{4E}\begin{pmatrix}-\cos 2\theta_\mu & \sin 2\theta_\mu\\ \sin 2\theta_\mu & \cos 2\theta_\mu\end{pmatrix}$$
$$=\frac{(m_1^2+m_2^2)}{4E}\mathbb{I}+\frac{\Delta m_{21}^2}{4E}\begin{pmatrix}-\cos 2\theta & \sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{pmatrix}+\begin{pmatrix}V_{cc} & 0\\ 0 & 0\end{pmatrix} \quad (1.57)$$

where θ_μ is the effective mixing angle in matter. Using Eq. 1.56 we can rewrite this as

$$\Delta m^2\begin{pmatrix}-\gamma\cos 2\theta_\mu & \gamma\sin 2\theta_\mu\\ \gamma\sin 2\theta_\mu & \gamma\cos 2\theta_\mu\end{pmatrix}+\begin{pmatrix}-A & 0\\ 0 & A\end{pmatrix}=\Delta m^2\begin{pmatrix}-\cos 2\theta & \sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{pmatrix}. \quad (1.58)$$

Comparing the entries we obtain the relation:

$$\sin 2\theta_\mu=\frac{\sin 2\theta}{\gamma} \quad (1.59)$$

The matter-mixing parameters and mass-splittings are therefore related to their vacuum counterparts via the γ parameter.

1.4.2. Matter resonance

In neutrino matter propagation, there is a criterion that, when satisfied, significantly enhances the impact of matter effects. This can be seen by reading off the from Eq. (1.58):

$$\cos 2\theta_\mu=\frac{\cos 2\theta-\frac{A}{\Delta m_{21}^2}}{\gamma} \quad (1.60)$$

$$\implies \tan 2\theta_\mu=\frac{\frac{\sin 2\theta}{\gamma}}{\frac{\cos 2\theta-\frac{A}{\Delta m_{21}^2}}{\gamma}}=\frac{\Delta m_{21}^2\sin 2\theta}{\Delta m_{21}^2\cos 2\theta-A}=\frac{\tan 2\theta}{1-\frac{A}{\Delta m_{21}^2\cos 2\theta}}.$$

If $A = \Delta m^2_{21} \cos 2\theta$ then the mixing angle θ_μ becomes maximal and transitions become greatly enhanced. This occurs when the electron density is,

$$N_e = \frac{\Delta m^2_{21} \cos 2\theta}{2\sqrt{2} E G_{\mathrm{F}}}. \tag{1.61}$$

A region in a medium that satisfies this condition is referred to as a resonance region. This enhanced oscillation in matter is of particular importance for neutrinos produced in the sun as the continuously varying density of the sun's medium ensures that resonance is always achieved. This phenomenon is known as the Mikheyev, Smirnov and Wolfenstein (MSW) effect [42]. Another context where this effect is important is in the propagation of atmospheric neutrinos through the Earth. In this circumstance, the larger of the two mass-slittings, Δm^2_{31} is the relevant one. Moreover, since the matter potential, A, has opposite sign for neutrinos and antineutrinos, the sign of Δm^2_{31} will depend on which of these channels (ν_μ or $\bar{\nu}_\mu$) the resonance appears in. Neutrino experiments sensitive to atmospheric neutrinos try to exploit this signal to determine the mass ordering.

1.4.3. 3ν oscillation in matter

The relevant channels for long-baseline experiments are the $\nu_\mu \to \nu_e$ and $\bar{\nu}_\mu \to \bar{\nu}_e$ appearance channels, together with the $\nu_\mu \to \nu_\mu$ and $\bar{\nu}_\mu \to \bar{\nu}_\mu$ disappearance channels. In what follows, we present useful approximate analytic expressions for each of these oscillation probabilities in the presence of matter effects.

Appearance channel

While analytical solutions for 3ν propagation in matter exist, they are too complex to reveal any useful insight. The usual approach is to instead perform a series expansion in terms that naturally appear small. A variety of schemes to do this have emerged (e.g., the Madrid expression, the AJLOS expression, the FL expression, the AKT expression, ect). A description of each and their comparative utility can be found in Ref. [43]. We will adopt the Madrid-like approach, which exploits the very different mass-splitting scales and the smallness of the θ_{13} that connects the two scales (i.e. $\sin^2\theta_{13} \approx 0.02$ and the ratio $\alpha \equiv \Delta m^2_{21}/\Delta m^2_{32} \approx 0.03$). In addition, we assume the electron density to

be approximately constant over the propagation distance. Under these assumptions, the full three-flavour oscillation probability can be decomposed into the sum of two effective two-flavour contributions, together with an interference term: 1) a Δm^2_{21}-driven component, controlling the slow oscillations with probability P_{sol}, and 2) a Δm^2_{31}-driven component, controlling the fast oscillation with probability, P_{atm} [24]:

$$P(\nu_\mu \to \nu_e) = P_{\text{atm}} + 2\sqrt{P_{\text{atm}}}\sqrt{P_{\text{sol}}}\cos(\Delta_{32} + \delta) + P_{\text{sol}}, \tag{1.62}$$

with,

$$\sqrt{P_{\text{atm}}} \equiv \sin\theta_{23}\sin 2\theta_{13}\frac{\sin(\Delta_{31}|_{\text{eff}})}{(\Delta_{31}|_{\text{eff}})}\Delta_{31} \tag{1.63}$$

$$\sqrt{P_{\text{sol}}} \equiv \cos\theta_{23}\sin 2\theta_{12}\frac{\sin(\Delta_{21}|_{\text{eff}})}{(\Delta_{21}|_{\text{eff}})}\Delta_{21}. \tag{1.64}$$

where $\Delta_\alpha \equiv \Delta m^2_\alpha L/4E$, and $\Delta|_{\text{eff}}$ is the effective mass-splitting in matter. The labels "solar" and "atmospheric" reflect their historical association with solar and atmospheric neutrino oscillations (in fact for many years these two-neutrino approximations where extensively used before it was possible to observed sub-leading effects).

The middle term of Eq (1.62) is particularly interesting. Its magnitude is proportional to $\sqrt{P_{\text{atm}}}\sqrt{P_{\text{sol}}}$, while $\cos(\Delta_{31} + \delta_{\mathcal{CP}})$ arises from the phase difference between the fast and slow oscillations. From a quantum mechanical perspective, the total phase difference is a combination of the real kinematic driven frequency difference and the complex terms that naturally arise in quantum interference: $(\Delta_{31} + \phi_1) - (\Delta_{21} + \phi_2)$. As a result, the $\mathcal{CP}$-violating Dirac phase of $\mathbf{U}_{\text{PMNS}}$ is observable through the contribution of this interference between the solar and atmospheric oscillations. The shift from the $\delta_{\mathcal{CP}}$ is more complex than a simple phase shift because $\sqrt{P_{\text{atm}}}\sqrt{P_{\text{sol}}}$ also depends on energy (so in a sense $\delta_{\mathcal{CP}}$ changes the "alignment" of the energy dependent cosine oscillations relative to the $\sqrt{P_{\text{atm}}}\sqrt{P_{\text{sol}}}$ amplitude). Moreover, because the argument of the cosine is reciprocal in energy, the shift due to $\delta_{\mathcal{CP}}$ not only moves the locations of the peaks but also slightly alters their shape by changing how rapidly the oscillations occur at a given energy (see Fig. 1.8).

For long-baseline experiments, where typical neutrino energies are below a few GeV, the relevant energy hierarchy is,

$$\underset{250}{\Delta m^2_{31}} \gg \underset{11.9}{2\sqrt{2}G_{\text{F}}N_e E} > \underset{7.4}{\Delta m^2_{21}}. \quad (\times 10^{-5}\text{eV}^2) \tag{1.65}$$

The values underneath are from T2K, which has a flux energy peak of $E_\nu \approx 0.6\,\text{GeV}$ [44]. When the vacuum splitting is much larger than the matter potential, the matter correction is small. In this regime, the matter effect dominates the solar term, but is only a relatively small perturbation to the dominant atmospheric term. Accordingly, the effective mass-splitting terms appearing in Eq. (1.63) can be approximated as

$$\frac{\sin\left(\Delta_{31}|_{\text{eff}}\right)}{\Delta_{31}|_{\text{eff}}} \sim \frac{\sin\left(\Delta_{31}-aL\right)}{\Delta_{31}-aL}, \qquad \frac{\sin\left(\Delta_{21}|_{\text{eff}}\right)}{\Delta_{21}|_{\text{eff}}} \sim \frac{\sin\left(aL\right)}{aL}. \tag{1.66}$$

with $a \equiv G_F N_e/\sqrt{2}$. The dominant atmospheric term is of order α^0, the interference term containing the $\delta_{\mathcal{CP}}$ is of order α, and the solar term is α^2. The interference term contributes approximately 20% of the atmospheric term at T2K values [44] (see Fig.1.9).

$P(\bar{\nu}_\mu \to \bar{\nu}_e)$ is obtained by interchanging $\delta_{\mathcal{CP}} \to -\delta_{\mathcal{CP}}$ and $a \to -|a|$. Because the matter parameter also changes sign, matter effects themselves introduce a $\mathcal{CP}$-asymmetry. It is therefore necessary to decouple intrinsic $\mathcal{CP}$v from those arising from matter effects. There are two main strategies to achieve this: either minimise the neutrino energy to reduce the matter induced $\mathcal{CP}$v to negligible levels, or sample different energies at the same baseline, as done in broad spectrum experiments, to break the degeneracy. It should be noted that in practice this probability is computed numerically by dedicated software, such as **Prob3++** [45], which provides exact values.

Disappearance channel

Propagation in matter alters the disappearance probability very little because it is dominated by the Δ_{31} splitting term. This probability to leading order can be produced from Eq. (1.38) by assuming $\sin^2\Delta_{21}$ to be negligible ($\Delta_{21} \ll \Delta_{31} \approx \Delta_{32}$):

$$\begin{aligned}
P(\nu_\mu \to \nu_\mu) &= 1 - 4\sum_{i>j}\text{Re}\left[|U_{\mu i}|^2|U_{\mu j}|^2\right]\sin^2\left(\frac{\Delta m^2_{ij}L}{4E}\right) \quad ; \quad (\text{Im}[|U_{\mu i}|^2|U_{\mu j}|^2] = 0) \\
&= 1 - 4\Big[|U_{\mu 1}|^2|U_{\mu 2}|^2\sin^2\Delta_{21} + |U_{\mu 1}|^2|U_{\mu 3}|^2\sin^2\Delta_{31} + |U_{\mu 2}|^2|U_{\mu 3}|^2\sin^2\Delta_{32}\Big] \\
&\approx 1 - 4\Big[\left(|U_{\mu 1}|^2|U_{\mu 3}|^2 + |U_{\mu 2}|^2|U_{\mu 3}|^2\right)\sin^2\Delta_{31}\Big] \\
&= 1 - 4|U_{\mu 3}|^2\left(|U_{\mu 1}|^2 + |U_{\mu 2}|^2\right)\sin^2\Delta_{31}
\end{aligned}$$

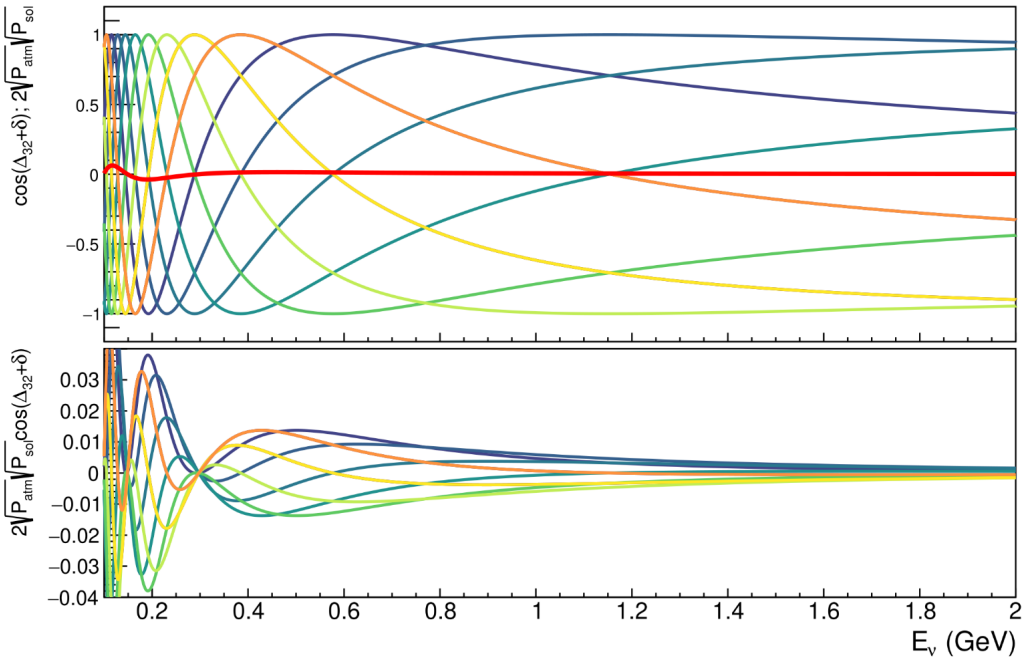
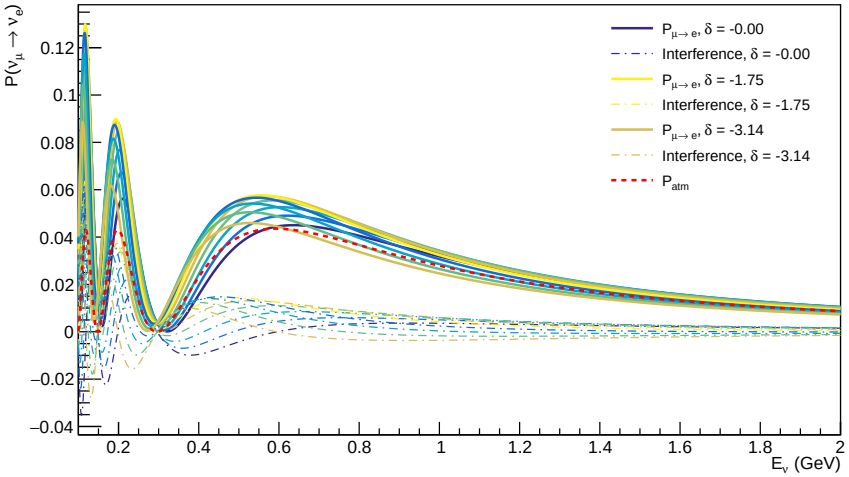
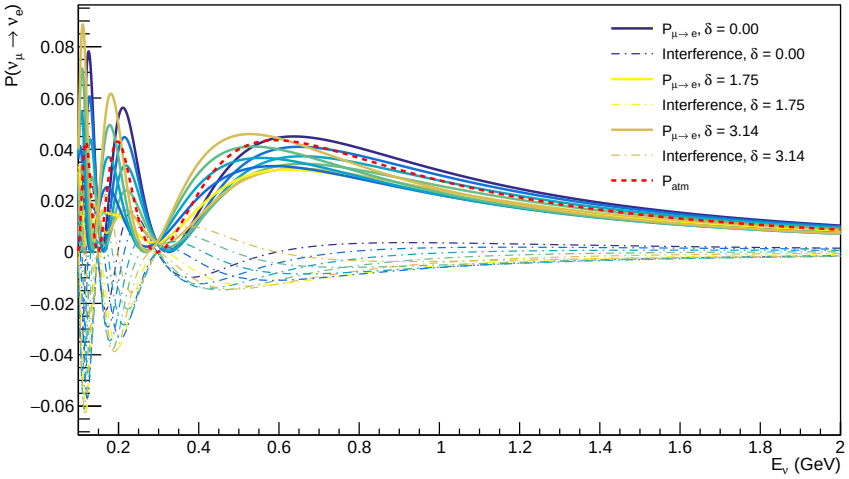


Figure 1.8.: The upper panel shows $2\sqrt{P_{\text{atm}}P_{\text{sol}}}$ (in red) superimposed over $\cos(\Delta_{32}+\delta)$ for a range of $\delta_{\mathcal{CP}}$ values over $[-\pi,\pi]$ (progressing from blue/purple through green and yellow to orange). All other oscillation parameters are fixed to their T2K best-fit values. Since the argument of the cosine is reciprocal in energy, the shift due to $\delta_{\mathcal{CP}}$ not only moves the locations of the peaks but also slightly alters the shape by changing how rapidly the oscillations occur at a given energy. The bottom panel shows the product $\cos(\Delta_{32}+\delta)2\sqrt{P_{\text{atm}}P_{\text{sol}}}$.



(a) $\delta_{\mathcal{CP}}$ plotted in range $[-\pi, 0]$



(b) $\delta_{\mathcal{CP}}$ plotted in range $[0, \pi]$

Figure 1.9.: The total and $\delta_{\mathcal{CP}}$ dependent interference terms are shown for a range of $\delta_{\mathcal{CP}}$ values. For reference, the leading-order $\delta_{\mathcal{CP}}$-independent atmospheric term is plotted in red.

using the unitarity relation $|U_{\mu 1}|^2 + |U_{\mu 2}|^2 = 1 - |U_{\mu 3}|^2$

$$\begin{aligned} P(\nu_\mu \to \nu_\mu) = P(\bar{\nu}_\mu \to \bar{\nu}_\mu) &\approx 1 - 4|U_{\mu 3}|^2(1 - |U_{\mu 3}|^2)\sin^2\Delta_{31} \\ &\approx 1 - 4\cos^2\theta_{13}\sin^2\theta_{23}\left(1 - \cos^2\theta_{13}\sin^2\theta_{23}\right)\sin^2\Delta_{31} \end{aligned} \tag{1.67}$$

1.5. Degeneracies

Values of the oscillation probabilities are not unique due to degenerate combinations of the oscillation parameters. These degeneracies translate to ambiguities in parameter constraints. Formally, this is an eight-fold degeneracy, which manifests through the θ_{13}–$\delta_{\mathcal{CP}}$, MO–$\delta_{\mathcal{CP}}$, and θ_{23}-octant degeneracies [46]. A common way to visualise this is through bi-probability plots, which display the appearance probabilities $P(\nu_\mu \to \nu_e)$ and $P(\bar{\nu}_\mu \to \bar{\nu}_e)$ on orthogonal axes for a fixed set of oscillation parameters. Because $\delta_{\mathcal{CP}}$ is cyclic, varying it over its full range traces out an ellipse. Different true values of other parameters displace and alter the size of this ellipse. A set of parameter values will correspond to a particular location on this bi-probability-space. However, because of this eight-fold degeneracy, it is possible to draw eight overlapping ellipses (called a cloverleaf diagram, as depicted in Fig. 1.10) that share the same location. If this is the only information, then clearly these true and clone points are indistinguishable. A few strategies are used to deal with this: 1) These degenerate points depend on energy, so experiments can use spectral information to distinguish them. 2) Leveraging strong matter effects can help to separate the MO$-\delta_{\mathcal{CP}}$ degenerate solutions. 3) Precision measurements from other oscillation experiments can be incorporated in a synergistic way to greatly reduce the region where ambiguous parameter combinations are possible. For example, reactor experiments are particularly sensitive to θ_{13}, which long-baseline experiments typically include as a strong prior. 4) Experiments may also be fortunate enough where the data favours advantageous regions that are less degenerate. For example, this is the case for T2K where the data favour more extreme vales of $\delta_{\mathcal{CP}}$.

1.6. Outstanding questions

The following are the unresolved questions accessible to neutrino oscillations:

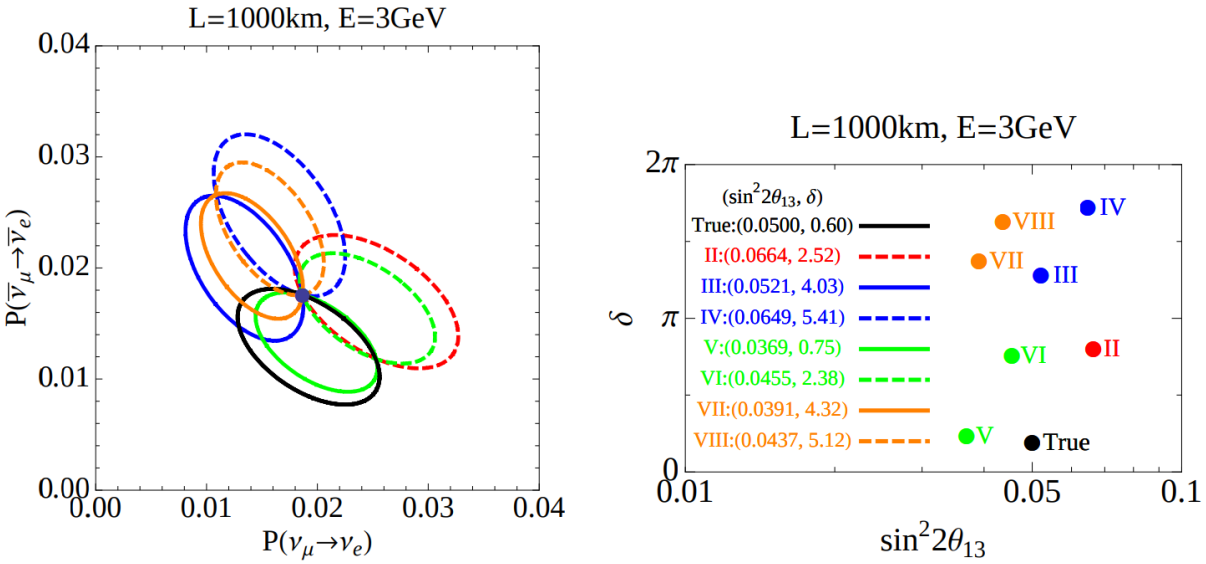
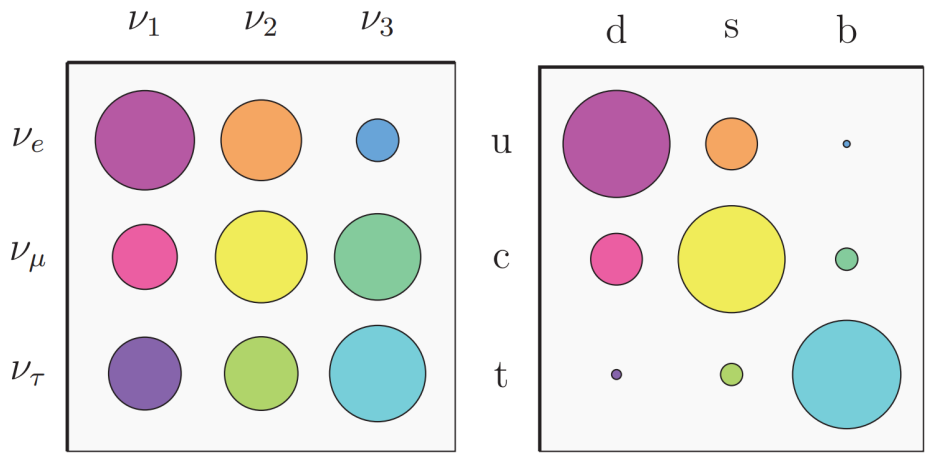


Figure 1.10.: The left plot shows the overlapping ellipses that share the same common solution as the true point. The right plot displays the locations of these true and degenerate solutions in the $\sin^2 2\theta_{13} - \sin\delta$ plane. Reproduced from Ref. [47].

- **Mass Ordering**
  By definition, $m_2 > m_1$, but it is still unknown whether ν_3 is light ($m_3 < m_1 < m_2$) or heavy ($m_1 < m_2 < m_3$). The former is known as "inverted ordering" (IO), while the latter is known as "normal ordering" (NO). The rationale is that if neutrino masses follow $m_3 < m_1 < m_2$, then lepton mixing follows an "inverted" pattern compared to that seen in quarks, where the lightest neutrino is not the primary partner of the lightest charged lepton.

- **Leptonic $\mathcal{CP}$v**
  It is still unknown whether lepton mixing violates $\mathcal{CP}$-symmetry. Since $\mathcal{CP}$-violation is one of the Sakharov conditions for baryogenesis [48], discovering an additional source of $\mathcal{CP}$v would have significant cosmological implications as the amount of $\mathcal{CP}$v from quark mixing is known to be small, ($J_{\text{CKM}} \approx 3 \times 10^{-5}$) and therefore insufficient to explain the observed baryon asymmetry of the Universe. Comparatively, the amount of $\mathcal{CP}$v from lepton-mixing could be significant $J_{\text{PMNS}}(MAX) \sim \mathcal{O}(10^3) \times J_{\text{CKM}}$.

- **Octant of θ_{23}**
  As the dominant constraint on θ_{23} comes from $\nu_\mu \to \nu_\mu$ disappearance, which is



(a) PMNS-matrix **(b)** CKM-matrix

Figure 1.11.: A comparison of the size of PMNS and CKM mixing matrix elements. The size of these circles represents the absolute size of the matrix elements Ref. [17].

quadratic in $\sin^2\theta_{23}$, it is unknown whether

$$\sin^2\theta_{23} > 0.5 \quad [\theta_{23} > \frac{\pi}{4}] \quad \textbf{or} \quad \sin^2\theta_{23} < 0.5 \quad [\theta_{23} < \frac{\pi}{4}]$$

- **Flavour problem**
  $\mathbf{U}_{\text{PMNS}}$ is highly non-diagonal in contrast to its quark counterpart, $\mathbf{V}_{\text{CKM}}$ (see Fig. 1.11), so determining the matrix elements to uncover its structure is interesting in its own right. This may reveal insight into the so called "flavour problem".

- **Test of the three flavour paradigm**
  The unitarity assumption of $\mathbf{U}_{\text{PMNS}}$ should be tested. Any departure from unitarity would indicate the mixing matrix is not complete and hint at the presence of undiscovered physics involving mixing. This will require over-constraining $\mathbf{U}_{\text{PMNS}}$ through precision measurements across multiple oscillation channels.

Chapter 2.

Long-baseline neutrino experiments

2.1. Fundamental principles & design considerations

Practical, technological, and unfortunately, financial constraints necessitate that neutrino experiments specialise in a set of measurements. This is largely driven by the significant difference in mass-ordering regimes, and by the fact that sensitivity to oscillation parameters is greatest near oscillation maxima (OM), where the signal relative to background is strongest. As these occur when the kinetic phase approximately satisfies $(\Delta m^2 L/E) = \pi/2$, oscillation experiments must compromise by optimising their baseline-to-energy ratio, L/E, to enhance sensitivity to specific mixing parameters at the expense of others. Experiments designed to be sensitive to the fast oscillations governed by Δ_{31} are relatively insensitive to contributions from Δ_{21}, as these slower oscillations do not have sufficient time to develop. Conversely, experiments tuned to Δ_{21}-driven oscillations are generally insensitive to the rapid Δ_{31} oscillations, which average out due to finite energy resolution and baseline smearing (a notable recent exception is the JUNO experiment [49], which, owing to its exceptional energy resolution, is sensitive to both oscillation scales). Fig. 2.1 summarises various oscillation trajectories and marks the locations where some notable experiments sit in this landscape.

Assuming $\mathbf{U}_{\text{PMNS}}$ is unitary, it is not necessary to have sensitivity to every oscillation channel to constrain its elements. However, to address important questions such as the presence of leptonic $\mathcal{CP}$-violation, sensitivity to at least one of the following

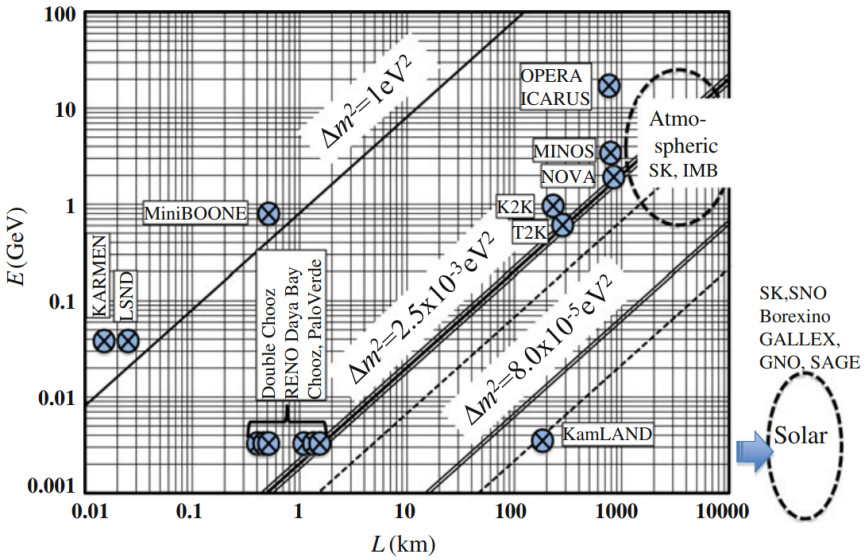


Figure 2.1.: Baseline-to-energy space showing various oscillation experiments marked. Solid lines indicate the first oscillation maximum, while dashed lines represent the second oscillation maximum. Reproduced from Ref. [50].

appearance channels is required:

$$\overline{\nu}_\mu \leftrightarrow \overline{\nu}_e, \quad \overline{\nu}_\mu \leftrightarrow \overline{\nu}_\tau, \quad \overline{\nu}_e \leftrightarrow \overline{\nu}_\tau. \tag{2.1}$$

Due to the experimental difficulties in producing and detecting pure sources of ν_τ, channels involving ν_τ are impractical and thus experimentally disfavoured. This makes $\overline{\nu}_\mu \leftrightarrow \overline{\nu}_e$ (referred to as the "golden channel" by the neutrino community) the most accessible and experimentally promising option for probing leptonic $\mathcal{CP}$v.

To measure oscillations near the OM, this channel still requires neutrino energies comfortably above the muon-production threshold (so of the order of 100s of MeV). At the solar mass-splitting scale, such energies correspond to oscillation lengths that are impractically large propagation distances for terrestrial artificial neutrinos (which is what is required for precision measurements). Therefore, probing this channel in the atmospheric oscillation sector has been identified as the most promising strategy. Even so, this still involves oscillation distances in the hundreds of kilometres, which presents a significant challenge: a detector positioned at such a distance from the source would see only a tiny fraction of the original flux, which would have diminished by a factor of at least $1/L^2$. Unfortunately, this remains true even for accelerator produced neutrinos, as neutrino beams cannot be perfectly collimated. This, in combination with extremely small neutrino cross-sections, means both an extremely intense neutrino source and

a detector with a very large target mass are essential to achieve sufficiently high statistics.

2.1.1. Long-baseline experiments

Long-baseline (LBL) experiments attempt to overcome these challenges using an intense, artificially generated neutrino beam whose properties can be tuned and studied. Several hypothetical methods exist for generating neutrino beams [51], however the prevailing design used by all present experiments employs a meson decay-in-flight concept. In this approach, high energy protons from accelerators are guided onto fixed nuclear targets, producing short-lived secondary charged mesons. These mesons are then focused, or defocused depending on the desired neutrino sign, by magnetic horns before decaying into neutrinos in flight. This results in a beam composed predominantly of either ν_μ or $\bar{\nu}_\mu$. LBL experiments thus measure $\nu_\mu(\bar{\nu}_\mu)$ disappearance and $\nu_e(\bar{\nu}_e)$ appearance in a predominantly $\nu_\mu(\bar{\nu}_\mu)$ beam. The ability to change the neutrino–antineutrino composition of the beam greatly enhances sensitivity to the $\mathcal{CP}$-violation phase.

Immediately downstream of the target region lies an empty space, known as a decay volume, which provides the necessary space for secondary mesons to decay. The majority of the ν_μ and $\bar{\nu}_\mu$ flux originates from the dominant pion decay mode, $\pi^\pm \to \mu^\pm \nu_\mu(\bar{\nu}_\mu)$. This mode dominates because the charged pion has spin 0, so the alternative decay, $\pi^+ \to e^+ \nu_e$ is strongly disfavoured by helicity suppression, with the ratio [52]:

$$\frac{\Gamma(\pi^+ \to e^+ \nu_e)}{\Gamma(\pi^+ \to \mu^+ \nu_\mu)} = \frac{m_e^2}{m_\mu^2} \cdot \left(\frac{m_\pi^2 - m_e^2}{m_\pi^2 - m_\mu^2} \right)^2 \approx 1.2 \times 10^{-4}. \tag{2.2}$$

The length of the decay volume is chosen to permit the primary pions to decay while minimising contamination from electron neutrinos arising from secondary muon decays (i.e. $\mu^\pm \to e^\pm + \nu_e\,(\bar{\nu}_e) + \bar{\nu}_\mu\,(\nu_\mu)$. Despite this, there will inevitably be some electron-neutrino component originating from both muon decays and K_{e3} decays ($K^+ \to \pi^0 e^+ \nu_e$) within the beam [53].

With a fixed baseline, the oscillation phase is solely an energy-dependent effect, removing errors associated with the propagation distance. Additionally, because the source location is precisely known, a second detector can be positioned near the source to

simultaneously sample the initial neutrino spectrum. Data collected at near detectors can be used to directly constrain model uncertainties and therefore reduce systematic errors. Neutrino beams are somewhat unique in this regard, as multiple detectors can be placed in the beam's path without altering the beam properties.

2.2. Systematic uncertainties in LBL experiments

Since neutrinos neither ionise matter nor deposit energy continuously as they traverse it, the beam spectrum cannot be accessed directly. Experiments must instead infer the flux from the observed rates of specific interaction processes[1]. The predicted event rate for a given process, targeting a particular neutrino flavour, as a function of reconstructed neutrino energy, E_R, and kinematic variables, $\vec{x}$, can be generically expressed as:

$$\langle R_{\nu_B}(\vec{x}, E_R)\rangle = \int \Big\{ \Big(\Phi_{\nu_A}(E_T) \times P(E_T, \nu_A \to \nu_B) + \Phi_{\nu_B}(E_T) \times P(E_T, \nu_B \to \nu_B) \Big) \times \sigma_{\nu_B}(E_T, \vec{x}) \times \epsilon_{\text{det}}(\vec{x}) \times n_{\text{Targets}} \times \delta(E_R, E_T) \Big\} \mathrm{d}E_T, \tag{2.3}$$

where, E_T is the true neutrino energy, $\Phi(E_T)$ is the neutrino flux, $\sigma(E_T, \vec{x})$ the neutrino cross-sections on the target nuclei, $\epsilon_{\text{det}}(\vec{x})$ is the detector efficiency, n_{Targets} is the number of target nuclei, $\delta(E_R, E_T)$ describes the correlation between reconstructed and true energy, and $P(E_T, \nu \to \nu)$ is the oscillation probability. The non-oscillation terms collectively constitute the systematic model. Given that the remaining questions addressable by oscillation experiments involve distinguishing small differences from sub-leading terms in the oscillation probability, it is important that these systematics are robustly understood to avoid either incorrectly attributing their effects to oscillation behaviour or rendering the systematic uncertainties so large that no conclusive statements can be drawn. In essence, this is the primary challenge of oscillation experiments; they are inferring oscillation parameters from proxies that depend on many processes beyond these parameters of interest. Here, the sources of uncertainty most relevant to LBL experiments are briefly discussed.

[1] Technically speaking, due to the finite energy-resolution of detectors, experiments actually measure the average interaction rate.

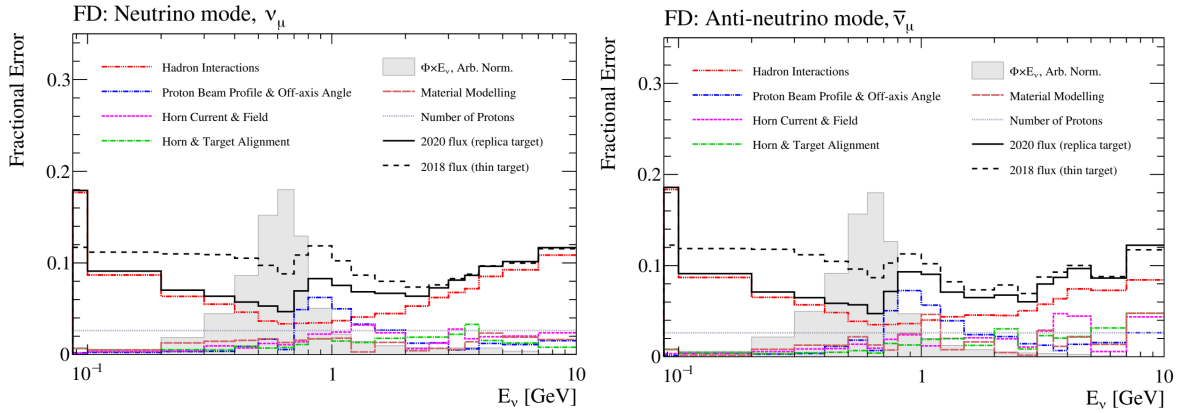
2.2.1. Flux uncertainties

To confidently attribute changes in the flavour and energy spectrum to oscillation phenomena, an accurate understanding of the initial neutrino flux is required. However, several challenges complicate this. Notably, LBL neutrino beams are not composed exclusively of ν_μ or $\bar{\nu}_\mu$. For example, the magnetic horns used to increase the yield of neutrinos available for oscillation and remove wrong-sign mesons from the secondary beam are not perfectly efficient: forward-going secondaries (those that traverse almost parallel to the beam axis) can pass through the necks of the horns where the transverse magnetic field is zero. These particles experience no Lorentz force and are therefore not deflected. If such particles carry the wrong charge, they will contribute to the wrong-sign background. Additionally, as previously mentioned, the beam will also contain impurities from secondary and tertiary neutrino-producing decays that generate an irreducible electron-neutrino contamination.

Another source of error in the flux measurement arises from the limited understanding of hadron production, including both the initial proton–carbon interactions and the subsequent reinteractions of secondaries within the target and surrounding structure. To improve modelling, experiments typically incorporate external hadron production data from specialised experiments. For instance, T2K tunes its flux model to data from NA61/SHINE, which is an experiment on CERN's Super Proton Synchrotron that measures hadron-production yields using a replica of T2K's target [54]. Despite this, hadron-production remains the dominant source of flux uncertainty. Other sources include proton-beam alignment (this can affect the off-axis angle seen at the far detector), target and horn alignment, and uncertainties in the horn magnetic field. The fractional contributions from these sources for T2K are shown in Fig. 2.2.

2.2.2. Interaction and cross-section model uncertainties

As previously noted, the neutrino flux is not directly measured but inferred from the rate of specific interaction processes defined by their final-state products. Neutrino–nucleus interactions, together with the complex nuclear processes that affect final-state observables, must therefore be modelled with care. This subject is vast and complex, representing one of the most active areas in neutrino research. We therefore limit the following to the minimum needed to understand the experiment and thesis topic. These are broadly broken into the following three subsections.



(a) Fractional uncertainty on the right-sign flux in neutrino mode

(b) Fractional uncertainty on the right-sign flux in antineutrino mode

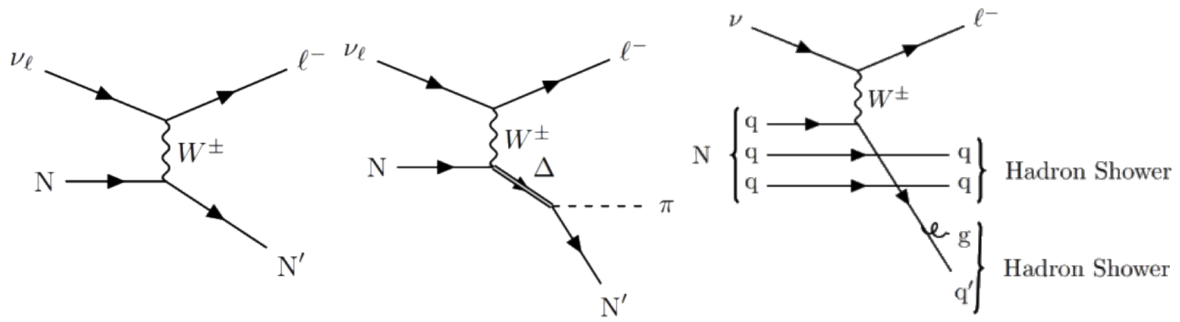
Figure 2.2.: Breakdown of T2K fractional flux systematic uncertainties. The solid black line shows the total flux uncertainty (before 2018, NA61/SHINE used a different thin target; the dashed line corresponds to the uncertainty when this thin target data was used). The grey shaded region shows the shape of the neutrino flux. This figure is reproduced from Ref. [55].

Base interaction modes

Primary neutrino–nucleon interactions can proceed via several mechanisms, depending on the amount of energy the interacting neutrino has available to transfer. Of particular relevance for flavour tagging (the identification of the incoming neutrino flavour via the charged lepton produced) are charged-current (CC) processes, which produce a charged lepton in the final state[2]. In the few-GeV range, the most relevant CC-modes are:

- **Charged-Current Quasi-Elastic (CCQE)**: This is where a neutrino (antineutrino) interacts approximately elastically with a single nucleon. The interaction converts the neutron (proton) into a proton (neutron), and simultaneously ejects it from the nucleus. The final state consists only of the charged lepton of the same flavour as the incoming neutrino and the ejected nucleon (depicted in Fig. 2.3a). This is termed "quasi-elastic" because, while the underlying neutrino-nucleon interaction may indeed be elastic, the emitted hadron may gain or lose energy as it exits the nuclear medium.

[2] Although, of course, a good understanding of NC processes is also important, as they contribute to background.



(a) CCQE **(b)** CCRES **(c)** CCDIS

Figure 2.3.: Charged-current interaction channels: CCQE, CCRES, and CCDIS.

- **Charged-Current Resonant (CCRES)**: At higher neutrino energies, the momentum transfer is sufficient to excite the interacting nucleon into a resonance state. The dominant intermediate state is the Δ-resonance, which decays almost immediately to produce additional pions in the final state (depicted in Fig. 2.3b).
- **Charged-Current Deep Inelastic Scattering (CCDIS)**: At even higher energies, the interaction can resolve individual quarks. This process causes the nucleon to disintegrate and initiates hadronisation. As a result, the final state is significantly more complex, typically consisting of multiple hadrons (depicted in Fig. 2.3c).

As shown in Fig. 2.4, these modes turn on and become dominant in different energy regions, each contributing to the total cross-section with their own associated uncertainties. Notably, these uncertainties are non-negligible across the relevant energy range, reaching the order of 10–40%. Furthermore, owing to the chiral structure of the weak interaction and angular-momentum conservation, neutrino–nucleon interactions are isotropic in the centre-of-mass frame, whereas antineutrino interactions exhibit a pronounced angular dependence. This means that antineutrino cross-sections are generally smaller than neutrino cross-sections, and so need to be studied separately. Moreover, since cross-sections scale with energy, the reconstructed neutrino energy affects not only the flux spectrum, but also the inferred cross-section of the interaction.

Initial-state and correlated nucleon effects

Detectors often use nuclear targets such as carbon and oxygen to maximise neutrino interaction rates and detection efficiency, as these nuclei can be used to construct high-

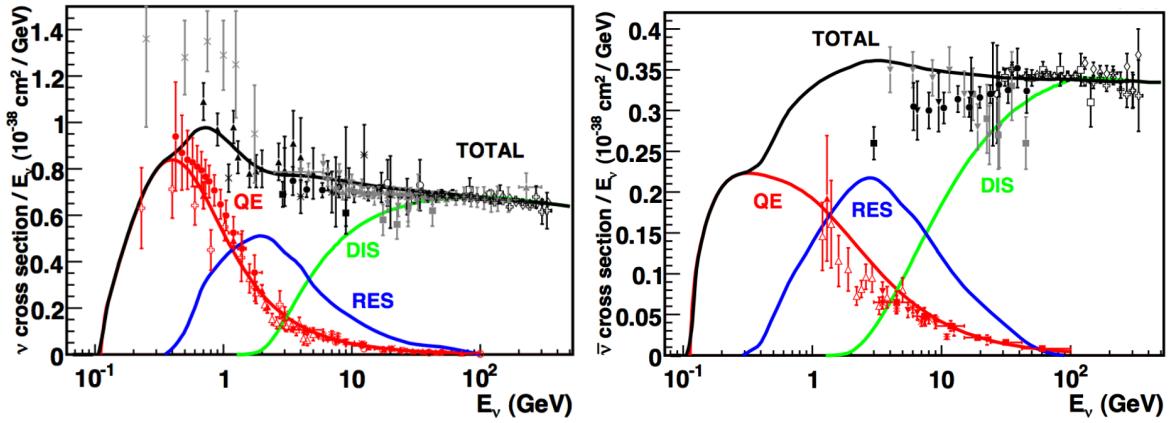


Figure 2.4.: Total inclusive muon-neutrino charged-current cross section as a function of neutrino energy, shown as the sum of quasi-elastic (QE, red), resonance production (RES, blue), and deep inelastic scattering (DIS, green) contributions. The black curve shows the total model prediction, while the data points represent experimental measurements of the inclusive cross section. Left: neutrino–nucleon interactions. Right: antineutrino–nucleon interactions. Reproduced from Ref. [56].

mass targets with favourable detection properties. While this provides higher statistics, it means interactions are not on free nucleons but take place in complex nuclear environments, introducing non-negligible nucleon-nucleon effects. For instance, before any interaction, nucleons have some initial momentum distribution due to Fermi motion. Because the reconstructed energy depends on this unknowable momentum, it introduces a smearing effect on the correlation between true and reconstructed energy. Similarly, the nuclear removal energy (the energy required to eject a bound nucleon) also distorts the reconstructed neutrino kinematics. Additionally, because the removal energy is naturally correlated with the initial nucleon momentum, these are intimately related and complex to model. Unfortunately these processes lack robust, theoretically motivated models and instead rely on a range of empirically driven approaches, which are difficult to apply consistently.

Multi-nucleon effects can also give rise to correlated groups of nucleons via virtual meson exchange. For instance, a neutrino can interact with a bound state of nucleons leading to multi-nucleon ejection (e.g. two-particle–two-hole, 2p-2h). Despite these interactions having significantly different kinematics from CCQE events, they have indistinguishable experimental signatures in water Čerenkov detectors, leading to poor reconstruction of the neutrino energy.

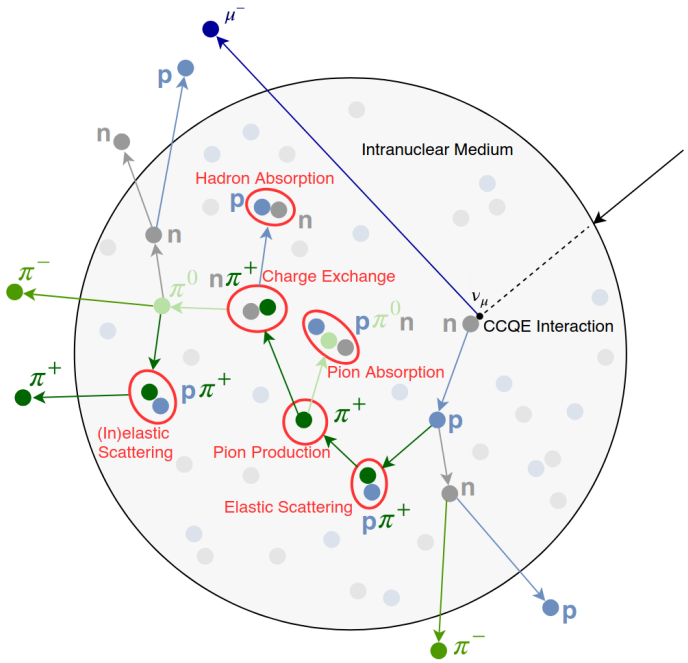


Figure 2.5.: An incoming ν_μ (black dashed line) undergoes a charged-current quasi-elastic (CCQE) interaction with a neutron, producing a muon and a proton. Various possible subsequent FSI are illustrated. Reproduced from Ref. [57].

Final-state interactions

Particles produced in initial neutrino-nucleon interactions need to propagate through a dense medium before exiting the nucleus. This can result in re-interactions that distort kinematics and alter the multiplicity and type of final state observables. These effects are collectively known as Final-State Interactions (FSI) and can manifest in various ways: particles vanishing in absorption, the production of additional particles from induced decays, charge exchanges, and modifications to the propagation directions (and energy) in (in)elastic scattering. Depending on the energy, even more complex combinations are possible, where FSIs produce a cascade of FSIs. These are schematically illustrated in Fig. 2.5. From an experimental standpoint, this is extremely challenging, as intra-nuclear processes and intermediate states are both invisible and unavoidable. Accordingly, experiments typically classify events not by the underlying interaction mode, but by topology, defined exclusively by final-state observables.

2.2.3. Detector systematics

This will be specific to the detector technology used, but broadly speaking, these refer to limitations such as finite energy resolution, errors in energy reconstruction, calibration errors, secondary interactions in the detector, limited solid angle coverage, and a limited understanding of the detector's modeled alignment and geometry. This category also usually includes the detector's efficiency in particle identification and the energy thresholds above which particles are visible (e.g. the inability to observe low-energy protons in water Čerenkov detectors).

2.2.4. modelling

Imperfect systematic understanding is incorporated into the fitted model through parameterised degrees of freedom. These parameters are varied alongside the oscillation parameters, ensuring that the fitted model remains sufficiently flexible to account for all processes influencing the far-detector prediction. Thus, the typical analysis strategy in LBL experiments takes the form,

$$\begin{aligned}\text{Total Information}(\Phi,\sigma,\theta) = \text{Theory}(\Phi,\sigma) \otimes \text{External Data}(\Phi,\sigma) \\ \otimes \text{Near Detector Data}(\Phi,\sigma) \otimes \text{Far Detector Data}(\Phi,\sigma,\theta)\end{aligned}$$

The Theory component comprises the parameterised flux model,Φ, the interaction model, σ, and the detector-response model. The flux model predicts the true neutrino-energy spectrum incident on the detectors, while the interaction and detector-response models together define the mapping from this true spectrum to the reconstructed observable. The details of how this is implemented in T2K's oscillation analysis is described Chapt. 3.

When it comes to the interaction model, however, multiple models exist that are equally well motivated by theory and consistent with existing data. Since discrete choices of interaction model like this directly affect the event rate prediction, but cannot generally be incorporated as tunable nuisance parameters, robustness studies are carried out to assess their impact. This typically involves fitting the nominal model to simulated datasets generated under the assumption of alternative interaction models. The idea is to evaluate how much this alters conclusions about the oscillation parameters. In T2K, these are internally referred to as "Fake Data Studies" (FDS).

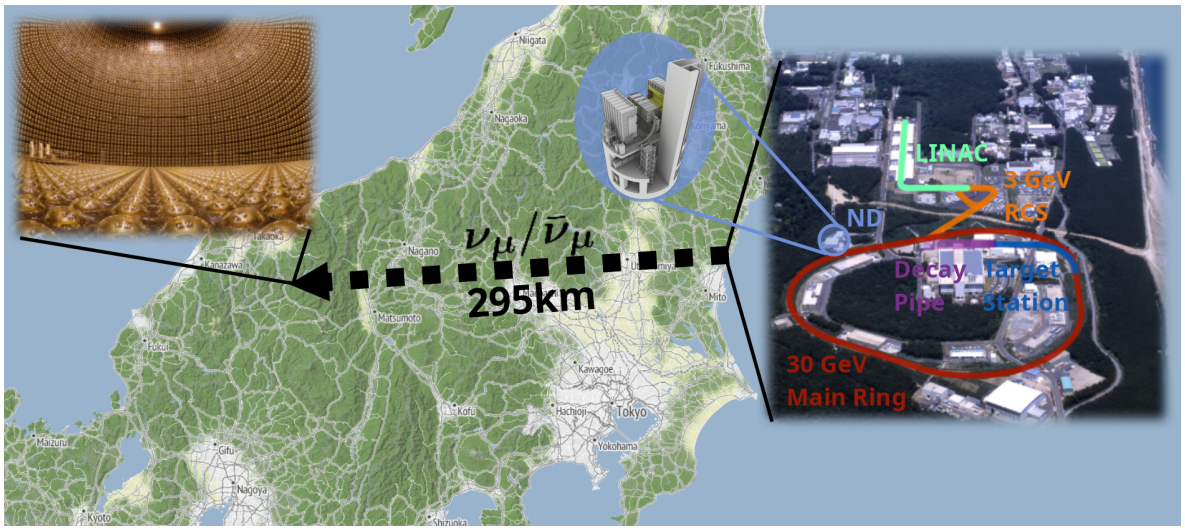


Figure 2.6.: Illustration of the neutrino beamline (black dashed line) across Japan, from the J-PARC facility on the east coast to the far detector, Super-Kamiokande, in the west of the country. Key features of the J-PARC beamline include the LINAC (green), Rapid Cycling Synchrotron (orange), Main Ring (red), beam preparation area, target station (blue), and the decay volume (purple). The location of the near detector complex is also marked in light blue. This figure is compiled from multiple schematics produced by the T2K and Super-Kamiokande collaborations.

2.3. The T2K Experiment

We now describe the specific configuration of T2K and the technologies shaped by these systematic considerations.

Tokai-to-Kamioka (T2K) is a narrow-band, long-baseline experiment in Japan. As depicted in Fig. 2.6, a beam of predominantly muon neutrinos is prepared at the Japan Proton Accelerator Research Complex (J-PARC) in the east-coast village of Tokai. The unoscillated flux is first measured by a suite of near detectors located approximately 280 m downstream of the production point. The beam then propagates 295 km across Japan to the Super-Kamiokande detector (SK) in Hida, Gifu, where oscillation effects are expected to be near maximal.

2.3.1. Accelerator beamline

The sub-image on the right of Fig. 2.6 highlights the key features of the J-PARC accelerator beamline. The accelerator chain starts with a LINAC accelerating H^- ions to 400 MeV. This semi-continuous ion beam then passes through charge-stripping foils,

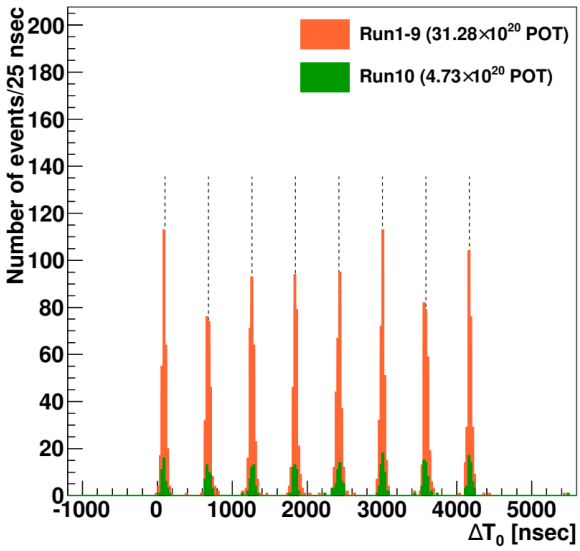


Figure 2.7.: GPS-synchronised $[-1.2, 5.6]$ μs event-timing window at the far detector for T2K Runs 1–9 and Run 10, shown relative to the expected beam-timing structure. Events are required to occur within the broader $[-2, 10]$ μs beam window to be included in the analysis. Reproduced from Ref. [55].

converting it to a proton (H^+) beam as it is injected into the Rapid Cycling Synchrotron (RCS) [58]. Once a sufficient number of protons has accumulated in the RCS, they are concentrated into two distinct bunches and accelerated to 3 GeV. The RCS operates at a cycle rate of 25 Hz, so a beam pulse with two proton bunches can be delivered to the Main Ring (MR) every 40 ms. The MR can accommodate up to eight bunches at a time, so four consecutive RCS pulses are injected into the MR in a single turn. These eight bunches, collectively known as a "spill", are then accelerated to 30 GeV and extracted in a single turn into the neutrino primary beamline using strong kicker magnets. The timing of these spills is transmitted to the near and far detectors using a fiber-optic link and GPS synchronisation respectively. This timing information is used to reject background events that occur outside the spill window. Specifically, only SK events reconstructed within a $[-2, 10]$ μs interval of the spill's arrival are included in the analysis. The eight-bunch structure of a typical spill is clearly visible in Fig. 2.7, which displays the distribution of SK event timings within the spill window.

2.3.2. Neutrino beamline

The T2K neutrino beamline consists of a target station and decay area, as illustrated in Fig. 2.8. The focusing system is configured into three magnetic horns, each operated

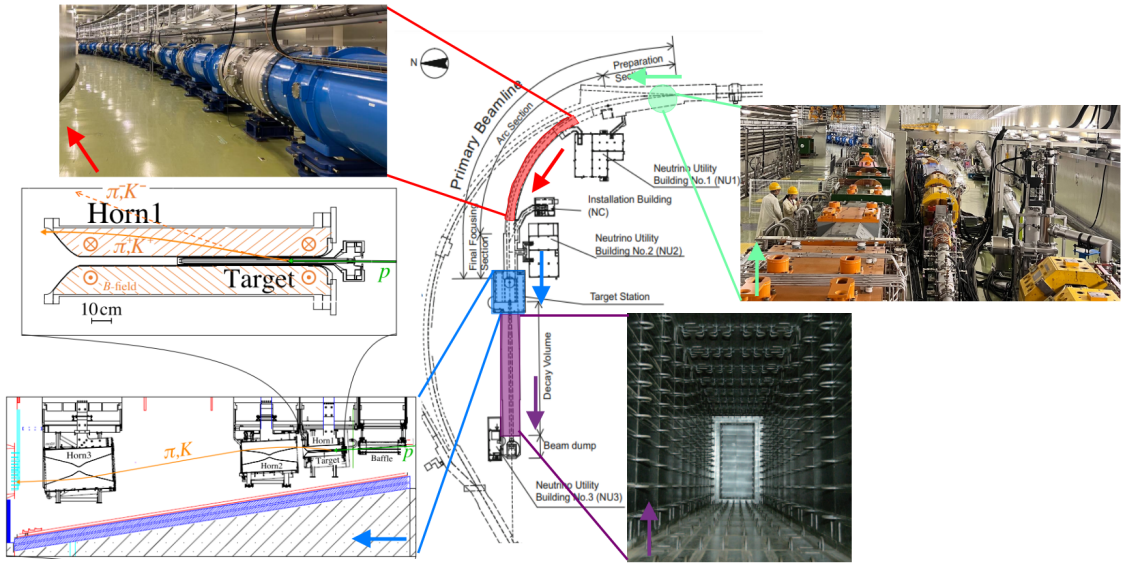


Figure 2.8.: Schematic of the J-PARC neutrino beamline. The arrows indicate the orientation of each image relative to how the corresponding section appears along the beamline. This figure is compiled from multiple schematics and images produced by the T2K collaboration.

with a pulsed current of 250 kA[3]. At this current, the horns generate a 1.7 T field, which increases the flux at the far detector by a factor of approximately 17 [59].

A graphite rod, 91.4 cm in length and 2.6 cm in diameter, is installed within the inner conductor of the first horn, which serves as T2K's target [60]. The beam spill is directed onto this target to produce a secondary beam of mesons (primarily pions and kaons). The first horn collects the mesons produced in the target, while the second and third horns focus or defocus them depending on their charge [61]. T2K refers to the polarity that enhances the neutrino component as Forward Horn Current (FHC), and the polarity that enhances the antineutrino component as Reverse Horn Current (RHC).

The target station is followed by a 96 m long, helium-filled decay volume, where the focused secondary mesons decay to produce the neutrino beam. At the end of this decay volume is the beam dump, comprising 15 iron plates and approximately 75 t of water-cooled graphite. This has two purposes: 1) to efficiently remove muons from the secondary beam before they decay and contribute to the wrong-sign background, and (2) to offer a robust trap for any undecayed particles and to dissipate the heat generated by their interactions.

[3]Upgrades to the beamlines have increased this to 320 kA; however, none of the data used in this thesis were collected with this configuration.

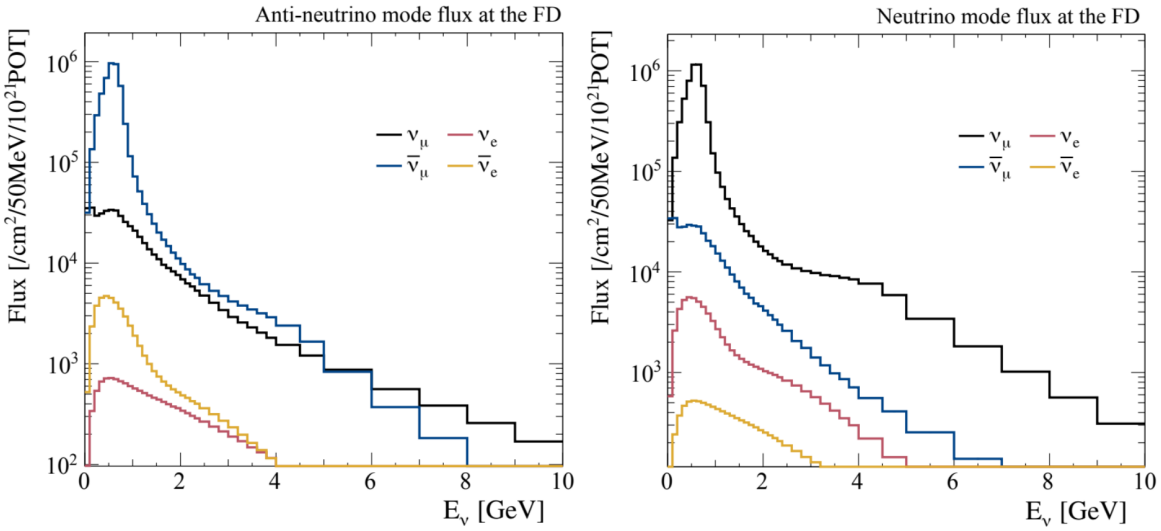
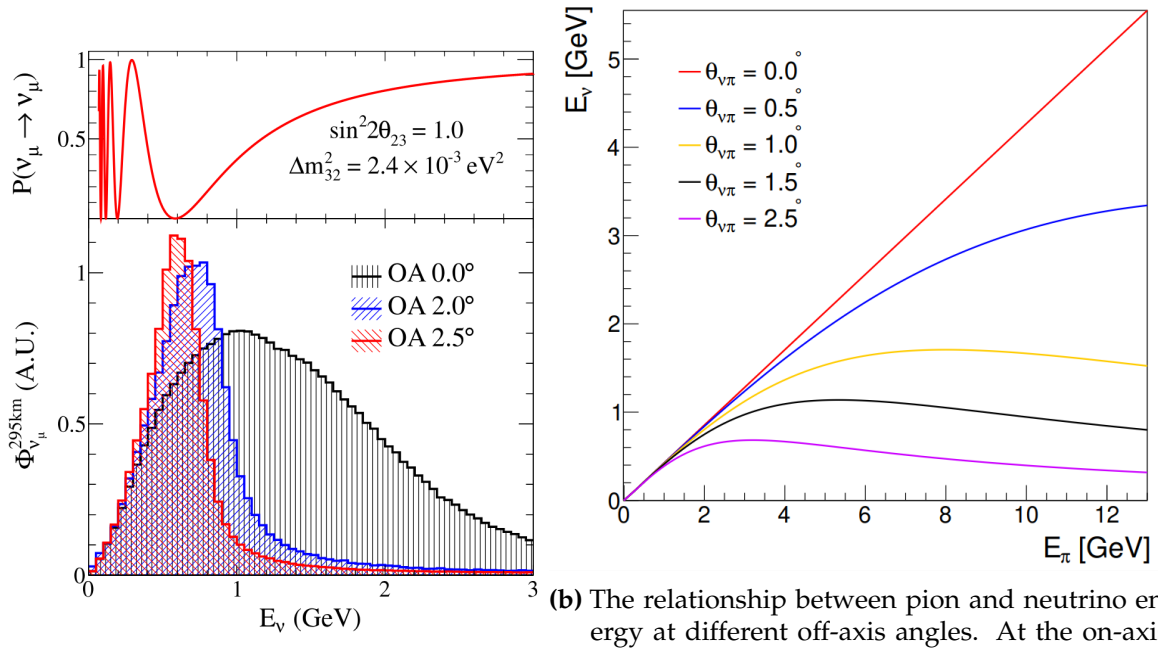


Figure 2.9.: The predicted unoscillated neutrino fluxes at the far detector in RHC (left) and FHC (right), shown on a logarithmic scale. Reproduced from [55].

The typical expected far detector (assuming absence of oscillations) flux composition for both horn modes is shown in Fig. 2.9. Because the beam is initiated via proton–nucleon interactions, there is an intrinsic production asymmetry favouring positively charged mesons over negatively charged ones. This results in a slightly higher ($\approx$ 15%) neutrino flux peak in FHC mode compared to the antineutrino flux peak in RHC mode, and a substantially higher wrong-sign background in RHC [4] [62].

T2K employs an off-axis beam technique to produce a narrow-band beam: just prior to the target, the proton beam is directed downwards at an angle of 3.6°. Since SK is located 1.3° below the horizontal relative to J-PARC, the additional offset means that the neutrino flux arriving at SK is 2.5° off the proton beam's central axis. The motivation for this is two-fold: (1) to peak the neutrino flux near the oscillation maximum of 0.6 GeV, and (2) to reduce the higher energy components of the spectrum that contribute to background. This technique does this by exploiting the kinematics of meson decay: neutrinos from high energy meson decays are Lorentz-boosted in the forward direction. By selecting neutrinos at an off-axis angle, the experiment preferentially samples those emitted transversely in the pion rest frame. These neutrinos receive a smaller Lorentz boost compared to forward emitted ones, resulting in lower energies and a weaker dependence on the pion energy. The relationship between neutrino

[4] Note that because the total number of neutrinos produced cannot be measured directly, T2K uses the number of protons on target (POT) as a proxy to estimate the total neutrino exposure.



(a) Neutrino flux at the T2K far detector for different off-axis angles, overlaid with the approximate disappearance probability.

(b) The relationship between pion and neutrino energy at different off-axis angles. At the on-axis limit, the neutrino energy is linearly related to the pion energy, whereas off-axis the neutrino energy decouples from the pion energy.

Figure 2.10.: Effect of the off-axis technique on the T2K neutrino flux at the far detector. Reproduced from [64].

energy, pion energy, and the off-axis angle is given by [63]:

$$E_\nu \approx \left(1 - \frac{m_\mu^2}{m_\pi^2}\right) \frac{m_\pi^2 E_\pi}{m_\pi^2 + E_\pi^2 \theta^2}. \tag{2.4}$$

In the on-axis limit ($\theta \to 0$), the neutrino energy scales linearly with the pion energy, while at higher angles the dependence is quenched (this is shown in Fig. 2.10) [36].

Being a tertiary beam, the neutrino beam requires continuous alignment monitoring. In addition to the systems that ensure proton beam alignment, there are two complementary monitoring systems: the Muon Monitor (MUMON), which monitors the secondary meson beam, and the Interactive Neutrino GRID (INGRID) detector, which monitors the direction of the neutrino beam directly.

MUMON is a muon detector located approximately 22 m downstream from the end of the decay volume. Muons above 5 GeV can penetrate the beam dump and are measured by this detector. It consists of two detector systems running in parallel: a series of ionisation chambers and an array of silicon PIN photodiodes. The profile of the muon beam detected here is used as a real-time proxy for monitoring the neutrino

beam's direction, stability, and intensity. A detailed description of this detector can be found in Ref. [65].

INGRID is the on-axis neutrino beam monitor, housed within the near-detector complex[5] located 280m downstream of the targets. A schematic of INGRID is shown on the right side of Fig. 2.11. It consists of 14 identical modules arranged in a cross shape configuration, with each module comprising eleven tracking scintillator planes sandwiched between nine iron plates, each 6.5 cm thick. The iron provides a high-mass target for neutrino interactions, while the long lever arms enable the beam to be sampled up to 1° off-axis with a precision of 0.4 mrad [44]. Direct monitoring is important as T2K's off-axis configuration means that a deviation of just 1 mrad would shift the peak neutrino energy by approximately 2% [62].

2.3.3. The ND280

ND280 is the primary near detector used to constrain the neutrino flux and interaction models. Located on the floor above INGRID within the near-detector complex, it comprises a series of subdetectors installed in a supporting structure known as the basket. The basket has rectangular dimensions 6.5 m × 2.6 m × 2.5 m and is enclosed by a magnet that generates a uniform 0.2 T field. The magnet's barrel is instrumented with lead-scintillator electromagnetic calorimeters (ECals) and the Side Muon Range Detector (SMRD). Additional ECals are mounted at the upstream and downstream ends of the basket (Us-ECal and Ds-ECal) to provide full coverage.

The central tracker region of ND280 consists of three Time Projection Chambers (TPCs) and two Fine-Grained Detectors (FGDs). The FGDs serve as the active target mass for neutrino interactions and track particles emerging from the interaction vertex. The TPCs provide 3D-tracking along with precise measurements of momentum and energy loss for charged particles. The most upstream section of ND280 includes components from the recently installed upgrade: two High-Angle TPCs (HA-TPCs) and the so-called Super Fine-Grained Detector (S-FGD). ND280 is positioned at the same off-axis angle as SK to replicate the neutrino spectrum observed at SK as closely as possible.The data used in this thesis, however, predate this upgrade and therefore do not include these new detectors.

[5] Confusingly, this complex shares its name with the main off-axis detector, ND280. To avoid ambiguity, throughout this text "ND280" will always refer specifically to the off-axis detector.

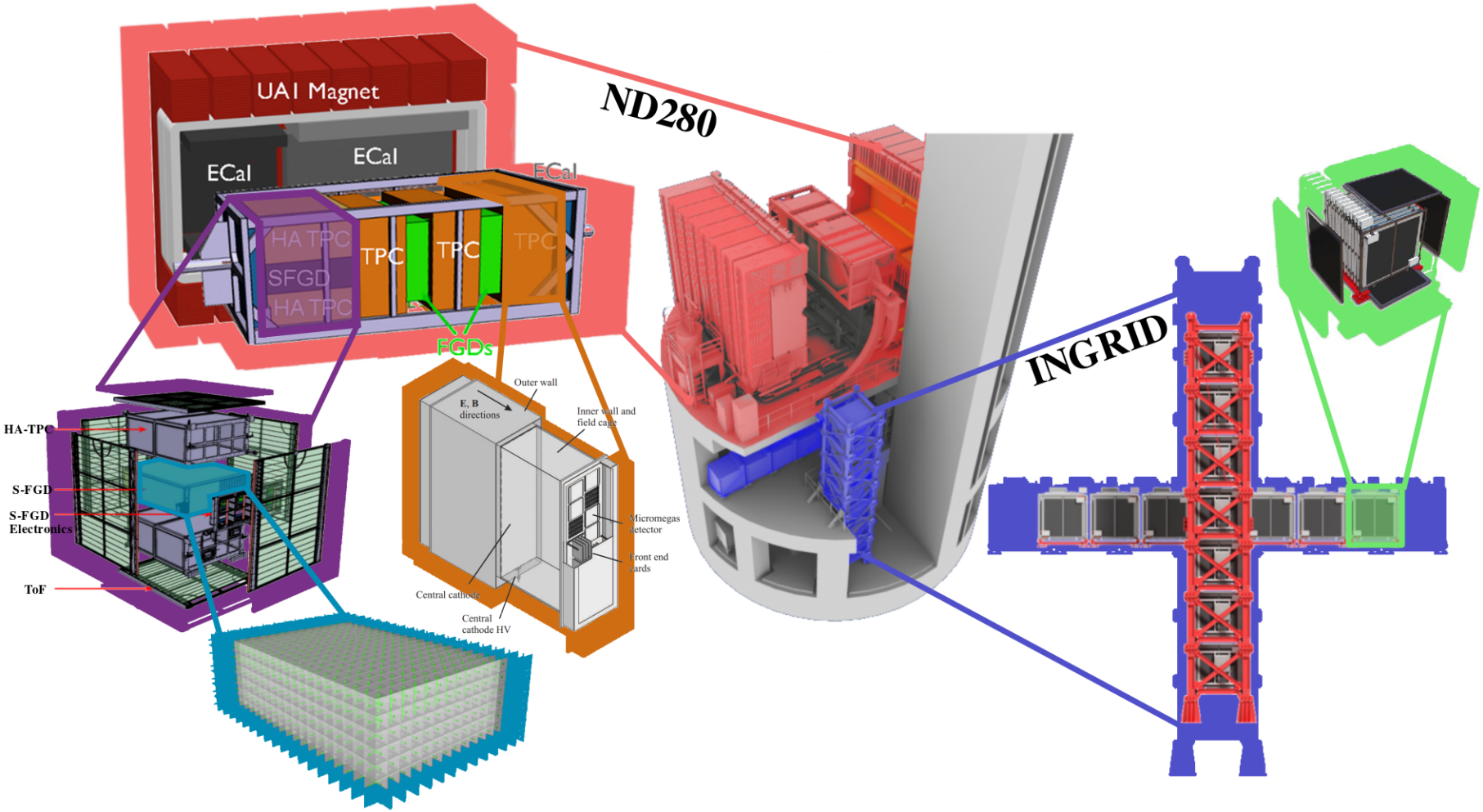


Figure 2.11.: Exploded view of the T2K near detectors. The central rendering shows the Near Detector Complex; on the left is ND280 and its corresponding subdetectors. On the right is INGRID.

Fine-grained detectors

The most upstream FGD is constructed entirely from plastic scintillator, using extruded polystyrene formed into rectangular bars with dimensions 9.61 mm × 9.61 mm × 1864.3 mm. These bars are assembled into layers, and these layers are stacked in orthogonal orientations to form the target volume. Alternating the bar orientation like this enables 3D tracking and vertex reconstruction by combining information from two orthogonal layers. A pair of these layers defines a single scintillator module, and FGD1 contains a total of 15 such modules. A 1.5 mm diameter Wavelength-shifting (WLS) fiber is threaded through the center of each bar and read out by a multi-pixel photon counter (MPPC) connected at the end. When neutrinos interact in the FGD, they produce electrons, muons, and charged hadrons. As these particles pass through the plastic bars, they produce scintillation light, which is guided by the WLS fiber onto the MPPC and converted into an electronic signal. The fine granularity of these bars, in addition to locating the neutrino interaction vertex, enables the tracking and identification of particles that do not enter the TPCs, such as low-momentum protons [66].

The more downstream FGD contains both plastic and inactive water targets. This allows neutrino interactions on water to be studied separately from those on carbon. Its design consists of six water modules sandwiched between seven scintillator modules. Since the far detector uses a water target, reducing the uncertainties related to extrapolating interaction rates across carbon and oxygen nuclear targets is desirable.

Time projection chambers

The TPCs are gaseous tracking detectors capable of resolving fine 3D particle trajectories. Each TPC consists of an inner box (the field cage), which contains an argon based drift gas mixture (specifically a Ar:CF_4:iC_4H_{10} mixture in a 95:3:2 concentration [55]), and a surrounding outer box filled with insulating CO_2 gas, whose walls serve as the grounding potential [67]. The field cage provides a uniform electric drift field oriented parallel to the magnetic field. Charged particles produced by neutrino interactions ionise the gas as they traverse the detector. The ionised electrons, under the influence of the electric field, drift toward the readout plane. The position at which the drift electrons land on the readout plane provides the X-Y coordinates with a resolution of 1 mm, while the drift time determines the Z-coordinate [55].

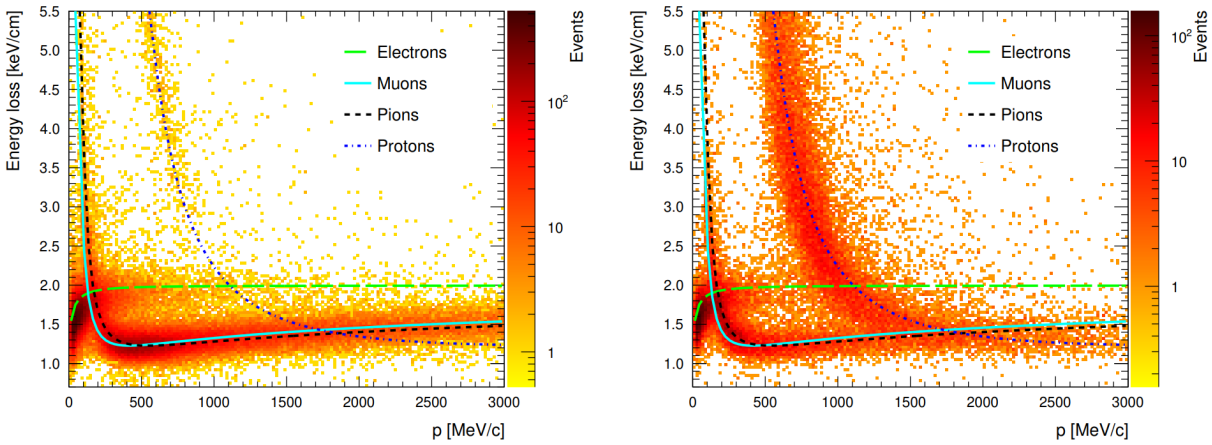


Figure 2.12.: Monte Carlo simulations of energy loss rate ($\mathrm{d}E/\mathrm{d}x$) vs. momentum (p) in the TPC for electrons, muons, pions, and protons exiting FGD1. Left: negatively charged tracks; right: positively charged tracks. Expected Bethe-Bloch curves are overlaid for reference. Reproduced from Ref. [68]

Being a magnetised detector, charged particles follow curved trajectories in the TPCs, enabling both charge determination and momentum measurements. Additionally, since the mean rate of energy loss ($\mathrm{d}E/\mathrm{d}x$) due to ionisation depends on the particle's velocity, rather than its momentum, species of particle with different mass-to-charge ratios can be distinguished. This is exploited to perform particle identification (PID) by comparing the observed energy loss rate to Monte Carlo generated Bethe–Bloch curves (see Fig. 2.12). However, these curves can be very similar for certain particle species. For example, although pions are slightly heavier than muons and therefore travel at slightly lower velocities at the same momentum, the difference falls well below the TPC resolution. As a result, muon and pion tracks are effectively indistinguishable in the TPCs.

Magnet and other subdetectors

ND280 uses the refurbished UA1/NOMAD magnet, which consists of a yoke and current-carrying magnetic coils arranged in two symmetrical halves, or "clamshells". These clamshells are mounted on movable carriages, allowing the magnet to be opened for access to the basket and the subdetectors inside. The yoke is divided into eight C-shaped sections, each made up of 16 layers of 5 cm thick iron plates. Each layer is separated by a 1.7 cm air gap, with the innermost gaps instrumented with plastic scintillator modules that forms the SMRD. The top and bottom have three scintillator

layers in all eight yokes, while on the sides, yokes 1-5 contain three scintillator layers each, yoke 6 contains four, and yokes 7 and 8 contain six layers each.

Because muons are highly penetrating, they are generally not contained within the tracker region, so the SMRD is used for tagging muons exiting at high angles. Additionally, it is used to veto cosmic ray muons that enter from outside [69] [58].

Four magnetic coils, each constructed from aluminum bars and wound into 208 turns, are aligned parallel to the beam direction. The current in these coils generates the magnetic field that is orthogonal to the beam. Inside the yoke barrel, are six ECal modules with multiple functions that complement the tracker. These include distinguishing between muon and pion tracks, as well as converting photons that are undetected by the tracker [70].

ND280 Upgrade

ND280 recently underwent a significant upgrade in which the PØD (a redundant subdetector originally designed to measure the rate of neutral pion production[6]) was replaced by two horizontal HA-TPCs, and the S-FGD. The HA-TPCs are improved versions of the existing TPCs, positioned to extend the ND280's angular acceptance. The S-FGD features a novel active target design composed of two million 1 cm^3 scintillating cubes. This new ultra-highly segmented detector is expected to enable better hadron reconstruction with very low energy thresholds. Surrounding this are six Time-of-Flight planes which will be used to determine whether particles crossing the detector volume are entering or exiting [71].

2.3.4. Near-Detector constraint

Near-detector events are classified into 22 samples. Events are categorised into reconstructed topologies according to the pions multiplicity in the final state. For RHC samples, events are divided into three topologies:

[6] In water Čerenkov detectors, two overlapping photons from neutral pion decay can mimic the signature of a single electron. At the time of T2K's construction, there was concern that this would constitute a significant background. However, a combination of greater than expected discriminating power at SK and the discovery that θ_{13} is relatively large meant that this background was smaller than initially anticipated.

- **CC-0π**: Events in which only the muon is reconstructed. These samples are enhanced in CCQE interactions.
- **CC-1π**: Events in which both a muon and a single pion are reconstructed. These samples are enhanced in CC single pion production.
- **CC-other**: Events in which a muon and multiple pions are reconstructed. These samples are enhanced in deep inelastic scattering.

For FHC, there are five topologies used. The additional samples arise from a finer subdivision of the CC-0π category. The FHC near-detector samples are:

- **CC-0π**, subdivided into:
  - **CC-0π-0p**: Events in which only the muon is reconstructed.
  - **CC-0π-Np**: Events in which a muon and at least one proton is reconstructed.
  - **CC-Photon**: Events in which a muon and at least one photon are reconstructed (implying the existence of an unseen π^0).
- **CC-1π**: Events in which both a muon and a single pion are reconstructed. These samples are enhanced in CC single pion production.
- **CC-other**: Events in which a muon and multiple pions are reconstructed. These samples are enhanced in deep inelastic scattering.

Events are further categorised by the FGD in which the interaction occurred (recall that comparisons between FGD1 and FGD2 are used to understand atomic number dependence). Candidate tracks for all samples must originate within one of the FGD volumes and include a track identified as muon-like by a TPC.

For each event, the highest-momentum track is selected as the muon candidate and is further categorised based on its charge and whether it is right- or wrong-sign (i.e. consistent with the horn polarity). For all FHC samples, the muon must be identified as a μ^- (i.e. the neutrino is consistent with the polarity). To help constrain the wrong-sign background in the antineutrino beam, events in which the muon charge is inconsistent with the horn polarity are also included as additional RHC samples. This sample selection scheme is illustrated in Fig. 2.13.

The event rates for these ND samples are governed by a distinct set of parameters from those of the far detector; however, the two sets are typically highly correlated. These

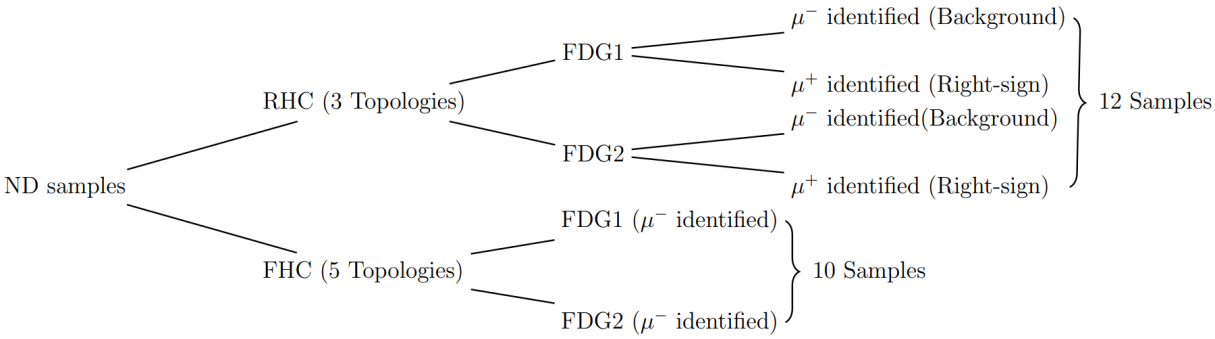


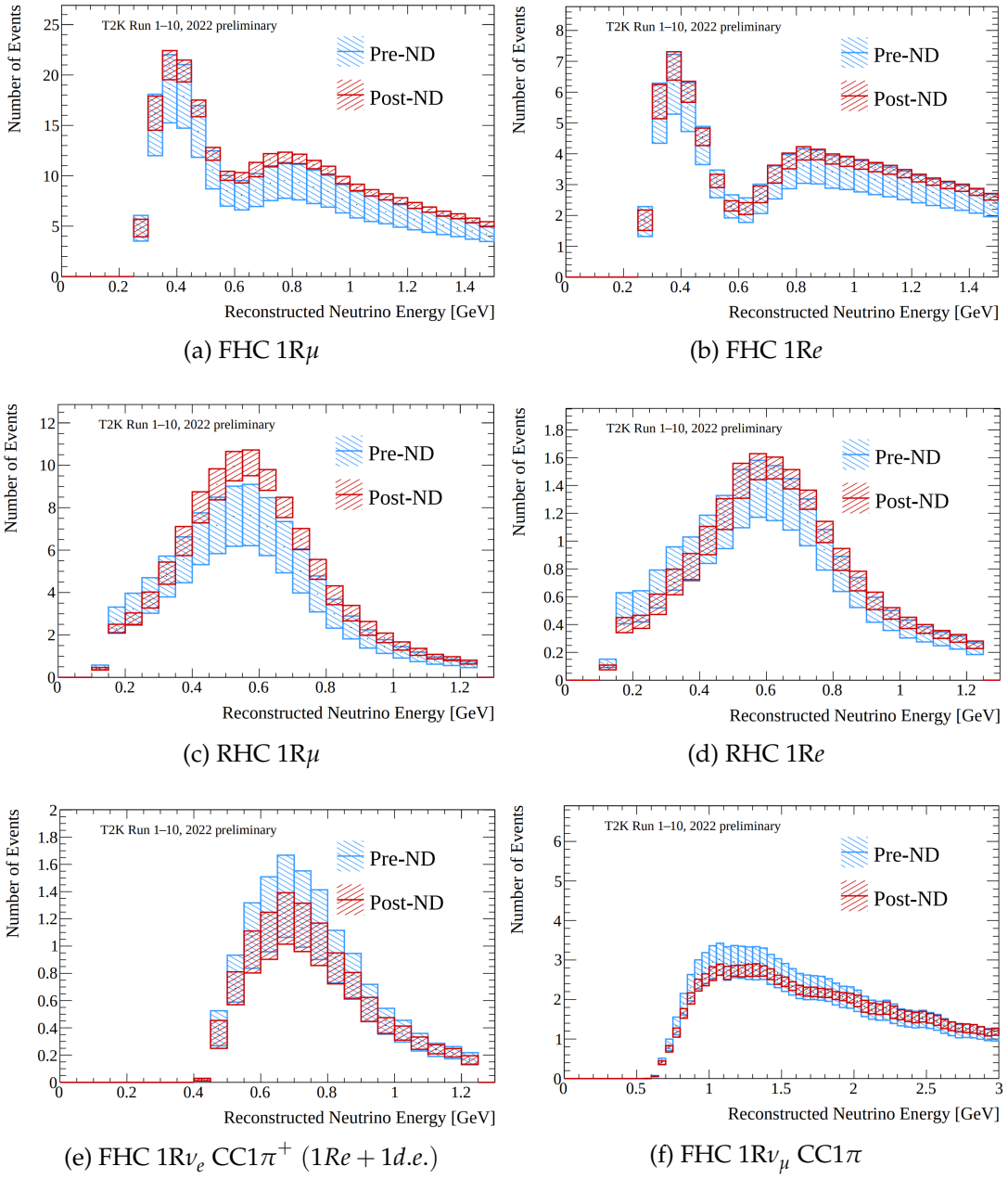
Figure 2.13.: Breakdown of the ND samples. Samples are classified based on topology and beam mode, the FGD in which the interaction occurred, and the charge of the muon.

correlations enable the ND fit to reduce uncertainties in the far detector prediction, which is ultimately what provides sensitivity to the oscillation parameters.

An illustration of this concept is shown in Fig. 2.15 for a pair of correlated ND and FD flux normalisation parameters, with the elliptical contours, in black, corresponding to the joint covariance of these parameters. Because the ND fit directly constrains only the ND flux parameter, the resulting likelihood is flat along the FD parameter direction, reflecting the ND's insensitivity to FD-specific effects. Nevertheless, the FD parameter is indirectly constrained through its correlation with the ND parameter, as evidenced by the reduction from the pre-fit (red) to post-fit (blue) uncertainty line on the FD parameter axis.

For the analyses presented in this thesis, the near–far detector systematic parameterisation, correlation modelling, and near-detector fit are performed by the T2K BANFF sub-group. The resulting post-fit far-detector systematic parameters are then propagated to the far-detector fitting group, P-Theta, in the form of a covariance matrix. The impact of the ND constraint on the Monte Carlo prediction at the far detector is shown in Fig. 2.14.

In practice, since observables at both detectors depend on the product of the neutrino flux and interaction models, the ND fit cannot independently constrain the parameters of either model. Instead, it constrains them simultaneously, introducing anti-correlations between flux and cross-section parameters. This interplay is advantageous, as deficiencies in one model can be partially compensated by adjustments in the other. For example, if the interaction model is poorly constrained in a particular kinematic region, the associated flux parameters may shift in a compensatory manner,



(a) FHC 1Rμ (b) FHC 1Re

(c) RHC 1Rμ (d) RHC 1Re

(e) FHC 1Rν_e CC1π^+ ($1Re + 1de.$) (f) FHC 1Rν_μ CC1π

Figure 2.14.: Total uncertainty in the reconstructed neutrino spectrum for the FD samples, shown with and without applying the near-detector constraint. Reproduced from Ref. [72].

leading to a reduced uncertainty in the predicted FD event-rate. A direct consequence of this behaviour is that the uncertainty on the total event rate is often smaller than the quadrature sum of the uncertainties of the individual parameters.

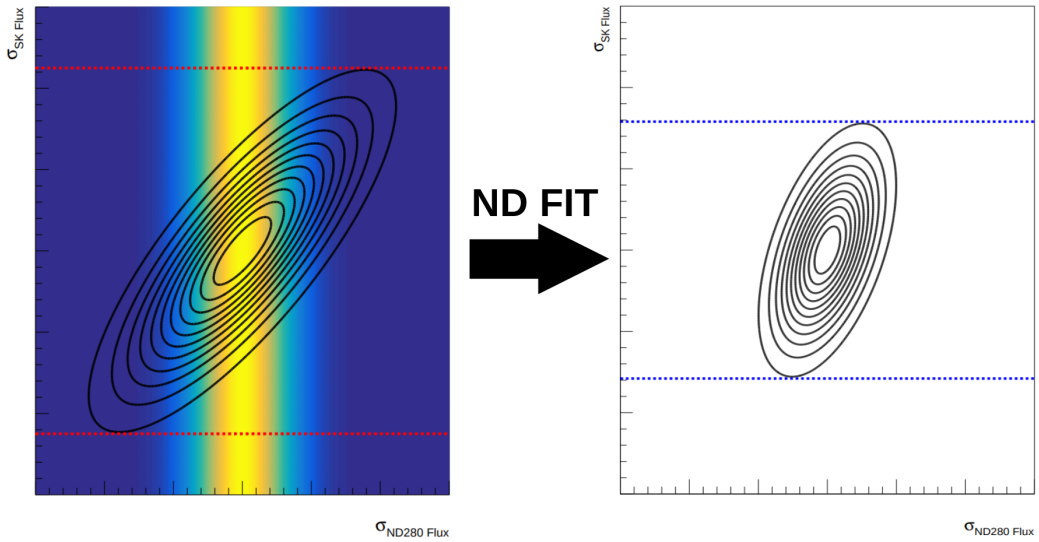


Figure 2.15.: Illustration of how the near-detector (ND) fit constrains correlated far-detector (FD) systematic parameters, shown here for two flux nuisance parameters. Left: The pre-fit uncertainty is represented by black error ellipses, while the coloured band indicates the ND constraint, which acts only along the direction of the ND parameter, as the ND is agnostic to the FD parameter. Right: The resulting post-fit uncertainty after applying the ND constraint. Although the ND fit does not directly constrain FD-specific parameters, correlations between parameters propagate the ND information, leading to a significant reduction in the FD uncertainties, as illustrated by the change between the red (pre-fit) and blue (post-fit) dotted bounds.

2.3.5. The far detector

Photomultiplier tubes

Photomultiplier tubes (PMTs) are highly sensitive light detectors that convert faint light signals into measurable electrical voltages. They operate as follows: incoming photons strike a photocathode, causing the emission of photoelectrons. An applied voltage collects and accelerates these photoelectrons into an electron-multiplier section consisting of a ladder of dynodes. When a photoelectron strikes the first dynode, it deposits enough energy to release multiple secondary electrons, which are then accelerated to the next dynode, which in turn releases more electrons. At each stage, more electrons are released, creating an amplified electron cascade [73].

The accumulated electron cloud strikes the anode at the end of the ladder, producing a measurable voltage spike. This amplification process allows PMTs to detect extremely faint light signals, such as those arising from particle interactions. These devices are designed so that the output voltage is proportional to the number of incident photons. Therefore, in light-producing processes where the number of photons is proportional to the energy deposited, such as Čerenkov radiation, PMTs can be used for energy reconstruction [73].

The Super-Kamiokande detector

Super-Kamiokande serves as T2K's far detector [74]. With a total mass of 50 kt of ultrapure water, Super-Kamiokande is the largest underground water Čerenkov detector built to date. The detector is situated 1 km underground in a cavern excavated within the Mozumi mine beneath Mt. Ikenoyama [75]. The overlying rock provides an overburden equivalent to approximately 2.7 km of water, significantly reducing the flux of cosmic rays that reach the detector.

The main structure of the detector is a stainless steel cylindrical frame measuring 39 m in diameter and 41 m in height. Mounted on the inner surface of this frame are 11,146 inward-facing 50 cm diameter Hamamatsu R3600 PMTs. These PMTs are densely arranged, providing an effective photocoverage of approximately 40%. On the outer surface, 1,885 outward-facing 20 cm Hamamatsu R1408 PMTs are installed. These two volumes are optically separated by light proof black sheets covering both sides of the frame. The interior of the structure is a buffer space that houses PMT cabling and

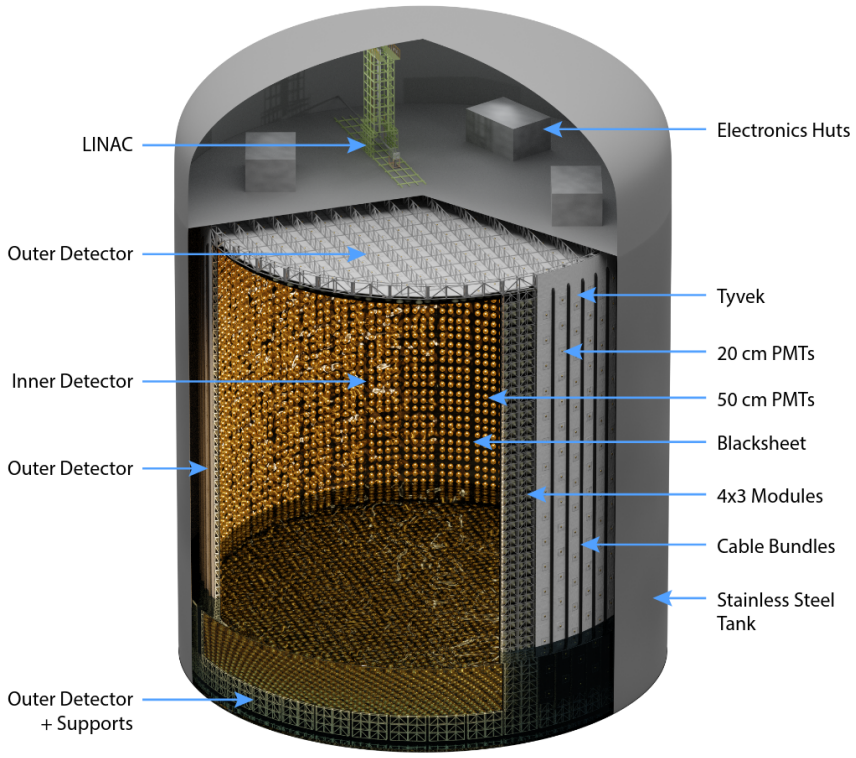
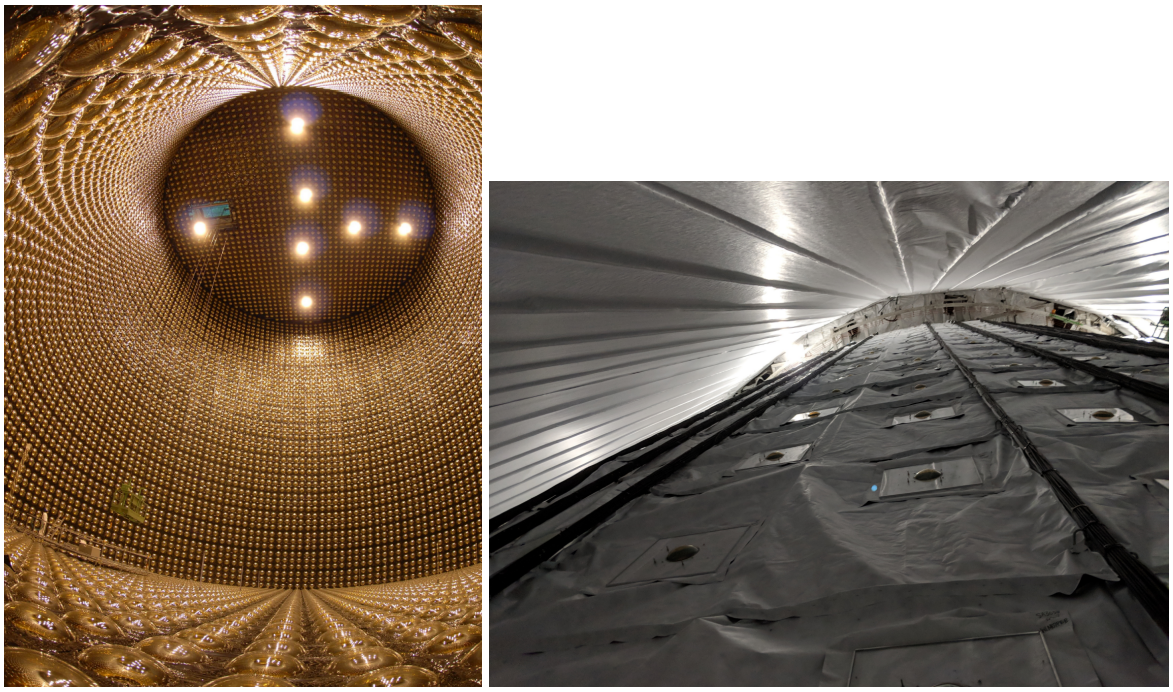


Figure 2.16.: Rendering of the Super-Kamiokande detector. Reproduced from Ref. [76]

electronics. This design effectively defines two detector volumes: the Inner Detector (ID), which encloses a 32 kt fiducial volume and is the principal detector region, and the Outer Detector (OD) located in the 2.2 m gap between the frame and the cavern wall (a schematic of the detector is provided in Fig. 2.16 and images of the ID and OD are given in Figs 2.17a and 2.17b). The primary purpose of the OD is to veto incoming cosmic rays and tag events that are not fully contained within the ID. The OD's interior is lined with reflective Tyvek material, which enables multiple reflections of light in the OD. Each OD PMT is mounted on a 50 × 50 cm acrylic WLS plate, which absorbs UV Čerenkov light and re-emits it as blue–green light, better matched to the spectral sensitivity of the PMTs. Additionally, by coupling these plates directly to the PMTs, the effective light collection area is also increased. The use of WLS plates and Tyvek lining significantly improves the OD's light collection efficiency, which helps compensate for the fact that the OD is more sparsely instrumented, with a low photocoverage of only 7%.



(a) Image of the inside of SK Inner Detector. Image taken and provided by the Super-Kamiokande collaboration.

(b) Image of the Outer Detector. Reproduced from Ref. [76].

Figure 2.17.: Images of the Outer and Inner Detectors of Super-Kamiokande.

Čerenkov radiation and detection principle

When a fast-moving particle interacts with the electromagnetic field of dielectric molecules, such as water, it induces the emission of light via rapid polarisation and depolarisation. Normally, these emissions would simply spread out incoherently, and because each emission is individually very small, it would not produce a signal distinguishable from dark noise in a PMT. However, if the particle's speed exceeds the phase velocity of light in the medium, coherent emission becomes possible. This happens because the particle emits radiation ahead of the wavefronts from previous emissions, and therefore there is guaranteed to be an angle at which all wavefronts overlap and interfere constructively. This is called Čerenkov radiation. While most of this radiation lies in the UV range, there is some visible blue light, giving Čerenkov radiation its characteristic blue glow.

The PID and energy reconstruction in SK are entirely based on detecting Čerenkov radiation. As depicted in Fig. 2.18, a charged particle traversing the water faster than the Čerenkov threshold, $v > c/n$ (where $n = 1.33$ is the refractive index of

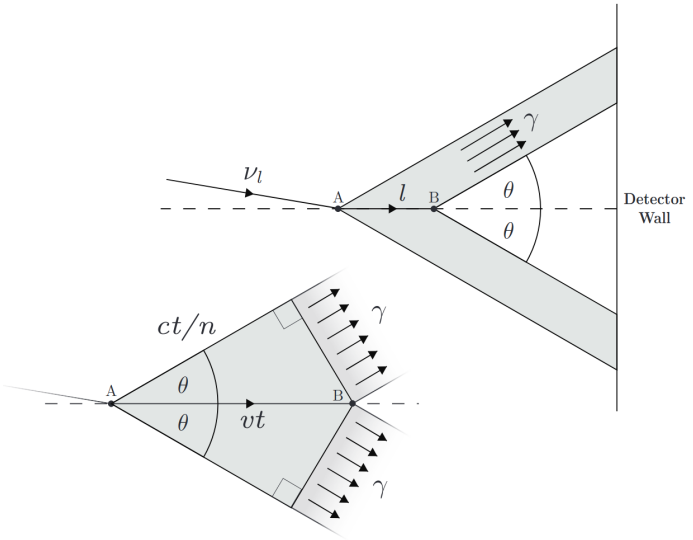


Figure 2.18.: Cartoon illustration of how a Čerenkov ring is formed in the detector wall. Reproduced from Ref. [76].

water), emits a cone of Čerenkov radiation. As the particle loses energy, either through excitation of the dielectric medium or through other processes such as scattering, its speed eventually falls below the Čerenkov threshold, and it stops radiating. This will leave a cone of light which, when incident on the detector wall, produces well-defined ring patterns. These rings are used for vertex location, incident-angle determination, energy reconstruction, and PID. Most important to SK are the lepton-like rings, as these are used to tag the interacting neutrino. Events are classified as electron-like or muon-like based on the pattern of rings they produce. Electrons, due to their lower mass, undergo multiple scattering and also generate electromagnetic showers, resulting in rings with diffuse edges. In contrast, muons produce much sharper, well-defined rings. This distinction gives SK a high-discriminating efficiency between ν_μ and ν_e interactions. Event-by-event identification of neutrino and antineutrino events is generally not possible due to the absence of a magnetic field (although there is technically some discrimination capability through neutron capture on gadolinium, but this was only introduced in 2020). Instead the polarity of the horn current is used as a proxy for neutrino-antineutrino tagging.

Energy reconstruction in water Čerenkov detectors is somewhat unique, as apart from the lepton, they do not use calorimetry but infer the neutrino energy entirely from the kinematics of the assumed interaction. For instance, if a charged lepton produced in a CCQE interaction has momentum p_ℓ (which is essentially just a proxy for energy and so is inferred from the charge collected by the PMTs), and incident angle, θ_ℓ, with

respect to the beam, then the reconstructed neutrino energy is given by,

$$E^{\text{Rec}}_{\text{CCQE}}(p_\ell, \theta_\ell) = \frac{2(M'_{N,i}E_\ell - M^2_\ell + M^2_{N,f} - (M'_{N,i})^2)}{2\left(M'_{N,i} - E_\ell + p_\ell \cos\theta_\ell\right)}, \tag{2.5}$$

where $M_{N,i}$ and $M'_{N,f}$ are the initial and final state nucleon masses, and $M'_{N,i} = M_{N,i} - (E_b \approx 27 \text{ MeV})$. This term obviously depends on the binding energy of a nucleon inside ^{16}O. Of course if the true value of E_b is different from 27 MeV for a given interaction this will bias this reconstruction.

SK also includes samples where more than 1 ring is observed. For these samples the reconstructed neutrino energy, assuming the pion comes from a Δ-resonance, is calculated using:

$$E^{\text{Rec}}_{\text{RES}}(p_\ell, \theta_\ell) = \frac{M^2_{\Delta^{++}} - M^2_p - M^2_x + 2M_p E_\ell}{2\left(M_p - E_\ell + p_\ell \cos\theta_\ell\right)}. \tag{2.6}$$

where $M^2_{\Delta^{++}}$ is the mass of the Δ^{++} resonance, and M_x denotes either the electron or charged pion mass depending on the sample. Since this sample contains only neutrino-induced events (and not antineutrino events), the initial nucleon is necessarily a proton.

2.3.6. Far detector samples

The current T2K oscillation analysis uses six far-detector samples. Events are classified as either coming from a muon or electron neutrino interaction by identifying the primary ring as either muon-like or electron-like by the fiTQun maximum likelihood estimation algorithm described in Ref. [77]. As recoil protons in CCQE interactions are produced below the Čerenkov threshold and are therefore not reconstructed, the CCQE-enriched electron neutrino samples target only a single reconstructed primary Čerenkov ring. For these samples, no decay electron is permitted when the primary ring is electron-like, and fewer than two are permitted when it is muon-like. Further separation is made based on beam operation mode, giving four single-ring (1R) lepton samples: FHC/RHC 1Rμ and FHC/RHC 1Re. For FHC data only, two additional CC$\pi 1^+$ enriched samples are used. The first is 1Re+1d.e., which contains events with a single reconstructed electron-like ring and a delayed decay electron consistent with

a muon from an unseen pion. The sixth sample is a multi-ring sample (ν_μ CC1π MR), in which the intermediate muon from the pion decay chain is reconstructed.

To reduce contributions from low-energy backgrounds, the reconstructed momentum for e-like (μ-like) samples is required to be greater than 100 MeV/c (200 MeV/c) [44]. To further reduce backgrounds from entering particles and improve event-reconstruction quality, additional cuts are made to determine whether interactions are within the bounds of SK's fiducial volume. These include minimum-distance requirements between the reconstructed vertex and the nearest ID wall, as well as between the vertex and section of wall where the emanating track was detected. Details about this criteria can be found in Ref. [44].

Chapter 3.

Statistical techniques

3.1. Probability interpretation

Hypothesised sets of oscillation parameters define testable predictions about the flavour composition and energy spectra of neutrinos after a given propagation distance. Contemporary oscillation experiments constrain these parameters by comparing observed and predicted neutrino spectra. Parameter sets whose predictions are incompatible with observations are rejected. However, because no measurement is infinitely precise and no predictive model free of systematic uncertainty, there is an inherent randomness in this process. As probability is the study of randomness, statistical statements are quantified in the language of probability.

While the Kolmogorov axioms provide a mathematically complete definition of probability [78], making useful experimental statements requires an additional operational interpretation. Generally speaking, two commonly accepted interpretations exist: the Frequentist and the Bayesian.

The Frequentist interpretation defines probability as the limiting frequency of observing data in repeated trials. In this framework the probability is assigned to data, while the hypothesis has no probabilistic properties at all (it is either true or not). In contrast, Bayesian probability is less prescriptive and interprets probability as a measure of the "degree of belief" that a hypothesis is true. In this framework, both the data and the hypothesis can be treated probabilistically. In the Bayesian posterior, once an observation is made, the data is considered fixed, and probability is then assigned to the hypothesis. Because this aligns more naturally with intuitive statements like "the probability that the true parameter is ..." or "the probability that the true parameter

lies within a particular interval is ...", Bayesian results are generally more appealing and easier to interpret. Doing this, though, requires specifying a prior probability for the hypothesis itself, which is where the subjective nature arises, since there are no universal rules for assigning priors. Despite this, there are some practical benefits to this approach as it allows prior knowledge about a model or existing data to be easily incorporated. The usual criticism is that because of the subjective nature of choosing a prior, Bayesian methods are less suitable for scientific contexts [79].

Ultimately, debates over which approach is "correct" are more a matter of philosophy, and in practice, both are widely used; what really matters is that results are produced and interpreted faithfully within the framework being used, and that they can be reasonably expected to converge toward an intuitive understanding of error.

T2K reports its FD oscillation results in both, with two independent working groups employing distinct fitting algorithms. P-Theta is the Frequentist fitter, which performs a grid-search over the parameter space and incorporates the ND-constraint in the form of a covariance matrix and central values derived from the ND-only "BANFF" fit. MaCh3 is the Bayesian fitter that uses a Markov Chain Monte Carlo (MCMC) technique to simultaneously fit near and far-detector data. This arrangement of two independent fitters provides an internal consistency check to ensure that results and physics conclusions are robust against choice of analysis strategy.

The results presented in this thesis are produced using the P-Theta fitter. Consequently, the techniques discussed here will primarily focus on those specific to the P-Theta framework.

3.2. The Likelihood

The most fundamental Frequentist question that can be asked is: What is the probability of observing an experimental outcome X under a hypothesis θ, $P(X|\theta)$? Once an observation is made, this statement can be reversed and expressed as a function of the model parameter, $\ell(\theta) = P(x = X|\theta)$, known as the likelihood function. Importantly, because the data is realised in this function, it is not a random variable, and therefore, the likelihood is not a probability distribution over θ. Rather, it simply labels each member of a parametric model with the probability of observing the given data under that model. The Likelihood Principle states that the likelihood function encapsulates

all information that a data set can provide about the model parameters [80]. Therefore, defining the likelihood function is typically the starting point of Frequentist inference.

Likelihood functions typically involve products of many probabilities, $\ell(\theta) = \prod_i p_i(\theta)$, which can become extremely small and may underflow to zero in floating-point arithmetic. To improve numerical stability and computational efficiency, it is therefore more convenient in practice to work with the statistically equivalent negative log-likelihood, which converts these products into sums, $\mathcal{L}_s(\theta) = -\ln \ell_s(\theta) = -\sum_i \log p_i(\theta)$.

The parameter value that maximises the likelihood, or in other words makes the observed data most probable, is used as an estimator, known as the maximum likelihood estimator (MLE). It is obtained by solving the score equation,

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0 \tag{3.1}$$

The second derivative of the log-likelihood, evaluated at the MLE, defines the local curvature of the log-likelihood surface. Intuitively, it quantifies how informative a measurement is about an unknown parameter: clearly, a surface with a steep peak around the maximum carries more "information" about an unknown parameter and therefore requires fewer statistics to attain the same precision than a flatter surface. The expected curvature at the true value is known as the Fisher information,

$$\mathbb{E}_\theta \left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2} \right] = -\mathcal{I}_\theta \tag{3.2}$$

The realisation of this quantity, obtained by evaluating at the MLE, is known as the observed information.

3.2.1. The T2K likelihood

Oscillation parameters are constrained using data by simultaneously varying both the oscillation and systematic parameters to construct the likelihood over the entire parameter space. Since T2K is insensitive to the solar parameters $\sin^2\theta_{12}$ and Δm^2_{21}, these are fixed in the P-Theta fitter to the world averages[1] reported in the RRP [81]. The systematic model parameters (i.e. the non-oscillation probability parameters that

[1]This thesis uses the PDG 2019 values: $\sin^2\theta_{12} = 0.307, \quad \Delta m^2_{21} = 7.53 \times 10^{-5}\ \text{eV}^2$.

characterise the neutrino flux, interaction cross-sections, and detector response) are treated as nuisance parameters[2], and are incorporated by reweighting the nominal systematic predictions. Far-detector data and MC predictions are binned in two-dimensional kinematic variables depending on the sample type: e-like samples are binned in reconstructed lepton angle and momentum ($p_l - \theta$), while μ-like samples are binned in reconstructed lepton angle and reconstructed neutrino energy ($E^{\text{Rec}} - \theta$). More detail about P-Theta's implementation, including a description of the analysis pipeline, can be found in Chapter 3 of Ref. [82].

The analysis considers contributions from six oscillation channels: $\nu_\mu \to \nu_\mu$, $\bar{\nu}_\mu \to \bar{\nu}_\mu$, $\nu_e \to \nu_e$, $\bar{\nu}_e \to \bar{\nu}_e$, $\nu_\mu \to \nu_e$, and $\bar{\nu}_\mu \to \bar{\nu}_e$. The $\nu_e \to \nu_\mu$ and $\bar{\nu}_e \to \bar{\nu}_\mu$ channels are not included, as their contributions are expected to be negligible due to their intrinsically low oscillation probabilities combined with the small ν_e and $\bar{\nu}_e$ beam contamination.

Using the fact that joint probabilities can be written as products of conditional probabilities, $P(A \cap B) = P(A)\,P(B \mid A)$, the total T2K likelihood is decomposed into two components: an extended likelihood describing both the total number of observed events and their kinematic distributions, and a separate term representing the contribution from systematic uncertainties.

$$\ell(\{N_s^{\text{obs}}, x_s^{\text{obs}}\}_{\forall s}; \theta, f) = \underbrace{\prod_{s \in \text{sample}} \left[\ell_{\text{Poiss}}(N_s^{\text{obs}}; \theta, f) \times \ell_{\text{shape}}(x_s^{\text{obs}}; \theta, f) \right]}_{\text{extended likelihood}} \times \ell_{\text{syst}}(f)\,, \tag{3.3}$$

where s runs through the SK samples, x denotes the reconstructed kinematic variables, θ are the oscillation parameters and f are the systematic nuisance parameters. Breaking this down: $\ell_{\text{Poiss}}(N^{\text{obs}}; \theta, f)$ is the Poisson likelihood to see N^{obs} events given a Poisson mean of $N^{\text{pred}}(\theta, f)$; $\ell_{\text{shape}}(x; \theta, f)$ is the likelihood associated with the distribution of kinematic variables, or in other words the likelihood for each event to have the observed distribution of kinematic variables (since signal and background events have different distributions, this term provides additional discriminating power between signal and background). In practice, this shape term is absorbed into the Poisson term by expressing the expected number of events as a function of the kinematic variables. Given that N^{obs} is fixed, the likelihood for the observed distribution across kinematic

[2] Nuisance parameters are defined as parameters that affect model predictions but are not of primary interest.

bins becomes multinomial. Therefore, the extended likelihood term can be written as,

$$\begin{aligned}\ell_{\text{Poiss}} \times \ell_{\text{shape}} &= \frac{(N^{\text{exp}})^{N^{\text{obs}}} e^{(N^{\text{exp}})}}{N^{\text{obs}}!} N^{\text{obs}}! \prod_{i \in \text{bins}} \frac{\left(\frac{n_i^{\text{exp}}}{N^{\text{exp}}}\right)^{n_i^{\text{obs}}}}{n_i^{\text{obs}}!} \\ &= \prod_{i \in \text{bins}} \frac{(n_i^{\text{exp}})^{n_i^{\text{obs}}} e^{(n_i^{\text{exp}})}}{n_i^{\text{obs}}!}\end{aligned} \tag{3.4}$$

where the total expected and observed event counts are given by $N^{\text{exp}} = \sum n_i^{\text{exp}}$ and $N^{\text{obs}} = \sum n_i^{\text{obs}}$, respectively. Thus, by subsuming the shape term, the likelihood becomes the product of independent Poisson likelihoods for each bin. The term $\ell_{\text{syst}}(f)$ corresponds to the likelihood term for the systematic uncertainties, which encodes our prior knowledge about nuisance parameters and their uncertainties.

As described earlier, P-Theta is part of the analysis chain that sequentially constrains data at the near and far detectors. Because these are propagated as a covariance matrix, systematic parameters, with the exception of a few with flat priors, are assumed to follow a multivariate normal distribution, represented by a mean vector and a covariance matrix:

$$\ell_{\text{syst}} = \exp\left(\frac{1}{2} \sum_{i,j} f_i M_{ij} f_j\right) \tag{3.5}$$

where f_i are systematic parameters and M_{ij} is the (i, j)-th element of the inverse covariance matrix propagated from the BANFF fit.

For technical reasons relating to its potential simultaneous use as a goodness-of-fit statistic, the likelihood function is calculated as the likelihood ratio, where the data term in the denominator is the model-independent maximiser of the likelihood, meaning n_i^{exp} is set equal to n_i^{obs} irrespective of any model restrictions. The likelihood

for sample s is therefore,

$$\begin{aligned}\ell_s(N_s^{\text{obs}}, x_s^{\text{obs}}\,\theta, f) = \prod_{i\in\text{bins}}\left(\frac{\ell(n_{s,i}^{\text{obs}}|n_{s,i}^{\text{exp}})}{\ell(n_{s,i}^{\text{obs}}|n_{s,i}^{\text{obs}})}\right) &= \prod_{i\in\text{bins}}\left[\frac{(n_{s,i}^{\text{exp}})^{n_{s,i}^{\text{obs}}}\,e^{-n_{s,i}^{\text{exp}}}}{n_{s,i}^{\text{obs}}!}\Bigg/\frac{(n_{s,i}^{\text{obs}})^{n_{s,i}^{\text{obs}}}\,e^{-n_{s,i}^{\text{obs}}}}{n_{s,i}^{\text{obs}}!}\right]\\ &= \prod_{i\in\text{bins}}\left[\frac{(n_{s,i}^{\text{exp}})^{n_{s,i}^{\text{obs}}}\,e^{-n_{s,i}^{\text{exp}}}}{(n_{s,i}^{\text{obs}})^{n_{s,i}^{\text{obs}}}\,e^{-n_{s,i}^{\text{obs}}}}\right]\\ &= \prod_{i\in\text{bins}}\left[\left(\frac{n_{s,i}^{\text{exp}}}{n_{s,i}^{\text{obs}}}\right)^{n_{s,i}^{\text{obs}}}\,e^{n_{s,i}^{\text{obs}}-n_{s,i}^{\text{exp}}}\right]\end{aligned} \tag{3.6}$$

Taking the negative log-likelihood,

$$\mathcal{L}_s = -\ln\ell_s = -\sum_{i\in\text{bins}}\ln\left[\left(\frac{n_{s,i}^{\text{exp}}}{n_{s,i}^{\text{obs}}}\right)^{n_{s,i}^{\text{obs}}}\,e^{n_{s,i}^{\text{obs}}-n_{s,i}^{\text{exp}}}\right] = \sum_{i\in\text{bins}}\left[n_{s,i}^{\text{exp}} - n_{s,i}^{\text{obs}} + n_{s,i}^{\text{obs}}\ln\left(\frac{n_{s,i}^{\text{obs}}}{n_{s,i}^{\text{exp}}}\right)\right]. \tag{3.7}$$

where implicitly $n_{s,i}^{\text{exp}} = n_{s,i}^{\text{exp}}(\theta, \mathbf{f})$,

3.2.2. Marginal likelihood

The likelihood can be expressed as a function of the parameters of interest by marginalising over nuisance parameters. In practice, this marginal likelihood is estimated by numerically summing over a set of pre-calculated parameter variations called throws,

$$\begin{aligned}\mathcal{L}_{\text{marg}}(N^{\text{obs}}, \mathbf{x}^{\text{obs}}, \theta) &= \int \mathcal{L}_{\text{Poiss}}(\mathbf{x}, \theta, \mathbf{f}) \times \mathcal{L}_{\text{sys}}(\mathbf{f})\, d\mathbf{f}\\ &\approx \frac{1}{N_{\text{throws}}}\sum_{k\sim\text{flat}}^{N_{\text{throws}}}\mathcal{L}_{\text{Poiss}}(\mathbf{x}, \theta, \mathbf{f}_k) \times \mathcal{L}_{\text{sys}}(\mathbf{f}_k)\end{aligned} \tag{3.8}$$

However, because the dimensionality of the total likelihood is so high, the majority of these throws explore irrelevant regions of phase space. Consequently, obtaining reasonable estimates of the marginal likelihood requires an impractically large number of throws, as most throws contribute minimally to the sum. To address this, importance sampling is used, where instead of drawing samples from a flat distribution, the nuisance parameters are drawn from ℓ_{sys} (conventionally this would be referred to as

the sampling weight). Since $\mathcal{L}_{\text{sys}}$ constrains the systematics, sampling from regions of low $|\mathcal{L}_{\text{sys}}|$ is likely to correspond to regions of low $|\mathcal{L}_{\text{main}}(x_i, \theta_j, f_k) \times \mathcal{L}_{\text{sys}}(f_k)|$. To compensate for the biased sampling (i.e. some values now appear more often than they otherwise would), the result is divided by the sampling weight:

$$\mathcal{L}_{\text{marg}}(N^{\text{obs}}, \mathbf{x}^{\text{obs}}, \theta) \approx \frac{1}{N_{\text{throws}}} \sum_{k \sim |\mathcal{L}_{\text{sys}}|}^{N_{\text{throws}}} \frac{\mathcal{L}_{\text{Poiss}}(\mathbf{x}, \theta, \mathbf{f}_k) \times \mathcal{L}_{\text{sys}}(\mathbf{f}_k)}{\mathcal{L}_{\text{sys}}(\mathbf{f}_k)}$$
$$= \frac{1}{N_{\text{throws}}} \sum_{k \sim |\mathcal{L}_{\text{sys}}|}^{N_{\text{throws}}} \mathcal{L}_{\text{Poiss}}(\mathbf{x}, \theta, \mathbf{f}_k) \quad (3.9)$$

Here, $\mathbf{f}_k$ are now sampled from the distributions where $|\ell_{\text{sys}}|$ is high.

Since it is impractical to display likelihoods in more than two dimensions, the dimensionality is further reduced by selecting one or two oscillation parameters while treating the remaining parameters as nuisance variables to be marginalised over. For the chosen parameters, random throws are drawn from uniform distributions across their allowed ranges, except for $\sin^2 2\theta_{13}$ when the reactor constraint is applied. In that case, $\sin^2 2\theta_{13}$ is varied according to a strong Gaussian prior[3] informed by reactor measurements [81].

3.2.3. Marginalisation artifacts

Sometimes called marginalisation errors, artifacts can arise when the parameters of interest have complex correlations such as flare, skew, or curved distributions. For instance, marginalisation can shift the marginal best-fit point away from the mode of the full joint distribution. This effect is particularly clear when compared with another method of handling nuisance parameters called profiling. In profiling, the likelihood is reduced by fixing the nuisance parameters to their values that maximise the likelihood for each parameter of interest, $\mathcal{L}_{\text{profile}}(\theta) = \max_{\mathbf{f}} \mathcal{L}(\theta, \mathbf{f})$, and as a result of this does not suffer from this kind of bias. This is not to say that profiling is a better method, as it also introduces artifacts (whenever data is reduced like this, artifacts are typically present, except in the simplest cases). However, this marginalisation bias can be particularly problematic in simulated data studies where the true parameter values

[3] This thesis uses the PDG 2019 value: $\sin^2 \theta_{13} = 2.18(\pm 0.07) \times 10^{-2}$.

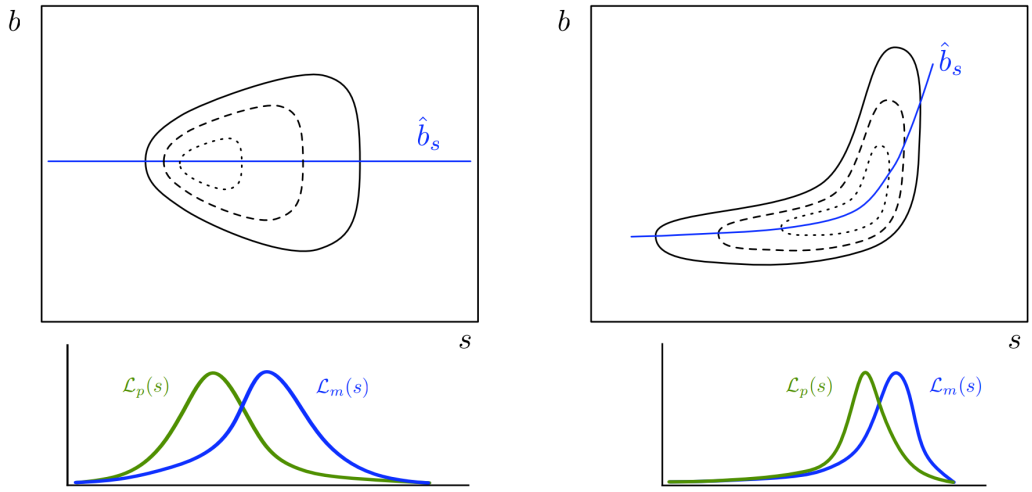


Figure 3.1.: Illustration of different distributions obtained by reducing nuisance parameters using profiling and marginalisation in the presence of non-Gaussian correlations. Reproduced from Ref. [83].

that generated the data are known, but not reproduced in the fit due to these biases. An illustration of this is shown in Fig. 3.1.

3.3. Hypothesis testing and Frequentist intervals

In hypothesis testing, hypotheses are classified into two complementary types, known as the null (H_0) and alternative (H_1). The null hypothesis is tested against the alternative and is rejected in favour of the alternative only if the evidence is sufficiently strong. Quantitatively, this is characterised by the significance level, or the rate of Type I error: a test of size α means that the probability of incorrectly rejecting the null is $100/\alpha$ [4]. However, there is also the probability of incorrectly accepting the null, called the Type II error. Since it is generally not possible to simultaneously minimise both error rates, controlling the Type I error is prioritised. Therefore, deciding which hypothesis refers to which amounts to allocating burden of proof. As the well-known aphorism goes, "extraordinary claims require extraordinary evidence". In line with this principle, the null hypothesis is traditionally regarded as the default position or the currently accepted theory, while the alternative hypothesis is a prediction that contradicts it.

[4] In particle physics, this is often expressed in terms of standard deviations of the standard normal: 1σ = 68.27% percentile, 2σ = 95.45% percentile, 3σ = 99.73% percentile.

In the context of parametric models, there is an ensemble of competing hypotheses (infinitely many in the case of a continuous parameter), each indexed by a parameter $\theta \in \Theta$. The null and alternative therefore correspond to mutually exclusive regions of the parameter space: $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$. To perform hypothesis testing, a test statistic is required. T2K adopts the likelihood-ratio based test statistic,

$$\Delta\chi^2(\theta_0) = \chi^2(\theta_0) - \min_{\theta\in\Theta} \chi^2(\theta) \tag{3.10}$$

where $\chi^2 = -2\mathcal{L}_{\text{marg}}$, and $\mathcal{L}_{\text{marg}}$ is the marginalised log-likelihood of Eq. 3.7. This test statistic essentially measures the "likelihood distance" between the MLE and the hypothesised parameter value θ_0. Because the model-independent maximiser $L(N_{s,i}^{\text{obs}}|N_{s,i}^{\text{obs}})$ is common to both $\chi^2(\theta_0)$ and $\min_{\theta\in\Theta}\chi^2(\theta)$, defining the original T2K likelihood as a ratio does not affect intervals or point estimation inferred from this test statistic.

This statistic is partially motivated by the Neyman–Pearson lemma, which states that the likelihood ratio is the most powerful test of the null for simple hypotheses [5], but it also has good properties for building frequentist confidence intervals.

3.3.1. Confidence intervals

From the perspective of hypothesis testing, intervals are constructed by inverting the test, where each point is treated as a potential true value and tested against the best-fit alternative. The interval includes the set of parameter values for which the null hypothesis (that the given parameter value is indeed the true value) cannot be rejected. Since in the Frequentist framework the true parameter value is considered fixed and unknown, it is the interval boundaries themselves, called confidence intervals, that are probabilistic and fluctuate with data.

Formally, a confidence interval with proper coverage is a random set, $C_\alpha(x) = \{\theta_a, \theta_b\} \subset \Theta$, with the property

$$P(\theta_0 \in C_\alpha(x)) \geq 1 - \alpha. \tag{3.11}$$

That is, the quoted interval contains the true value, θ_0, with probability of at least $1 - \alpha$, where α is the significance of the test. In this sense, any given confidence interval is

[5] For composite hypotheses there is no universally most powerful test.

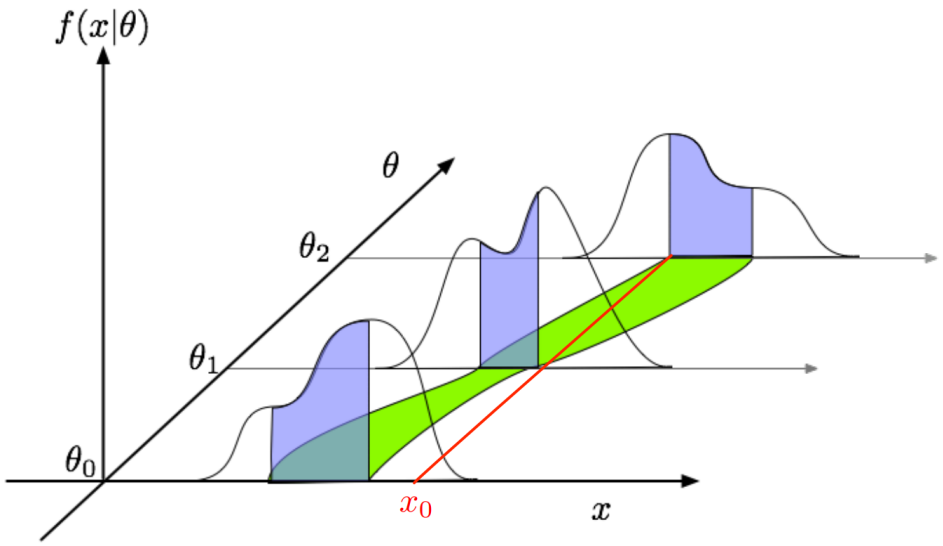


Figure 3.2.: Schematic of the confidence belt of the Neyman construction used to produce confidence intervals with proper coverage. Reproduced from Ref. [84].

just an element of a 'confidence set', which collectively has the property that $1 - \alpha$ of its elements contains θ_0. Constructing such intervals involves calibrating a critical rejection region for the test statistic at every point in the parameter space.

The Neyman construction is the prototypical procedure for doing this: each parameter point is in turn taken as the true value, and a confidence belt is defined from the distribution of the test statistic over the ensemble of hypothetical data sets that could be observed. Once real data is observed, the slice of this belt consistent with the observed test statistic defines the confidence interval (see Fig. 3.2 for a schematic illustration). Intervals built this way possess the correct coverage by construction.

To ensure the uniqueness of this construction, an ordering principle is required. This is essentially the rule used to rank hypothetical outcomes in terms of how extreme they are. Common historical choices include central intervals, and upper and lower limits. However, these classical choices can lead to issues, such as empty intervals when data fluctuates to nonphysical regions unsupported by any model parameter. Another complication arises from flip-flopping phenomena in which this post-hoc choice of ordering principle acts as an additional unseen degree of freedom, potentially leading to undercoverage[6]. To avoid these issues, a unified approach is adopted that uses the $\Delta\chi^2$ test statistic of Eq. 3.10 as the ordering principle.

[6] If unrestricted, there would be a tendency to select the ordering principle based on what would make the most profound statement.

3.4. Asymptotics

Models often exhibit highly complex parameter-dependencies, resulting in test statistics with analytically intractable distributions. In this context, asymptotic relations are particularly powerful, as they permit analysis in terms of simpler limiting distributions, independent of the underlying parameter complexity. A widely used framework for this purpose is the theory of Local Asymptotic Normality (LAN), which we briefly outline here. The following closely follows that presented in Ref. [85].

3.4.1. Local asymptotic normality

Statistical models are said to be locally asymptotically similar if, under local parameter perturbations, their likelihood processes converge asymptotically to the same form.

Consider a Taylor expansion of the log-likelihood in the local parameter $h = \sqrt{n}(\theta - \theta_0)$ around a fixed, known parameter θ_0, and $n \in \mathbb{R}$. That is, we are comparing the value of the likelihood surface at $\mathcal{L}_{\theta_0}(x)$ and a nearby point $\mathcal{L}_{\theta}(x)$, defined by h. For a single observation, x, this yields:

$$\mathcal{L}_{\theta_0+h}(x) - \mathcal{L}_{\theta_0}(x) = h\,\mathcal{L}'_{\theta_0}(x) + \frac{1}{2}h^2\,\mathcal{L}''_{\theta_0}(x) + \mathcal{O}_x\left(h^2\right). \tag{3.12}$$

Summing over n independent and identically distributed observations gives,

$$\ln \prod_{i=1}^{n}\left[\ell_{\theta_0+h/\sqrt{n}}(X_i) - \ell_{\theta_0}(X_i)\right] = \frac{h}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{L}'_{\theta_0}(X_i) + \frac{1}{2}\frac{h^2}{n}\sum_{i=1}^{n}\mathcal{L}''_{\theta_0}(X_i) + \mathrm{Rem}_n. \tag{3.13}$$

Given that $\mathbb{E}[\mathcal{L}'_{\theta_0}(X_i)] = 0$ and $\mathrm{Var}[\mathcal{L}'_{\theta_0}(X_i)] = -\mathbb{E}_\theta[\mathcal{L}''_{\theta_0}(X_i)] = \mathcal{I}_\theta$, the first term, by the central limit theorem, converges in distribution to a normal random variable with mean zero and variance $\mathcal{I}_\theta$ (denoted by $\Delta_{n,\theta}$), while the second converges in probability, by the law of large numbers to $-\frac{1}{2}h^2\mathcal{I}_\theta$. The higher-order remainder terms are bounded in probability and thus asymptotically negligible. Under suitable regularity conditions, the convergence rate is optimal, of order $1/\sqrt{n}$. Consequently, in large samples the log-likelihood surface assumes a quadratic form,

$$\ln \prod_{i=1}^{n}\left[\ell_{\theta+h/\sqrt{n}}(X_i) - \ell_{\theta_0}(X_i)\right] = h\Delta_{n,\theta} - \frac{1}{2}\mathcal{I}_\theta h^2 + o_{P_\theta}(1). \tag{3.14}$$

The key point is that this expansion has the same quadratic form as the log-likelihood of a Gaussian model.

$$\ln\left[N(h,\mathcal{I}_\theta^{-1}) - N(0,\mathcal{I}_\theta^{-1})\right] = h^T\mathcal{I}_\theta X - \frac{1}{2}h^T\mathcal{I}_\theta h. \tag{3.15}$$

Therefore, the theory of LAN establishes that any model whose likelihood admits a local quadratic form shares the same asymptotic statistical properties as a Gaussian model. This is the basis for the term "Local Asymptotic Normal".

These expansions consider the behaviour of the likelihood function in a local neighbourhood around θ_0. The size of the sample statistics, indexed by n, determines the meaning of "locality": for increasing n, larger regions of the parameter space can be considered "local" for the same h. Equivalently, with increasing n, one has to consider increasingly smaller neighborhoods around θ_0 to observe the same curvature of the likelihood function. This notion of increasing locality, will be useful in the following chapters, where an alternative parameterisation of the neutrino oscillation probability is considered.

3.4.2. Wilks' theorem and asymptotic distribution of $\Delta\chi^2$

By similar arguments used to establish LAN, it can be shown that the distribution of the $\Delta\chi^2$ test statistic is asymptotically well approximated by a chi-squared distribution. This result is formalised by Wilks' theorem, which states that, under suitable regularity conditions, the log-likelihood ratio under the null hypothesis converges in distribution to a chi-squared distribution with k degrees of freedom, where k is the difference between the dimensionalities of the null and alternative parameter spaces.[7] This can be seen explicitly by expanding the log-likelihood ratio around the null hypothesis:

$$\Delta\chi^2(\theta_0) = 2\mathcal{L}(\hat{\theta}) - 2\left[\mathcal{L}(\hat{\theta}) + (\theta_0-\hat{\theta})\nabla\mathcal{L}(\hat{\theta}) + \tfrac{1}{2}(\theta_0-\hat{\theta})^T H\{\mathcal{L}(\hat{\theta})\}(\theta_0-\hat{\theta}) + o_{P_\theta}(1)\right]. \tag{3.16}$$

The constant terms $\mathcal{L}(\hat{\theta})$ cancels. In the asymptotic regime ($n\to\infty$), the linear term vanishes since the MLE $\hat{\theta}$ satisfies the score equation. Higher-order terms are asymp-

[7] In this thesis, the null hypothesis will always be a point hypothesis of dimension zero.

totically negligible, leaving

$$\Delta\chi^2 \approx (\theta_0)(\theta_0 - \hat{\theta})^T H\{\mathcal{L}(\hat{\theta})\}(\theta_0 - \hat{\theta}). \tag{3.17}$$

Under the null, the Hessian converges to n times the Fisher information, $n\mathcal{I}(\theta_0)$. Moreover, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and covariance $\mathcal{I}(\theta_0)^{-1}$. Therefore, in the asymptotic limit, the sequence converges to the squared Gaussian distance,

$$\Delta\chi^2(\theta_0) \to \mathbf{N}^T\mathbf{N} = \sum_{i=1}^{k} N_i^2 \sim \chi_k^2 \tag{3.18}$$

where k is the dimension of the normal vector, $\mathbf{N}$.

This result is particularly powerful for constructing confidence intervals, since producing critical values requires prior knowledge of the test statistic's distribution under the null hypothesis. Wilks' theorem allows these critical values, $\Delta\chi_c^2$, to be obtained directly from the corresponding percentiles of the χ_k^2 distribution. In situations where Wilks' theorem is not applicable, critical values must instead be empirically determined by generating an ensemble of pseudo-experiments and extracting thresholds from the resulting distribution.

3.4.3. Regularity condition

LAN, and by extension Wilks' Theorem, relies on considering the behaviour of the log-likelihood under small perturbations of the form $h/\sqrt{n}$. As already alluded to, this principle assumes a set of technical properties, known as regularity conditions, which the parametric model must satisfy. When these conditions are violated, higher-order terms in the log-likelihood expansion may be insufficiently small to justify the necessary quadratic approximation. The most relevant regularity conditions include [86, 87]:

- **Locally open parameter space:** This condition essentially concerns boundary effects within the parameter space. Specifically, there may exist regions where local variations of the parameters do not change the test statistic because the model is already maximised in those directions. Even if θ_0 lies in the interior of the model, being close to a boundary can still pose problems as allowable

perturbations may still be able to reach outside the parameter space, forcing a truncation of the quadratic approximation in those directions.

- **Identifiability:** Model parameters determine model predictions. If this relationship is one-to-one ($\mathcal{L}(\theta_1) \neq \mathcal{L}(\theta_2)$ whenever $\theta_1 \neq \theta_2$), then the model is said to be identifiable. Local identifiability is a necessary condition for LAN, as without it, the log-likelihood surface may be flat or exhibit more complexity than the required quadratic form.
- **Nested:** The null and alternative hypotheses must be nested, i.e. share the same parameter space, with the null corresponding to a restricted case of the alternative. This condition is implicitly assumed in LAN; however, when applying Wilks' theorem to the $\Delta\chi^2$ test statistic, where the alternative is evaluated at the MLE, the hypotheses may not share the same parameter space. For example, this can occur if the sign of a parameter is allowed to vary between the null and alternative hypotheses.
- **Differentiability:** The log-likelihood must be locally continuously differentiable so that the score function and Fisher information are well defined.

3.4.4. Statistical properties of the standard PMNS parametrisation

While the PMNS parametrisation has nice physics properties, at current statistics it possesses many of the aforementioned irregularities, most notably in the $\delta_{\mathcal{CP}}$, the θ_{23} and the mass ordering parameters.

Irregularities due to $\delta_{\mathcal{CP}}$ dependence

The oscillation model is parameterised in terms of trigonometric functions that are bounded between -1 and 1 (e.g. $-1 < \sin(\delta_{\mathcal{CP}}) < 1$). This introduces boundary regions in phase space where the amount of ν_e and $\bar{\nu}_e$ appearance is maximised. The precise location of these boundaries depends on all the oscillation parameters and the neutrino energy. Focusing on the energy dependence, consider a small energy perturbation, ϵ, around the oscillation maximum, such that $\Delta \approx -\frac{\pi}{2} + \epsilon$. The $\mathcal{CP}$v

term of Eq. 1.62 then becomes:

$$A(\epsilon)\cos\left(\delta_{\mathcal{CP}}-\frac{\pi}{2}+\epsilon\right)=A(\epsilon)\cos\left(\delta_{\mathcal{CP}}-\frac{\pi}{2}\right)\cos(\epsilon)+A(\epsilon)\sin\left(\delta_{\mathcal{CP}}-\frac{\pi}{2}\right)\sin(\epsilon). \tag{3.19}$$

Applying a first-order expansion around small ϵ yields,

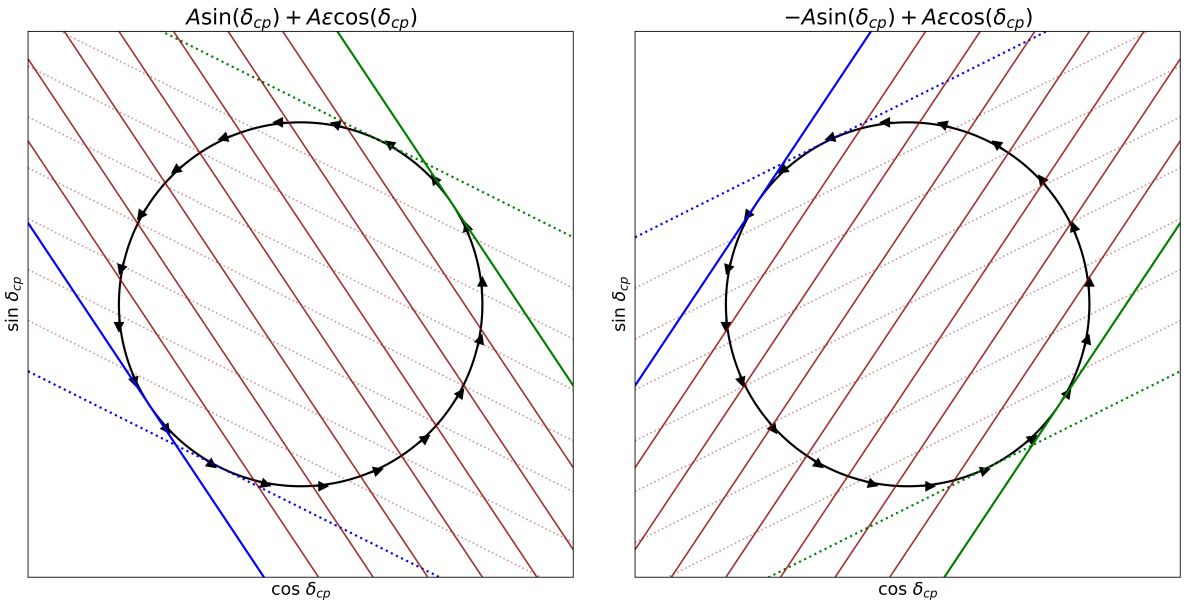
$$A'\cos\left(\delta_{\mathcal{CP}}-\frac{\pi}{2}\right)\cdot 1+A'\sin\left(\delta_{\mathcal{CP}}-\frac{\pi}{2}\right)\cdot\epsilon=A'\sin(\delta_{\mathcal{CP}})+A'\epsilon\cos(\delta_{\mathcal{CP}}), \tag{3.20}$$

where A refers to the non-$\mathcal{CP}$v dependent terms. Clearly for fixed A and ϵ, there exists a value of $\delta_{\mathcal{CP}}$ that maximises the appearance probability.

One useful way to conceptualise this (that will be helpful later) is to plot this expression in $\cos\delta_{\mathcal{CP}}-\sin\delta_{\mathcal{CP}}$ space, as shown in Fig. 3.3. In this representation, the appearance probability corresponds to a straight line, with each combination of A' and ϵ defining the line's gradient. All points along a given line represent the same appearance probability, and therefore the same expected event rate. $\delta_{\mathcal{CP}}$ defines the line's offset (intercept), but does not affect its gradient. This defines a set of parallel lines, each corresponding to a different event rate.

Varying $\delta_{\mathcal{CP}}$ traces out a unit circle in this space. The boundary points correspond to those where a line just tangentially intersect the circle. Changing A' and ϵ alters the gradient, producing a different set of parallel lines and thus a different set of tangents defining new boundaries. In reality, the $\delta_{\mathcal{CP}}$ boundary is more complex, as the likelihood includes both FHC and RHC samples. These samples differ in statistical power and their underlying models exhibit different pure boundary locations due to the sign flip of $\sin\delta_{\mathcal{CP}}$.

Additionally, because these functions of $\delta_{\mathcal{CP}}$ are cyclic, there are inherent degeneracies that violate the identifiability requirement. The $\cos\delta_{\mathcal{CP}}$ and $\sin\delta_{\mathcal{CP}}$ terms are constrained by different aspects of the data: the $\mathcal{CP}$-conserving term, $\cos\delta_{\mathcal{CP}}$, since it appears with ϵ, primarily carries the energy dependence and is therefore constrained by the shape of the energy spectrum. Conversely, the $\sin\delta_{\mathcal{CP}}$ term is directly $\mathcal{CP}$-violating as it changes sign between neutrino and antineutrino samples, making it more strongly constrained by the relative number of events of the FHC and RHC samples. At current statistics, $\mathrm{sgn}(\cos\delta_{\mathcal{CP}})$ is insufficiently constrained to satisfy identifiability, even at modest confidence levels.



(a) neutrino oscillation probability ($+\delta_{\mathcal{CP}}$) **(b)** antineutrino oscillation probability ($-\delta_{\mathcal{CP}}$)

Figure 3.3.: Parallel lines corresponding to different values of $\delta_{\mathcal{CP}}$ for fixed A' and ϵ. The solid lines represent one set of (A', ϵ), while the dashed lines represent another set. The blue and green lines represent the tangent boundary lines. The black arrows indicate the trajectory traced by varying $\delta_{\mathcal{CP}}$.

Irregularities due to θ_{23} dependence

Similar to $\delta_{\mathcal{CP}}$, the mixing angle θ_{23} appears in two distinct terms within the oscillation probability: $\sin^2 2\theta_{23}$ and $\sin^2\theta_{23}$. The difference here though is that while both $\cos\delta_{\mathcal{CP}}$ and $\sin\delta_{\mathcal{CP}}$, are constrained by different aspects of the same appearance channel, the constraints on θ_{23} arise from different channels. Specifically, $\sin^2 2\theta_{23}$ is constrained by the disappearance channel, while $\sin^2\theta_{23}$ is constrained by the appearance channel. Remember that $\sin^2 2\theta_{23} = 4\sin^2\theta_{23}\left[1 - \sin^2\theta_{23}\right]$ is quadratic in $\sin^2\theta_{23}$, and therefore an octant degeneracy arises about maximal mixing, $\sin^2\theta_{23} = 0.5$. Although the disappearance channels have much higher statistics and therefore provide the highest statistical power, the constraint on θ_{23} is typically reported in terms of the subdominant $\sin^2\theta_{23}$ term, as this permits the octant preference to be displayed on a single plot. At current precision, however, the intrinsic degeneracy of this term violates the identifiability condition.

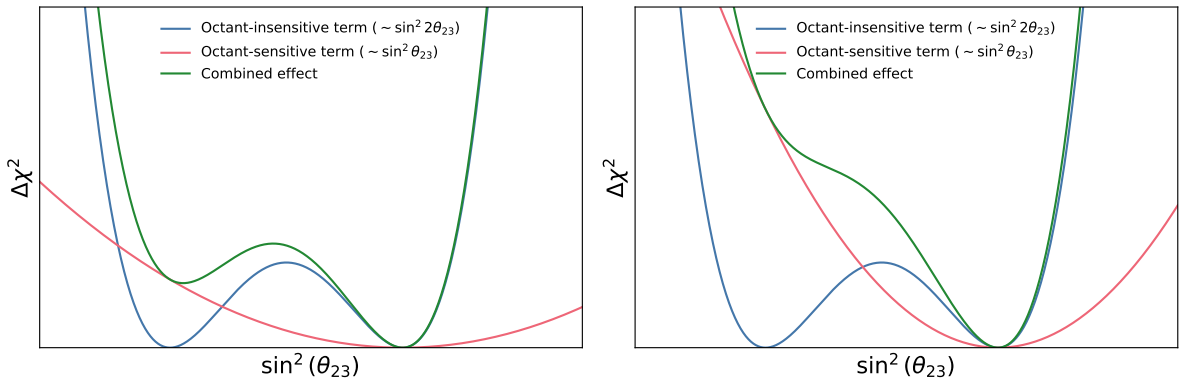
Irregularities due to Mass-ordering dependence

Given that LAN relies on smooth, continuous parameter spaces to evaluate local perturbations, it cannot be applied to discrete hypotheses such as determining the sign of the atmospheric mass splitting. Consequently, asymptotic methods break down entirely when both mass-ordering hypotheses are fitted. Moreover, the two orderings do not share a common parameter space, violating the nesting requirement of Wilks' theorem.

3.4.5. Transience and subtleties of PMNS irregularities

There is a subtlety to these irregularities: with the exception of the mass ordering (MO), their presence, and therefore their impact, diminishes with increasing statistics, vanishing entirely in the limit $n \to \infty$. Take, for example, θ_{23}: with a large enough e-like sample, the likelihood across the parameter space becomes sufficiently distinguishable to approximately satisfy identifiability. This idea is illustrated in Fig. 3.4: at low statistics, the e-like sample lacks the power to distinguish between the octants in a joint fit, but with enough data, the sensitivity improves, and the degeneracy is eventually broken.

The boundaries of these models are also both complex and transient. For conventional boundaries (such as those described in Feldman-Cousins [88]), although the region around the boundary where the likelihood deviates from the Gaussianity shrinks with increasing statistics, the exact boundary point will always be poorly approximated by Gaussian properties. Such simple boundaries can often be modeled with corrective terms, e.g. combinations of χ^2 and δ functions [89]. However, the boundaries that arise in oscillation models are far more complex since the likelihood is joint across multiple samples and, crucially, multiple energy bins. For example, in the case of $\delta_{\mathcal{CP}}$, a true boundary exists only at a single energy value (corresponding to the tangent line in Fig. 3.3 for a given ϵ); neighbouring energy values yield different boundary locations. As statistics increase, the relative influence of this single point boundary diminishes, and the deviation from Gaussianity becomes progressively milder. In fact, in the limit $n \to \infty$, provided the binning becomes sufficiently fine, the influence of this boundary vanishes entirely. Importantly, though, this convergence to Gaussianity is much slower than the standard $1/\sqrt{n}$ rate; it depends not only on the rate at which



(a) Low e-like sample statistics

(b) High e-like sample statistics

Figure 3.4.: Illustration of how increasing the e-like sample statistics breaks the octant degeneracy and leads to model identifiability. This assumes a quadratic log-likelihood in $\sin^2\theta_{23}$.

higher-order asymptotics vanish at regular points, but also on the rate at which the region converges to regularity.

3.4.6. Asimov analysis

Sensitivity is typically defined as the median significance of an experiment under the assumption of a true signal predicted by a given assumed model parameter. In principle, estimating this requires calculating the average significance across an ensemble of simulated experiments. However, it has been shown through Monte Carlo studies that the same result can be extremely well approximated by fitting a single representative data set, known as the "Asimov data set" [90]. In the context of this thesis, this refers to fitting a simulated data set generated with all systematic parameters fixed at their nominal values. Because only a single parameter set is used, statistical fluctuations are suppressed, effectively mimicking the infinite-statistics limit. Moreover, since the parameters used to generate this are known exactly, the Asimov data set provides an efficient diagnostic tool used extensively both in this thesis and in T2K analysis development more generally. We refer to fits produced using this data set as Asimov fits.

3.5. Poisson properties

3.5.1. Poisson statistics

The Fisher information for a Poisson distribution with expected number of events λ is given by,

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta\left[\ddot{\mathcal{L}}_\theta\right] = \frac{1}{\lambda(\theta)}\left(\frac{d\lambda(\theta)}{d\theta}\right)^2. \tag{3.21}$$

Notice that this depends not only on the rate at which the model parameter alters the model prediction, but also on the absolute value of the model prediction itself. This is an important feature, as it means that a region of phase space can be more informative either because the model prediction changes more rapidly with respect to the model parameter, or because the expected event rate and therefore the Poisson variance is larger. In this thesis, it will be important to distinguish between these two effects, with the latter informally referred to as "Poisson statistics".

3.5.2. Notion of increasing *n*

One potential point of confusion concerns the language around 'increasing statistics' or 'increasing n'. In the literature, this is typically framed as repeated sampling from a fixed underlying probability distribution; that is, aggregating results from n independent experiments. In T2K and particle physics experiments more broadly, the interpretation differs slightly, with increasing n corresponding to increasing exposure, so that the underlying distribution itself evolves, and a single observation is made from this updated distribution. However, because of the infinite divisibility of the Poisson distribution, these perspectives are equivalent. To stay consistent with the language used in the literature, and because this is a more intuitive way to consider these concepts, the former interpretation, in which the notion of increasing statistics means more sampling will be used.

A related clarification concerns the term "event rate". While we often speak of observing an event rate, in practice what is fitted is the total number of observed events at a given exposure. Again any inference from these are equivalent because the difference amounts to a trivial scaling factor corresponding to the lifetime, t: consider again the

Poisson information for an event rate is given by $\lambda(\theta) = \rho(\theta)t$,

$$\left(\frac{\partial}{\partial\theta}\left(\rho(\theta)t\right)\right)^2 \frac{1}{\rho(\theta)t} = \left(\frac{\partial\rho(\theta)}{\partial\theta}\right)^2 \frac{t^2}{\rho(\theta)^2 t} = t\left[\left(\frac{\partial\rho(\theta)}{\partial\theta}\right)^2 \frac{1}{\rho(\theta)^2}\right] = tI_\rho = I_\lambda, \quad (3.22)$$

where $\rho(\theta)$ is the event rate.

Chapter 4.

Linearised oscillation models

4.1. Linear models

Linear models are a class of statistical model in which the predicted values depend linearly on the model parameters. These models are highly desirable because they naturally satisfy regularity conditions and are uniform in variance and bias across the parameter space. This uniformity arises from the fact that all regions of the parameter space influence the model output symmetrically. Consequently, when considering the sensitivity of an experiment, it does not matter which region of parameter space the data prefers.

For example, consider a nonlinear model that is quadratic in its parameters, with a Poisson mean, $\lambda(\theta) = a\theta^2 + b$, where, for simplicity, the parameter is one-dimensional. In this case, the Poisson Fisher information (Eq. 3.21) exhibits a non-trivial dependence on the local rate of change of the model prediction.

$$\mathcal{I}(\theta) \propto \left(\frac{d\lambda(\theta)}{d\theta}\right)^2 = 4a^2\theta^2 \tag{4.1}$$

That is, how informative a measurement is depends on the region of parameter space preferred by the data.

When irregularities such as the boundaries of the PMNS parameter space are also included, this data dependence can be even more pronounced and complex: if the observed data corresponds to spectra more extreme than the model can accommodate, the fit will pull against the boundary, and the boundary itself begins to provide statistical information, effectively constraining what values the parameter is unlikely

to be. Moreover, because such irregularities inherently violate the assumptions that give rise to Gaussianity, the resulting likelihood function can be extremely complex. As discussed in Chap. 3, these irregularities manifest as higher-order contributions whose convergence to zero is non-uniform across the parameter space. Resultantly, the influence of these irregularities are themselves data-dependent: the complexity of the likelihood function depends on the number of events and the local geometry of the parameter space around the best-fit point.

Linear models, by contrast, are free from these complications. For such models, $\lambda(\theta) = a\theta + b$, the derivatives (or the Hessian in higher dimensions) are parameter-independent. This means that the curvature of the likelihood is fixed across the entire parameter space, with the Poisson Fisher information given by a constant,

$$\mathcal{I}(\theta) \propto \left(\frac{d\lambda(\theta)}{d\theta}\right)^2 = a^2. \tag{4.2}$$

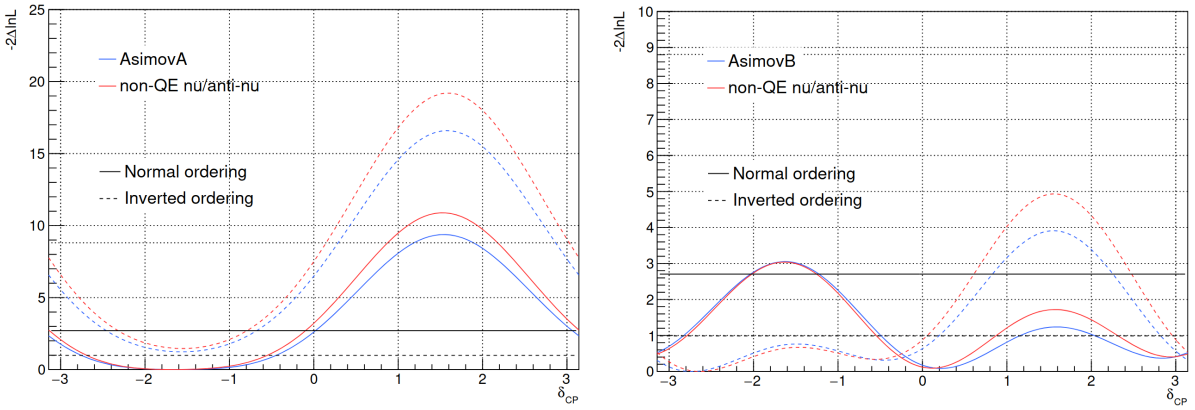
Additionally, this means that all regions of the parameter space converge uniformly to LAN conditions at the fastest possible rate [85].

4.2. Statement of the problem

The combination of parameter-space irregularities and data-dependent sensitivities undermines the performance and efficiency of commonly used methods for sensitivity and bias studies. In this section, we briefly explore the implications of these effects and how they can compromise techniques such as Asimov fits [90].

4.2.1. Efficiency

T2K's sensitivity studies make heavy use of Asimov fits due to their computational efficiency [72]. This efficiency is predicated on the idea that fitting a representative data set provides information about the general sensitivity of the experiment (or, in the context of FDS, the impact of alternative models). Conceptually, this assumption is closely connected to the properties of linear models discussed above. However, in nonlinear, non-Gaussian models, the sensitivity shape is complex and depends on the local parameter space, meaning multiple studies are required to understand



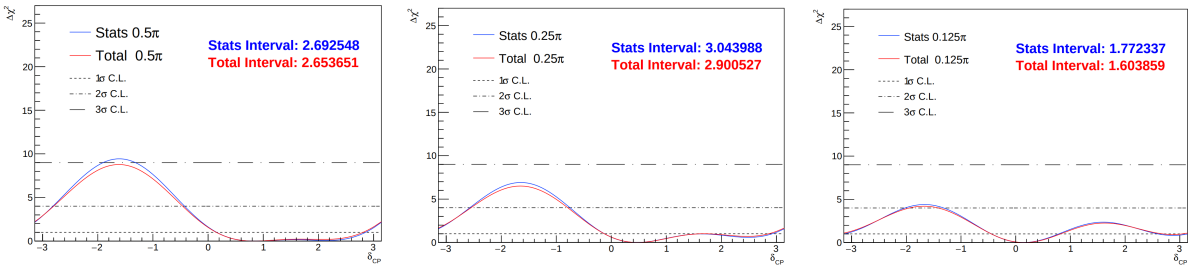
(a) Asimov A **(b)** Asimov B

Figure 4.1.: An example of a T2K fake data study performed at two different Asimov points: the blue contour is an Asimov fit using the nominal interaction model, while the red contour shows the fit using data generated from an alternative interaction model. From Ref. [91]

the fit behaviour in different regions. Fig. 4.1 illustrates this point, showing $\delta_{\mathcal{CP}}$ contours from a typical T2K FDS performed at distinct Asimov points: Asimov A, located near maximal $\mathcal{CP}$v ($\delta_{\mathcal{CP}} = -1.601$), and Asimov B, corresponding to $\mathcal{CP}$c ($\delta_{\mathcal{CP}} = 0$) [91]. Note how the contour shapes and interval sizes in these fits differ substantially. Specifically, at the 1σ level, Asimov B has two distinct intervals, whereas Asimov A has only one. The effect of the alternative models also differs between these regions: at $\delta_{\mathcal{CP}} = -1.601$, the ratio of the nominal 1σ Asimov interval to the corresponding fake-data interval is 0.951, while at $\delta_{\mathcal{CP}} = 0$ the ratios for the two intervals are 0.884 and 0.811. This shows that the impact of systematic uncertainty, represented by these alternative models, varies across parameter space, which makes interpreting their impact challenging.

4.2.2. Interpretation problems

Due to model irregularities, fake data and sensitivity studies can often be difficult to interpret, and in some cases, seemingly nonsensical. For example, intervals from statistics-only fits (in which systematic parameters are fixed at their nominal values and only the oscillation parameters are varied) can sometimes be larger than those from corresponding fits in which both statistical and systematic parameters are allowed to vary (see Fig. 4.1). This appears confusing and somewhat paradoxical, because by adding additional sources of uncertainty by allowing the systematic parameters to vary,



(a) Asimov point: $\delta_{\mathcal{CP}} = 0.5\pi$ **(b)** Asimov point: $\delta_{\mathcal{CP}} = 0.25\pi$ **(c)** Asimov point: $\delta_{\mathcal{CP}} = 0.125\pi$

Figure 4.2.: Examples of Asimov fits where the stats-only interval is larger than the total interval.

the constraint on the parameter of interest appears has improved. This counterintuitive behaviour can arise when the complete fit pulls more strongly against a boundary than the statistics-only fit. In practice, however, this effect can be obscured by the parameterisation.

When this happens, the systematic uncertainty cannot be meaningfully estimated by assuming the statistical and systematic uncertainties are related in quadrature, via $\sigma_{\text{Syst}} = \sqrt{\sigma^2_{\text{Total}} - \sigma^2_{\text{Stat}}}$. Consequently, T2K's FDSs have routinely needed to adopt alternative, and perhaps poorly motivated, metrics to accommodate these effects. For example, the following ad hoc definitions are extracted from a T2K technical note:

Interval changes
Due to the cyclic nature and current statistical sensitivity to the value of $\delta_{\mathcal{CP}}$, the previously defined metrics are not well suited for assessing changes to the $\delta_{\mathcal{CP}}$ intervals. Instead of forming summary metrics from pairs of interval boundaries, we consider shifts induced by each fake data study in each edge, $b^{s,\downarrow}_{\delta_{\mathcal{CP}}}, b^{s,\uparrow}_{\delta_{\mathcal{CP}}}$, of four significance intervals, $s = [1\sigma, 90\%, 2\sigma, 3\sigma]$, separately, $\Delta b^{s,\downarrow}_{\delta_{\mathcal{CP}}}$ and $\Delta b^{s,\uparrow}_{\delta_{\mathcal{CP}}}$.

Bias
The bias is defined as the fractional shift in the center of the 2σ intervals between the full Asimov and the fake data fits.

4.2.3. The linearisation concept

The central idea of this thesis is to avoid many of the aforementioned issues by implementing a linearised oscillation model. The strategy involves constructing an extended parameter space (Ω), such that the physical one (Θ) is embedded within it ($\Omega \supset \Theta$), while ensuring that the predicted event rates depend linearly on the extended parameters. This idea will become clearer in the sections that follow.

Principally, linearisation achieves two things:

- It eliminates the parameter space irregularities, and restores the desired Gaussian behaviour.
- It makes the variance data-independent, up to simple scaling factor from the Poisson statistics.
- It reveals and simplifies physics related features that are obscured by the standard non-linear PMNS parameterisation.

Given that the extended model has more parameters, we anticipate that the best-fit point will lie somewhere outside the physical sub-space. This is not a problem; indeed it is by construction, since the purpose is to understand which regions the data prefers when not restricted by the complex PMNS parameterisation, and to study how the extended likelihood surface intersects with the physical subspace.

4.2.4. An analogue

The underlying idea is quite similar to that of a kernel trick in machine-learning, so it is useful to use this as an analogue to help understand the concept [92].

Classification is a common machine learning task in which raw data is assigned to categories based on their measurable features. For example, determining whether an email is spam or not based on variables such as total word count, time sent, language used, etc. Many classification algorithms perform best when the data can be separated by a linear decision boundary. However, when the data is not linearly separable, a technique called the kernel trick can be used, in which these data representations are projected into additional dimensions where a linear cut can be applied. Once a separating hyperplane is applied in the transformed space, the data is projected back into its original coordinates.

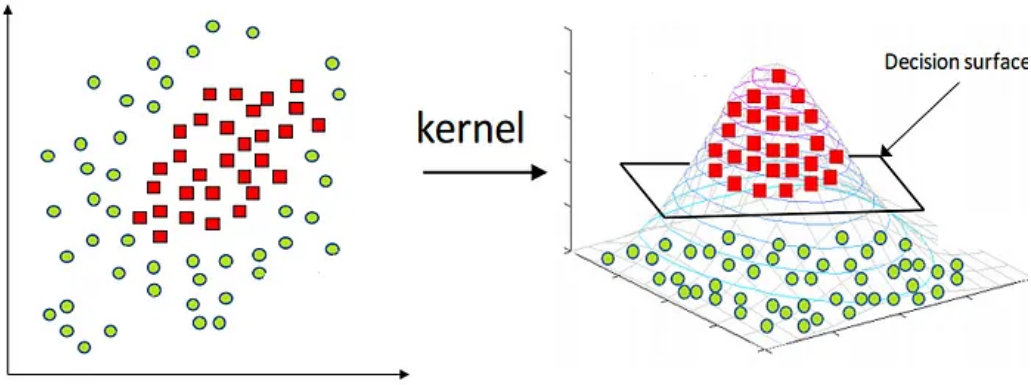


Figure 4.3.: Cartoon illustration of the kernel trick in which data is projected into a third dimension and then a decision surface is used to perform a selection.

A cartoon example is shown in Fig. 4.3, where the data is not linearly separable in the original parameterisation, but by projecting it into an additional dimension, a linear cut can be made. The key point here is that this extra dimension does not need to be inherently meaningful, but rather is just a useful mathematical reparameterisation that makes the data more processable.

This analogy serves to motivate a similar strategy for the oscillation probability: the physical PMNS parameterisation is transformed into a more convenient, extended one by introducing additional dimensions. These extra dimensions are constructed such that they contain the physical space, preserving physical inference in the extended space and enabling statements on the physical parameters to be recovered by projecting back into the original PMNS parameterisation.

4.2.5. Reparameterising criteria

The minimum linearity criteria are quite mild, though somewhat abstract. A suitable linear model will satisfy the following conditions:

- Be linear in its parameters.
- Contain the physical model as a subset, such that the standard physical parameterisation is embedded within it.

While these two conditions are fundamentally sufficient, we also introduce a third desirable criterion:

- The extended variables should be as simple as possible and remain closely related to the standard parameterisation.

This additional requirement is valuable because, beyond restoring uniform Gaussianity, the core philosophy of this linearisation concept is to maximise interoperability of sensitivity and bias related studies. By re-expressing the model in terms of simpler but PMNS-related parameters, some of the complexities inherent in the original parameterisation are removed and some features that would otherwise be obscured by it are more obvious.

4.2.6. Strategy

PMNS parameters appear in distinct functional forms within the standard probability (i.e. $\delta_{\mathcal{CP}}$ appears in $\sin\delta_{\mathcal{CP}}$ and $\cos\delta_{\mathcal{CP}}$). In line with the third criterion introduced above, our strategy is to treat the leading-order terms governed by these parameters as independent linear variables that can vary freely.

More generally, if θ is an oscillation parameter that appears in bounded functions $\omega_1(\theta), \omega_2(\theta) \in [a, b]$, then the extended parameterisation would be defined in terms of new, linear variables (ω_1, ω_2) where $(\omega_1, \omega_2) \in (-\infty, \infty) \times (-\infty, \infty)$. The corresponding extended probability becomes:

$$P(\theta) = A\omega_1(\theta) + B\omega_2(\theta) \rightarrow P_{Ext}(\omega_1, \omega_2) = A'\omega_1 + B'\omega_2. \tag{4.3}$$

The physical subspace is then defined by a constraint (e.g. $f(\omega_1, \omega_2) = C$), and the physical points correspond to those that satisfy this rule.

The subscript "Ext" is used to distinguish the extended probability from the physical one, as they are obviously not the same function and only coincide on the physical subspace (i.e. when $(\omega_1 = \omega_1(\theta), \omega_2 = \omega_2(\theta))$). In fact, although the extended oscillation probability has the desired mathematical properties, away from the physical subspace, it is strictly not a probability as crucially it is not bounded between 0 and 1. Nevertheless, for convenience it will be interpreted as such, since in practice the probability functions only as a weight applied to the neutrino spectrum. The appearance of negative "probabilities" in unphysical regions is therefore not inherently problematic; however, if they translate into negative event rates, this would correspond to an

undefined value of the Poisson likelihood. The practical implications of this, along with mitigating strategies, will be addressed in Chap. 5.

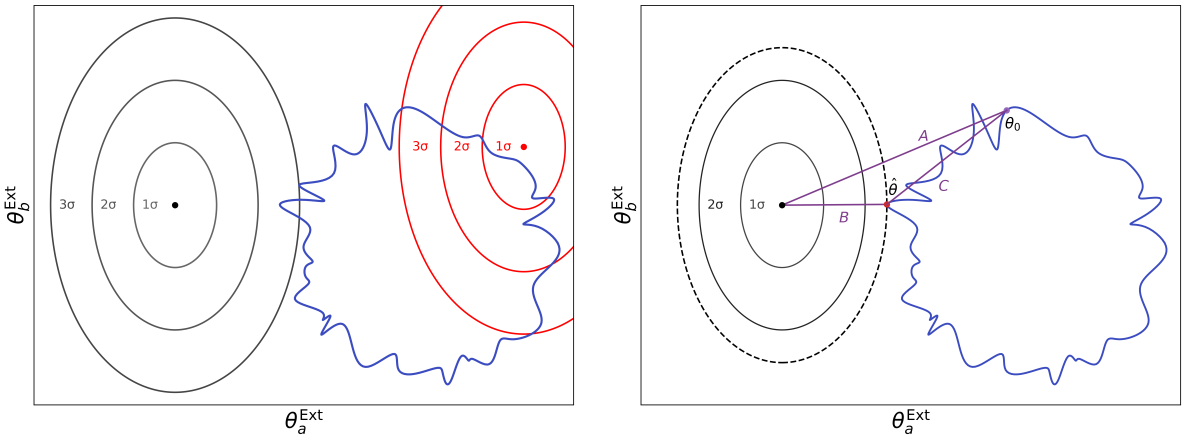
In this thesis we focus on $\delta_{\mathcal{CP}}$ as T2K's most high-profile measurement. However we note that this scheme can be applied to other non-linear PMNS parameters that introduce similar complexities (in Appendix A we propose an extension that includes θ_{23} and mass-ordering).

The $\mathcal{CP}$c and $\mathcal{CP}$v terms control different aspects of the probability, so by fitting in $\delta_{\mathcal{CP}}$, the individual influence of both of these terms becomes entangled, making some features of this fit appear complex, essentially requiring the analyser to mentally perform a coordinate transform to understand the implications. By fitting directly in terms of $\cos\delta_{\mathcal{CP}}$ and $\sin\delta_{\mathcal{CP}}$, the origins of these features become immediately apparent, and the experimental sensitivity decouples from the complexities of the physical parameterisation.

Additionally, because the extended probability defines a linear model, the corresponding likelihood surface is approximately Gaussian, with a shape that is independent of the observed data. In this formulation, the experimental design fixes the Gaussian sensitivity: different data realisations do not change the shape of the surface, but instead shift the physical parameter subspace relative to it, thereby modifying how the physical subspace intersects and maps onto the likelihood surface. Fig. 4.4a illustrates this schematically for an arbitrary physical subspace. The black and red Gaussian ellipses represent identical 1, 2, and 3σ contours in the extended space (noting that the likelihood is a three-dimensional surface and that these ellipses correspond to contours at fixed significance), differing only in their best-fit locations. Although the underlying surfaces are the same size and shape, the preferred region of the extended space changes, altering how the physical subspace (shown as the blue path) is positioned relative to them. It is this relative displacement that generates differences in the log-likelihood function of the physical parameter.

The likelihood surfaces in the physical and extended spaces are related through projections (where "projection" is the height in the extended parameter space evaluated at the point on the extended surface). Quantitatively, $\Delta\chi^2$ values in the extended and physical spaces are related by:

$$\begin{aligned}\Delta\chi^2_{Ext}(\epsilon(\theta_0)) - \Delta\chi^2_{Ext}(\epsilon(\hat{\theta})) &= \chi^2_{\epsilon(\theta_0)} - \chi^2_{\hat{\epsilon}} - (\chi^2_{\epsilon(\hat{\theta})} - \chi^2_{\hat{\epsilon}}) \\ &= \Delta\chi^2_{Phys}(\theta_0).\end{aligned} \tag{4.4}$$



(a) The physical subspace is shown interacting differently with Gaussians of equal size and shape, but with different best-fit points in the extended space.

(b) The dotted contour marks the minimum χ^2 within the physical subspace. Line A corresponds to the extended $\Delta\chi^2$ at $\epsilon(\theta_0)$; line B represents the χ^2 distance to the dotted line, i.e., $\Delta\chi^2_{Ext}(\epsilon(\hat{\theta}))$. Their difference gives C: the physical $\Delta\chi^2_{\text{Phys}}(\theta_0)$.

Figure 4.4.: Cartoon schematic illustrating how an arbitrarily complicated physical subspace (blue contour) appears when embedded in an extended parameter space.

Here, ϵ denotes the embedding map from the physical parameter space into the extended parameter space, such that $\epsilon(\theta_0)$ is a point on the physical subspace evaluated in the extended space, $\epsilon(\hat{\theta})$ is the best fit point on the physical subspace evaluated in the extended space, and $\hat{\epsilon}$ is the best point in the extended space. This is illustrated in Fig. 4.4b, where $\epsilon(\hat{\theta})$ correspond to the point on the physical blue contour that is lowest on the extended likelihood surface; $\min_\theta \Delta\chi^2_{Ext}(\epsilon(\theta))$.

4.3. $X_S - X_C$ probability model

To construct these linear models, we rely extensively on results familiar to neutrino physicists, but placed on a firmer footing by Yokomakura *et al.* [93]. There, they show that the matter potential is invariant under $O_{23}\Gamma_\delta$ transformations, which permits the neutrino evolution equation to be decomposed into two components: one that includes the matter potential and one that does not. Crucially, since θ_{23} and the $\delta_{\mathcal{CP}}$ phase appear in the term without the matter potential, the oscillation probabilities can be expressed exactly in particularly simple analytic forms with respect to these parameters. Specifically, in the case of $\delta_{\mathcal{CP}}$, they are shown to be linear in $\sin\delta_{\mathcal{CP}}$,

$\cos\delta_{\mathcal{CP}}$, and $\cos 2\delta_{\mathcal{CP}}$:

$$\begin{aligned} P(\nu_\mu \to \nu_e) &= A_{e\mu}\cos\delta_{\mathcal{CP}} + B_{e\mu}\sin\delta_{\mathcal{CP}} + C_{e\mu} \\ P(\nu_\mu \to \nu_\mu) &= A_{\mu\mu}\cos\delta_{\mathcal{CP}} + C_{\mu\mu} + D_{\mu\mu}\cos 2\delta_{\mathcal{CP}} \\ P(\nu_e \to \nu_e) &= C_{ee}\,, \end{aligned} \tag{4.5}$$

where the coefficients hold all of the non-$\mathcal{CP}$ phase dependent terms. As described in the previous section, we wish to replace the $\mathcal{CP}$v, $\mathcal{CP}$c terms with new linearised versions: X_S and X_C, respectively. Therefore, the parameterisation becomes:

$$\theta = \delta_{\mathcal{CP}} \in [-\pi, \pi] \longrightarrow \epsilon = (X_S, X_C) \in \mathbb{R}^2. \tag{4.6}$$

and the extended probability is:

$$\begin{aligned} P(\nu_\mu \to \nu_e) &= A_{e\mu}X_C + B_{e\mu}X_S + C_{e\mu} \\ P(\nu_\mu \to \nu_\mu) &\approx A_{\mu\mu}X_C + C_{\mu\mu} \\ P(\nu_e \to \nu_e) &= C_{ee}\,, \end{aligned} \tag{4.7}$$

where the $\cos 2\delta_{\mathcal{CP}}$ term has been neglected due to its demonstrated small contribution at T2K statistics (we show this in Sec. 4.3.1).

In this extended parameterisation, the physical subspace corresponds to the points satisfying $f(X_S, X_C) = X_S^2 + X_C^2 = 1$, meaning that the physical subspace, shown generically as a blue path in Fig. 4.4, appears simply as the unit circle.

4.3.1. Validation

Here, the $P(\nu_\mu \to \nu_\mu) \approx A_{\mu\mu}\cos\delta + C_{\mu\mu}$ approximation is justified by demonstrating that the ratio $(A_{\mu\mu}\cos\delta + C_{\mu\mu})/D_{\mu\mu}\cos 2\delta$ is large across the whole parameter space. This is verified via a simple numerical analysis in which all the oscillation parameters are randomly varied, and the corresponding $A_{\mu\mu}, B_{\mu\mu}, C_{\mu\mu}$ coefficients are calculated for each throw.

Performing this, only 1,000 samples out of 10 million had a ratio less than 10. The energies at which these instances occurred are shown in Fig. 4.5. Focusing on those above 100 MeV (this is reasonable as rapid oscillations occur below this energy, and in any case T2K's efficiency is negligible below $E_\nu = 200$ MeV) we see that these

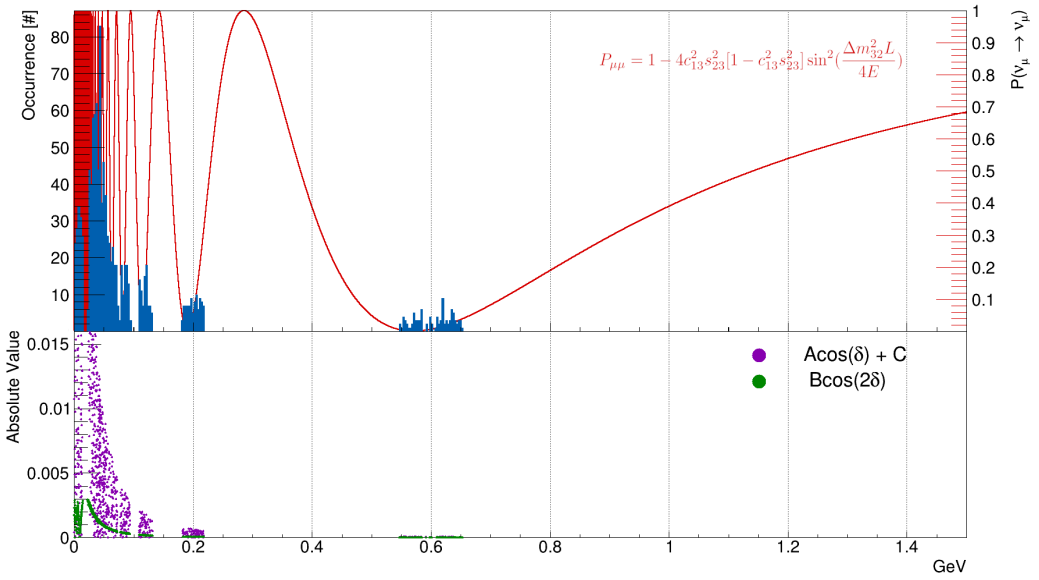


Figure 4.5.: Top panel: Energy histogram of throws where $(A_{\mu\mu}\cos\delta + C_{\mu\mu})/D_{\mu\mu}\cos 2\delta < 10$, superimposed over the oscillation probability. Bottom panel: Absolute value of the small ratio instances.

instances coincide with regions of low probability, indicating that the small ratios there are due to both $(A_{\mu\mu}\cos\delta + C_{\mu\mu})$ and $D_{\mu\mu}\cos 2\delta$ being simultaneously small, rather than due to a breakdown of the approximation. Supporting this, Fig. 4.5 shows that the absolute values of these small ratios are indeed very small for $E_\nu > 100$ MeV.

4.3.2. Implementation

Our implementation of the extended probability model uses a useful trick that exploits the rotational symmetry of the $\mathcal{CP}$-dependent component by working in polar coordinates. Doing this allows us to calculate the probability at a point in the standard parameter space and project it onto an arbitrary test point in the extended space, using ρ as an interpolating variable,

$$P_{\alpha\beta}^{Ext}(\rho, \delta, \vec{\theta}) = \frac{1+\rho}{2} P_{\alpha\beta}(\delta, \vec{\theta}) + \frac{1-\rho}{2} P_{\alpha\beta}(\delta + \pi, \vec{\theta}). \tag{4.8}$$

where α and β denote the neutrino flavours, δ is the corresponding physical $\mathcal{CP}$ phase, and ρ is the interpolation variable defined respectively as,

$$\rho(X_S, X_C) = \sqrt{(X_S^2 + X_C^2)}, \quad \delta(X_S, X_C) = \text{atan2}\,(X_S, X_C)\,. \tag{4.9}$$

Using $P(\nu_e \to \nu_\mu)$ as an example, it is easy to show that this gives the desired result:

$$\begin{aligned}
P^{\text{Ext}} &= \frac{1+\rho}{2}(A\cos(\delta_{\mathcal{CP}}) + B\sin(\delta_{\mathcal{CP}}) + C) + \frac{1-\rho}{2}(A\cos(\delta_{\mathcal{CP}}+\pi) + B\sin(\delta_{\mathcal{CP}}+\pi) + C) \\
&= \frac{1}{2}(A\cos(\delta_{\mathcal{CP}}) + B\sin(\delta_{\mathcal{CP}}) + C) + \frac{\rho}{2}(A\cos(\delta_{\mathcal{CP}}) + B\sin(\delta_{\mathcal{CP}}) + C) \\
&\quad + \frac{1}{2}(-A\cos(\delta_{\mathcal{CP}}) - B\sin(\delta_{\mathcal{CP}}) + C) - \frac{\rho}{2}(-A\cos(\delta_{\mathcal{CP}}) - B\sin(\delta_{\mathcal{CP}}) + C) \\
&= A\rho\cos(\delta_{\mathcal{CP}}) + B\rho\sin(\delta_{\mathcal{CP}}) + C \\
&= AX_C + BX_S + C.
\end{aligned}$$

Remember that one of the requirements of the extended parameterisation is that it reproduces the probability that would be obtained in the standard parameterisation on the physical subspace. A major advantage of this method is that it satisfies this requirement by construction. Moreover, it is also computationally efficient as despite there being three unknown coefficients, the symmetry acts as an additional constraint,

$$P(\delta,\theta,\rho) + P(\delta+\pi,\theta,\rho) = 2C, \tag{4.10}$$

allowing the extended probability to be fully determined using only two evaluations of the standard probability. A graphical illustration of this method is given in Fig. 4.6. In this diagram, the red dot represents a hypothetical test point in the extended space. A line is drawn from this point through the centre of the unit circle. The points at which this line intersects the circle (shown as green and purple dots) correspond to the physical evaluations at (δ,θ) and $(\delta+\pi,\theta)$. The distance from the origin to the red point (shown in blue) is ρ.

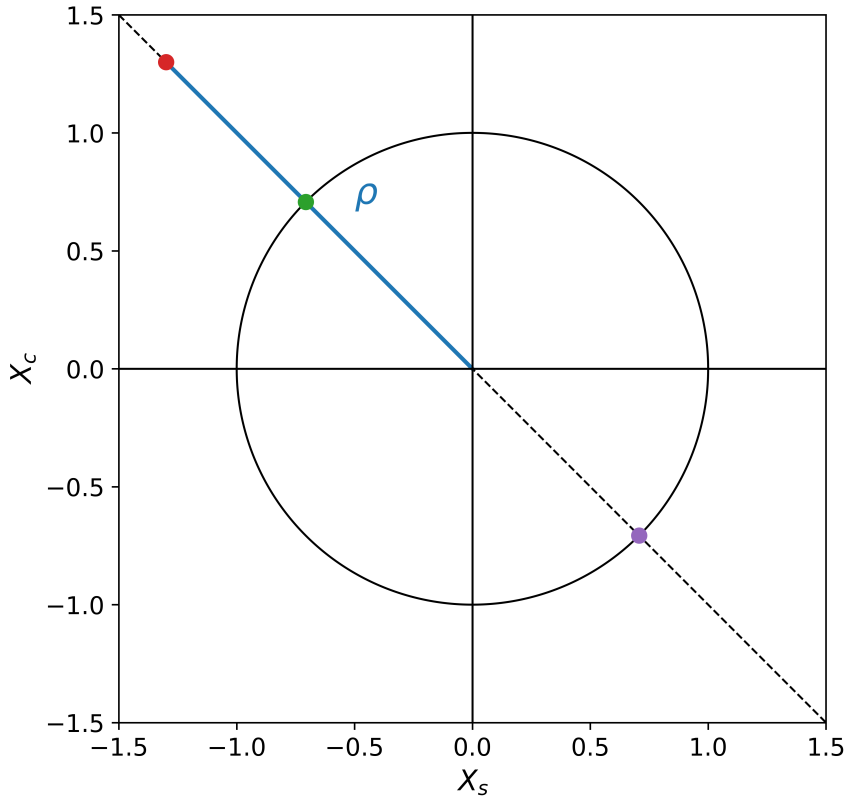


Figure 4.6.: Cartoon illustration of how the extended X_S-X_C probability is calculated for the red point by interpolating the green and purple points.

Chapter 5.

Advantages of linearising $\delta_{\mathcal{CP}}$

In the previous chapter, we demonstrated that fits in the standard PMNS parameterisation can exhibit complex and sometimes anomalous behaviour. Linearisation recasts the problem into a much simpler statistical space while preserving meaningful physical inference.

The symmetry of the linearised $\delta_{\mathcal{CP}}$ space (the X_S–X_C parameterisation) provides an intuitive framework for considering this concept. We therefore dedicate this chapter to the X_S–X_C parameterisation to highlight the attractive features and utility of this approach.

5.1. Sensitivity & bias studies

5.1.1. Expected X_S–X_C likelihood surface

We expect the likelihood surface in the extended parameterisation to be well approximated by a Gaussian, and therefore the corresponding log-likelihood surface to be parabolic, with a width that is independent of the true parameter value. Applying the ideas of Sec. 4.2.6 to the X_S–X_C parameterisation, we show that the one-dimensional $\delta_{\mathcal{CP}}$ log-likelihood function can be understood geometrically as the mapping of the physical subspace, defined by the unitarity constraint $(X_S^2 + X_C^2 = 1)$, onto this parabolic surface.

Because the experimental sensitivity to the $\mathcal{CP}$v term, X_S, is significantly stronger than the $\mathcal{CP}$c term, X_C, the log-likelihood surface exhibits a pronounced valley-like

structure: it is steeply constrained along the X_S-direction, while varying only weakly along the X_C-direction. Points on the physical circle that lie nearest to the valley floor are therefore in better agreement with the data.

Assuming a different best-fit trial point corresponds to simply translating the centre of this paraboloid without altering its overall shape. As a result, the trajectory traced by the physical subspace on the valley of the log-likelihood surface changes accordingly.

One nice way to visualise this is to imagine applying a cylindrical "cookie-cutter" orthogonal to the X_S–X_C plane, carving out the physical subspace from the paraboloid. The resulting intersection curve directly yields the 1D $\chi^2(\delta_{\mathcal{CP}})$ profile, as illustrated in Fig. 5.1.

In the following, using several fit scenarios characterised by different best-fit points, we demonstrate how the extended parametrisation renders the 1D $\delta_{\mathcal{CP}}$ log-likelihood function more intuitive. After only a few examples, it is even possible for one to predict the form of the corresponding 1D $\chi^2(\delta_{\mathcal{CP}})$ function simply by considering the best-fit location within the extended parameter space.

Case 1: Physical circle sits away from the valley bottom ($|X_S| \geq 1$)

The simplest case to consider is when the physical subspace lies away from the valley bottom (i.e $|X_S| \geq 1$). In this scenario, the circle sits on the steep X_S valley sides, and the one-dimensional $\delta_{\mathcal{CP}}$ log-likelihood has a single maximum and a single minimum (see Fig. 5.2). As the absolute X_S best-fit value increases, the physical circle moves progressively higher on the valley side, where the local gradient is steeper. Consequently, the separation between the minimum and maximum values of $\chi^2(\delta_{\mathcal{CP}})$ increases.

A special case occurs when the best-fit point lies exactly at $|X_S| = 1$, corresponding to $\delta_{\mathcal{CP}} = \pm\pi/2$. In this limit, the minimum of the physical trajectory sits at the bottom of the likelihood valley, and the resulting $\chi^2(\delta_{\mathcal{CP}})$ function becomes extremely flat locally, as shown in Fig. 5.3, where the flat region has been highlighted in red.

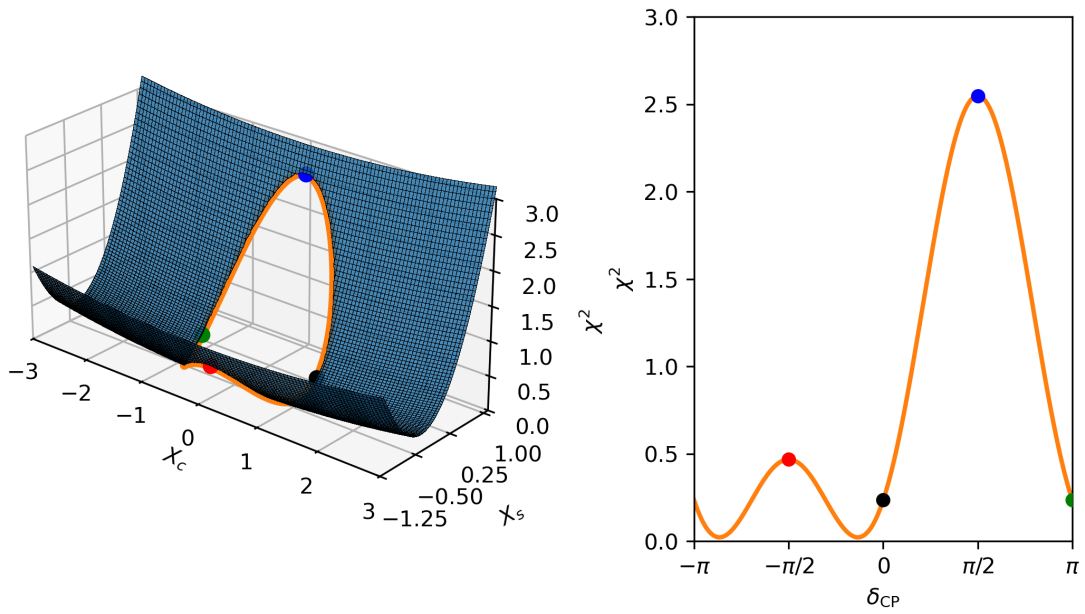


Figure 5.1.: Schematic illustration of the physical circle $X_S^2 + X_C^2 = 1$ mapped onto the parabolic log-likelihood surface expected in the extended parameterisation. Left: The physical parameter space can be visualised as being "cut out" by a cylindrical cookie-cutter applied orthogonally to the X_S—X_C plane. The orange edge is the resulting map of the trajectory of the physical $\delta_{\mathcal{CP}}$ space, with the height corresponding to the $\chi^2(\delta_{\mathcal{CP}})$ value. Right: Unwrapping this orange trajectory into the $\delta_{\mathcal{CP}}$ coordinate directly yields the one-dimensional $\chi^2(\delta_{\mathcal{CP}})$ profile. The $\mathcal{CP}$c points, $\delta_{\mathcal{CP}} = 0$ (black) and $\delta_{\mathcal{CP}} = \pi$ (green), and the $\mathcal{CP}$v points, $\delta_{\mathcal{CP}} = \pm\pi/2$ (red and blue), are marked on both panels to show their corresponding locations on the extended likelihood surface and in the resulting 1D $\chi^2(\delta_{\mathcal{CP}})$ function.

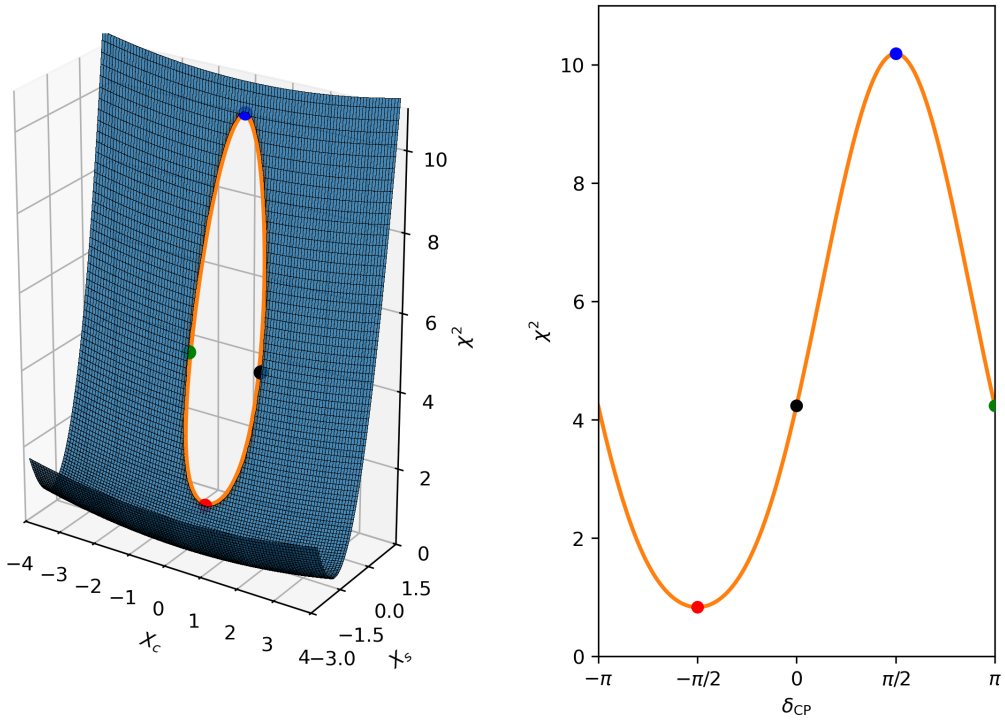


Figure 5.2.: Schematic illustration of the physical circle $X_S^2 + X_C^2 = 1$ mapped onto the parabolic log-likelihood surface expected in the extended parameterisation. In this example, a best-fit value with $|X_S| > 1$ is assumed, such that the cylindrical "cut-out" lies entirely along one of the steep X_S sides of the surface. As in the previous figure, the orange curve traces the intersection of the physical subspace on likelihood surface, and its height defines the corresponding 1D $\chi^2(\delta_{\mathcal{CP}})$ function shown on the right. The $\mathcal{CP}$c points, $\delta_{\mathcal{CP}} = 0$ (black) and $\delta_{\mathcal{CP}} = \pi$ (green), and the $\mathcal{CP}$v points, $\delta_{\mathcal{CP}} = \pm\pi/2$ (red and blue), are marked on both panels to show their corresponding locations on the extended likelihood surface and in the resulting 1D $\chi^2(\delta_{\mathcal{CP}})$ function.

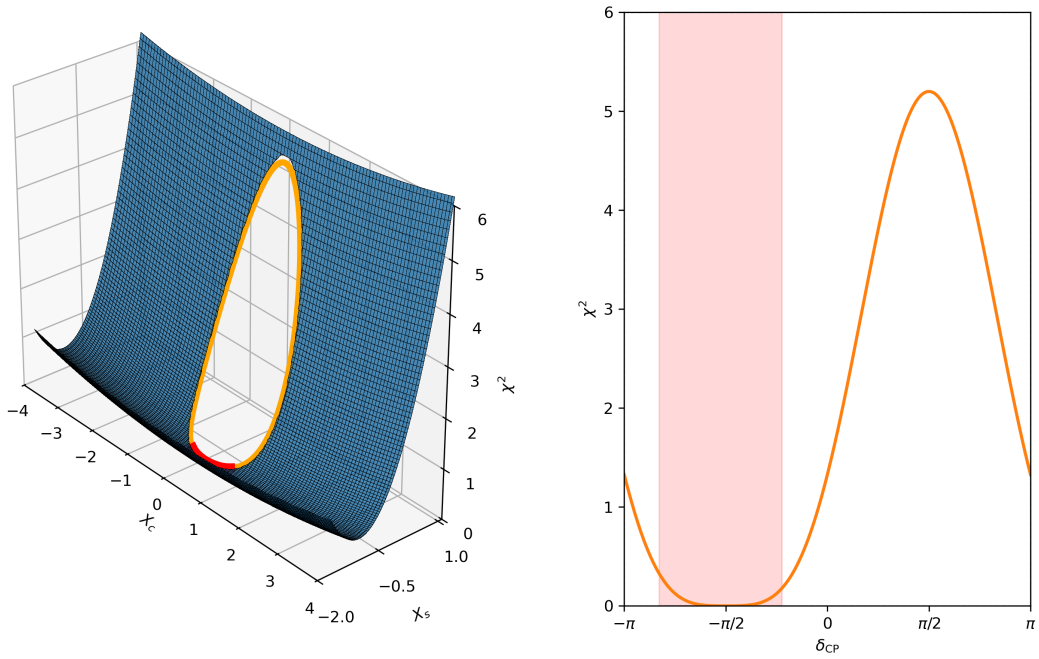


Figure 5.3.: Schematic illustration of the physical circle $X_S^2 + X_C^2 = 1$ mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| = 1$ is assumed, such that the edge of the physical subspace lies precisely along the floor of the log-likelihood valley. The resulting flat region in both panels, arising from the physical circle following the valley floor, is highlighted by the red shaded band.

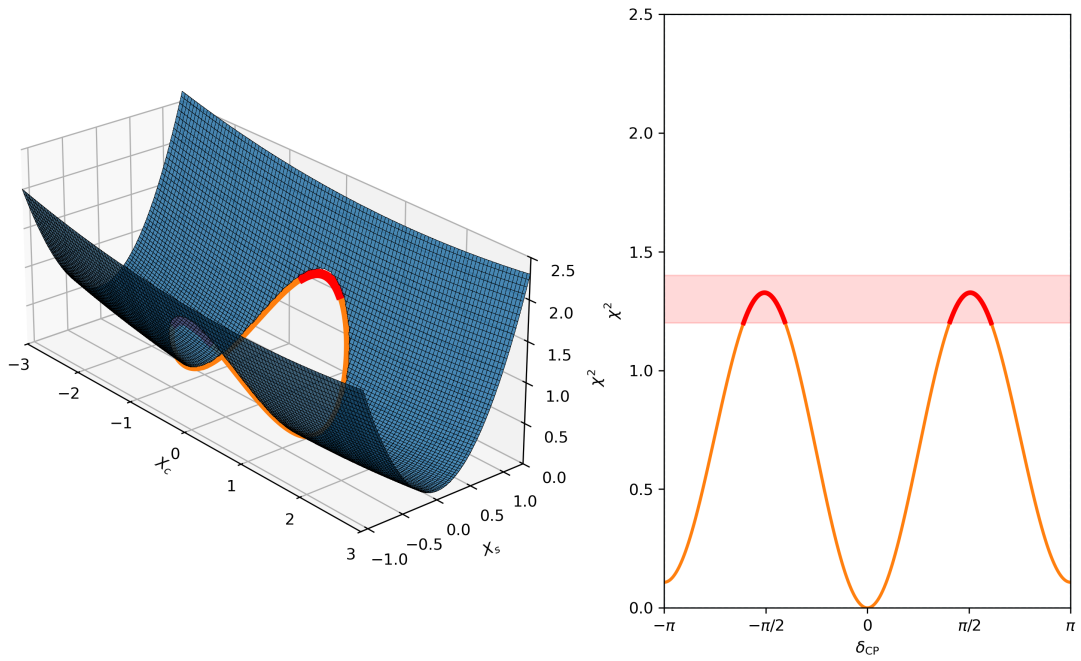


Figure 5.4.: Schematic illustration of the physical circle $X_S^2 + X_C^2 = 1$ mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| < 1$ is assumed, such that the physical subspace traces a saddle-like trajectory across the paraboloid. The corresponding local maxima are highlighted by the red shaded band in both panels.

Case 2: Physical circle sits in the valley ($|X_S| < 1$)

If $|X_S| < 1$, then the physical circle sits across the valley (traversing both steep paraboloid sides), and therefore the $\chi^2(\delta_{\mathcal{CP}})$ function exhibits multiple local maxima and minima. Geometrically, the physical subspace traces a saddle-like trajectory across the paraboloid in this regime.

The two maxima correspond to the locations of the highest points on the steep X_S sides (see Fig. 5.4), while the three minima (there are technically only two, as π and $-\pi$ represent the same point) correspond to the highest and lowest points on the shallow X_C sides, as highlighted in Fig. 5.5. Holding X_C fixed while varying X_S shifts the center of the paraboloid along the steep direction, thereby modifying the asymmetry between the maxima. The X_S best-fit therefore determines the relative $\chi^2(\delta_{\mathcal{CP}})$ heights of the two maxima, as shown in Fig. 5.6, in which the saddle shape is retained, but there is now a preference for one $\mathcal{CP}$v term ($\delta_{\mathcal{CP}} = -\pi/2$ in this case).

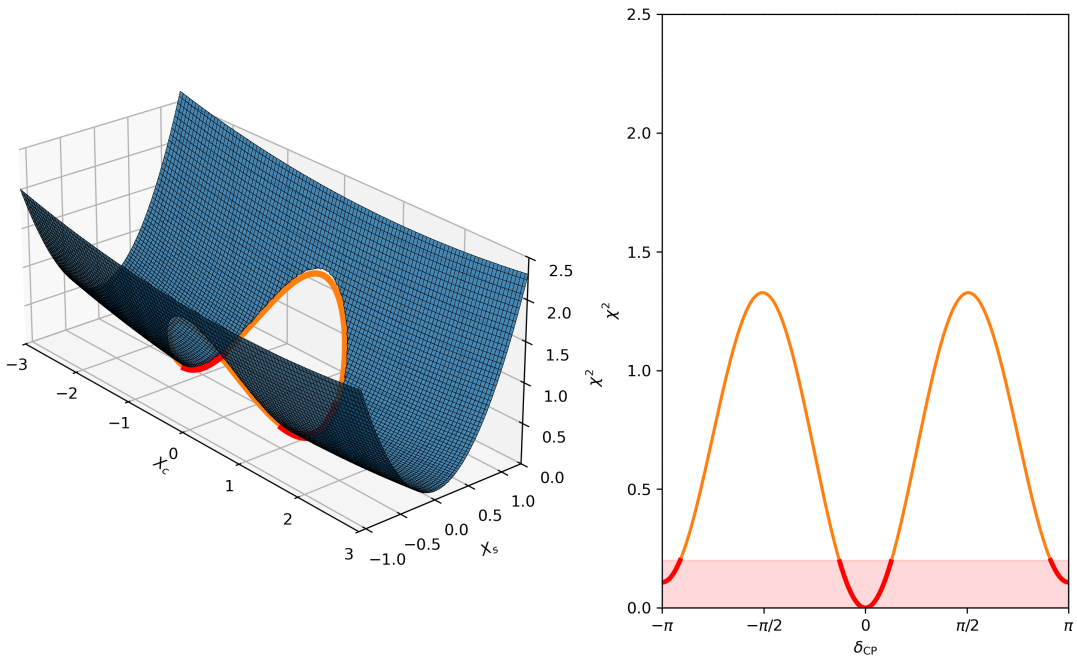


Figure 5.5.: Schematic illustration of the physical circle $X_S^2 + X_C^2 = 1$ mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| < 1$ is assumed, such that the physical subspace traces a saddle-like trajectory across the paraboloid. The corresponding local minima are highlighted by the red shaded band in both panels.

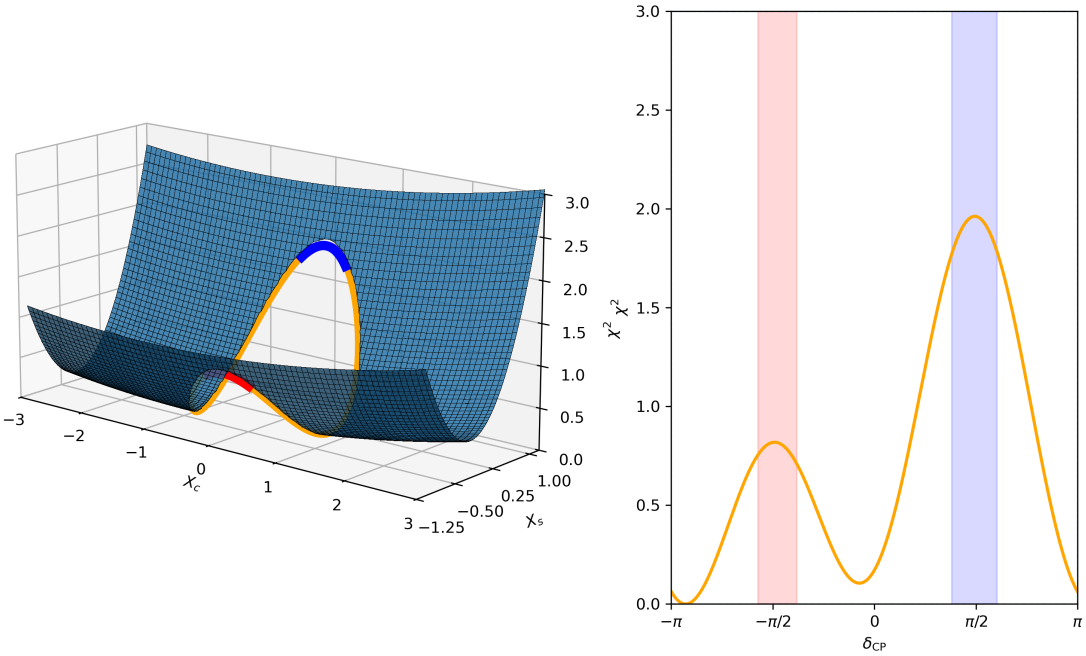


Figure 5.6.: Schematic illustration of the physical circle mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| = 0.3$ is assumed, such that the physical subspace traces a saddle-like trajectory across the paraboloid, with a preference for one of the $\mathcal{CP}$v points. The corresponding major and minor local maxima are highlighted by the red and blue shaded bands in both panels.

Role of X_C best-fit value

As the dominant constraint on $\delta_{\mathcal{CP}}$, the principal features of the 1D $\chi^2(\delta_{\mathcal{CP}})$ function are determined by how the physical circle maps onto the steep X_S sides. By contrast, shifting the best-fit value of X_C primarily alters which of the $\mathcal{CP}$c solutions, $\cos(\delta_{\mathcal{CP}} = 0) = +1$ or $\cos(\delta_{\mathcal{CP}} = \pm\pi) = -1$, is preferred.

Consider again the case $|X_S| = 1$, but now with a more extreme best-fit value of X_C (see Fig. 5.7). Shifting the X_C best-fit value enhances the influence of the shallower X_C side, which is locally steeper in this region of the valley. Consequently, the region that previously appeared flat (highlighted in red in Fig. 5.3) develops a tilt, skewing toward the favoured $\mathcal{CP}$c solution.

Likewise, in the $|X_S| < 1$ case, where the physical circle lies across the valley, changing the magnitude of X_C alters the relative heights of the two minima. In Fig. 5.8, where a positive value of X_C is assumed, the $\delta_{\mathcal{CP}} = 0$ (highlighted in red in this figure) solution lies closer to the extended log-likelihood minimum than the $\delta_{\mathcal{CP}} = \pm\pi$ solution (highlighted in blue).

Conversely, in Fig. 5.9, a negative value of X_C is assumed. In this case, the $\delta_{\mathcal{CP}} = \pm\pi$ solution lies closer to the extended log-likelihood minimum than the $\delta_{\mathcal{CP}} = 0$ solution, and is therefore preferred (again highlighted in red, with the disfavoured solution shown in blue).

Remarks

Hopefully is now clear how straightforward the features of the 1D $\delta_{\mathcal{CP}}$ log-likelihood become in this extended space. By combining the behaviours described above, essentially any configuration of the 1D $\chi^2(\delta_{\mathcal{CP}})$ can be anticipated geometrically simply by shifting the centre (the best-fit point) of the paraboloid within the extended parameter space.

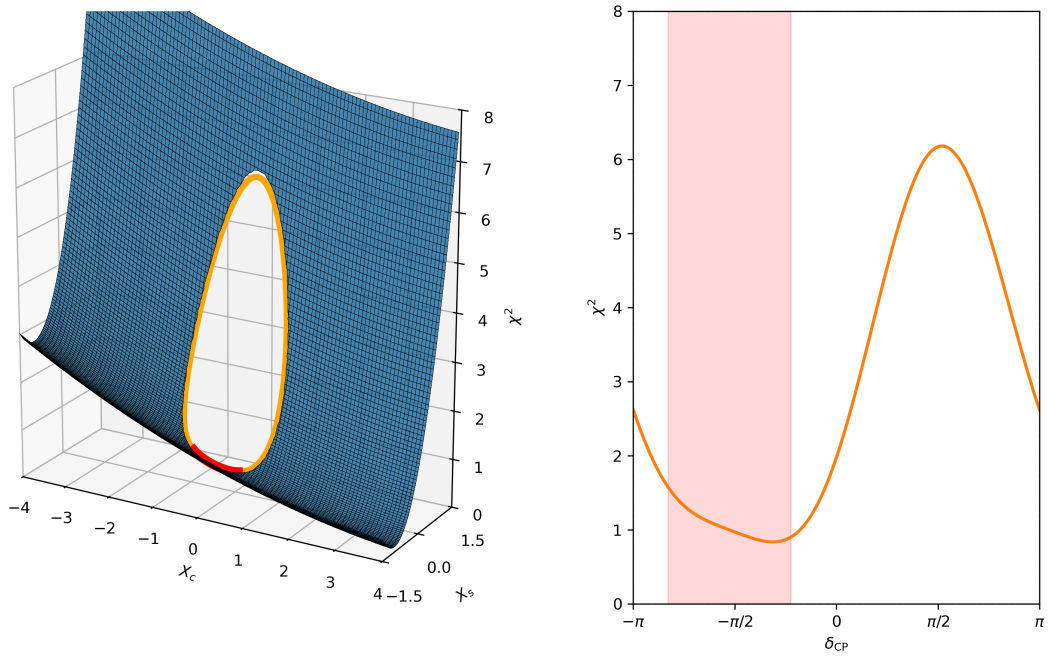


Figure 5.7.: Schematic illustration of the physical circle mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| = 1$ and a relatively extreme value of X_C is assumed, such that the influence of the shallower X_C side becomes pronounced. As a result, the region that would otherwise appear flat (for $X_C = 0$) becomes tilted, with the skew pointing towards the favoured $\mathcal{CP}$c solution. The resulting tilted region is highlighted by the red shaded band in both panels.

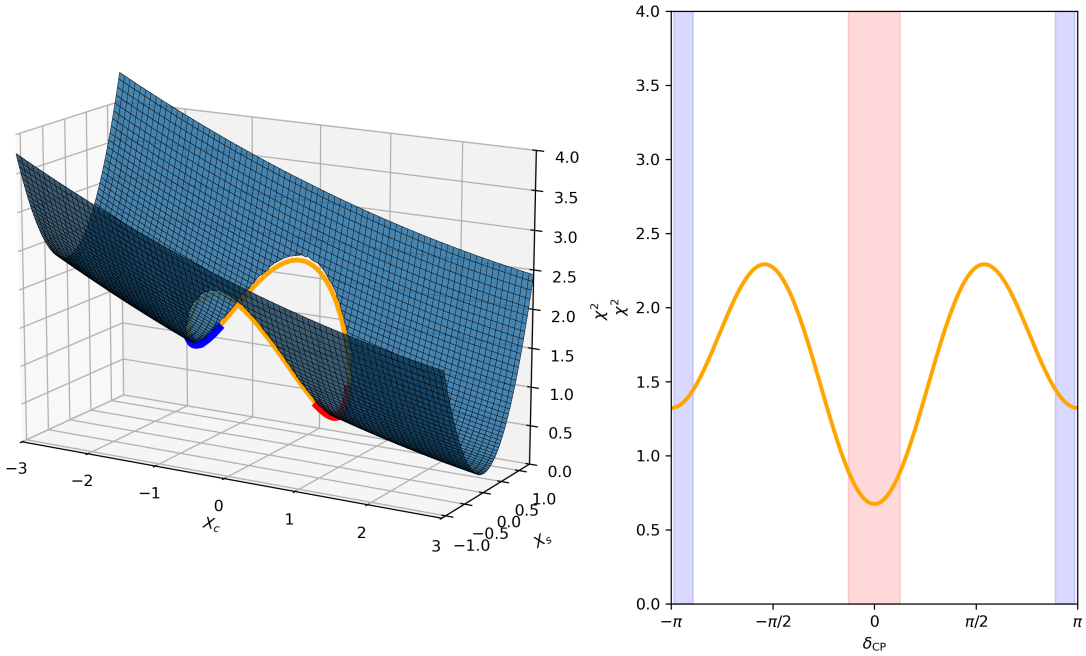


Figure 5.8.: Schematic illustration of the physical circle mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| < 1$ and a relatively extreme *positive* value of X_C is assumed, such that the influence of the shallower X_C side becomes pronounced, which has the effect of selecting which one of the $\mathcal{CP}$c values is preferred (in this case $\delta_{\mathcal{CP}} = 0$, which is highlighted in red).

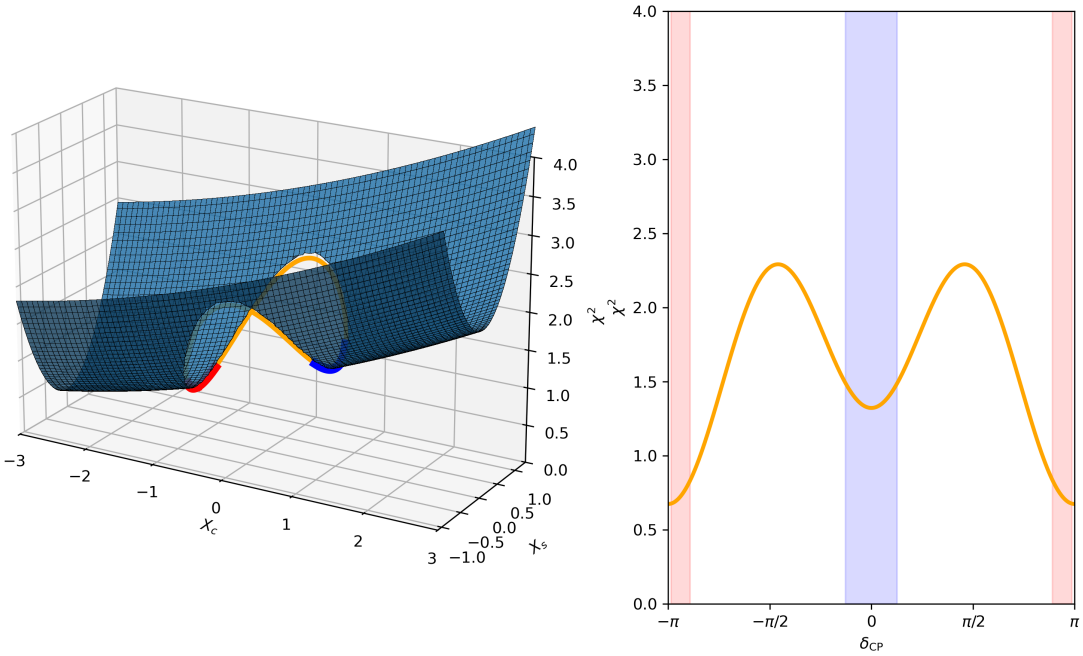


Figure 5.9.: Schematic illustration of the physical circle mapped onto the parabolic log-likelihood surface. In this example, a best-fit value with $|X_S| < 1$ and a relatively extreme *negative* value of X_C is assumed, such that the influence of the shallower X_C sides becomes pronounced, which has the affect of selecting which one of the $\mathcal{CP}$c values is preferred (in this case $\delta_{\mathcal{CP}} = \pm\pi$, which is highlighted in red).

5.1.2. Asimov fits

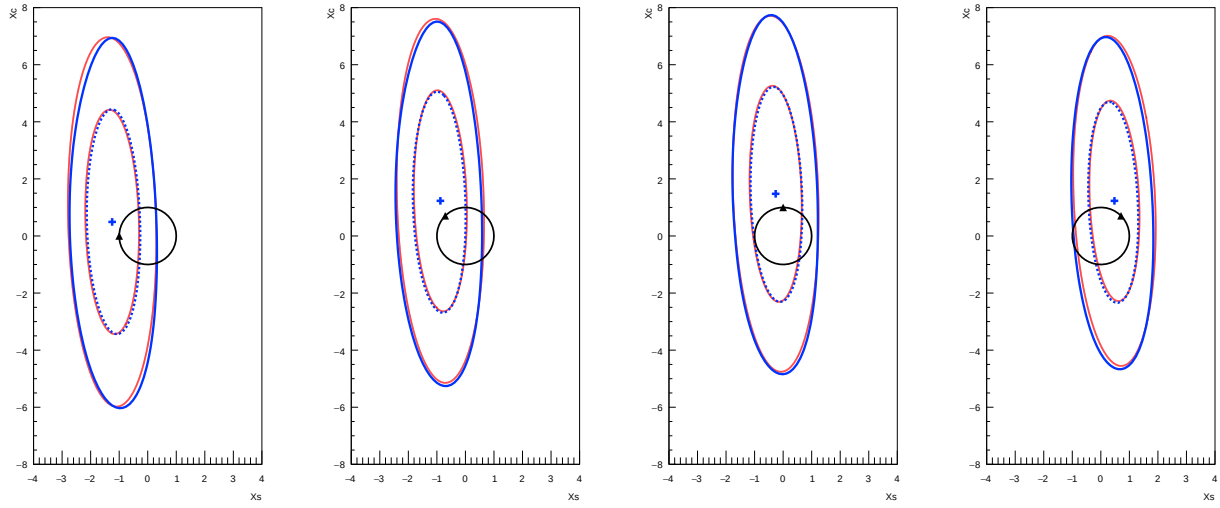
Here, we present and comment on the appearance of sensitivity studies performed with the P-Theta T2K fitting software, modified to operate in the extended $\delta_{\mathcal{CP}}$ parameterisation. Figure 5.10 shows several Asimov fits generated for different true values of $\delta_{\mathcal{CP}}$. The mass-ordering ambiguity is unimportant to the main message of this chapter, so to keep the figures less busy, true Normal Ordering is assumed throughout.

As desired, these contours are elliptical (Gaussian-like) and therefore have comparatively simple shapes irrespective of the true value of $\delta_{\mathcal{CP}}$. As a visual reference, a "true Gaussian" generated by evaluating the local Hessian at the best-fit point has been superimposed (in red) on each of the Asimov fits in Fig. 5.10. To directly compare their shapes, Fig. 5.11a shows the contours of these Asimov fits, recentered at $(0, 0)$, with the difference understood to arise from Poisson statistics (see Sect. 3.5.1). We will demonstrate this to be the case in Sect. 5.1.6, but for now, we can look at a Poisson-corrected version of Fig. 5.11b to compare the shape of the contours. Since only a scaling has been applied to produce this correction, the ellipse shapes are preserved, so Fig. 5.11b serves to demonstrate the agreement in contour shapes. This means that it is easy to predict the experimental sensitivity for different values of $\delta_{\mathcal{CP}}$.

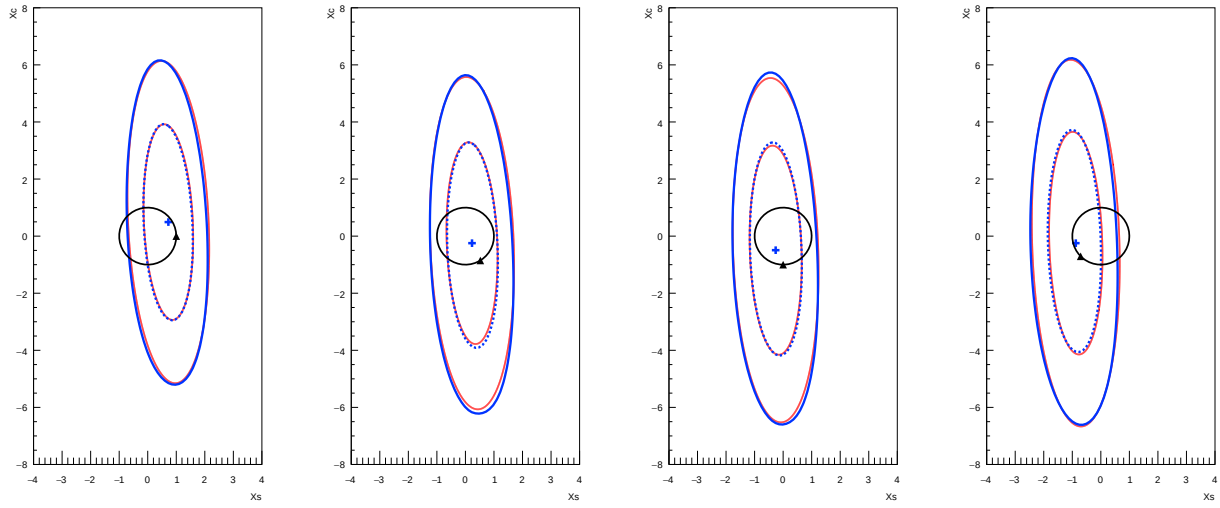
These contours are approximately four times larger in the X_C direction than in X_S, which is consistent with T2K's relative sensitivities to the $\mathcal{CP}$c and $\mathcal{CP}$v terms. The way these contours intersect with the physical subspace (the black circle) is also consistent with how T2K generates sensitivity to $\delta_{\mathcal{CP}}$: when the data prefers values of $\delta_{\mathcal{CP}}$ near maximal $\mathcal{CP}$v (which has consistently been the case for T2K), the ellipse sits to the left of the physical circle, where its overlap with the circle is minimal and the sensitivity is greatest; this is why T2K has the tightest constraint on $\delta_{\mathcal{CP}}$. It also shows why the relative insensitivity to $\cos\delta_{\mathcal{CP}}$ is less important for rejecting the $\mathcal{CP}$c values $\delta_{\mathcal{CP}} = 0, \pi$ (which correspond in this space to the points $(X_S = 0, X_C = 1)$ and $(X_S = 0, X_C = -1)$). Clearly, to maximise the power of excluding $X_S = 0$, it is the narrowness of the ellipses in the X_S direction that is critical.

Additional remarks on improved properties

We have established that, unlike in the standard parameterisation, fits performed in this linearised space exhibit the desired simplicity and, up to a trivial scaling factor due to Poisson statistics, are uniform across the parameter space. Linearisation therefore

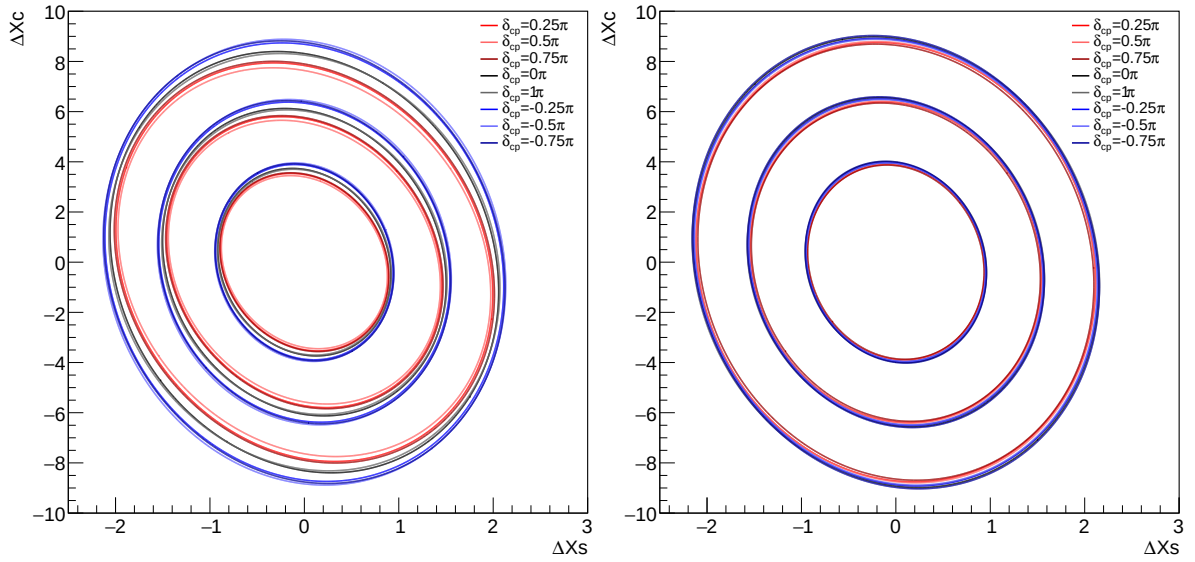


(a) $\delta_{\mathcal{CP}} = -\pi/2$ **(b)** $\delta_{\mathcal{CP}} = -0.25\pi$ **(c)** $\delta_{\mathcal{CP}} = 0$ **(d)** $\delta_{\mathcal{CP}} = 0.25\pi$



(e) $\delta_{\mathcal{CP}} = \pi/2$ **(f)** $\delta_{\mathcal{CP}} = 0.75\pi$ **(g)** $\delta_{\mathcal{CP}} = \pi$ **(h)** $\delta_{\mathcal{CP}} = -0.75\pi$

Figure 5.10.: Asimov fits at selected true $\delta_{\mathcal{CP}}$ values. Shown are the 1σ and 2σ contours of the log-likelihood surface in the extended (X_S, X_C) space. The blue cross indicates the best-fit point. The blue contours correspond to the full likelihood surface obtained from the data fit, while the red contours show the parabolic (Gaussian) approximation derived from the Hessian evaluated at the best-fit point, providing a visual demonstration of how elliptical the blue contours are. The black circle denotes the physical subspace, and the black triangle marks the true parameter value used to generate the Asimov dataset.



(a) Direct comparison. **(b)** Poisson corrected.

Figure 5.11.: Comparison of 1σ, 2σ, and 3σ confidence contours from Asimov fits generated at different true $\delta_{\mathcal{CP}}$ values. The red shaded contours denote those for $\cos\delta_{\mathcal{CP}} > 0$, the blue for $\cos\delta_{\mathcal{CP}} < 0$, and the green for $\cos\delta_{\mathcal{CP}} = 0$.

restores the efficiency principles underlying the Asimov concept and the statistical intuition of having an invariant shape (determined by experimental sensitivity) that shifts according to data and shrinks uniformly with increased statistics.

Separately examining the $\mathcal{CP}$v and $\mathcal{CP}$c contributions is also instructive for understanding how biases shape the $\delta_{\mathcal{CP}}$ constraint. In the Asimov fits shown in Fig. 5.10, marginalisation over $\sin^2 2\theta_{23}$ induces a systematic shift of the best-fit values, pushing X_S to smaller values and X_C to larger values relative to the true point. In the true $\delta_{\mathcal{CP}} = -\pi/2$ scenario, this bias mostly moves the contour into the unphysical region $X_S < -1$, thereby shrinking the interval. Conversely, in the true $\delta_{\mathcal{CP}} = 0$ scenario, the bias mostly shifts the contour into the unphysical region $X_C > 1$, reducing the $\mathrm{sgn}(\cos\delta_{\mathcal{CP}})$ degeneracy.

These Xs–Xc contours can be entirely specified by just the X_S, X_C mean values, their widths, and their covariance (at least to the 2σ level). The likelihood of any point on the plane is thus well described by the appropriately scaled distance from the best-fit point,

$$r^2 = \frac{1}{1-\rho^2}\left[\frac{(X_S - X_{S,\mathrm{bf}})^2}{\sigma_s^2} + \frac{(X_C - X_{C,\mathrm{bf}})^2}{\sigma_c^2} - \frac{2\rho}{\sigma_s\sigma_c}(X_S - X_{S,\mathrm{bf}})(X_C - X_{C,\mathrm{bf}}),\right] \quad (5.1)$$

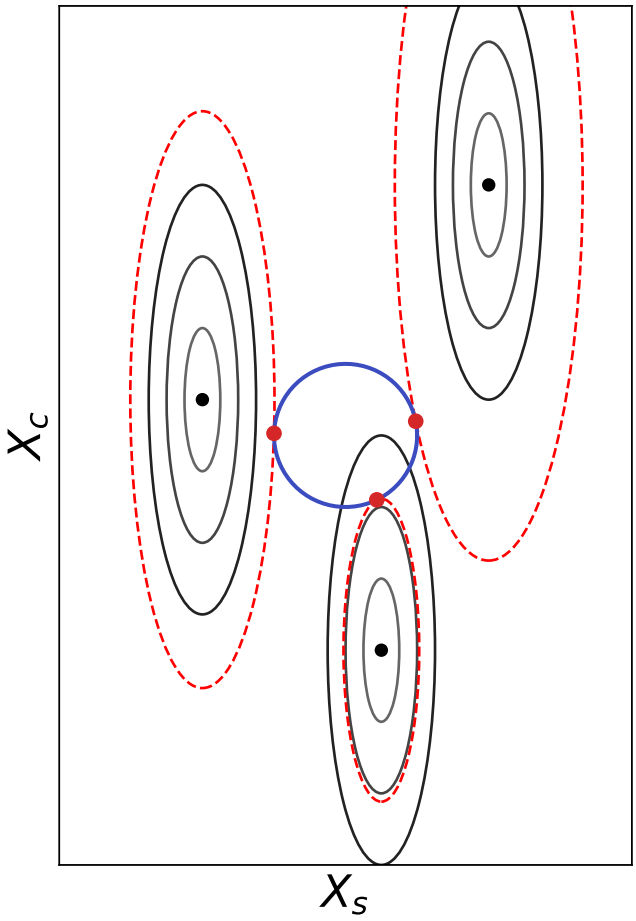
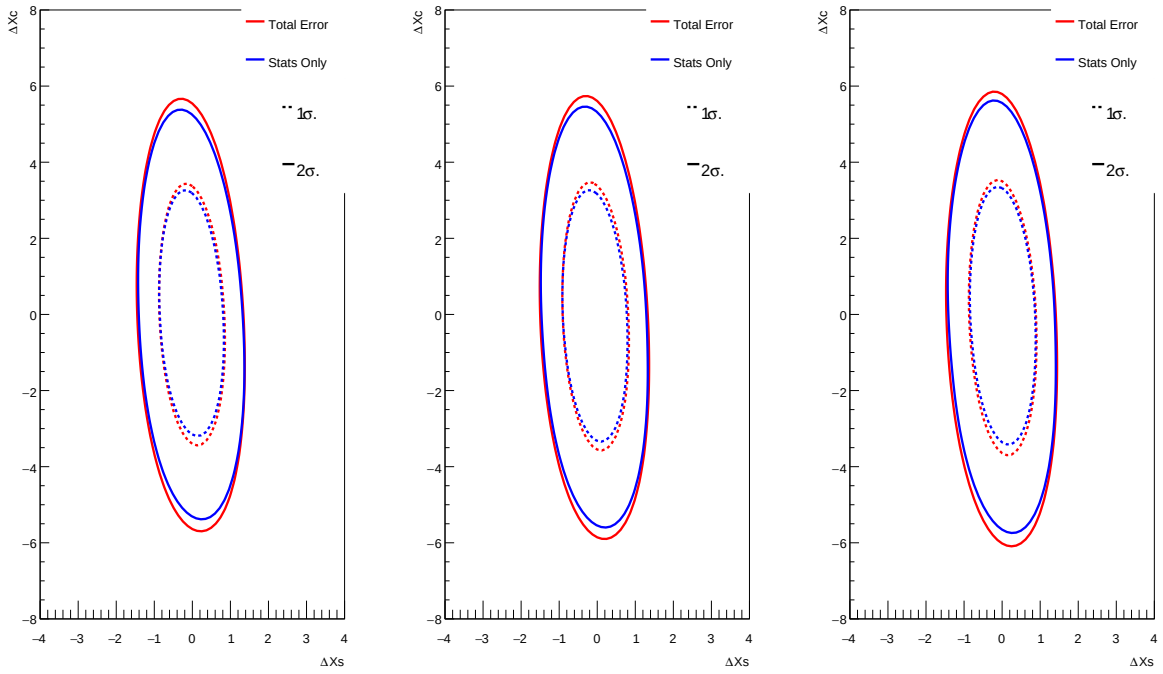


Figure 5.12.: Example of sensitivity accounted for projections onto the physical region from different best-fit points in the (X_S, X_C) space. The dotted red ellipse represents the contour that just tangentially touches the unit circle, thereby defining the location of the best-fit point on the physical boundary.

where $\rho = \sigma_{sc}/(\sigma_s\sigma_c)$ is the correlation coefficient, $X_{S,\text{bf}}$ denotes the best-fit value of X_S, and $X_{C,\text{bf}}$ denotes the best-fit value of X_C. Since this is the equation of an ellipse, any point in the physical subspace (the physical circle) has a χ^2 proportional to r^2. The best-fit physical point therefore corresponds to the location on the physical circle with the minimum r^2. This concept was introduced in Sec. 4.2.6 and further developed in Sect. 5.1.1 using the full three-dimensional likelihood surface. However, it is instructive to visualise the same idea directly in the two-dimensional X_S–X_C plane. Conceptually, one considers ellipses of increasing significance until the smallest ellipse that is tangent to the physical circle is reached. Fig. 5.12 illustrates this projection for several hypothetical best-fit points.



(a) Asimov point: $\delta_{\mathcal{CP}} = 0.5\pi$ **(b)** Asimov point: $\delta_{\mathcal{CP}} = 0.25\pi$ **(c)** Asimov point: $\delta_{\mathcal{CP}} = 0.125\pi$

Figure 5.13.: Asimov fits where $\delta_{\mathcal{CP}}$ Stats-only>Total interval reproduced in X_S–X_C extended space.

The limitations of this will be discussed in Sect. 5.2, but ostensibly, this is a remarkable simplification: the previously complex contours, as in Fig. 4.1, have been reduced to just five numbers.

Finally, we quickly highlight here that in this space, familiar statistical operations such as adding uncertainties in quadrature are again well defined. In particular, the pathological behaviour highlighted in Sect. 4.2.2, where the statistics-only uncertainty was larger than the total uncertainty, does not occur in the linearised space. Reproducing these fits in the X_S–X_C parameterisation (see Fig. 5.13), we observe that the statistics-only contour lies entirely within the total error contour.

5.1.3. Common features of the PMNS parametrisation in extended space

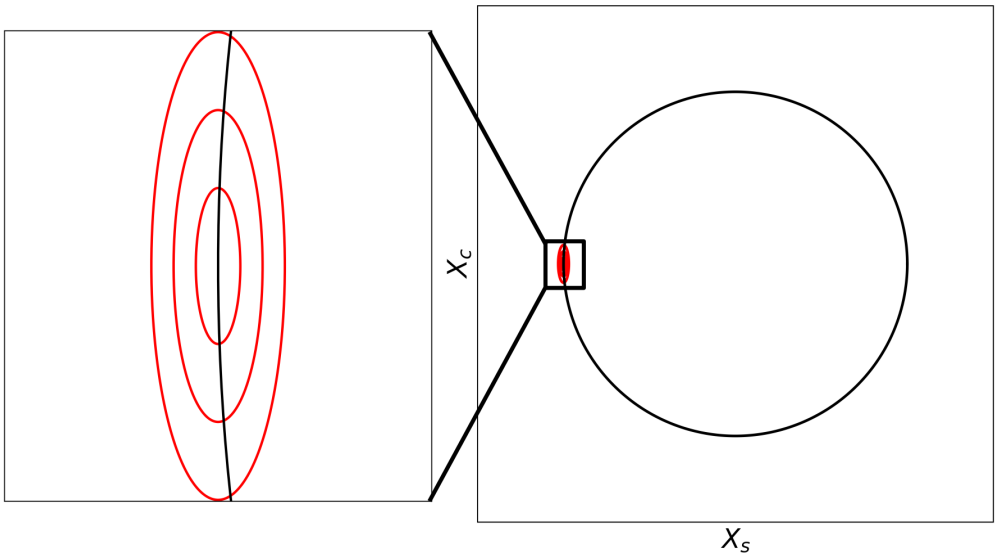
At this point, it is useful to discuss how some common features of the PMNS parameterisation manifest in the linearised space. We start by discussing how non-Gaussian behaviour generically appears as a curved physical subspace, and then specifically consider how boundaries, degeneracies, and pulls against boundaries are represented in this space.

Curved physical subspace

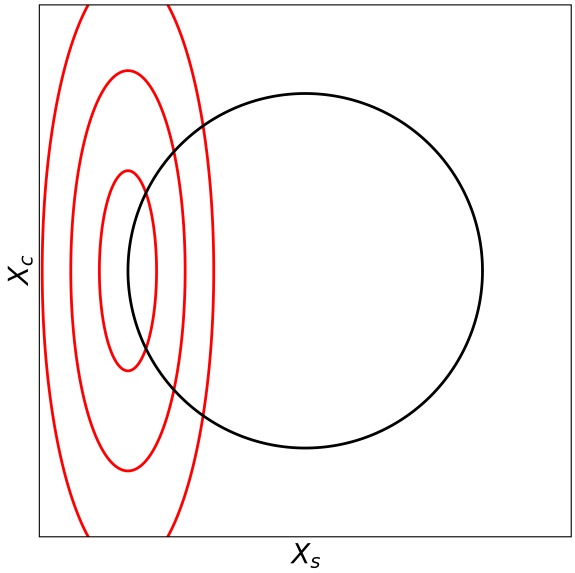
In Sect. 3.4.5, we discussed how the irregularities responsible for non-Gaussianity disappear for sufficiently large statistics. The extended space intuitively captures this behaviour: as statistics increase, the size of the contours shrinks, and, in the limit, the curvature of the physical subspace is no longer perceivable. When this happens, the physical subspace segments the 2D Gaussian contour approximately as a straight line (see Fig. 5.14a). Since projections onto a straight line define a 1D Gaussian regardless of the shape of the 2D Gaussian, it follows that visible curvature represents deviation from Gaussianity when results are reported in terms of $\delta_{\mathcal{CP}}$.

Boundaries

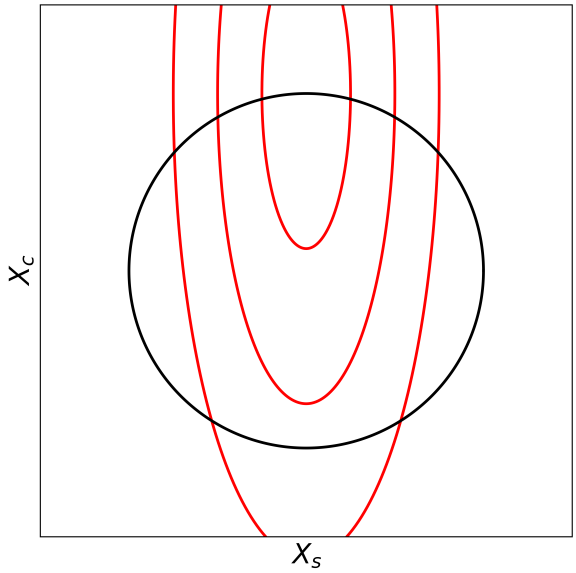
Starting from the large-sample limit discussed above, as the contours grow, the nonlinear structure of the physical subspace becomes visible: varying the model parameter, in this case $\delta_{\mathcal{CP}}$, no longer corresponds simply to slicing out a 1D Gaussian but instead traces a more complex trajectory across the extended surface (see Fig. 5.14b). These nonlinear trajectories relative to the extended surface are how boundaries, commonly treated using the prescription in Ref. [88], manifest in the linearised space. For T2K, since the sensitivity to X_S is much greater than that to X_C, it is immediately clear that the visible curvature (and therefore the effects of boundaries) is not uniform across the parameter space. From this, we can conclude that boundary effects are greatest at $\delta_{\mathcal{CP}} = -\pi/2$ and $\delta_{\mathcal{CP}} = \pi/2$.



(a) Large sample limit.



(b) Boundary example.



(c) Degenerate example.

Figure 5.14.: Appearance of common PMNS parameter features. Red lines correspond to fixed significance contours of a Gaussian measurement (e.g. 1σ, 2σ, 3σ), black path is the physical sub space where $(X_S, X_C) = (\sin\delta_{\mathcal{CP}}, \cos\delta_{\mathcal{CP}})$

Degeneracies

Degeneracies in X_S–X_C space manifest as multiple, disconnected regions of the physical subspace appearing within a given likelihood contour. The regions for which this occurs are degenerate at that significance level. In the extended framework, disconnected intervals in $\delta_{\mathcal{CP}}$ can therefore emerge from a single, smooth Gaussian surface, purely due to the way the physical constraint maps onto the extended likelihood space.

An illustration is shown in Fig. 5.14c. Because the log-likelihood surface is elongated along the X_C direction, both the near side of the physical circle (the side containing the best-fit location) and the far side (the side opposite the physical best-fit point) lie inside the contour, while portions of the circle between them do not, despite being closer in terms of $\delta_{\mathcal{CP}}$ to the best-fit value.

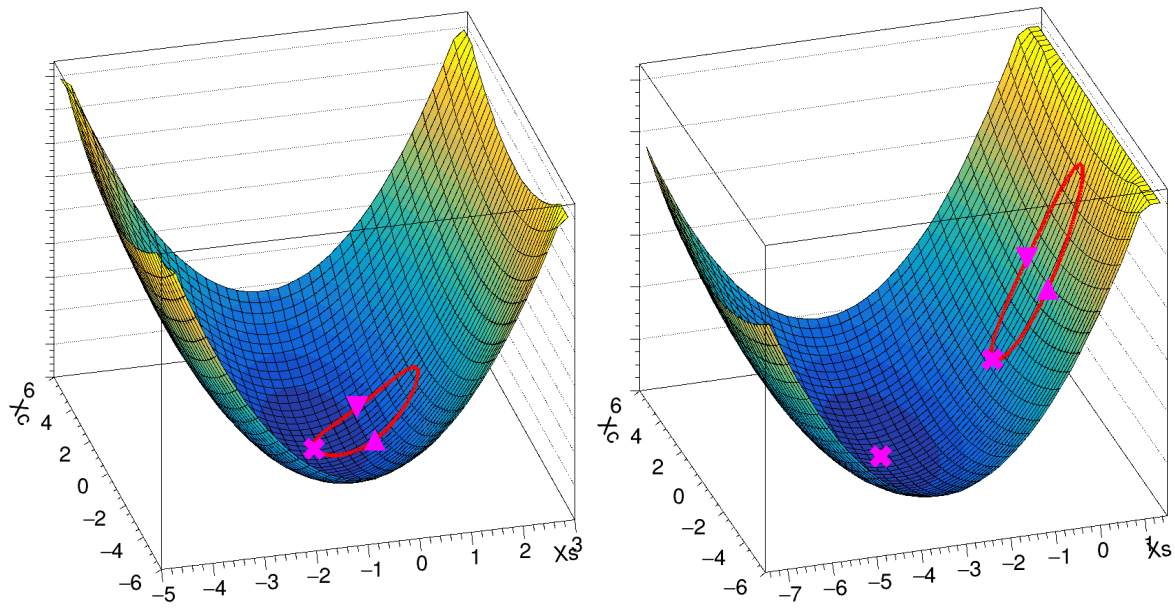
Pulls against boundaries

Recall that when data prefers a region more extreme than the oscillation model can accommodate, the resulting parameter constraint becomes tighter than it otherwise would. This effect is referred to as a pull against a boundary.

In the extended space, this manifests as the best-fit point lying off of the physical circle. When this occurs, the minimum along the physical circle no longer coincides with the global minimum of the extended surface. Instead, the circle occupies a steeper region of the log-likelihood surface and, consequently, the χ^2 distance between the physical minimum on the circle and all other points along it is increased (see Fig. 5.15), thereby generating a tighter constraint.

Remarks

Again, all of this highlights the simplicity of the linearised framework: a single fit is sufficient to characterise the shape of the Gaussian contour. Once this shape is known, one can conceptually shift the contour's centre throughout the extended parameter space. The usual features of the 1D $\chi^2(\delta_{\mathcal{CP}})$ are then easily understood geometrically by considering how the physical circle maps onto the log-likelihood surface.



(a) Extended best-fit on physical circle.

(b) Best-fit off circle (steeper region example)

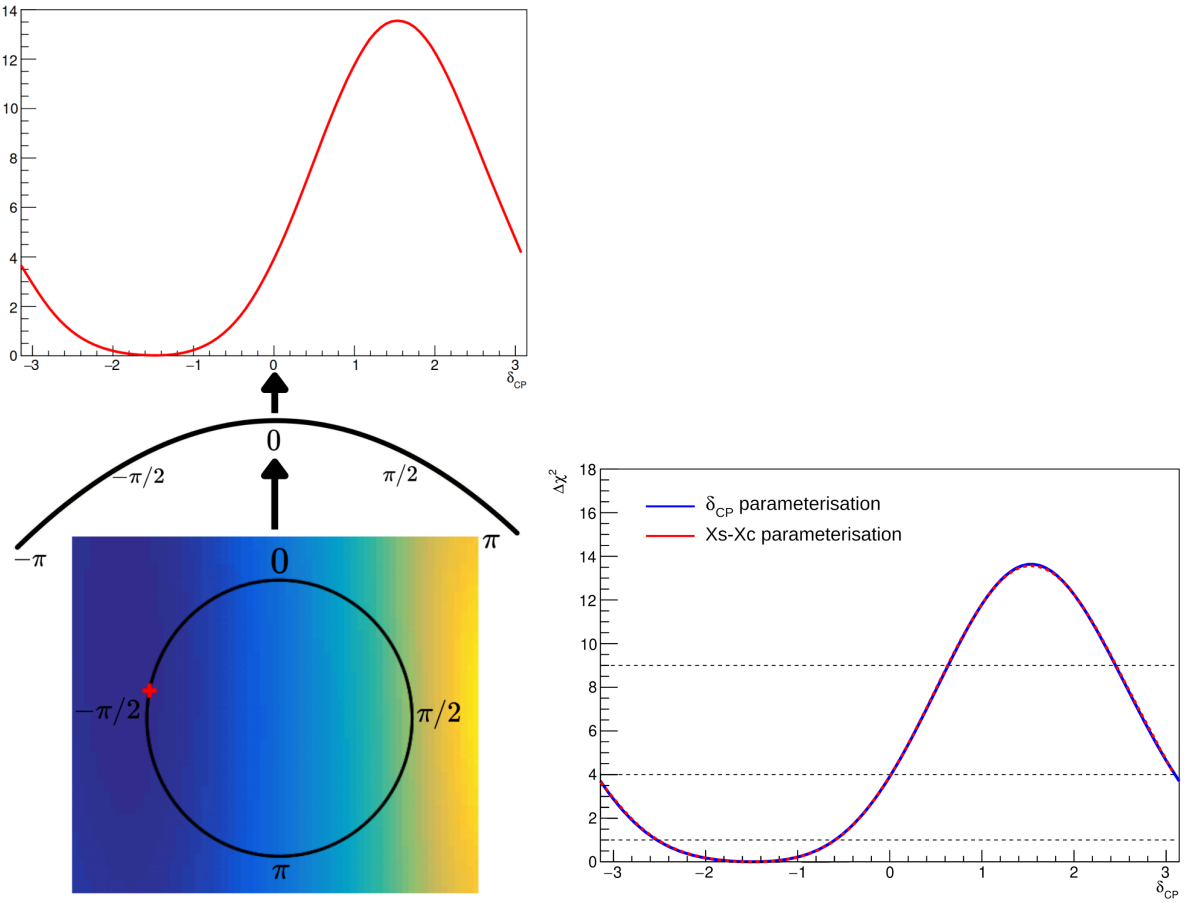
Figure 5.15.: Cartoon illustration of how a pull against a boundary appears. Shown is parabolic log-likelihood surface, with the physical subspace superimposed in red. In (a), the best-fit point lies on the circle, while in (b) it lies off of the circle meaning the circle occupies a much steeper region of the surface.

5.1.4. Recovering the physical constraint

Since, by construction, the extended space contains the physical parameter space as a subset, the constraint on $\delta_{\mathcal{CP}}$ can be exactly recovered by evaluating the extended likelihood surface along the physical circle. This principle was described in Sect. 4.2.6, but as a more explicit example, Fig. 5.16 shows this for an Asimov fit performed in the X_S–X_C parameterisation. The physical best-fit is obtained by finding the lowest point on the circle, and the χ^2 value at that point is used as the reference minimum to be subtracted from all other values. The 1D $\Delta\chi^2(\delta_{\mathcal{CP}})$ function is then recovered by "unwrapping" the circle and plotting the values of the extended surface in one dimension (Fig. 5.16a). We demonstrate that it precisely matches the function obtained by performing the fit directly in the $\delta_{\mathcal{CP}}$ parameterisation (see Fig. 5.16).

5.1.5. Biases

Since all regions of the parameter space are locally equivalent in the extended framework, biases are straightforward to interpret. We illustrate this with the example



(a) Physical constraint recovered by evaluating the extended surface on the physical circle and "unwrapping".

(b) Recovered contour compared to contour produced in $\delta_{\mathcal{CP}}$ parameterisations.

Figure 5.16.: Illustration of how the physical $\delta_{\mathcal{CP}}$ constraint it recovered. This fit was generated using Asimov A values for the non-$\delta_{\mathcal{CP}}$ parameters.

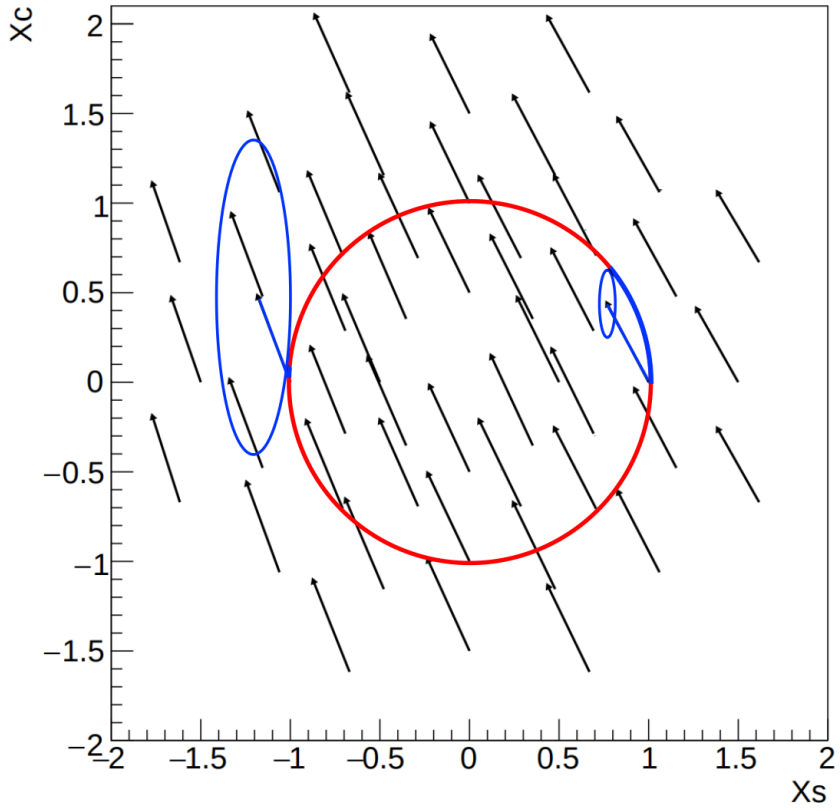
of marginalisation bias arising from marginalising over $\sin^2\theta_{23}$, however, the same reasoning applies more generally to all sources of bias, including those arising from systematic mismodelling in fake-data studies. Fig. 5.17a shows this marginalisation bias for different true values of $\delta_{\mathcal{CP}}$, where the tail of each arrow marks the assumed true value, while the head marks the corresponding best-fit point. Note that the bias vectors are uniform across the parameter space, independent of the true value of $\delta_{\mathcal{CP}}$. By contrast, when the same fits are performed directly in $\delta_{\mathcal{CP}}$, the resulting biases appear irregular, with varying magnitudes and directions and no obvious global structure (see Fig. 5.17b).

Again, the extended space provides insight into the reason for these apparently irregular bias patterns in $\delta_{\mathcal{CP}}$. In a biased fit to $\delta_{\mathcal{CP}}$, the simple shift implied by the extended space is not possible; instead, the best-fit moves to the point on the physical circle corresponding to the lowest physical χ^2, i.e., where the circle is tangent to the smallest ellipse. As an example, Fig. 5.17b highlights biases at $\delta_{\mathcal{CP}} = -\pi/2$ and $\delta_{\mathcal{CP}} = \pi/2$, which differ in both direction and magnitude and can be understood within this framework. To illustrate this, Fig. 5.17a superimposes the physical circle on the bias vectors in the X_S–X_C space. At $\delta_{\mathcal{CP}} = -\pi/2$, the parameter space curves away from the bias direction; because the sensitivity to X_C is poor, the bias in this region cannot move the physical best-fit point far around the circle. In contrast, at $\delta_{\mathcal{CP}} = \pi/2$, although the X_S–X_C bias remains the same, the physical circle curves along the bias direction, so the physical best-fit point shifts significantly around the circle.

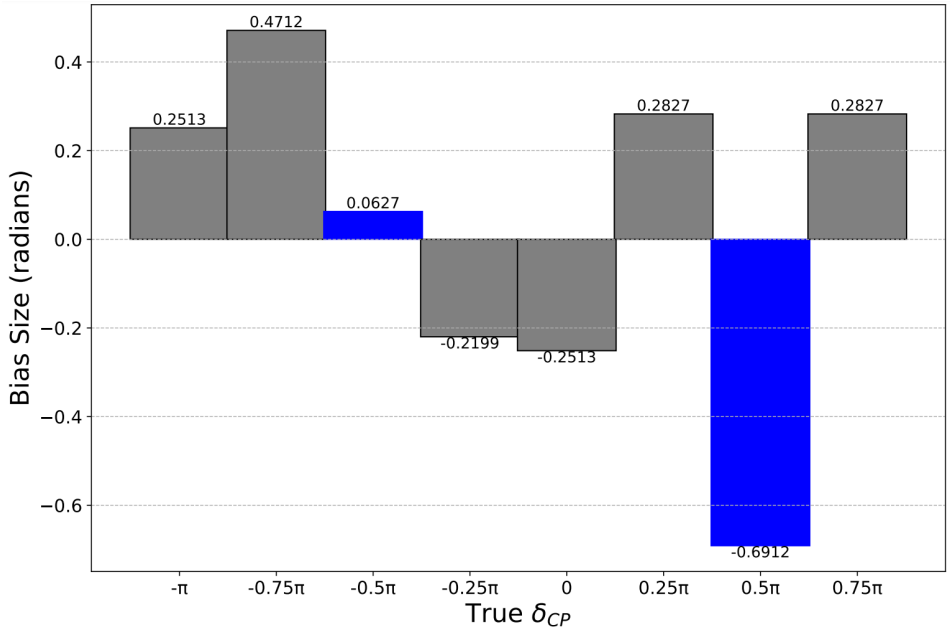
X_S—X_C independent biases

It is also instructive to consider how biases manifest in the $\mathcal{CP}$v and $\mathcal{CP}$c terms when treated independently. Fig. 5.18a shows that there are distinct sources of bias acting on X_S and X_C which, when expressed solely in terms of the $\delta_{\mathcal{CP}}$ parameterisation, would be blended together and largely obscured. By separating these contributions, biases that would otherwise appear highly abstract can be understood more transparently.

For example, the origin of the bias in X_S can be readily understood as arising from the octant degeneracy. An Asimov fit necessarily assumes a particular octant for θ_{23}; however, for any chosen Asimov point, there exists a degenerate value of $\sin^2 2\theta_{23}$,



(a) Example of uniform (marginalisation) bias in the extended space. The tails of the black arrows indicate the true points, while the arrowheads correspond to the best-fit point. The red circle represents the physical subspace. Biases at $\delta_{\mathcal{CP}} = -\pi/2$ and $\delta_{\mathcal{CP}} = \pi/2$ are highlighted to show that, due to the sensitivity shape and bias direction, they correspond to very different apparent biases, which in turn correspond to very different physical biases once projected onto the circle.



(b) Marginalisation in $\delta_{\mathcal{CP}}$, using the same true points as those shown on the physical circle above.

Figure 5.17.: Comparison of marginalisation bias from the same fit shown in the $\delta_{\mathcal{CP}}$ parameterisation and in the X_S–X_C space.

which is also marginalised over in the fit. The appearance probability takes the form

$$P_{\mu e} \sim A \sin^2 \theta_{23} - (B \sin 2\theta_{23}) X_S \,. \tag{5.2}$$

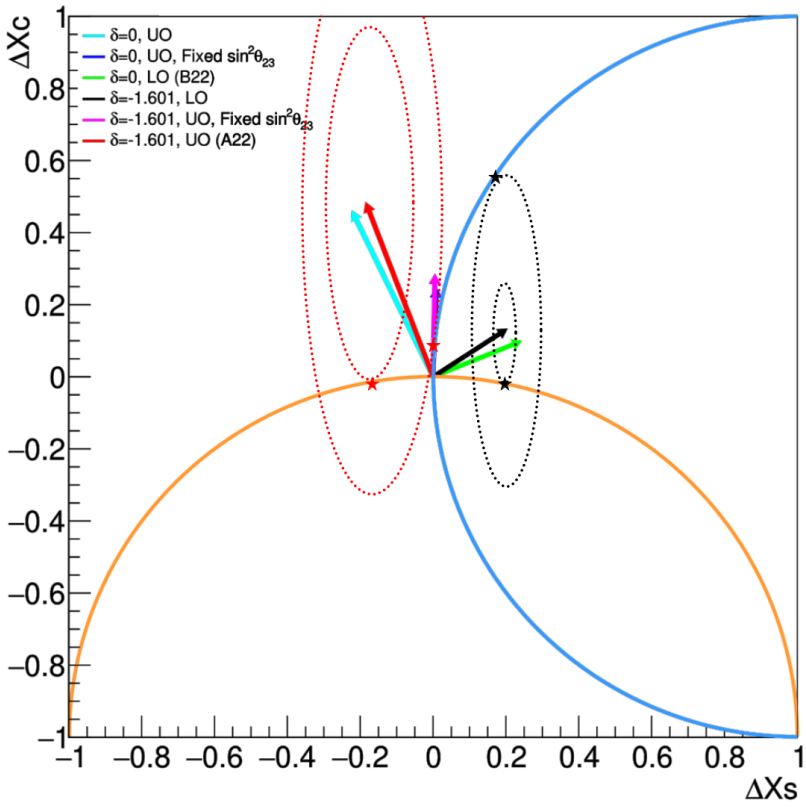
If, for instance, the Asimov data set is generated assuming the true value lies in the upper octant, then marginalisation points near the lower octant leaves $\sin 2\theta_{23}$ largely unchanged while reducing $\sin^2 \theta_{23}$. To reproduce the required event rate, the fitter therefore favours smaller values of X_S to compensate. Conversely, if the true value lies in the lower octant, the opposite occurs and the fitter is biased towards larger values of X_S. This relationship is reversed for the RHC sample; however, because the FHC samples dominate statistically, the net bias is determined primarily by the FHC contribution.

A strength of the X_S–X_C construction is the easy visual interpretation. However, there remains a choice in how to scale the axes, which becomes particularly important when interpreting biases. As in Fig. 5.18a, one option is to keep X_S and X_C on an equal scale so that the circular appearance of the physical subspace is preserved. In this approach, distances around the circle correspond to equal increments in $\delta_{\mathcal{CP}}$ everywhere; the tradeoff being that the unequal sensitivities to X_S and X_C must be taken into account when projecting onto the closest point. The alternative option is to stretch the space so that the error is circular and the notion of the closest point retains its usual Euclidean meaning, but now equal steps around the physical circle no longer correspond to equal increments in $\delta_{\mathcal{CP}}$ which can make it difficult to interpret how far the bias moved around the circle (see Fig. 5.18b). Given that both approaches have value, we present both.

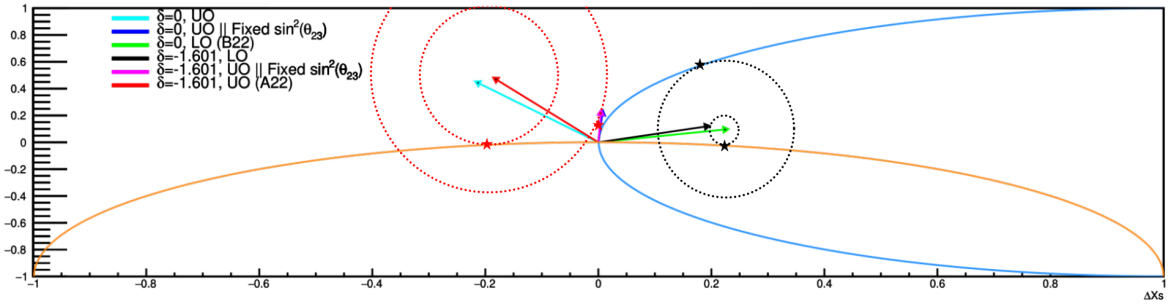
5.1.6. Poisson-statistics corrections

For linear models, the Fisher information acquires a parameter dependence through the Poisson mean,

$$\mathcal{I}(\theta) = \frac{a}{a\theta + b}. \tag{5.3}$$



(a) X_S and X_C on equal scale.



(b) Stretched X_C scale so that sensitivity is circular.

Figure 5.18.: Absolute size of the marginalisation bias for different true points, superimposed on the local physical subspace at $\delta_{\mathcal{CP}} = -1.601\pi$ and $\delta_{\mathcal{CP}} = 0$. For an assumed upper octant (UO) the bias points to the left, while for a lower octant (LO) assumption the bias points to the right. With a fixed octant there is no X_S bias. Stars indicate the projection onto the physical circle (red: true UO at $\delta_{\mathcal{CP}} = -1.601\pi$; black: true LO at $\delta_{\mathcal{CP}} = -1.601\pi$).

This means that the observed variation in size of ellipses should be inversely proportional to the parameter values,

$$\mathcal{I}^{-1}(\theta) = \frac{a\theta + b}{a} \sim \theta. \tag{5.4}$$

In the X_S–X_C model, this behaviour can be expressed as

$$\mathcal{I}^{-1}(\theta) \sim \Sigma_{Total} = \Sigma_0 + X_S\Sigma_s + X_C\Sigma_c, \tag{5.5}$$

where Σ_0 denotes the baseline covariance, and Σ_s and Σ_c encode the contributions associated with X_S and X_C, respectively. Since the $\mathcal{CP}$c term is subdominant, we expect this Poisson statistics effect to be largely driven by Σ_s. We demonstrate this by plotting the average predicted event rate in the X_S–X_C plane (Fig. 5.19a), which shows the event rate is highest at the most negative values of X_S and decreases with increasing X_S (Figs. 5.19b, 5.19c show that RHC samples have the opposite relationship compared to the FHS samples, however because of the dominant statistics of FHC, the overall trend is determined by these samples).

If we also plot the corresponding X_S and X_C variances (the diagonal elements of the inverse Hessian) as a function of X_S, we can directly relate the ellipse size to Poisson statistics. Indeed, as shown in Fig. 5.20, the regions of highest variance coincide with those of highest event rate and decrease approximately linearly as the event rate decreases (that is, as X_S increases, as previously established).

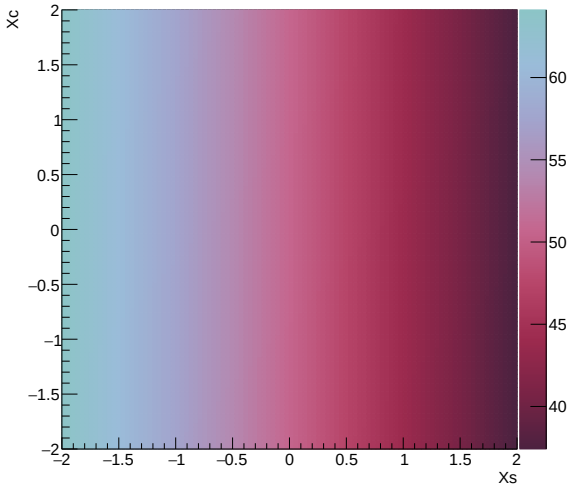
As a final proof of this, we produced Asimov fits where either X_S is fixed and X_C is varied, or X_C is fixed and X_S is varied (Fig. 5.21). There is comparatively little variation in the contour sizes when X_S is fixed compared to when X_C is fixed, which is entirely consistent with our interpretation that Poisson statistics are responsible for the observed size differences.

Given that this effect is driven almost entirely by the $X_S\Sigma_s$ term, we absorb the Σ_c contribution into the baseline covariance, yielding

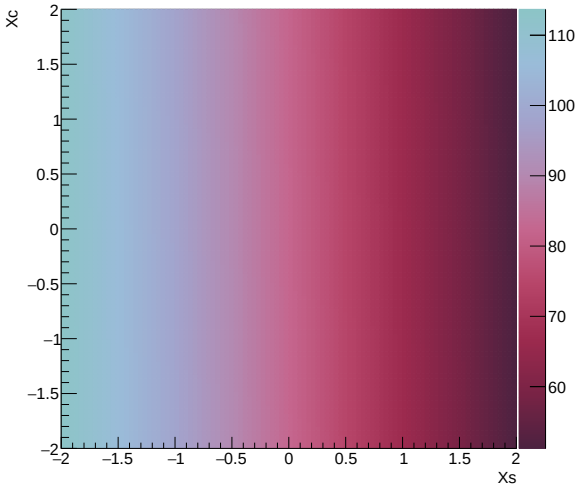
$$\Sigma_{\text{Total}} \sim \Sigma_0 + X_S\Sigma_s. \tag{5.6}$$

We can then compensate for the resulting scaling behaviour by re-expressing Σ_s as

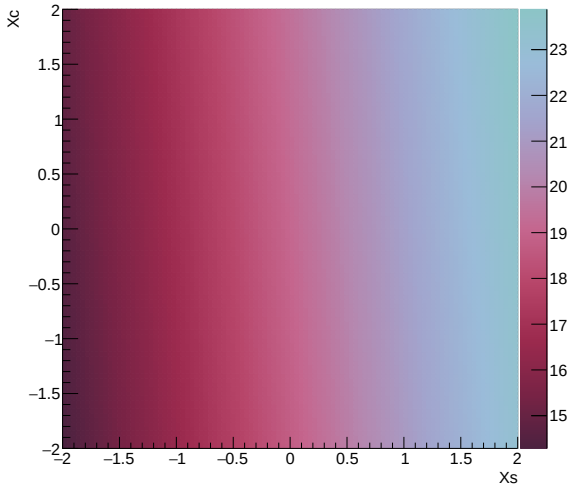
$$\Sigma_s = \frac{\Sigma_{\text{Total}} - \Sigma_0}{X_S}, \tag{5.7}$$



(a) Combined FHC-RHC 1R*e* samples.

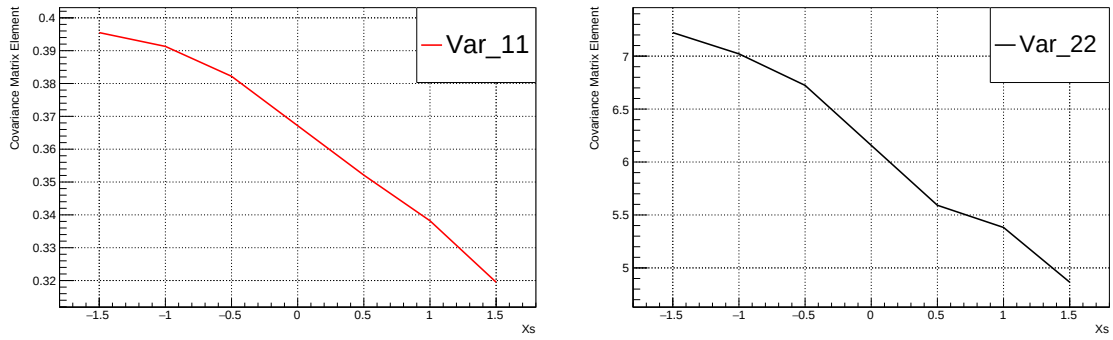


(b) FHC 1R*e* sample.



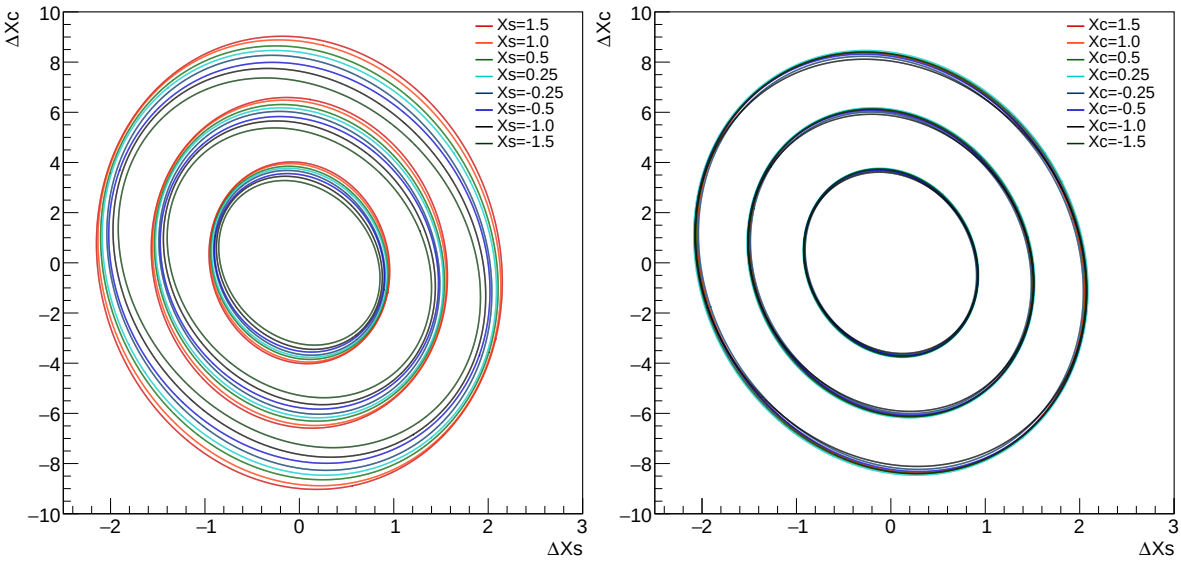
(c) RHC 1R*e* sample.

Figure 5.19.: Average event rates for different samples in the X_S–X_C plane. The event rate is highest at the most negative values of X_S and decreases as X_S increases. The RHC samples exhibit the opposite dependence compared to the FHC samples; however, because the FHC samples dominate statistically, the overall trend is driven by their behaviour.



(a) X_S variance. **(b)** X_C variance.

Figure 5.20.: X_S and X_C variance as a function of X_S, taken from the diagonal elements of the Hessian matrix, calculated at local best-fit point.



(a) Fixed X_C. **(b)** Fixed X_S.

Figure 5.21.: Comparison of contours from Asimov fits with either X_S or X_C held fixed. Highlight of the variance in contour size is mostly driven by X_S.

so that the effective covariance at each true $\delta_{\mathcal{CP}}$ point becomes

$$\Sigma(X_S) = \frac{\Sigma_{\text{Total}}(X_S) - \Sigma_0}{X_S} + \Sigma_0, \tag{5.8}$$

where $\Sigma_{\text{Total}}(X_S)$ is the covariance matrix obtained assuming a true value X_S. Using $\Sigma(X_S)$ from Eq. (5.8) as the covariance matrix produces the contours shown in Fig. 5.11b, which demonstrate substantially improved agreement.

5.2. Negative event rates

Given the extended oscillation probability does not enforce unitarity, negative probabilities can occur at extreme distances from the physical circle. While not inherently problematic, these require careful consideration, as such regions can lead to negative predicted event rates for which no valid likelihood function can be defined.

Of course, since Eq. 4.8 reduces to the physical oscillation probability on the physical circle (when $\rho = 1$), there is no issue with negative predicted events on the circle. Additionally, because the interpolating coefficients $\frac{1+\rho}{2}$ and $\frac{1-\rho}{2}$ are both positive for $|\rho| \leq 1$, and $P_{\alpha\beta}(\delta, \vec{\theta}), P_{\alpha\beta}(\delta + \pi, \vec{\theta}) \in [0, 1]$, there are no negative probabilities in the interior of the physical circle. Thus, negative probabilities can only arise for $|\rho| > 1$. Specifically, the extended probability

$$P'_{\alpha\beta}(\rho, \delta, \vec{\theta}) = \frac{1+\rho}{2} P_{\alpha\beta}(\delta, \vec{\theta}) + \frac{1-\rho}{2} P_{\alpha\beta}(\delta + \pi, \vec{\theta}) < 0 \tag{5.9}$$

if

$$\rho < -\frac{P_\delta + P_{\delta+\pi}}{P_\delta - P_{\delta+\pi}}, \quad P_\delta > P_{\delta+\pi}, \quad \text{or} \quad \rho > \frac{P_\delta + P_{\delta+\pi}}{P_{\delta+\pi} - P_\delta}, \quad P_{\delta+\pi} > P_\delta. \tag{5.10}$$

In practice, for any given reconstructed bin there are typically sufficient $P_{Ext} > 0$ contributions, along with mis-reconstructed ν_e events and non-ν_e backgrounds (to which this extended "probability" is not applied), to compensate. However, if too many negative-probability events accumulate in a single bin, the predicted rate for that bin can become negative. For these extremely unphysical throws, the negative log-likelihood is set to infinity, so when exponentiated, they do not contribute to the marginal likelihood other than by increasing the sample size by one.

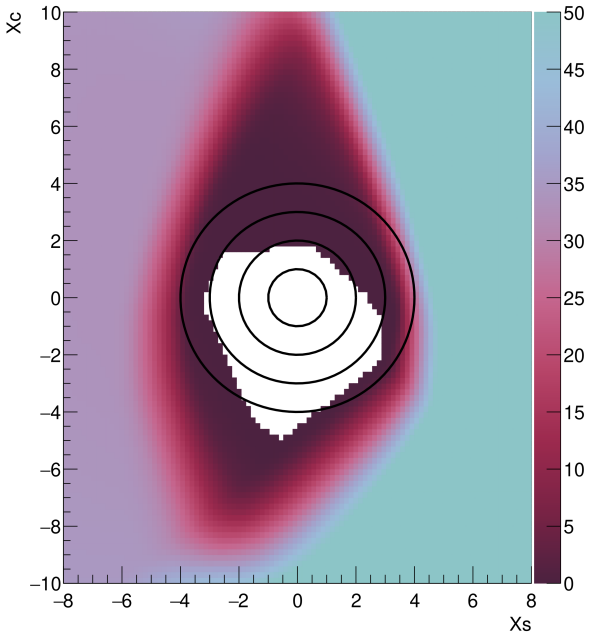
Nevertheless, we need to be critically aware of where these instances occur, since they will distort the log-likelihood surface if they are too prevalent and thus significantly degrade the Gaussian approximation. Fig. 5.22 shows the frequency of negative predicted events across the X_S–X_C plane, expressed as a percentage of total throws. As expected, no negative events are observed within or on the physical circle. Moreover, their occurrence remains below 5% within at least three radii of the physical subspace.

Fig. 5.23 shows their effect on the likelihood surface: with T2Ks current statistics, past 2σ the agreement with Gaussian surface begins to breakdown, and beyond 3σ catastrophically diverges. For this reason, we suggest that analysis should, for the time being, limit its considerations to 2σ effects. Note that because the regions where these occur are fixed, this is not an inherent limitation of the framework. With more statistics, the ellipses will shrink, and the significance levels affected by these negative event rates will shift to higher values. In practice, the significance level at which the extended parametrisation breaks down because of them will typically coincide with the level at which the experiment has no sensitivity to $\delta_{\mathcal{CP}}$ in any case.

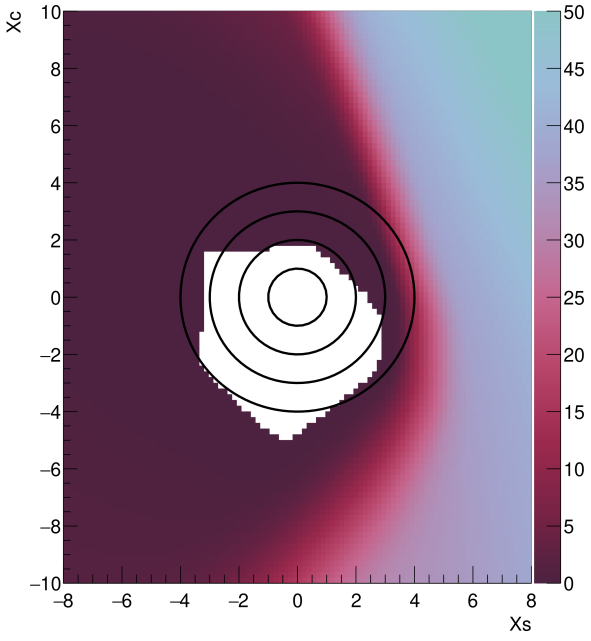
5.3. Precision measurements in extended space

If, at some point in the future, leptonic $\mathcal{CP}$-violation is confirmed by the exclusion of $\delta_{\mathcal{CP}} \neq 0, \pi$, interest will shift to determining the size of these $\mathcal{CP}$-violating effects through precision measurements of $\delta_{\mathcal{CP}}$. As outlined by the Snowmass working group [94], this will have important theoretical implications (where they set a target precision of 5° for $\delta_{\mathcal{CP}}$).

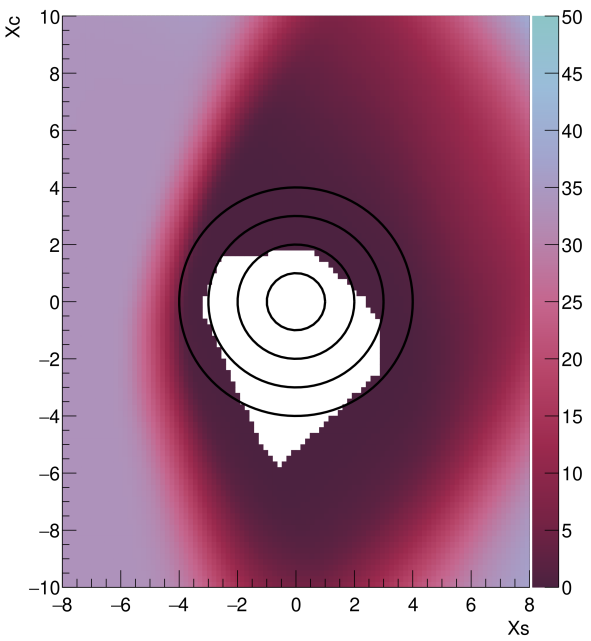
Independent of simpler sensitivity studies, linear oscillation models can also be useful for developing intuition in this context. Since sensitivity to the $\mathcal{CP}$c term comes from the energy distortion, T2K's narrow beam configuration, while suited to measurements of $\mathcal{CP}$v, will be unsuited for precision measurements of $\delta_{\mathcal{CP}}$ if its true value is near maximal $\mathcal{CP}$-violation. The usual explanation for this is that $\sin\delta_{\mathcal{CP}}$ is very flat in this region, while $\cos\delta_{\mathcal{CP}}$ (to which these experiments have comparatively poor sensitivity) changes steeply. In the extended linearised framework, this is easy to understand. If the experimental sensitivity to the $\mathcal{CP}$v term is many times greater than that to the $\mathcal{CP}$c term (so the ellipse is much taller than it is wide), and the true value lies near $\delta_{\mathcal{CP}} \sim -\pi/2$, then as the statistics increase the trajectory of the physical circle increasingly aligns with the X_C axis.



(a) Combined FHC-RHC 1Re samples.

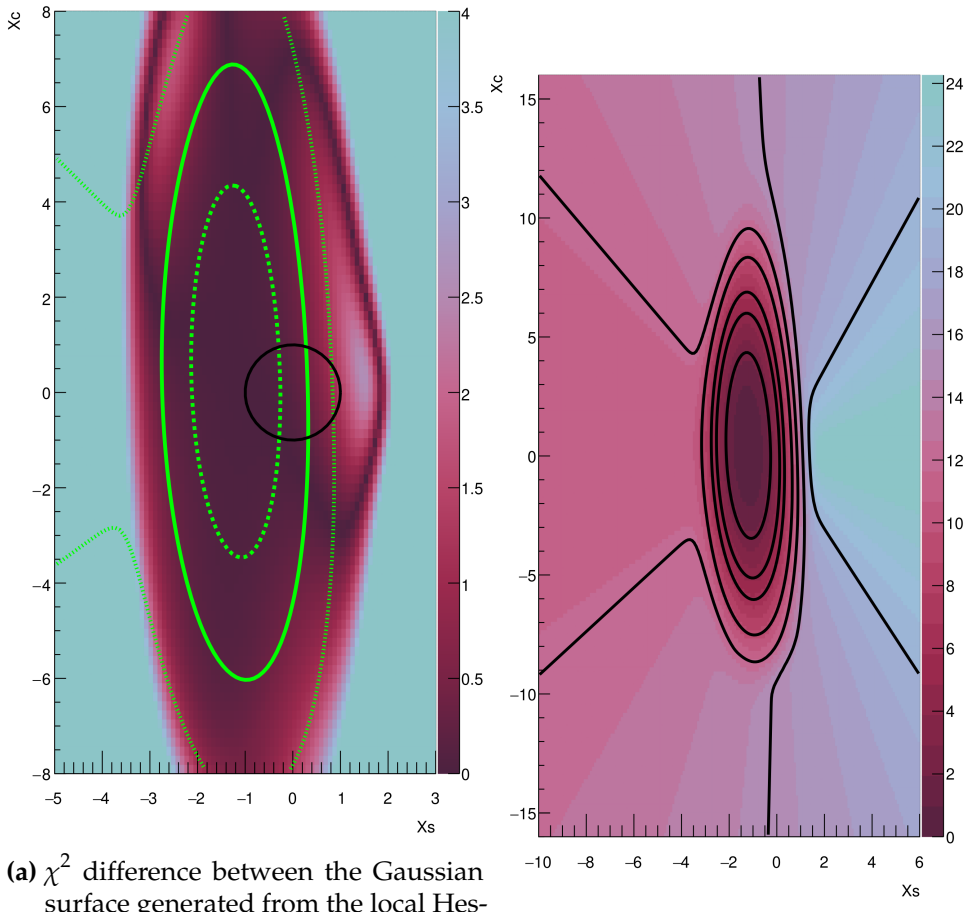


(b) FHC 1Re sample.



(c) RHC 1Re sample.

Figure 5.22.: Percentage of throws that result in negative predicted events across the X_S–X_C plane. Black circles with radii 1, 2, 3, and 4 are superimposed for reference.



(a) χ^2 difference between the Gaussian surface generated from the local Hessian (true Gaussian approximation) and the surface obtained from the Asimov fit.

(b) χ^2 surface from Asimov fit showing divergence far away from the physical region.

Figure 5.23.: Effect of negative predicted events on the χ^2 surface.

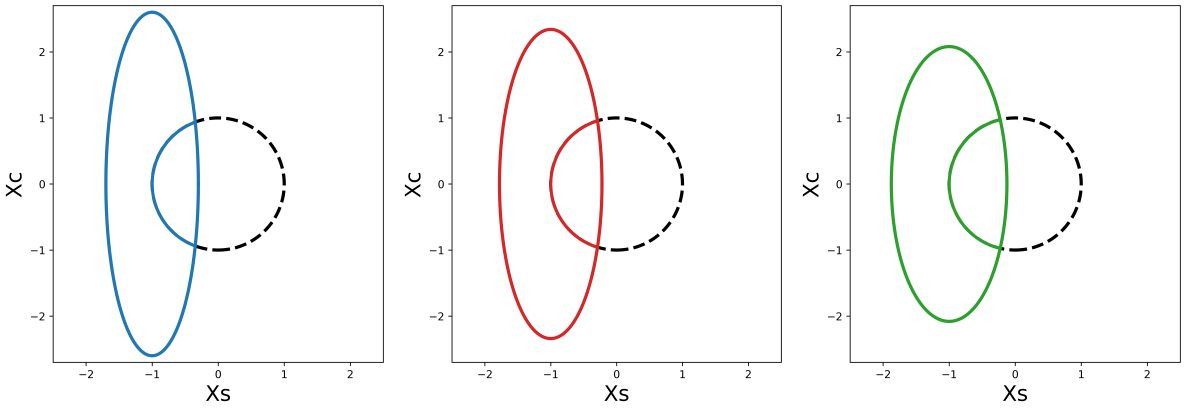
To show this, we look at several ellipse shapes (shown in Fig. 5.24) and consider the resulting fraction of the physical circle overlapping (used as a simple proxy for $\delta_{\mathcal{CP}}$ sensitivity) as they are incrementally shrunk to 10% of their original size. We observe that the most T2K-like ellipse (Fig. 5.24a) initially excludes the largest fraction of the circle, but as statistics increase and the ellipses shrink, the physical circle becomes increasingly perpendicular to the X_C axis. Thus, when the ellipses are smaller, the original relationship flips, and the wider, shorter ellipses (such as might be expected for a wide band experiment like DUNE), now have the least overlap. If, however, $\mathcal{CP}$-violation is small (i.e. $\delta_{\mathcal{CP}}$ is near 0 or π), a T2K- or HK-like configuration will continue to provide superior precision on $\delta_{\mathcal{CP}}$.

For completeness, we also fit several toy spectra of varying widths, obtained by reweighting the nominal T2K flux. Fig. 5.25 shows that the wider spectra, with more events above and below the oscillation maximum, have greater sensitivity to X_C. The X_S constraint is also improved slightly due to the overall increase in events. Therefore, wider-band beams produce narrow ellipse in the X_C direction which are better for precision measurement if $\delta_{\mathcal{CP}}$ is near maximal.

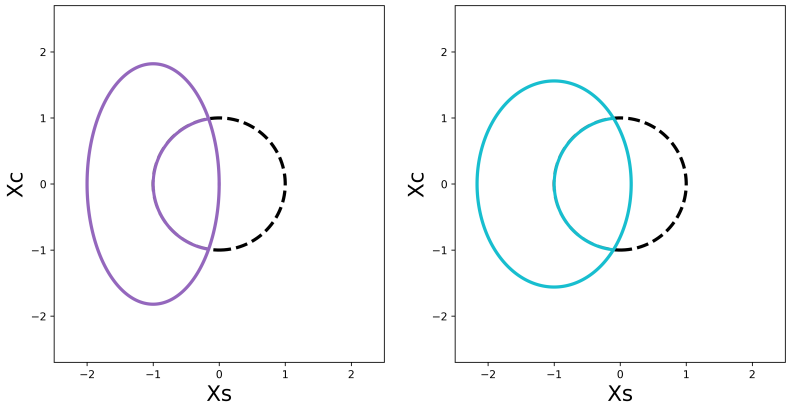
5.4. Utility for fake data studies

As a brief reminder, Fake Data Studies (FDS) are used to assess the robustness of fits against non-parameterisable systematics, such as those that arise from discrete model choices. They are performed by fitting simulated data generated under an alternative model to the nominal model. The idea is that if the tension between the models does not substantially change the contours or the reported sensitivity to $\mathcal{CP}$v, then the model differences are unimportant. Since contours in the linearised parameterisation are simple, FDS can be greatly simplified. This is because the impact of an alternative model can be fully specified by the changes in the five parameters that describe the ellipse.

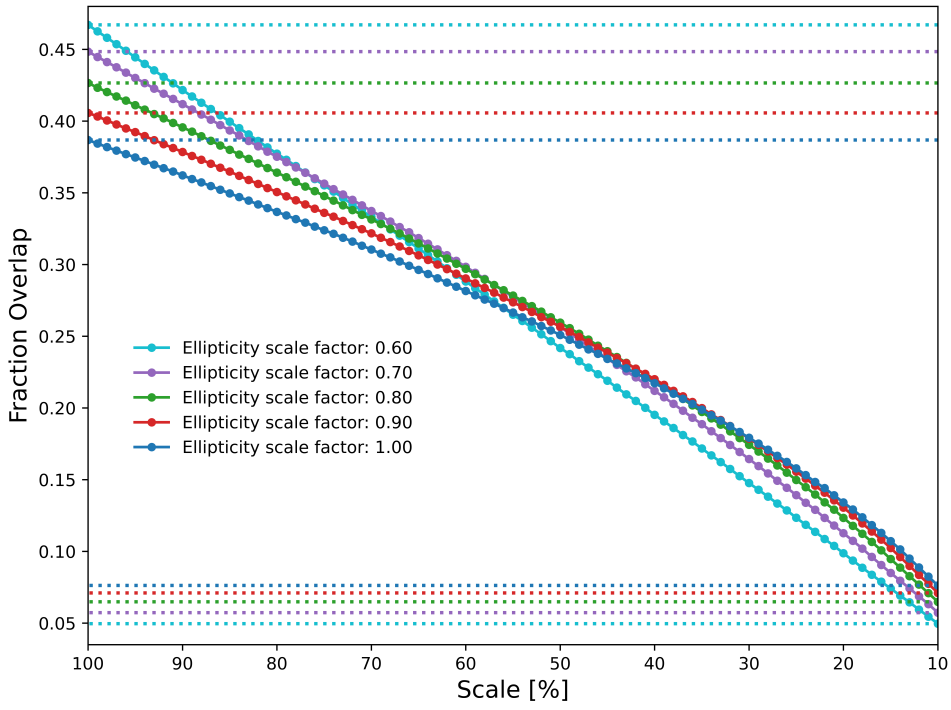
One common example where the standard parameterisation complicates FDS is in pulls against boundaries (PAB). Because PAB effects shorten intervals, distinguishing between a PAB and a genuine increase in statistical sensitivity is difficult. The point is that these are two separate effects in the linearised space (i.e. the former corresponds to a change in bias, while the latter corresponds to a change in ellipse size). Fig. 5.26



(a) T2K-like (scale factor: 1) **(b)** Scale factor: 0.9 **(c)** Scale factor: 0.8



(d) Scale factor: 0.7 **(e)** Scale factor: 0.6



(f) Fraction of the physical circle overlapped by the ellipse as a function of size (down to 10% of the original). The dotted horizontal lines indicate the initial overlap fraction and the final overlap fraction.

Figure 5.24.: Toy study illustrating how sensitivity scales for different ellipse shapes at true value of $\delta_{\mathcal{CP}} = -\pi/2$.

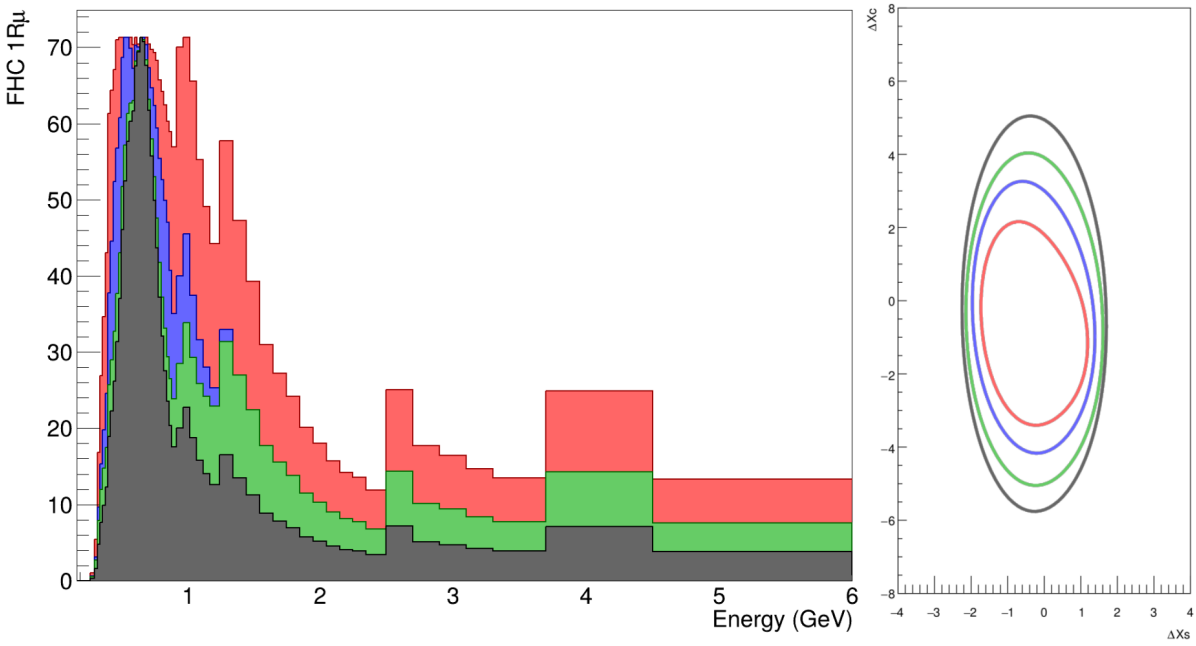


Figure 5.25.: Toy study illustrating how the effect of fitting spectra of varying broadness affects the ellipse shape. Left: example FHC 1Rμ spectra (with the same scaling applied to the other samples). Right: the corresponding ellipse obtained from fitting these spectra.

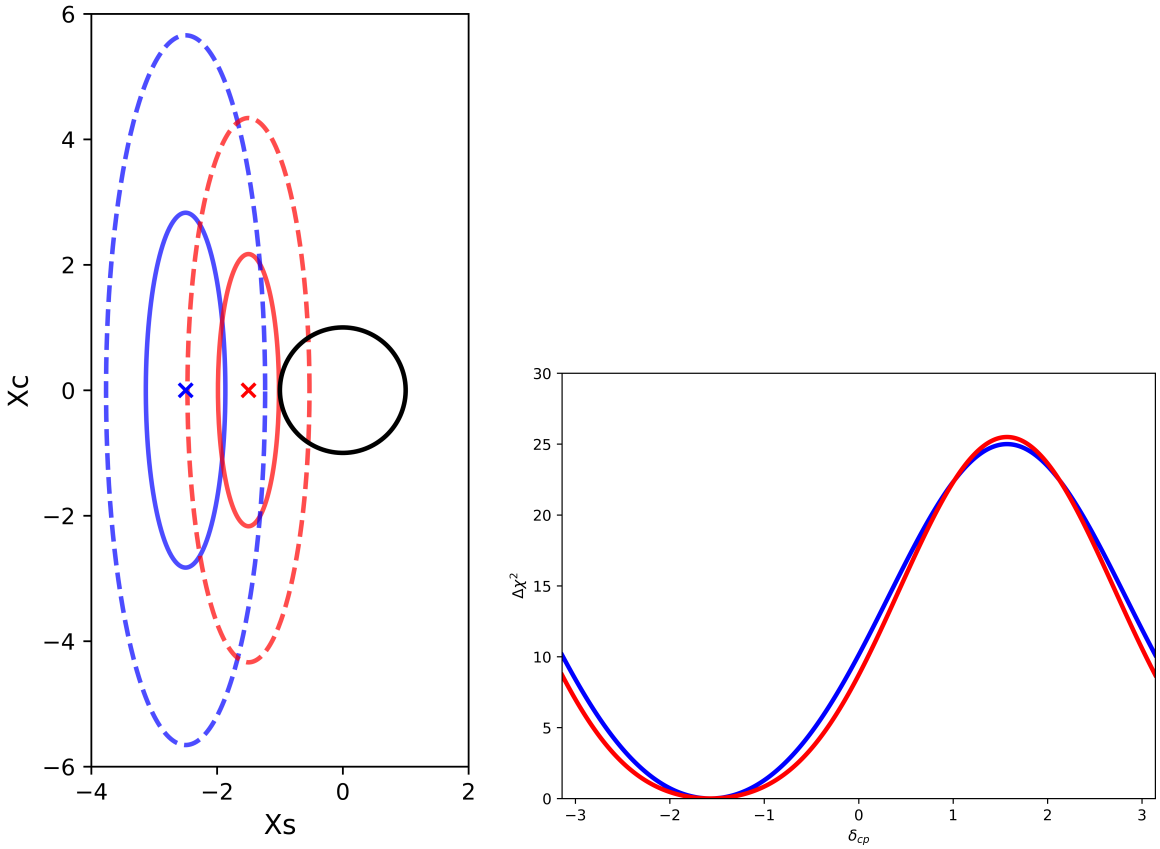
illustrates this by showing how two very different ellipses, corresponding to very different sensitivities, can produce similar looking contours in $\delta_{\mathcal{CP}}$.

5.4.1. Simulated FDS's

Rigorously performing FDS for T2K's non-parametric model variations is quite involved and probably unnecessary for a proof of concept. Instead, we will consider two simplified FDSs inspired by real T2K FDS.

- A low-momentum excess in the 1Re+1d.e sample, produced by scaling the lowest-momentum bins (< 200 MeV) by a factor of 2.
- A mismodelling of the 2p2h rate, with data produced using the 2p2h-ν_μ normalisation parameter set to three times its nominal value.

The impact of these alternative models on the $\delta_{\mathcal{CP}}$ contours is shown in Fig. 5.27. In this space, the FDS exhibits complicated behaviours that are difficult to interpret and



(a) Contours in X_S–X_C parameterisation appears very different.

(b) Contours in $\delta_{\mathcal{CP}}$ parameterisation appears much more similar.

Figure 5.26.: Illustration of how very different inherent sensitivities (indicated by the ellipse sizes) can produce similar $\Delta\chi^2$ contours due to pulls against boundaries.

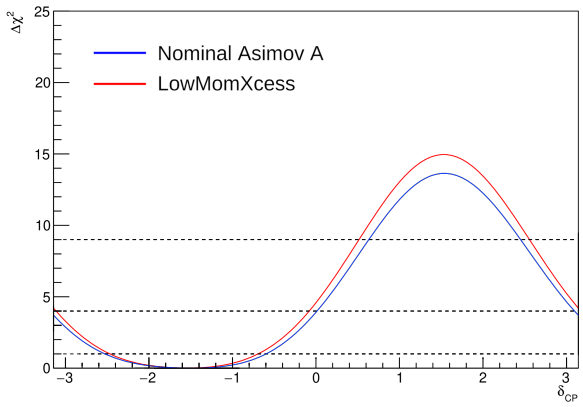
appear to depend strongly on the true value of $\delta_{\mathcal{CP}}$. In particular, for the 2p2h case the differences are dramatic.

Repeating the study in the X_S–X_C parametrisation clarifies the source of these changes. The changes in the X_S–X_C ellipses between fits using data generated with the nominal and alternative models are shown in Figs. 5.28a and 5.28b. In both, the difference in these ellipses is very small, with a near-identical X_S sensitivity and only a slightly poorer X_C sensitivity. On their own, these changes would not translate into the observed $\delta_{\mathcal{CP}}$ contour changes of Fig. 5.27.

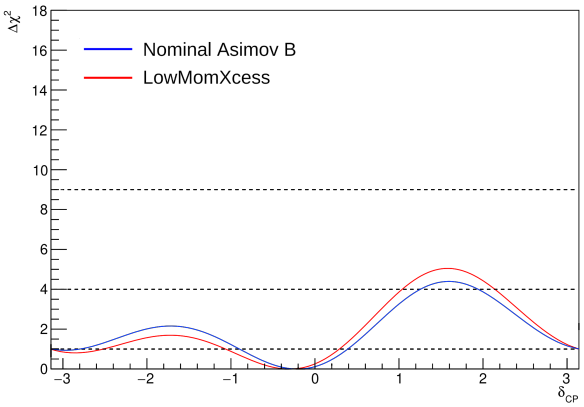
However, examining the biases in this space (Fig. 5.28c) reveals significant shifts, particularly for the 2p2h case. We therefore conclude that the observed changes in the $\delta_{\mathcal{CP}}$ contours are predominantly driven by pulls against boundaries (PAB).

To again emphasise the uniformity of this framework, a vector plot (Fig. 5.29) showing the associated bias for the 1Re+1d.e. low-momentum FDS has been generated in this space. One nuanced point here is that there is a slight true-value dependence on the arrow length (this was also true of Fig. 5.17a, but it was glossed over at the time so as not to detract from the main point). Because this seems entirely to work as a function of X_S, we attribute the behaviour to Poisson statistics. In any case, the bias on and around the unit circle differs negligibly, meaning that a single fake-data study is sufficient to characterise the bias for any value of $\delta_{\mathcal{CP}}$.

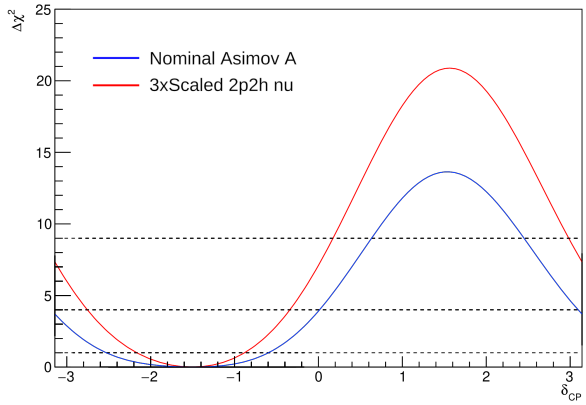
This, complete with the true parameter independent ellipse, greatly simplifies both the generation and interpretation of FDS, which are currently a major fraction of the effort in each update of T2K analyses. Moreover, because the physical likelihood can always be recovered by evaluating on the physical subspace, no information is lost. In this sense, fits performed in the extended space are more informative as they carry, in addition to the conventional constraint, information that the standard parametrisation does not.



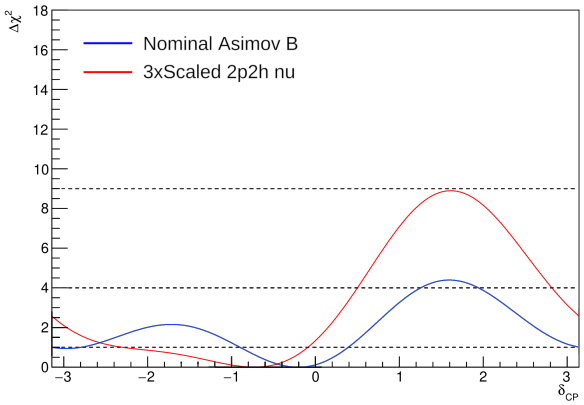
(a) Low-momentum excess FDS at Asimov A



(b) Low-momentum excess FDS at Asimov B

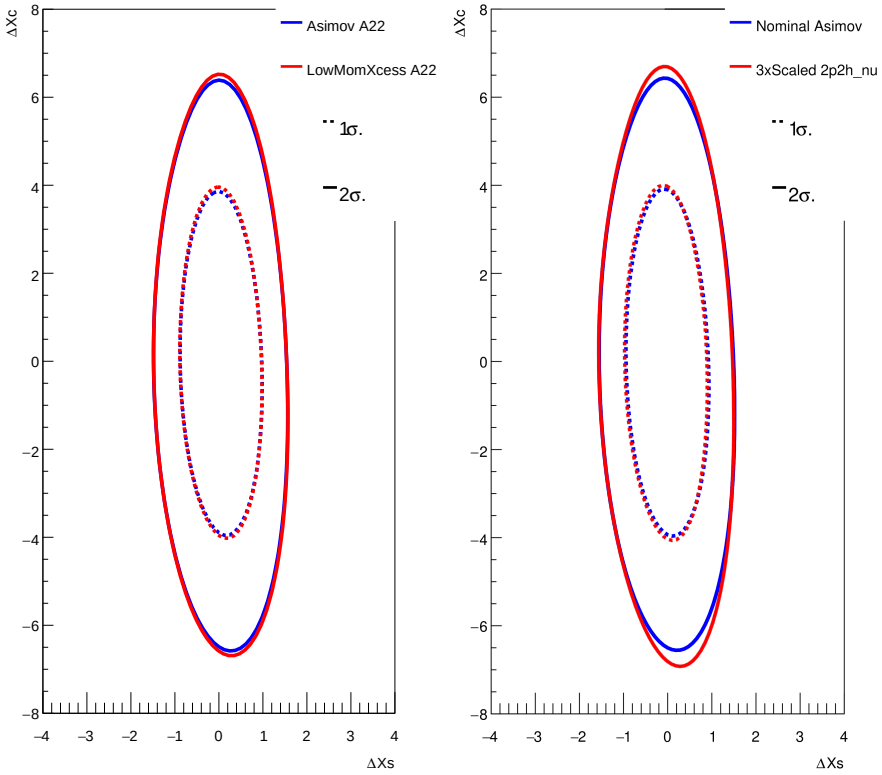


(c) 2p2h normalisation FDS at Asimov A

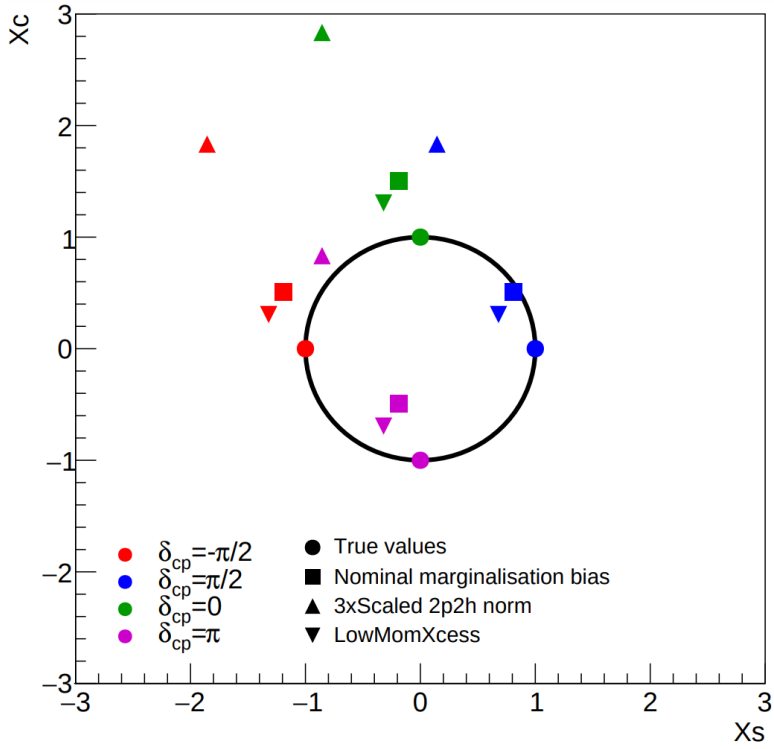


(d) 2p2h normalisation FDS at Asimov B

Figure 5.27.: 1D likelihood contours for the 1Re+1d.e. low-momentum excess and the 2p2h normalisation FDS, evaluated in the conventional $\delta_{\mathcal{CP}}$ parametrisation. Results are shown for Asimov A ($\delta_{\mathcal{CP}} = -1.601$) and Asimov B ($\delta_{\mathcal{CP}} = 0$).



(a) Low-momentum excess FDS. **(b)** 2p2h FDS.



(c) Best-fit points shown to demonstrate shift due to FDS.

Figure 5.28.: Comparison of the 1 and 2 σ contours and biases for the simplified FDS in X_S–X_C space. The contours and biases are compared to those obtained using the nominal Asimov. For the 2p2h FDS there was a slight covariance difference which was corrected to show change in the contours only.

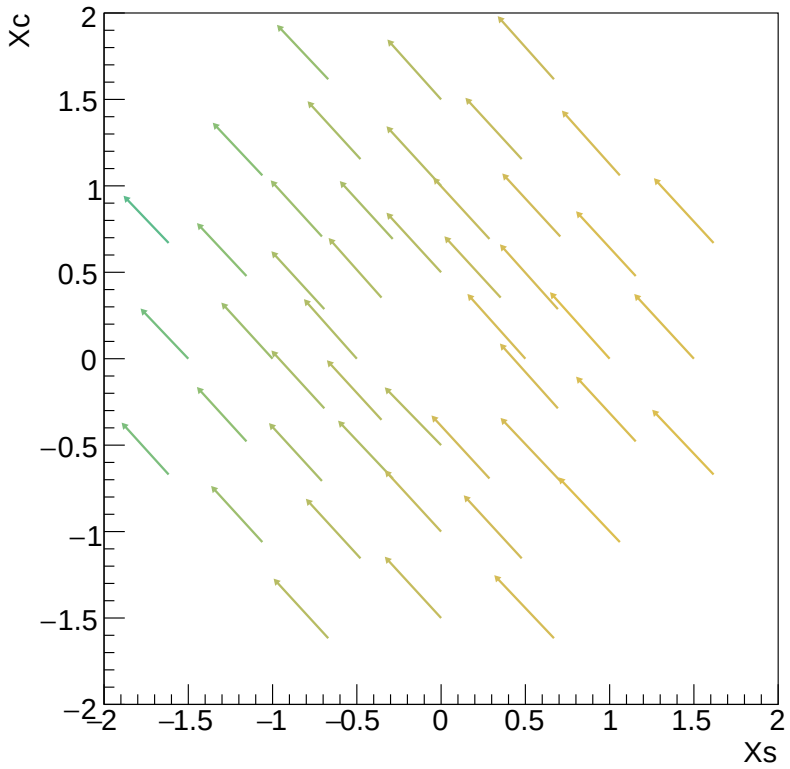


Figure 5.29.: Bias in the 1Re+1d.e low-momentum excess across the X_S–X_C plane. The base of the arrow indicates the true point, and the head of the arrow indicates the best-fit point. A slight dependence on the true X_S is visible in the arrow sizes, as shown by the colour gradient.

Chapter 6.

Reproducing the OA21 results

As a final test of the method, we reproduce the 2021 Oscillation Analysis (OA21) using the linearised X_S–X_C parametrisation. The application to the real data set should not affect the results when constrained to the physical circle $X_S^2 + X_C^2 = 1$, indeed reproduction of the existing likelihood should be guaranteed by the construction and is an important closure test. But, the expanded space fit allows us to better understand the T2K results and exactly how sensitivity to $\delta_{\mathcal{CP}}$ is generated. This is not the latest T2K dataset and systematic model; however, at the time this analysis started, it was the most recent stable OA. For a description of this OA, see the Neutrino 2022 talk [95], and for a more detailed, comprehensive description see Ref. [72].

6.1. SK data and Monte Carlo

The OA21 results correspond to data collected during the first ten SK run periods (2010–2020), with a total exposure of 19.7×10^{20} POT in FHC mode and 16.3×10^{20} POT in RHC mode. A detailed breakdown of beam power and POT collected for each run is shown in Fig. 6.1. The full dataset comprises 94 1Re and 318 1Rμ events in FHC mode, 16 1Re and 137 1Rμ events in RHC mode, and 14 Re+1d.e and 134 ν_μ CC1π MR events (this was also the first analysis to include this sixth ν_μ CC1π MR sample).

As explained previously, P-Theta first generates sets of unoscillated FD Monte Carlo predictions by varying the systematic parameters. The PMNS parameters are then constrained by applying oscillation probabilities as weights to these precalculated predictions. The inputs for this, including the ND constraints, are provided by the T2K-SK working group (for a detailed description, see Ref. [96]). We note that this

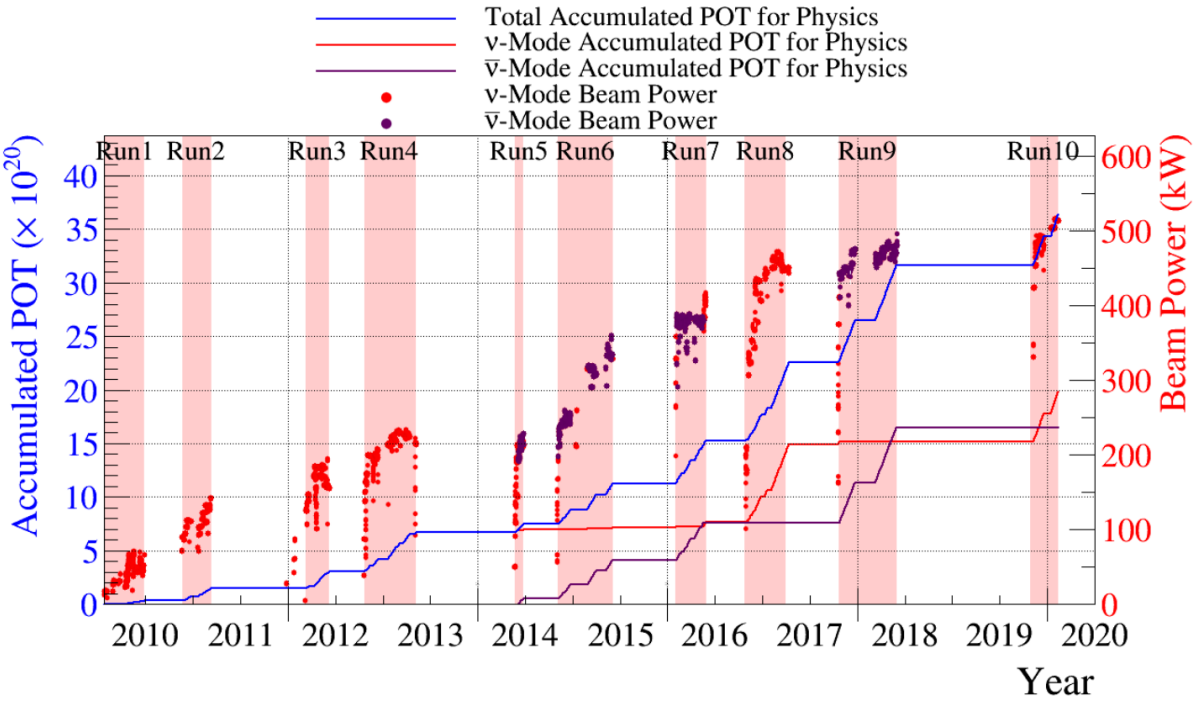


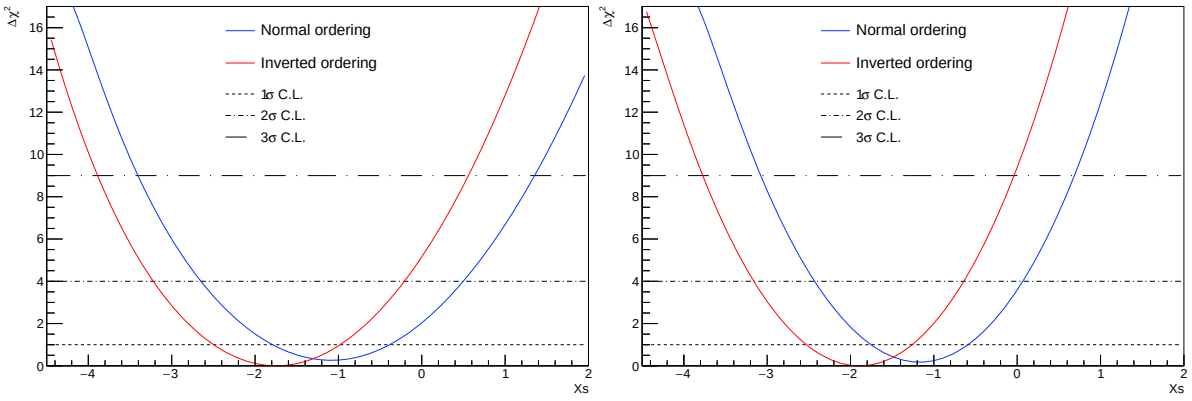
Figure 6.1.: Beam power, run duration, and accumulated POT for each J-PARC beamline runs 1-10. Reproduced from Ref. [55].

is the same systematic model and set of inputs used to generate all of the studies presented in this thesis, including the sensitivity studies of the previous chapter.

6.2. 1D intervals

Here we present 1D intervals for θ_{13}, θ_{23}, X_S, and X_C, produced for both MO and with and without the reactor constraint applied (the latter referred to as "T2K only"). For the T2K only case, these intervals are obtained by marginalising over all other parameters with flat priors. When the reactor constraint is used, a Gaussian prior on θ_{13} with mean 0.0853 and width 0.0027 (1σ) is applied to break the θ_{13}–$\delta_{\mathcal{CP}}$ degeneracy. The contours for θ_{13} and θ_{23} are compared with those obtained in the standard $\delta_{\mathcal{CP}}$ parametrisation.

Figs. 6.2 and 6.3 show the 1D contours for X_S and X_C, while Figs. 6.4 and 6.5 display the corresponding results for θ_{23} and θ_{13}. Note that we follow the P-Theta convention of defining the χ^2 minimum as the global minimum across both MO. For both θ_{23} and θ_{13}, the contours are wider in the X_S–X_C parametrisation than those in the standard



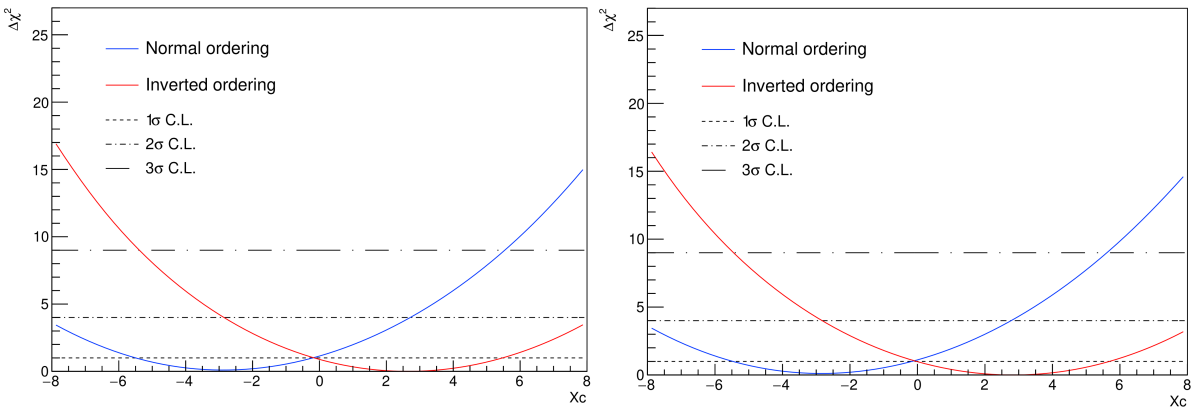
(a) T2K only **(b)** T2K + reactor

Figure 6.2.: 1D data fit for X_S in both normal and inverted orderings. The marked significance levels correspond to 1σ, 2σ, and 3σ intervals of a χ^2 distribution with 1 degree of freedom.

parametrisation, as expected given the additional degree of freedom introduced by treating the $\mathcal{CP}$v and $\mathcal{CP}$c terms independently. A reduction in octant and MO discriminating power is also observed, which we likewise attribute to this additional degree of freedom. In the standard parametrisation, degeneracies between different parameter combinations cannot persist across the full parameter space (e.g. the prediction for NO at $\delta_{\mathcal{CP}} = -\pi/2$ is not degenerate with any IO prediction). Because T2K's data favours such extreme regions where degeneracies are partially decoupled, T2K has some sensitivity to MO and θ_{23}-octant [55]. In the X_S–X_C parametrization, however, the $\delta_{\mathcal{CP}}$-related boundary is removed by construction, meaning these degeneracies remain active and the discriminatory power associated with this boundary is lost.

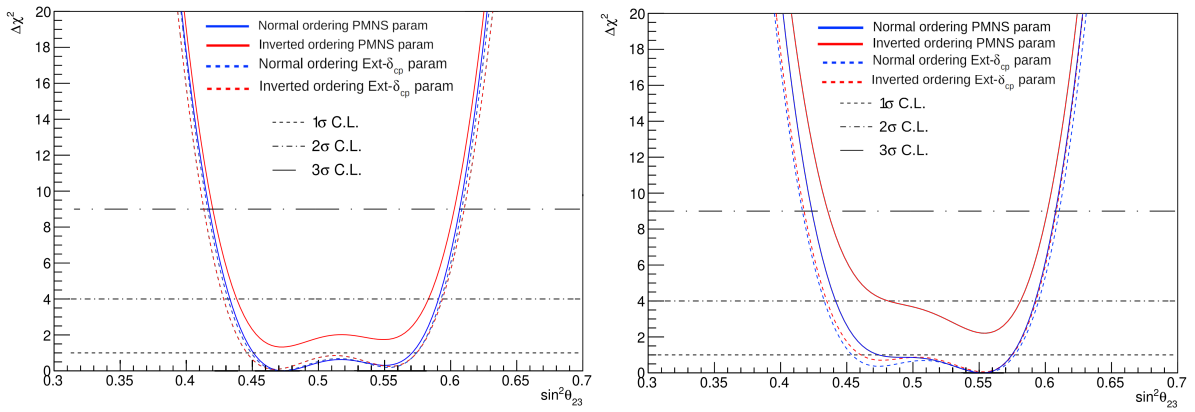
6.3. 2D intervals

T2K typically also presents results as 2D intervals for selected parameter pairs in order to study their correlations. Following this approach, we present in Figs. 6.6–6.11 intervals for θ_{23}–X_S, θ_{23}–X_C θ_{13}–X_S, θ_{13}–X_C, X_S–Δm^2, and X_C–Δm^2, where either X_S or X_C serves as the proxy for correlations usually expressed in terms of $\delta_{\mathcal{CP}}$. Again, results are shown for both MO with and without the reactor constraint applied. In these cases, the physical subspace corresponds to the region $X_S \in [-1, 1]$, such that tracing a path through the whole $\delta_{\mathcal{CP}}$ range ($\delta_{\mathcal{CP}} \in [-\pi, \pi]$) crosses the ellipse multiple times. Contours are drawn using the fixed $\Delta\chi^2$ method of Wilks' theorem.



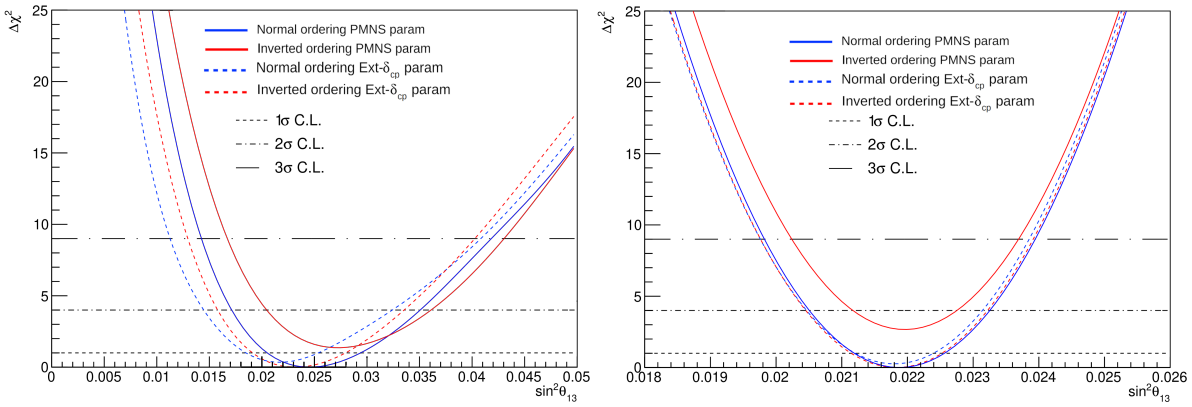
(a) T2K only **(b)** T2K + reactor

Figure 6.3.: 1D data fit for X_C in both normal and inverted orderings. The marked significance levels correspond to 1σ, 2σ, and 3σ intervals of a χ^2 distribution with 1 degree of freedom.



(a) T2K only **(b)** T2K + reactor

Figure 6.4.: 1D data fits for $\sin^2\theta_{23}$ in normal and inverted orderings. Superimposed are the corresponding contours from the fit performed using the conventional parametrisation, where the $\mathcal{CP}$v and $\mathcal{CP}$c terms are not independent but both functions of $\delta_{\mathcal{CP}}$. The marked significance levels correspond to 1σ, 2σ, and 3σ intervals of a χ^2 distribution with 1 degree of freedom.



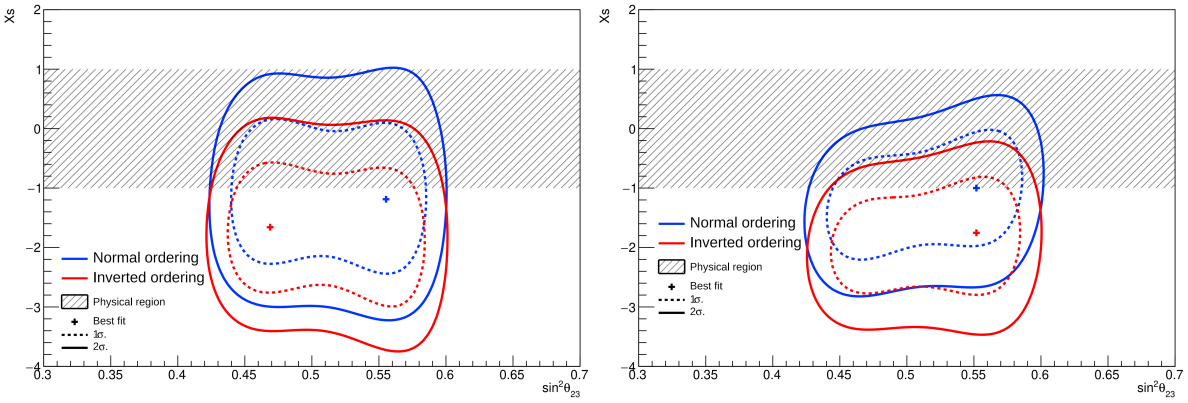
(a) T2K only **(b)** T2K + reactor

Figure 6.5.: 1D data fits for $\sin^2\theta_{13}$ in normal and inverted orderings. Superimposed are the corresponding contours from the fit performed using the conventional parametrisation, where the $\mathcal{CP}$v and $\mathcal{CP}$c terms are not independent but both functions of $\delta_{\mathcal{CP}}$. The marked significance levels correspond to 1σ, 2σ, and 3σ intervals of a χ^2 distribution with 1 degree of freedom.

For the θ_{23} likelihood, similar non-Gaussian (i.e. non-elliptical) behaviour to that seen in the conventional analysis is visible, which is expected given the inherent nonlinearity of this parameter. A separate linear model incorporating θ_{23} is discussed in Appendix A.

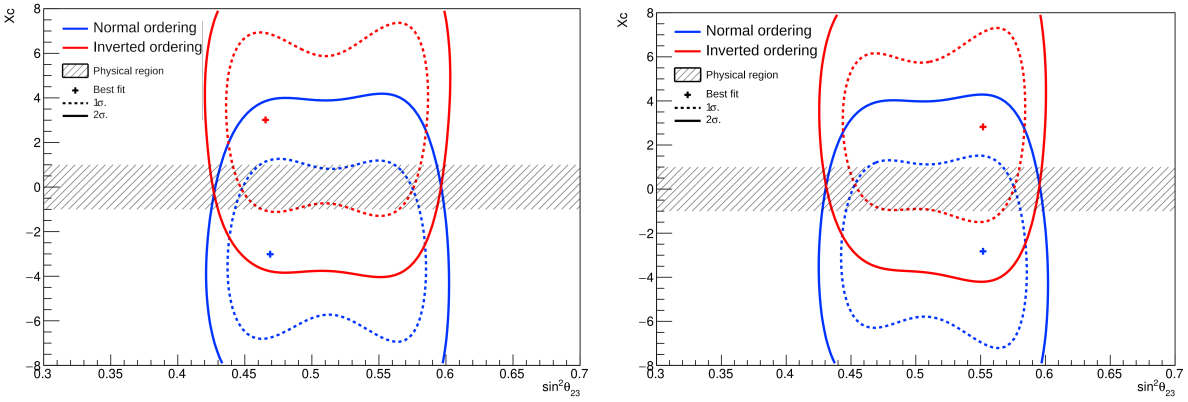
We observe a correlation between X_C and Δm^2 that is not present in the corresponding X_S likelihood (Figs.6.10-6.11). We speculate that this arises because both X_C and Δm^2 are constrained by the shape of the neutrino energy spectra, and thus correlations between them naturally emerge. Consistent with this interpretation, we also observe that the sign of the correlation flips depending on the assumed MO. This behaviour is expected, since in the standard oscillation probability the prefactor of the $\mathcal{CP}$v terms is odd under a change of MO, so therefore, changing the sign of Δm^2 reverses the correlation.

The joint interval in the X_S–$X_C$space is shown in Fig. 6.12, which shows that, as in the sensitivity studies, the contours have the desired Gaussian form. The best-fit value of X_C lies in a substantially more unphysical region than in the studies of Chap. 5, and its sign flips between the two mass orderings. This behaviour is expected since the sensitivity to X_C is driven by the shape of the energy spectrum, so the sign of X_C will correspondingly flip when the model assumes the opposite sign of Δm^2.



(a) T2K only **(b)** T2K + reactor

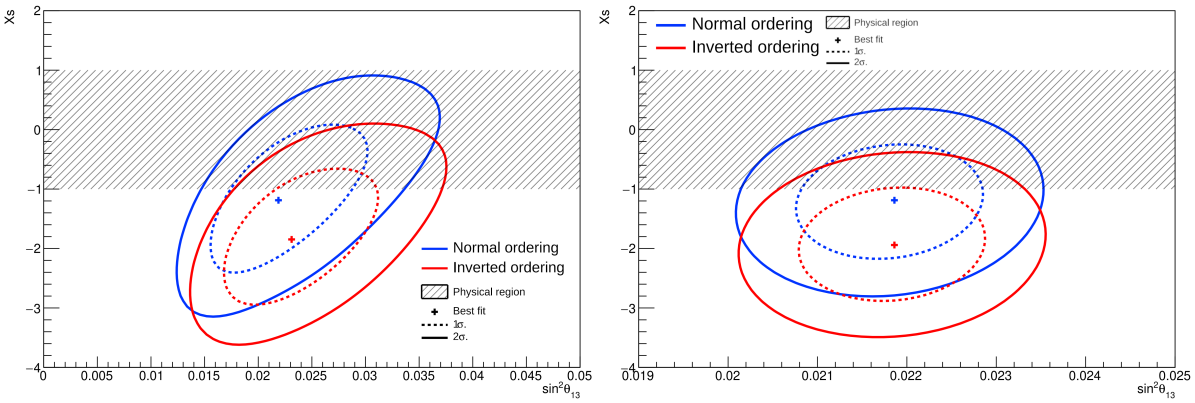
Figure 6.6.: 2D likelihood surfaces for $\sin^2\theta_{23}$–X_S in normal and inverted orderings, with and without the reactor constraint applied. Contours are drawn at fixed $\Delta\chi^2$ values corresponding to the 1σ and 2σ confidence levels. The shaded regions indicate the physical subspace.



(a) T2K only **(b)** T2K + reactor

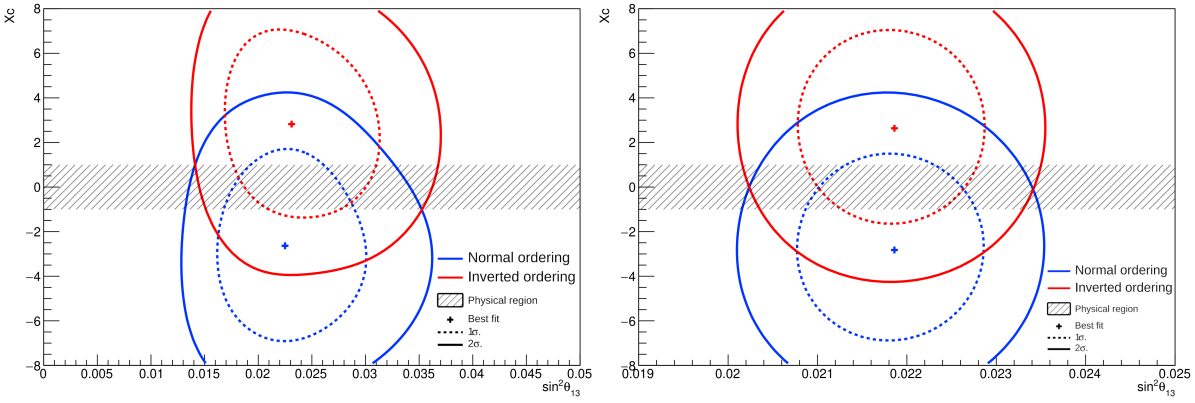
Figure 6.7.: 2D likelihood surfaces for $\sin^2\theta_{23}$–X_C in normal and inverted orderings, with and without the reactor constraint applied. Contours are drawn at fixed $\Delta\chi^2$ values corresponding to the 1σ and 2σ confidence levels. The shaded regions indicate the physical subspace.

We also find that the T2K only fit, while maintaining similar sensitivity to X_C, exhibits degraded sensitivity to X_S compared to the T2K+reactor fit. As a result, the 1σ and 2σ contours cover a larger fraction of the physical circle, meaning a poorer effective sensitivity to $\delta_{\mathcal{CP}}$.



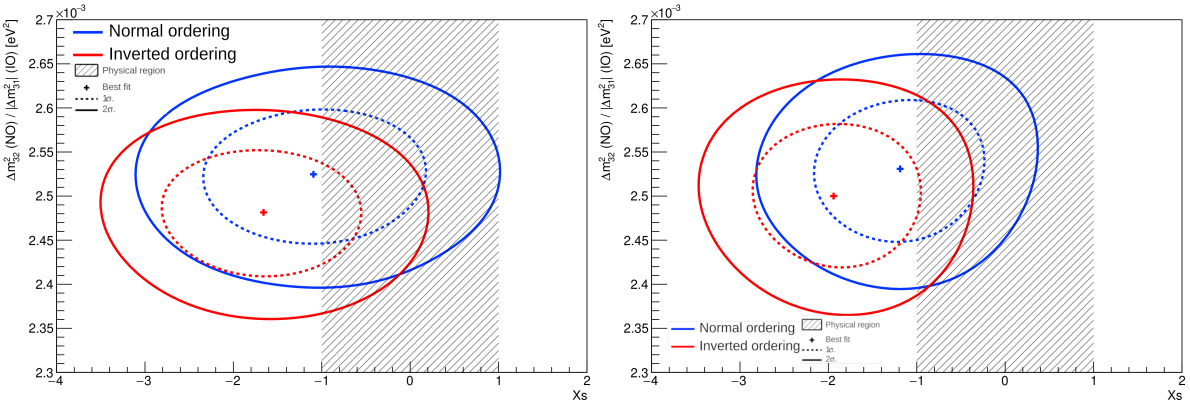
(a) T2K only **(b)** T2K + reactor

Figure 6.8.: 2D likelihood surfaces for $\sin^2\theta_{13}$–X_S in normal and inverted orderings, with and without the reactor constraint applied. Contours are drawn at fixed $\Delta\chi^2$ values corresponding to the 1σ and 2σ confidence levels. The shaded regions indicate the physical subspace.



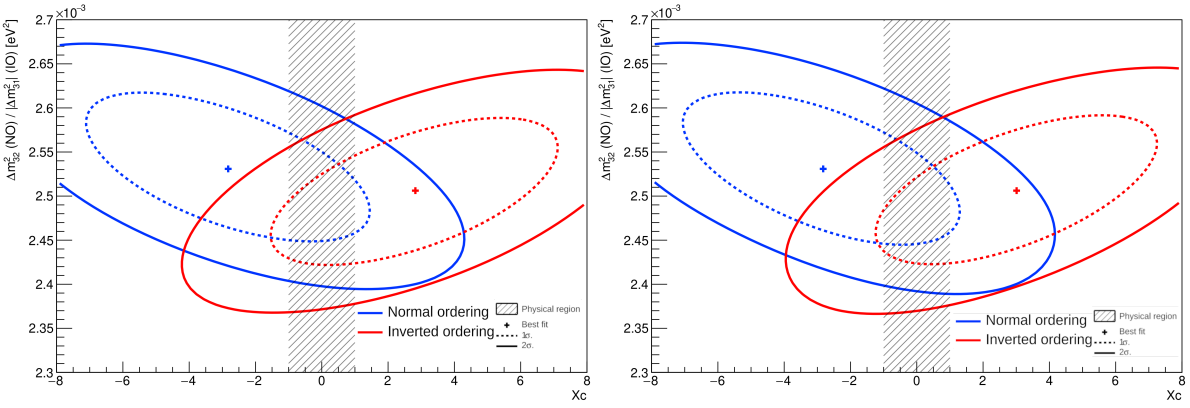
(a) T2K only **(b)** T2K + reactor

Figure 6.9.: 2D likelihood surfaces for $\sin^2\theta_{13}$–X_C in normal and inverted orderings, with and without the reactor constraint applied. Contours are drawn at fixed $\Delta\chi^2$ values corresponding to the 1σ and 2σ confidence levels. The shaded regions indicate the physical subspace.



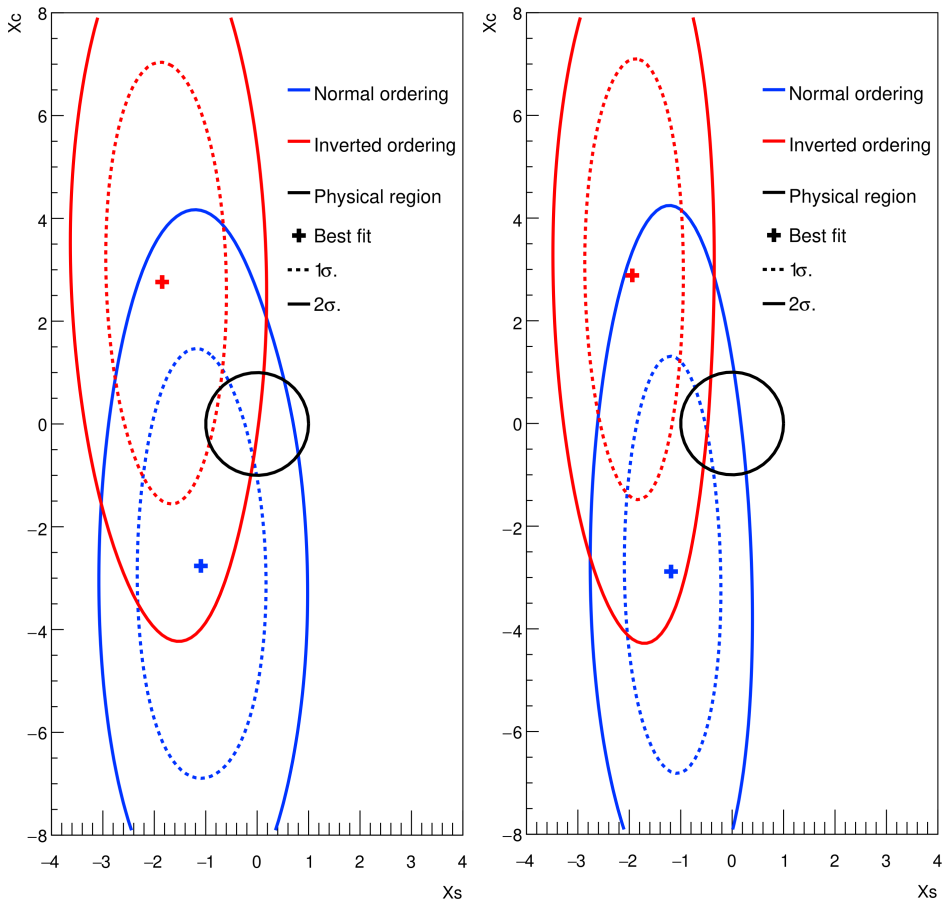
(a) T2K only **(b)** T2K + reactor

Figure 6.10.: 2D likelihood surfaces for X_S–$\Delta m^2_{32}(|\Delta m^2_{13}|$ for inverted$)$ in normal and inverted orderings, with and without the reactor constraint applied. Contours are drawn at fixed $\Delta\chi^2$ values corresponding to the 1σ and 2σ confidence levels. The shaded regions indicate the physical subspace.



(a) T2K only **(b)** T2K + reactor

Figure 6.11.: 2D likelihood surfaces for X_C–$\Delta m^2_{32}(|\Delta m^2_{13}|$ for inverted$)$ in normal and inverted orderings, with and without the reactor constraint applied. Contours are drawn at fixed $\Delta\chi^2$ values corresponding to the 1σ and 2σ confidence levels. The shaded regions indicate the physical subspace.



(a) T2K only **(b)** T2K + reactor

Figure 6.12.: 2D data fit for X_S–X_C in both normal and inverted orderings.

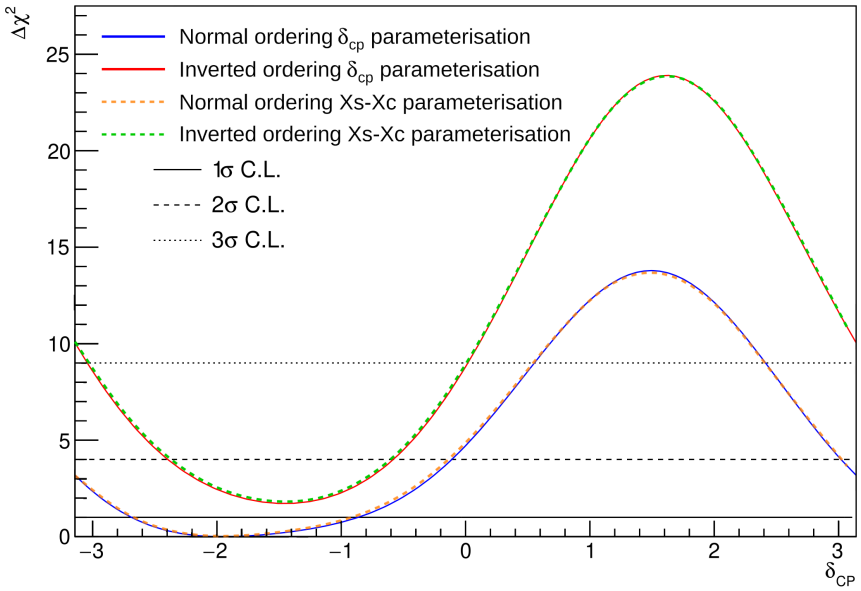


Figure 6.13.: Comparison of the 1D $\delta_{\mathcal{CP}}$ $\Delta\chi^2$ function obtained by evaluating the X_S–X_C surface of the data fit on the physical circle with the $\Delta\chi^2$ function from the fit using the standard parametrisation. The two contours are identical, demonstrating that the official constraint on $\delta_{\mathcal{CP}}$ is reproduced exactly by the fit performed in the extended parameterisation.

Recovering the $\delta_{\mathcal{CP}}$ constraint

For completeness, we repeat the procedure described in Sec. 5.1.4 to recover the physical $\delta_{\mathcal{CP}}$ constraint. Fig. 6.13 shows the result, superimposed on the contours obtained using the standard parametrisation. As in the sensitivity studies, the agreement is very good, providing further validation of the linearised framework.

6.4. Application to PMNS model-testing

In parametric hypothesis testing, an implicit assumption is that the parameter space contains the true hypothesis. In our case, this means that $\delta_{\mathcal{CP}}$, and the PMNS parametrisation more broadly, represents the data and therefore oscillation phenomena well. While dedicated goodness-of-fit tests exist to probe this, a common strategy to test things like this is to fit an alternative model with additional degrees of freedom and then test the tension between the data fitted in the alternative model and the original parametrisation. We can therefore use the X_S–X_C model for this purpose. As noted earlier, models with more adjustable parameters will, by construction, generally fit the

Ordering	χ^2_{bf}	Physical χ^2_{bf}	$\Delta\chi^2$	p-value
NO	667.457	666.538	0.919	0.38
IO	669.200	666.407	2.793	0.095

Table 6.1.: Comparison of best-fit negative log-likelihood values in the extended and physical spaces for each MO. The $\Delta\chi^2$ is computed as the difference between the two, with the corresponding p-value evaluated assuming a χ^2 distribution with one degree of freedom.

data better. However, if the standard $\delta_{\mathcal{CP}}$ space is found to be very strongly disfavored, this may indicate that the PMNS parameterisation poorly represents the data.

To test this, we calculated the χ^2 difference between the best-fit point in the extended space and the physical best-fit point on the circle, independently for each ordering. These values are then compared to the χ^2 distribution with one degree of freedom, and the corresponding p-values are obtained (see Tab. 6.1). The p-value for the test of "excluding" the physical parametrisation is 0.38 for NO and 0.095 for IO. Since neither ordering exhibits significant tension with the physical parametrisation, we conclude that the $\delta_{\mathcal{CP}}$ parametrisation represents T2K oscillation data well.

6.5. Just five numbers

In Chap. 5, we demonstrated, using sensitivity studies, that in the extended parametrisation the contours are approximately Gaussian and therefore simple to characterise. As a final step, we now examine how far this reduction can be taken by testing whether the 1D $\delta_{\mathcal{CP}}$ contour can be described using only five numbers. The idea is that if the Gaussian approximation remains valid, at least in the region overlapping with the physical circle, then T2K's $\delta_{\mathcal{CP}}$ constraint can be fully specified by the mean, center, and covariance of the corresponding X_S–X_C 2D Gaussian. The more familiar $\delta_{\mathcal{CP}}$ likelihood function is then encoded entirely in the geometry of the physical circle mapping on the X_S–X_C surface. In its purest form, one could imagine providing only these five numbers, from which the $\delta_{\mathcal{CP}}$ contour could be reconstructed using a simple program that evaluates the Gaussian surface along the physical circle.

We do not expect perfect agreement, owing to negative event rates and the limitations of the approximation at finite statistics. To test its performance under current condi-

tions, we adopt a simple procedure in which the Hessian is evaluated at the best-fit point of Fig. 6.12 (for simplicity, this was performed only for NO), obtaining,

$$H = \begin{bmatrix} 2.40705 & 0.125045 \\ 0.125045 & 0.027241 \end{bmatrix}. \tag{6.1}$$

Again, we then repeat the procedure of Sec. 5.1.4, this time not for the fitted surface but for the exact Gaussian defined by Eq. 6.1. The results are shown in Fig. 6.14, which compares the resulting $\delta_{\mathcal{CP}}$ contour to that obtained directly with the $\delta_{\mathcal{CP}}$ parametrisation. At high significance levels, the agreement is poor, as expected from the discussion in Sec. 5.2, but below 2σ, which is about the level T2K can make non-trivial statements about $\delta_{\mathcal{CP}}$, the agreement is very good. Fig. 6.15a compares the ellipses of the fitted X_S–X_C surface with this exact Gaussian, while Fig. 6.15c shows the absolute differences between the two surfaces, highlighting that the largest deviations are near $\delta_{\mathcal{CP}} \sim -\pi/2$.

Although the deviations here are possibly too large for the headline result of a T2K analysis, they are certainly good enough for use in auxiliary information to answer common planning questions such as making sensitivity projections, deciding on neutrino vs antineutrino running, and making fast combination with other experiments (see Appendix C).

To finish we reiterate again that the real strength of this framework lies in the intuition it provides: one can quickly interpret how the shapes and shifts of the ellipses map onto the $\delta_{\mathcal{CP}}$ likelihood. In doing so, the method offers a universal, data-independent means of characterising sensitivity to $\delta_{\mathcal{CP}}$, a capability currently absent from T2K analyses.

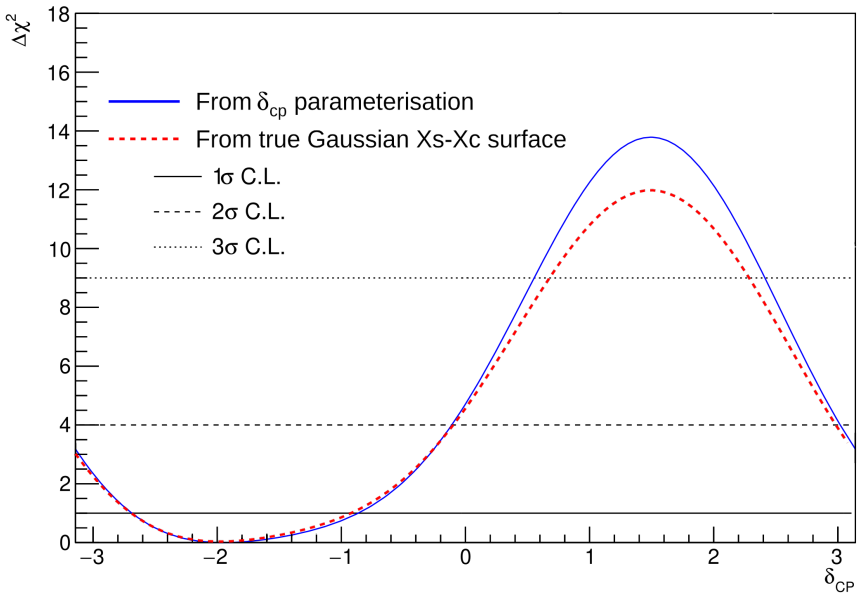
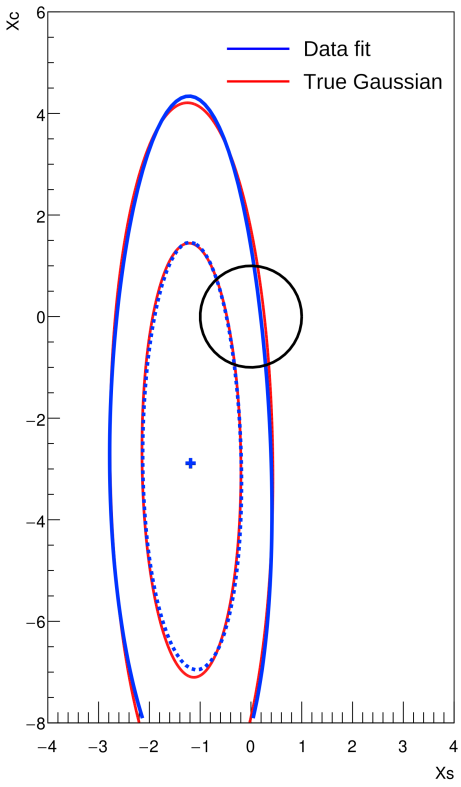
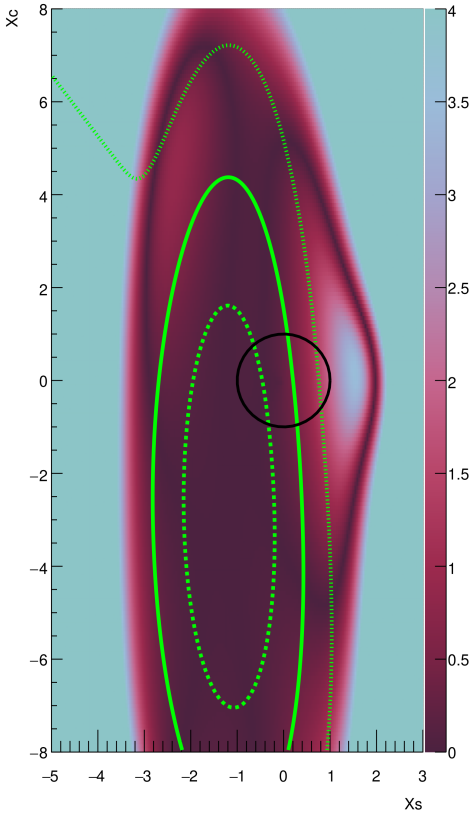


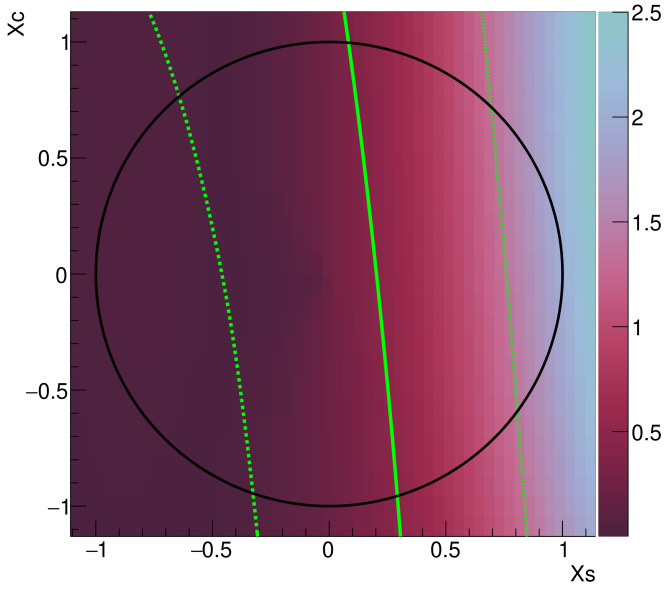
Figure 6.14.: Comparison of the 1D $\delta_{\mathcal{CP}}$ $\Delta\chi^2$ function obtained using the "just five numbers" approximate surface (defined by the covariance matrix in Eq. 6.1) evaluated on the physical circle, with the $\Delta\chi^2$ function obtained from the fit using the standard parametrisation. As discussed in the text, the agreement is good at low significance ($< 2.5\sigma$), but deviations appear at higher confidence levels, with a maximum difference of approximately 3 units in χ^2.



(a) Contour comparison.



(b) Absolute difference between the surfaces, expressed in units of χ^2.



(c) Absolute difference between the surfaces around the physical surface, expressed in units of χ^2.

Figure 6.15.: Comparison of the surface from the data fit with the surface from the true Gaussian.

Conclusion and outlook

In this thesis, we have developed and presented a strategy to implement a linearised oscillation model for $\delta_{\mathcal{CP}}$. We demonstrated that this model can be used in a supplementary fashion to improve the interoperability of T2K sensitivity studies by enabling analyses in a statistically simpler and more intuitive Gaussian space. This provides a universal, data-independent method for assessing sensitivity and interpreting changes arising from the adoption of alternative models in robustness tests; something that is missing in current T2K analyses. We have also demonstrated that while fits in this space are more insightful and can reveal information usually obscured by the $\delta_{\mathcal{CP}}$ parameterisation, no information is lost, since the physical likelihood can always be recovered. Finally, we applied the framework to real T2K data, checked that the results are sensible in the extended space, and confirmed that the OA21 analysis can be successfully reproduced.

There remains considerable scope for further development. An immediate next step would be to integrate this framework into the full oscillation analyses as a P-Theta tool. Extending the framework to include other non-linear oscillation parameters, such as θ_{23} and the mass ordering (see Appendix. A), is also an important next step. Moreover, this framework has many potential applications beyond those discussed in this thesis. For example, we have identified a promising strategy to obtain approximate critical values for the $\Delta\chi^2$ test statistic with a fraction of the computational cost and time normally required (see Appendix. B). It also has promising applications for rapidly combining constraints from different experiments in the absence of a dedicated joint fitter (see Appendix. C).

As a final thought, we note that neutrino physics is at an exciting juncture. Over the coming decade, the next generation of long-baseline experiments (DUNE and Hyper-Kamiokande) will begin operations and are expected to answer many of the field's outstanding questions. However, as the statistical challenges associated with the PMNS parametrisation will persist, these experiments are likely to face many of

the same complications discussed in this thesis, with the added complication that they hope to present "discovery" (5σ) results, which are prohibitively slow to compute using existing tools. Therefore, we expect that the linearised framework detailed in this thesis will remain relevant to these future experiments, and we hope it can be of use to their analyses.

Appendix A.

Proposed extensions

The $\mathcal{CP}$-dependence is, of course, not the only source of non-linearity in the oscillation model. In this chapter we propose and explore a few methods for extending this linearisation concept to also include MO and θ_{23}. Sect. A.1 describes an oscillation model that achieves linearity in both θ_{23} and $\delta_{\mathcal{CP}}$, while Sect. A.3 presents a strategy for linearising the mass ordering.

A.1. The $X_S - X_C - \Sigma_{\theta_{23}} - \Sigma_{2\theta_{23}}$ model

Analogous to $\delta_{\mathcal{CP}}$, we select as basis functions the leading θ_{23}-dependent terms that appear in the standard oscillation probability, namely $\sin^2 2\theta_{23}$ and $\sin^2 \theta_{23}$. Their extended linear model analogue parameters will be denoted as $\Sigma_{2\theta_{23}}$, and $\Sigma_{\theta_{23}}$ respectively.

However, unlike $\delta_{\mathcal{CP}}$, there is no readily exploitable symmetry for θ_{23}, and so a method for computing probabilities on the physical region and interpolating into the extended space cannot be easily devised for this parameter. Consequently, a direct numerical approach is adopted, in which the Fourier coefficients (i.e. the A, B, C prefactors) are computed explicitly. Since this method is easily scalable, we will generalise it to construct a model that is linear in both θ_{23} and $\delta_{\mathcal{CP}}$, as such a model will have more general future utility.

The most general linear model that can be constructed using these leading-order functions spans the combined space:

$$(1, \sin\delta_{\mathcal{CP}}, \cos\delta_{\mathcal{CP}}) \otimes (1, \sin^2 2\theta_{23}, \sin^2\theta_{23}) \tag{A.1}$$

That is, the space of basis functions and their interaction. Although this expression defines the most general form, not all of the nine basis directions (or "Fourier modes") necessarily contribute to the oscillation probability. Additionally, contributions from higher-order terms may be necessary. In the following section, we investigate this by referring again to the results of Yokomakura *et al.* [93], and by applying an analysis similar to that previously used to verify $\cos 2\delta_{\mathcal{CP}}$ was small.

A.1.1. Implementation

Generalising what was described in Chapt. 4, we have an oscillation model expressed as a sum of N basis functions with corresponding Fourier coefficients C_i:

$$P(\theta_{23}, \delta_{\mathcal{CP}}) = C_1\omega_1(\theta_{23}, \delta_{\mathcal{CP}}) + C_2\omega_2(\theta_{23}, \delta_{\mathcal{CP}}) \ + ...C_N\omega_N(\theta_{23}, \delta_{\mathcal{CP}}) + Const. \tag{A.2}$$

Where each of the omegas correspond to a specific dependence on θ_{23} and $\delta_{\mathcal{CP}}$. For a throw of non-$(\theta_{23}, \delta_{\mathcal{CP}})$, we need to calculate the vector of Fourier coefficients $\mathbf{C} = C_1..C_N$. The probability in the corresponding extended space is then obtained by replacing the physical functions $\omega(\theta)$ with their linearised analogues:

$$P_{Ext}(\bar{\omega} = (\omega_1, ..., \omega_N)) = C_1\omega_1 + C_2\omega_2 + C_3\omega_3 + C_4\omega_4 + ...C\omega_N + Const. \tag{A.3}$$

To determine $\mathbf{C}$, we perform a least-squares regression using the probabilities computed at a predefined set of $(\theta_{23}, \delta_{\mathcal{CP}})$: $P_i = C^j\phi_{ij}$, where ϕ is the design matrix whose entries are the value of the $(\theta_{23}, \delta_{\mathcal{CP}})$ basis function evaluated at the predefined values,

$$\phi = \begin{bmatrix} \omega_0(\theta^1_{23}, \delta^1_{\mathcal{CP}}) & \omega_1(\theta^1_{23}, \delta^1_{\mathcal{CP}}) & \cdots & \omega_N(\theta^1_{23}, \delta^1_{\mathcal{CP}}) \\ \omega_0(\theta^2_{23}, \delta^2_{\mathcal{CP}}) & \omega_1(\theta^2_{23}, \delta^2_{\mathcal{CP}}) & \cdots & \omega_N(\theta^2_{23}, \delta^2_{\mathcal{CP}}) \\ \vdots & \vdots & \ddots & \vdots \\ \omega_0(\theta^j_{23}, \delta^j_{\mathcal{CP}}) & \omega_1(\theta^j_{23}, \delta^j_{\mathcal{CP}}) & \cdots & \omega_N(\theta^j_{23}, \delta^j_{\mathcal{CP}}) \end{bmatrix} \tag{A.4}$$

The coefficients are then given by

$$\bar{C} = (\phi^T \phi)^{-1} \phi^T P \tag{A.5}$$

Naively, one might expect that determining N coefficients would require sampling the probability at N points. However, as we are fitting an approximate model to the full oscillation probability computed with Prob3++, where for instance higher order terms like $\cos 2\delta$ have been ignored, using only N samples can result in a good local fit to the sampled points but a poor global fit elsewhere. To mitigate for this we deliberately oversample, and for the model discussed here we conservatively use 16 points:

$$\delta_{\mathcal{CP}} = (-0.872,\ -2.918,\ 2.469,\ 0.253) \times \sin^2\theta_{23} = (0.44,\ 0.48,\ 0.52,\ 0.56) \tag{A.6}$$

A.1.2. Evaluating model error

To determine which Fourier modes in Eq. (A.1) are necessary, we will consider a series of approximate models for the appearance and disappearance probabilities separately, and quantify the model error using the normalised percentage difference metric:

$$\frac{|P_{\text{Prob3++}} - P_{\text{Linear}}|}{\frac{1}{2}\left(P_{\text{Prob3++}} + P_{\text{Linear}}\right)} \times 100 \tag{A.7}$$

where $P_{\text{Prob3++}}$ is the oscillation probability calculated using the oscillation engine of P-Theta, and P_{Linear} is the probability calculated using the candidate linearised model. We require that this metric remains below 1% for all sampled points.

Appearance

The full appearance probability in Yokomakura *et al.* is,

$$P(\nu_e \to \nu_\mu) = A_{e\mu} \cos\delta + B_{e\mu} \sin\delta + C_{e\mu}, \tag{A.8}$$

with

$$A_{e\mu} = 2\,\mathrm{Re}\left[S'^{*}_{\mu e}S'_{\tau e}\right] c_{23}s_{23},$$
$$B_{e\mu} = -2\,\mathrm{Im}\left[S'^{*}_{\mu e}S'_{\tau e}\right] c_{23}s_{23},$$
$$C_{e\mu} = \left|S'_{\mu e}\right|^2 c_{23}^2 + \left|S'_{\tau e}\right|^2 s_{23}^2.$$

where $S_{\alpha\beta}$ are the $\{\theta_{23}, \delta_{\mathcal{CP}}\}$ independent transition amplitudes. Given that $\sin^2 x$ and $\cos^2 x$ are not linearly independent ($\sin^2 x = -\cos^2 x + 1$), and $\sin x \cos x = 0.5 \sin 2x$, the appearance probability can be re-expressed as:

$$P_{\mu e} = A \sin 2\theta_{23} \sin \delta_{\mathcal{CP}} + B \sin 2\theta_{23} \cos \delta_{\mathcal{CP}} + C \sin^2 \theta_{23} + \text{const.} \tag{A.9}$$

and therefore the series describing the appearance probability terminates with these four terms. As this oscillation probability is exact, it comfortably satisfies the 1% model error requirement, however, for consistency, it would be preferable for the $\sin 2\theta_{23}$ term here to appear as $\sin^2 2\theta_{23}$. We achieve this by expanding around $\sin^2 \theta_{23} = 1$ with the first-order approximation,

$$\sin 2\theta_{23} = \sqrt{\sin^2 2\theta_{23}} \approx 1 + \frac{1}{2}\left(\sin^2 2\theta_{23} - 1\right).$$

Applying this to the interaction terms gives,

$$A \sin 2\theta_{23} \sin \delta_{\mathcal{CP}} + \text{const} \approx \left[1 + \frac{1}{2}(\sin^2 2\theta_{23} - 1)\right] A \sin \delta_{\mathcal{CP}} + \text{const} \tag{A.10}$$
$$= A \sin \delta_{\mathcal{CP}} + B \sin^2 2\theta_{23} \sin \delta_{\mathcal{CP}} + \text{const} \tag{A.11}$$

This yields a probability with the more recognisable basis functions,

$$P_{\mu e} = A \sin \delta_{\mathcal{CP}} + B \cos \delta_{\mathcal{CP}} + C \sin^2 \theta_{23} + D \sin \delta_{\mathcal{CP}} \sin^2 2\theta_{23} + E \cos \delta_{\mathcal{CP}} \sin^2 2\theta_{23} + F \tag{A.12}$$

The loss of precision from this approximation is negligible, and Eq. (A.12) still comfortably satisfies the 1% model error requirement. Fig. A.3 shows the average absolute value of the model error metric is small for both these, while Fig. A.2 verifies that the behaviour of this approximation is as expected, breaking down first when $\sin^2 \theta_{23}$ is far from 1.

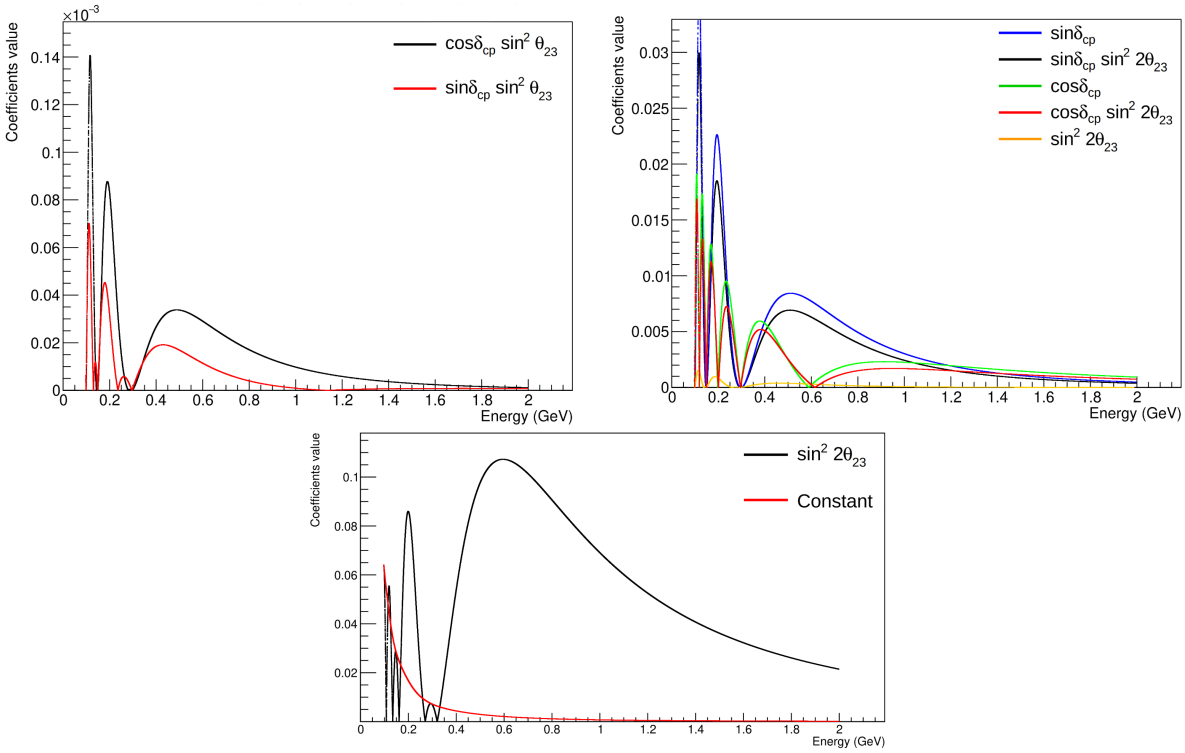


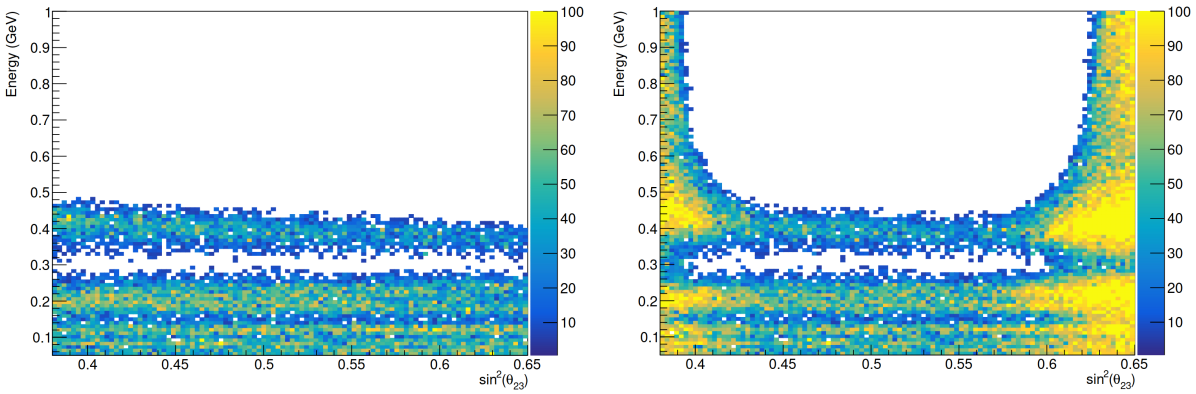
Figure A.1.: Fourier coefficient values for fit performed at Asimov A.

To assess the relative importance of each term in Eq. (A.12), the corresponding coefficient values were evaluated at the Asimov A values. The idea is that the magnitude of a coefficient serves as a proxy for the influence of that Fourier mode; the larger the coefficient, the larger the influence. As an additional consistency check, we also included a few extra modes ($\sin^2 2\theta_{23}$, $\sin\delta_{\mathcal{CP}} \sin^2\theta_{23}$, $\cos\delta_{\mathcal{CP}} \sin^2\theta_{23}$) which, based on the arguments above, are expected to be small. Fig. A.1 confirms this and shows that $sin^2\theta_{23}$ is dominant over the $\mathcal{CP}$v, $\mathcal{CP}$c terms as anticipated.

Disappearance

The full disappearance probability in Yokomakura *et al.* [93] is given by,

$$P(\nu_\mu \to \nu_\mu) = A_{\mu\mu}\cos\delta + B_{\mu\mu}\cos 2\delta + C_{\mu\mu} \tag{A.13}$$



(a) Exact Yokomakura *et al.* appearance probability.

(b) Appearance probability with $\sin 2\theta_{23}$ approximation.

Figure A.2.: Regions where agreement between the oscillation model and Prob3++ breaks down, shown as the percentage of bins in which the model error metric of Eq. (A.7) exceeds 0.005%. The threshold is set deliberately low, as both models are generally in good agreement.

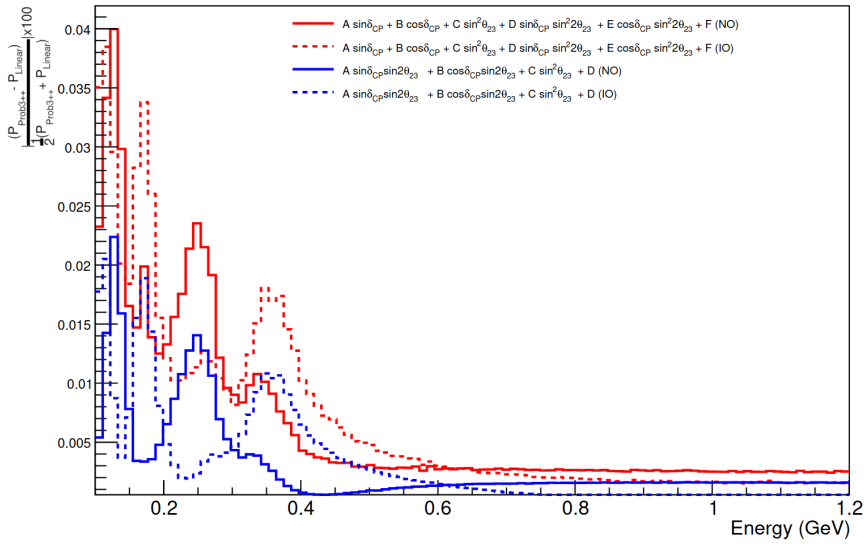


Figure A.3.: Comparison of the average absolute model error metric (as defined in Eq. (A.7)) for the exact appearance oscillation probability of Yokomakura *et al.* [93] and for the $\sin 2\theta_{23}$ approximation.

with

$$
\begin{aligned}
A_{\mu\mu} &= 2\mathrm{Re}\left[(S'_{\mu\mu}c_{23}^2 + S'_{\tau\tau}s_{23}^2)^*(S'_{\tau\mu} + S'_{\mu\tau})\right] c_{23}s_{23} \\
B_{\mu\mu} &= 2\mathrm{Re}[S'^*_{\tau\mu}S'_{\mu\tau}]c_{23}^2 s_{23}^2 \\
C_{\mu\mu} &= |S'_{\mu\mu}|^2 c_{23}^4 + (|S'_{\mu\tau}|^2 + |S'_{\tau\mu}|^2)c_{23}^2 s_{23}^2 + |S'_{\tau\tau}|^2 s_{23}^4 + 2\mathrm{Re}[S'^*_{\mu\mu}S'_{\tau\tau}]c_{23}^2 s_{23}^2
\end{aligned}
$$

Using standard trigonometric identities, this can be re-expressed as

$$
P_{\mu\mu} = A\sin 2\theta_{23}\cos\delta_{\mathcal{CP}} + B\sin 4\theta_{23}\cos\delta_{\mathcal{CP}} + C\sin^2 2\theta_{23}\cos 2\delta_{\mathcal{CP}} + D\sin^2\theta_{23} + E\sin^2 2\theta_{23} + F \tag{A.14}
$$

For instance, $C_{\mu\mu}$ can be expressed in terms of $\cos 2\theta_{23}$ and $\cos 4\theta_{23}$ as,

$$
\begin{aligned}
\sin^2\theta_{23}\cos^2\theta_{23} &= \frac{1}{4}\sin^2 2\theta_{23}, \\
\sin^4\theta_{23} &= \frac{1}{8}\left[3 - 4\cos 2\theta_{23} + \cos 4\theta_{23}\right], \\
\cos^4\theta_{23} &= \frac{1}{8}\left[3 + 4\cos 2\theta_{23} + \cos 4\theta_{23}\right].
\end{aligned}
$$

These identities can be further reduced using:

$$
\cos 2\theta_{23} = 1 - 2\sin^2\theta_{23} \quad : \quad \cos 4\theta_{23} = 1 - 2\sin^2 2\theta_{23}.
$$

Thus, ultimately $C_{\mu\mu}$ can be expressed as a linear combination of the familiar $\sin^2\theta_{23}$ and $\sin^2 2\theta_{23}$ terms.

Likewise, $A_{\mu\mu}$ is proportional to a linear combination of odd powers of $\cos\theta_{23}$ and $\sin\theta_{23}$,

$$
A_{\mu\mu} \propto K_1\cos^3\theta_{23}\sin\theta_{23} + K_2\cos\theta_{23}\sin^3\theta_{23}
$$

which themselves can be decomposed into $2\theta_{23}, 4\theta_{23}$ sine Fourier modes,

$$
\cos\theta_{23}\sin^3\theta_{23} = \frac{1}{4}\sin 2\theta_{23} - \frac{1}{8}\sin 4\theta_{23} \quad : \quad \cos^3\theta_{23}\sin\theta_{23} = \frac{1}{4}\sin 2\theta_{23} + \frac{1}{8}\sin 4\theta_{23}
$$

Fig. A.4 shows that Eq. (A.14) agrees precisely with the Prob3++ calculation of the probability. However, as with the appearance probability, we aim to simplify the expression by retaining only the recognisable leading-order terms of Eq. (A.1). Ulti-

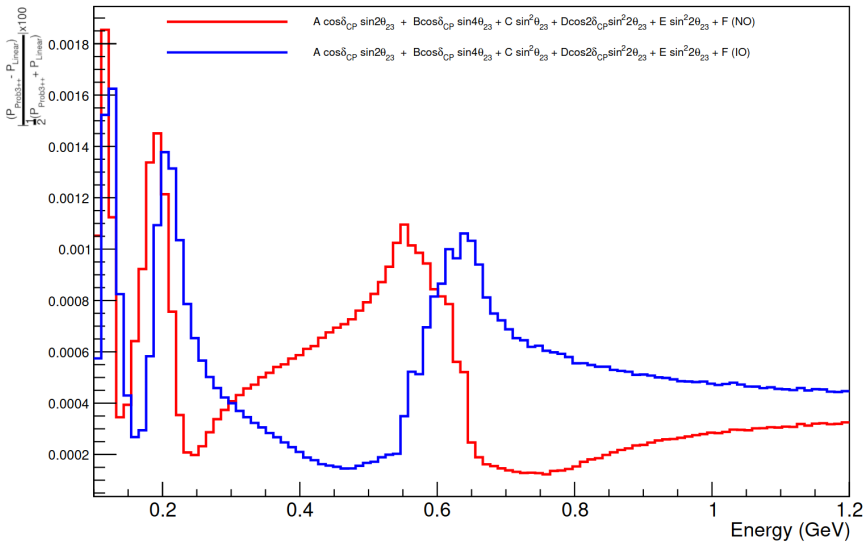


Figure A.4.: Comparison of the average absolute model error metric (as defined in Eq. (A.7)) for the exact disappearance oscillation probability of Yokomakura *et al.* [93] (Eq. (A.14)).

mately, this is a matter of judgment: while including more terms generally reduces model error, it also degrades interpretability and makes the model more cumbersome. Several models were considered, each testing increasingly aggressive approximations:

Expression	**Terms removed**
$A\cos\delta_{\mathcal{CP}}\sin 2\theta_{23} + B\cos\delta_{\mathcal{CP}}\sin 4\theta_{23} + C\sin^2\theta_{23} + D\sin^2 2\theta_{23} + E$	$\cos 2\delta_{\mathcal{CP}}$
$A\cos\delta_{\mathcal{CP}}\sin 2\theta_{23} + B\sin^2\theta_{23} + C\sin^2 2\theta_{23} + D$	$\cos 2\delta_{\mathcal{CP}},\ \sin 4\theta_{23}$
$A\sin^2\theta_{23} + B\sin^2 2\theta_{23} + C$	$\sin\delta_{\mathcal{CP}},\ \cos\delta_{\mathcal{CP}},\ \sin 4\theta_{23}$

An MC study was again used to compare the accuracy of the candidate models with the results shown in Fig. A.5. The highlighted model was selected as the preferred choice, and again the $\sin 2\theta_{23} = \sqrt{\sin^2 2\theta_{23}} \approx 1 + \frac{1}{2}\left(\sin^2 2\theta_{23} - 1\right)$ approximation was used to yield the final oscillation model:

$$P_{\mu\mu} = A\cos\delta_{\mathcal{CP}} + B\sin^2\theta_{23} + C\sin^2 2\theta_{23} + D\cos\delta_{\mathcal{CP}}\sin^2 2\theta_{23} + E \tag{A.15}$$

We also examined how this model error translates into event rate differences. The results, summarised in Tab. A.1a-A.1d for a range of true oscillation parameter values were found to be acceptable.

	$\delta_{\mathcal{CP}} = 0$	0.25π	0.5π	0.75π	1π	-0.75π	-0.5π	-0.25π
FHC 1Re	0.0027	0.0051	0.0053	0.0026	0.0006	0.0019	0.0018	0.0000
FHC 1Rμ	0.0592	0.0366	0.0082	0.0386	0.0474	0.0388	0.0079	0.0368
RHC 1Re	0.0006	0.0011	0.0022	0.0029	0.0019	0.0000	0.0015	0.0021
RHC 1Rμ	0.0243	0.0140	0.0051	0.0159	0.0178	0.0159	0.0051	0.0140
FHC 1R ν_e CC1π	0.0072	0.0084	0.0047	0.0011	0.0044	0.0044	0.0016	0.0028
FHC MR ν_μ CC1π	0.0816	0.0577	0.0008	0.0525	0.0737	0.0525	0.0000	0.0577

(a) $\text{Sin}^2\theta_{23} = 0.40$

	$\delta_{\mathcal{CP}} = 0$	0.25π	0.5π	0.75π	1π	-0.75π	-0.5π	-0.25π
FHC 1Re	0.0013	0.0014	0.0011	0.0008	0.0005	0.0006	0.0006	0.0009
FHC 1Rμ	0.0188	0.0065	0.0116	0.0126	0.0083	0.0126	0.0115	0.0062
fHC	0.0000	0.0005	0.0010	0.0010	0.0011	0.0006	0.0000	0.0000
RHC 1Rμ	0.0077	0.0014	0.0070	0.0056	0.0028	0.0056	0.0063	0.0021
FHC 1R ν_e CC1π	0.0023	0.0021	0.0008	0.0002	0.0001	0.0000	0.0004	0.0015
FHC MR ν_μ CC1π	0.0182	0.0121	0.0017	0.0120	0.0155	0.0120	0.0017	0.0121

(b) $\text{Sin}^2\theta_{23} = 0.48$

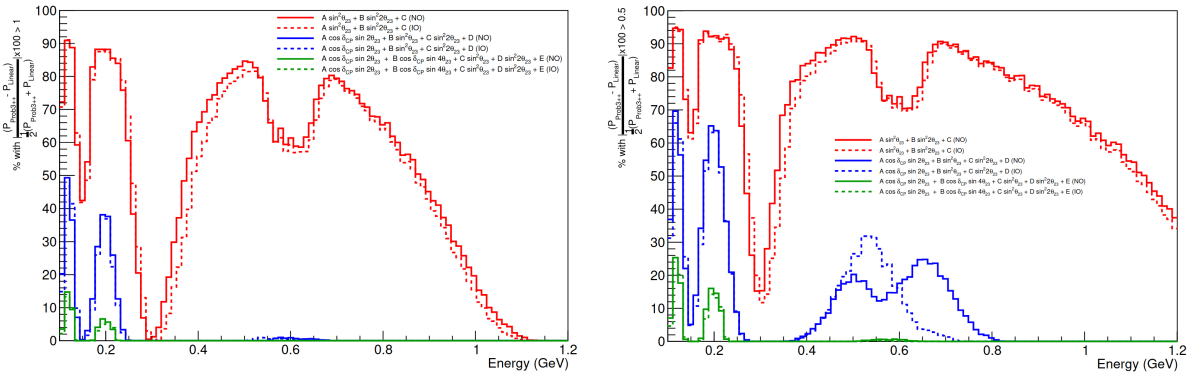
	$\delta_{\mathcal{CP}} = 0$	0.25π	0.5π	0.75π	1π	-0.75π	-0.5π	-0.25π
FHC 1Re	0.0006	0.0010	0.0011	0.0011	0.0010	0.0009	0.0005	0.0005
FHC 1Rμ	0.0071	0.0120	0.0125	0.0041	0.0155	0.0038	0.0125	0.0120
RHC 1Re	0.0005	0.0010	0.0005	0.0010	0.0005	0.0000	0.0006	0.0000
RHC 1Rμ	0.0007	0.0042	0.0071	0.0007	0.0070	0.0007	0.0071	0.0049
FHC 1R ν_e CC1π	0.0001	0.0001	0.0008	0.0019	0.0021	0.0012	0.0003	0.0000
FHC MR ν_μ CC1π	0.0157	0.0122	0.0035	0.0095	0.0164	0.0104	0.0035	0.0122

(c) $\text{Sin}^2\theta_{23} = 0.52$

	$\delta_{\mathcal{CP}} = 0$	0.25π	0.5π	0.75π	1π	-0.75π	-0.5π	-0.25π
FHC 1Re	0.0002	0.0019	0.0036	0.0035	0.0016	0.0010	0.0009	0.0010
FHC 1Rμ	0.0524	0.0437	0.0126	0.0332	0.0563	0.0332	0.0126	0.0437
RHC 1Re	0.0015	0.0023	0.0022	0.0009	0.0005	0.0016	0.0011	0.0000
RHC 1Rμ	0.0185	0.0172	0.0066	0.0112	0.0217	0.0118	0.0073	0.0165
FHC 1R ν_e CC1π	0.0039	0.0016	0.0029	0.0062	0.0057	0.0022	0.0010	0.0039
FHC MR ν_μ CC1π	0.0799	0.0588	0.0059	0.0510	0.0760	0.0510	0.0050	0.0588

(d) $\text{Sin}^2\theta_{23} = 0.60$

Table A.1.: Percentage event rate difference between predictions obtained with Prob3++ and those computed using the probability model of Eq. (A.15) at different true points.



(a) metric threshold: 1% **(b)** metric threshold: 0.5%

Figure A.5.: Comparison of candidate disappearance probability models. Shown is the percentage of bins in which the model error metric (see Eq. (A.7)) exceeds the displayed threshold.

A.2. Fitting modes

This model can be operated in three distinct modes (for simplicity, all fits presented in this section assume true NO):

1. Perform the fit in the (X_S, X_C) subspace while holding $\Sigma_{2\theta_{23}}$ and $\Sigma_{\theta_{23}}$ fixed at their physical values. Here, we use the term "on-shell" to denote a parameter fixed at its physical point; in this case, θ_{23} is on-shell.
2. Fix $\delta_{\mathcal{CP}}$ on-shell and perform the fit in the $(\Sigma_{2\theta_{23}}, \Sigma_{\theta_{23}})$ subspace.
3. Allow all parameters to vary freely.

A.2.1. θ_{23} on-shell mode

In this mode, the fitter is effectively a less efficient implementation of the approach presented in Chapt. 4, which on the surface, it may not appear particularly useful. However, it serves as an additional cross-check for this coefficient fitting approach by producing X_S–X_C surfaces and verifying the results are consistent with those obtained using the interpolation based method. As shown in Fig. A.6, the fits are in agreement within the precision limits of the fitter.

From a software perspective this may also be useful, depending on how the analysis framework is packaged, as having an implementation in which the user can toggle between linearised parameters without drastically changing how the extended probability is computed may be cleaner.

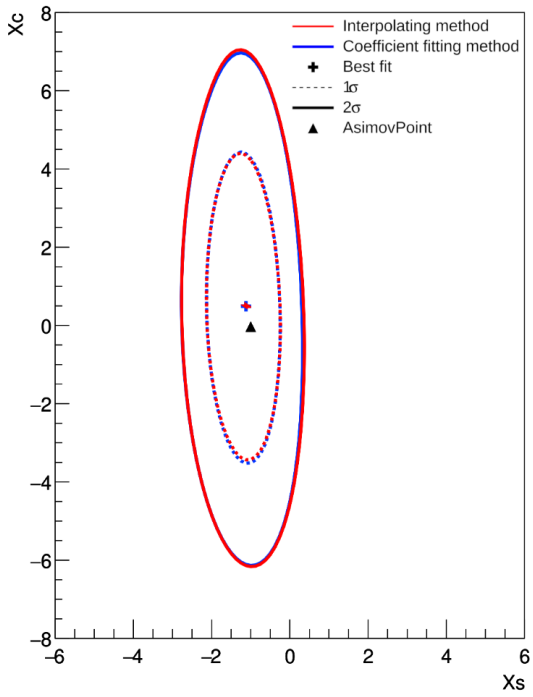


Figure A.6.: Comparison of X_S–X_C contours obtained using the Fourier coefficient fitting and the interpolation-based method described in Chapt. 4. The two approaches show excellent agreement.

A.2.2. $\delta_{\mathcal{CP}}$ on-shell mode

In this mode, $\Sigma_{2\theta_{23}}$–$\Sigma_{\theta_{23}}$ surfaces can be generated. A few illustrative Asimov fits are shown in Fig. A.9, for different true $\sin^2\theta_{23}$ values. In this parameter space, the physical region traces out a parabola defined by the rule,

$$f(\Sigma_{2\theta_{23}}, \Sigma_{\theta_{23}}) = \Sigma_{2\theta_{23}} - 4\Sigma_{\theta_{23}}(1 - \Sigma_{\theta_{23}}) = 0.$$

The results are quite similar in character to the X_S–X_C case, with the elliptical shapes in the extended space largely consistent across different Asimov true values. The marginalisation biases also show the same consistency in both direction and magnitude across the parameter space. Known features of the conventional $\sin^2 2\theta_{23}$ (which corresponds to the horizontal axis in Fig. A.9) likelihood such as the W-shape and preference for upper octant can be understood as the intersection of the parabolic physical subspace with the likelihood surface. As with δ_{CP}, the constraint on the physical parameter $\sin^2\theta_{23}$ can be recovered by evaluating the 2D extended surface along the physical region defined by this parabola. As shown in Fig A.8, the agreement is generally very good, but diverges slightly at higher significance levels. This is understood to be caused by finite throw statistics.

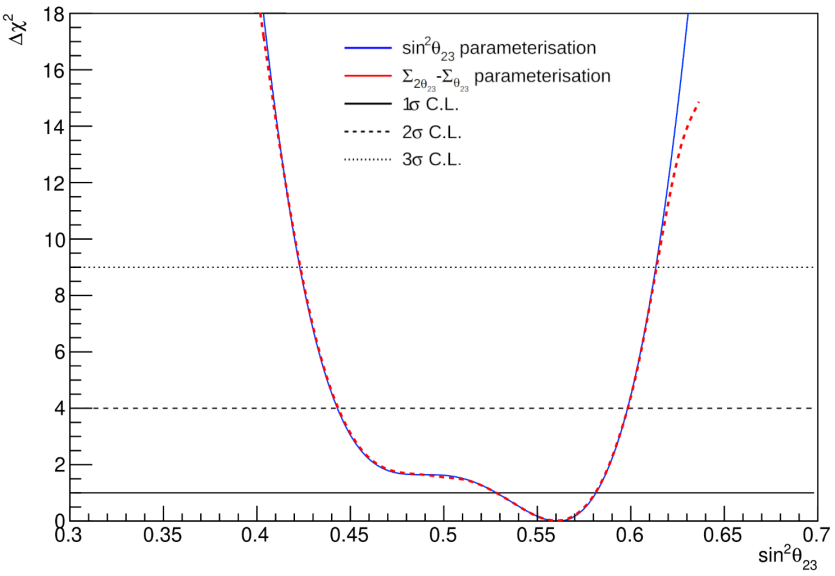


Figure A.7.: Comparison of $\sin^2\theta_{23}$ contour generated using the conventional parameterisation and that recovered by evaluating the physical region on the $\Sigma_{2\theta_{23}}$–$\Sigma_{\theta_{23}}$ fit.

Because of the subtlety that T2K does not directly measure $\sin^2 2\theta_{23}$, but rather the effective term $\sin^2 2\theta_{\mu\mu}$, which contains a θ_{13} correction,

$$\sin^2 2\theta_{\mu\mu} = 4c_{13}^2 s_{23}^2 \left(1 - c_{13}^2 s_{23}^2\right) \approx \sin^2 2\theta_{23}.$$

there is a slight correlation between $\Sigma_{\theta_{23}}$ and $\Sigma_{2\theta_{23}}$. We also note that the Gaussian contours in these fits are less elliptical than in the X_S—X_C case. We attribute this to a greater relative influence of higher-order terms in the asymptotic likelihood expansion. Denoting the $\Sigma_{\theta_{23}}$ model Poisson mean as $\lambda^0_{\Sigma_{\theta_{23}}}$ (the nought just means evaluated at the Asimov point), and similarly for either X_S or X_C as $\lambda^0_{X_S(X_C)}$, we write

$$\lambda^0_{\Sigma_{\theta_{23}}} = a\Sigma^0_{\theta_{23}} + b, \qquad \lambda^0_{X_S(X_C)} = cX^0_S(X^0_C) + d.$$

The cubic term in the expansion then takes the form,

$$\text{cubic terms} \propto \frac{a^3}{3(\lambda^0_{\Sigma_{\theta_{23}}})^3}\left[\Sigma_{\theta_{23}} - \Sigma^0_{\theta_{23}}\right]^3, \qquad \frac{c^3}{3(\lambda^0_{X_S(X_C)})^3}\left[X_S(X_C) - X^0_S(X^0_C)\right]^3.$$

In the e-like samples, $\sin^2\theta_{32}$ is the dominant term while the $\mathcal{CP}$-terms are subdominant. Consequently, $a > c$, and so for the same amount of statistics, higher-order asymptotics are larger when producing fits in $\Sigma^0_{\theta_{23}}$. To test this interpretation, we generated an Asimov fit for $\Sigma^0_{\theta_{23}}$ with both nominal statistics and ten times the statistics. The result, shown in Fig. A.7, demonstrates that with increased statistics, the contour becomes more parabolic.

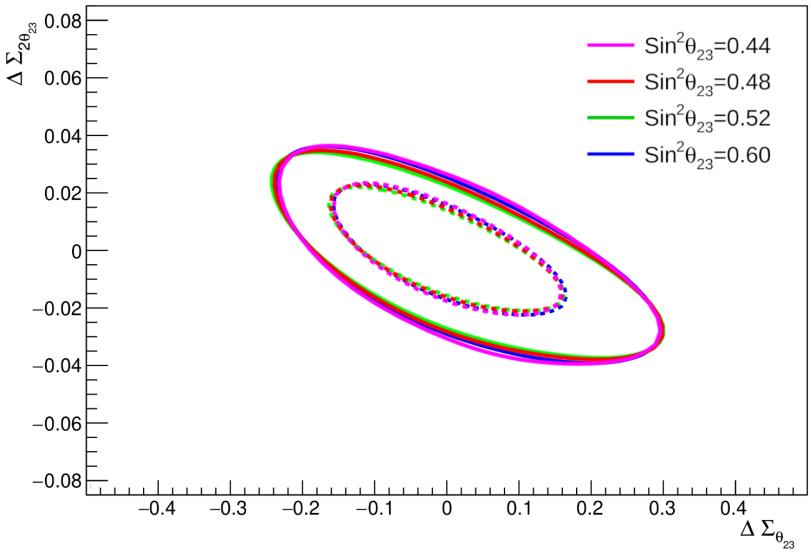
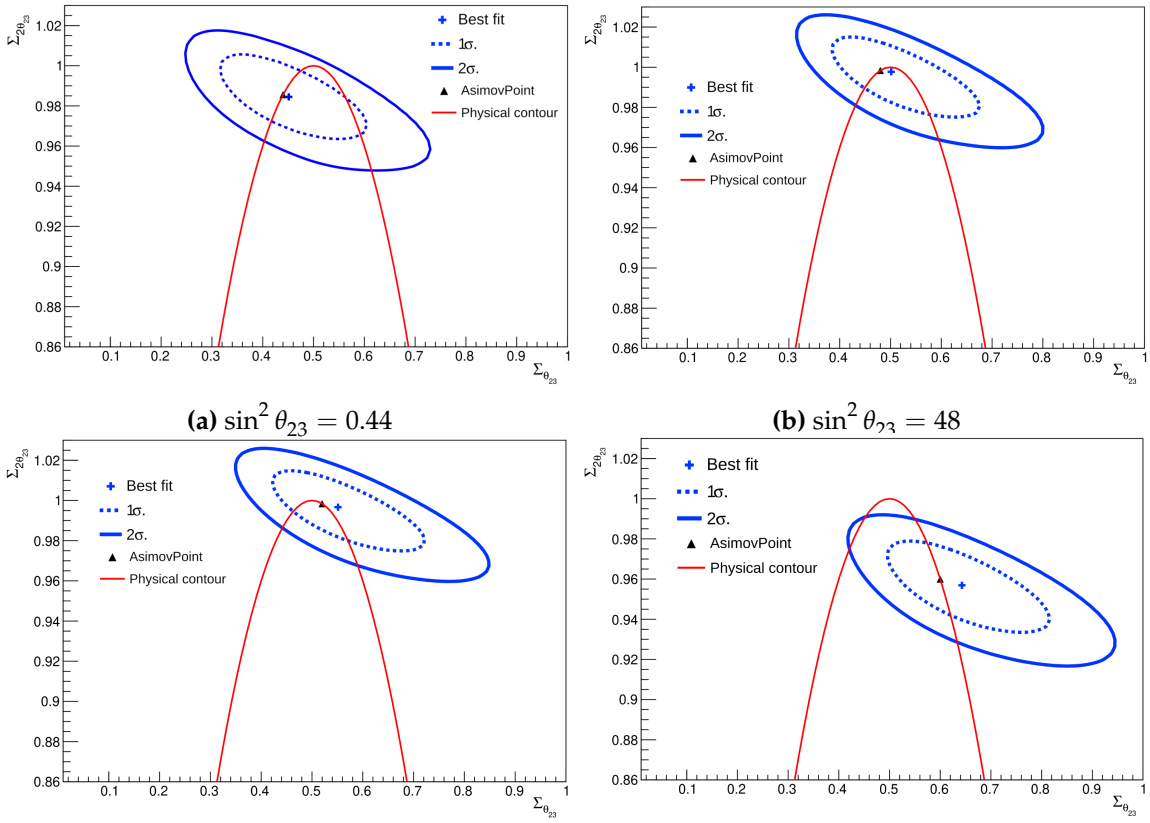


Figure A.8.: Shape comparison of contours generated at different true $\sin^2\theta_{23}$ values after correction for Poisson statistics. All other oscillation parameters at Asimov A values.



(a) $\sin^2\theta_{23} = 0.44$ **(b)** $\sin^2\theta_{23} = 48$

(c) $\sin^2\theta_{23} = 0.52$ **(d)** $\sin^2\theta_{23} = 0.6$

Figure A.9.: Asimov fits in $\Sigma_{2\theta_{23}}$–$\Sigma_{\theta_{23}}$ space for different true $\sin^2\theta_{23}$ values, with all other oscillation parameters fixed to their Asimov A values.

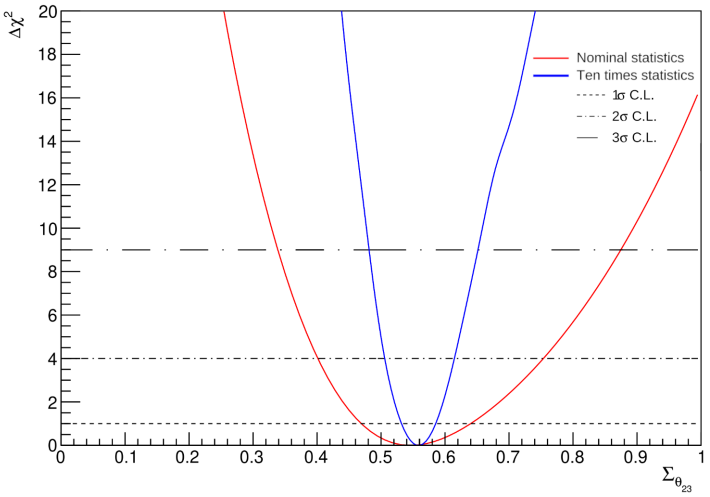


Figure A.10.: Comparison of $\Sigma_{\theta_{23}}$ contours at nominal and $\times 10$ statistics. As expected, the likelihood surface becomes significantly more parabolic at higher statistics.

A.2.3. All off-shell mode

In this mode, interpreting confidence contours is challenging due to the additional degrees of freedom. However, one interesting idea, although at this stage underdeveloped, is to produce the fit in six-dimensional space, and then depict the result in terms of a covariance matrix and six-dimensional best fit location. The effect of the fake data studies could be expressed as the shifts in the best fit values of six parameters, and the change in their associated covariance matrix (although as seen in the FDS chapter the shifts are expected to be the main effect). The dimensionality is six because the terms $\cos\delta_{\mathrm{CP}}\sin^2 2\theta_{23}$ and $\sin\delta_{\mathrm{CP}}\sin^2 2\theta_{23}$ must be treated as independent basis functions in order to preserve the linearity of the model. This is because for a probability of the form,

$$P_{Ext}(\epsilon(\omega_1,\omega_2)) = A\omega_1 + B\omega_2 + C\omega_1\omega_2,$$

the rate at which the prediction changes with respect to one parameter (e.g., ω_1) explicitly depends on the value of the other (ω_2). This would violate the foundational principle that the model should have a phase space that locally looks the same everywhere.

Additionally, in this mode the model is fully linear which would make it possible to examine correlations between oscillation parameters and individual systematic parameters.

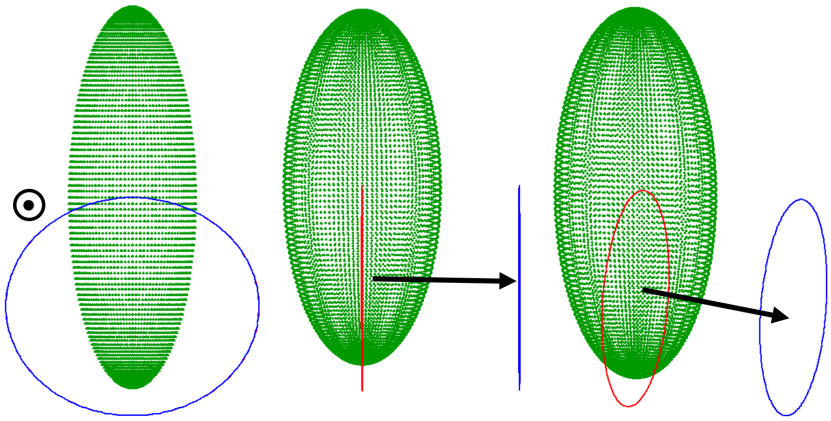
A.3. Mass ordering linearisation

Attempts were made to incorporate a linearised treatment of the mass ordering by introducing an extended MO parameter, $m \in \mathbb{R}$ and using the Fourier coefficient fitting method described above. However, we were unable to identify a probability model, using this coefficient fitting strategy, capable of reproducing the Prob3++ probability at the required 1% precision. Further work is required to understand this, however here we briefly outline a possible extension using an interpolation-based approach for future consideration.

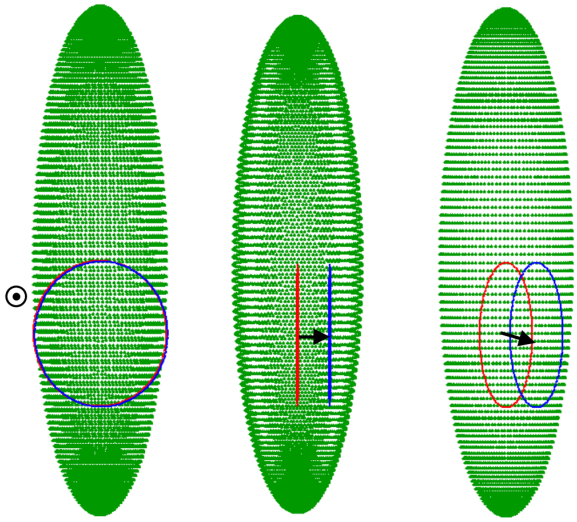
In some sense, MO linearisation is simpler than $\delta_{\mathcal{CP}}$ and θ_{23}. In those cases, the parameter space is extended from one to two dimensions by decoupling leading-order functions. For MO, we move from a discrete parameter (with zero dimensions) to a continuous one-dimensional space, for which an interpolation scheme can be easily devised: the probability assuming both orderings is computed and then projected into the extended space using,

$$P_{Ext} = \frac{1+m}{2} P(NO) + \frac{1-m}{2} P(IO), \tag{A.16}$$

with the physical regions corresponding to $m = -1$ (IO) and $m = 1$ (NO).Additional parameters can be incorporated by using probabilities that already include those extensions. For example, one may define $P_{X_S\text{–}X_C}(\text{NO})$ and $P_{X_S\text{–}X_C}(\text{IO})$, where $P_{X_S\text{–}X_C}$ denotes the familiar extended X_S–X_C probability. An illustrative schematic of the resulting 3-dimensional X_S–X_C–m space is shown in Fig. A.11 where the physical regions for the two mass orderings appear as two distinct rings, with the confidence-level boundary indicated by the green surface. The upper panel shows a confidence level for which the fit is compatible with one ordering but not the other (that is, the rings are separable). The lower panel shows a confidence level for which both rings lie within the Gaussian volume, meaning the data at this level is consistent with either ordering.



(a) Low confidence surface.



(b) High confidence surface.

Figure A.11.: Sketch of a hypothetical contour in X_S–X_C–m space. The green 3D surface (gherkin-shaped) denotes the boundary of the allowed region at a chosen confidence level. The three panels show rotated views of the same surface. The black arrow indicates the m parameter axis (the linearised MO parameter). The red and blue rings represent the physical regions within this space; one for each MO

Appendix B.

$\Delta\chi^2_c$ thresholds from linear models

In order to make frequentist statements, the critical threshold for the $\Delta\chi^2$ test statistic must be known. In simple cases such as a 1D gaussian these correspond to fixed differences ($'n\sigma' = \sqrt{(\Delta\chi^2)}$), but a common procedure in more complicated cases where it is computationally feasible is to calculate these thresholds with many randomised 'toy' experiments. In neutrino physics this is most commonly known as the Feldman-Cousins [88] prescription. In this chapter, we present an efficient procedure for obtaining these thresholds from the linearised oscillation probability.

B.1. Breakdown of Wilks' theorem

Due to the irregularities of the PMNS parameterisation at current statistics, Wilks' theorem fails, and so setting critical values using the standard χ^2 percentiles may lead to overcoverage or unacceptable undercoverage.

Consider, for instance, the case of $\delta_{\mathcal{CP}}$. Suppose the null hypothesis is set at a parameter value corresponding to maximal appearance. By definition, half of all possible statistical fluctuations at this point will produce appearance rates more extreme than the model can accommodate. Consequently, 50% of realisations will be fully compatible with the null hypothesis, leading to a pile-up at $\Delta\chi^2 = 0$. This shifts the probability mass downward and lowers the critical values compared to the Gaussian case (see Fig. B.1). Note that, as per the discussion of Sect. 3.4.4, this explanation is incomplete, as it does not consider the multi-bin energy structure of the analysis, the degeneracies with $\delta_{\mathcal{CP}}$, or the influence of other irregular parameters such as the mass ordering (MO) and θ_{23}, whose deficiencies also affect the $\Delta\chi^2$ critical values. It is, however, a useful simplified mental model for the discussions that follow

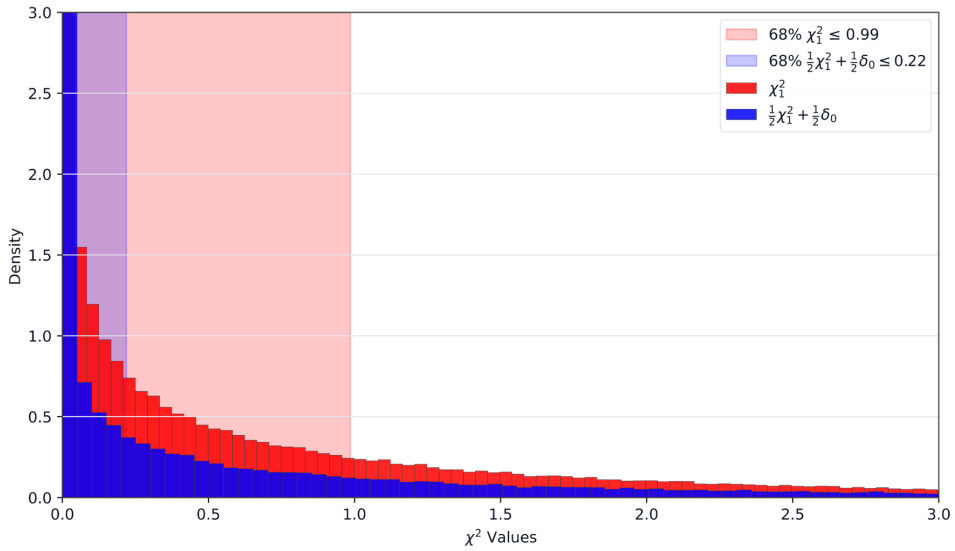


Figure B.1.: Density values drawn from two distributions: Red – χ^2 with 1 degree of freedom; Blue – mixture distribution with 50% at zero and 50% from $\chi^2(1)$. The sharp peak at $\chi^2 = 0$ in the blue histogram shifts the probability mass downward and shortens the 68% interval compared with the pure χ^2 case, as indicated by the blue and red bands.

B.2. Empirical intervals from toy experiments

In lieu of a suitable asymptotic approximation, T2K's frequentist analysis uses the Feldman-Cousins [88] method to obtain intervals with proper coverage. This is a brute-force strategy in which an ensemble of pseudo-experiments is generated by fixing $\delta_{\mathcal{CP}}$ while varying all other oscillation and systematic parameters. Necessarily, an MO must also be chosen, so separate thresholds are produced for each ordering. The critical values are then determined empirically from the distribution of $\Delta\chi^2$ obtained from these pseudo-experiments. The selected fixed points are $\delta_{\mathcal{CP}}/\pi = \{-0.75, -0.5, -0.25, 0, 0.125, 0.25, 0.5, 0.75, 0.825, 1\}$, with the critical values for the intermediate values inferred by linearly interpolating these points [44].

Unfortunately, this procedure is both computationally expensive and time consuming. For this reason, the number of pseudo-experiments at each fixed value is limited to 50,000. This is sufficient to accurately map the tails out to 3σ CL, which corresponds to a Type II error rate of 1 in 370. However, for higher CL, for instance the 5σ discovery threshold, which corresponds to a type-II error rate of 1 in 1.7 million this is completely insufficient. In fact, at such high significance levels, literally millions of pseudo-experiments would be required to obtain an accurate approximation of the critical value. This presents a serious practical challenge that future experiments, such as the successor to T2K, Hyper-Kamiokande, will need to address.

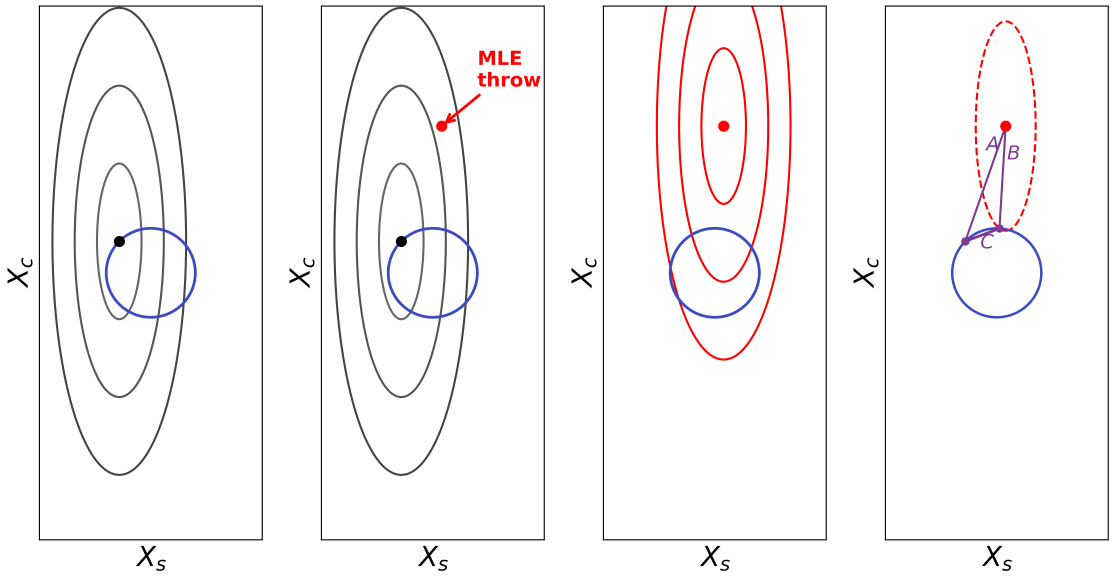


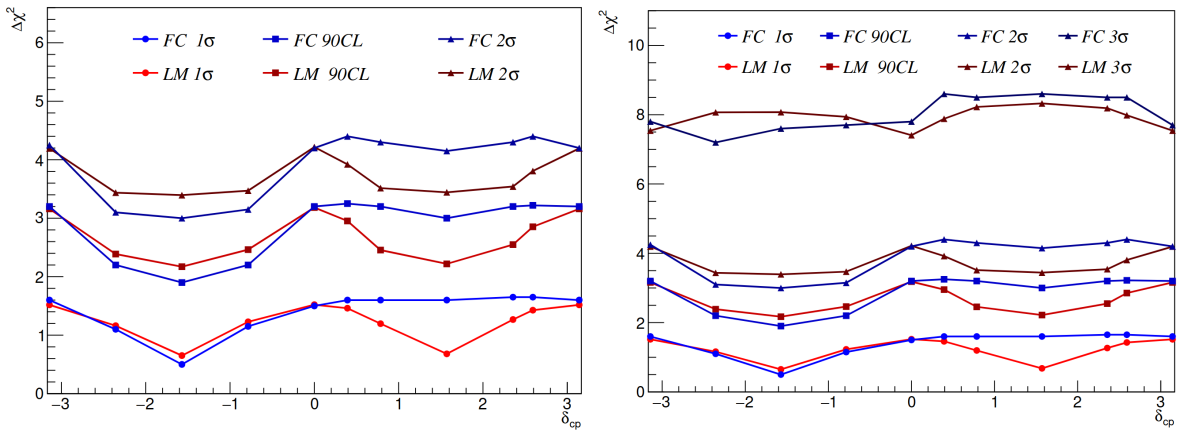
Figure B.2.: Procedure for obtaining the physical $\Delta\chi^2$ critical values. From left to right: 1) a Gaussian with predetermined width is centered on the true value; 2) an MLE candidate is sampled from this surface; 3) a Gaussian of the same width is then centered on the sampled MLE; 4) the physical best-fit point on the unit circle is identified, and the physical $\Delta\chi^2$ (the "C" distance) is obtained using Eq. (4.4).

B.3. Critical values from linear model

Here we show that the linearised oscillation model can be used as an efficient alternative to FC. The idea is that, in the extended parametrisation, local asymptotic normality (see Sec. 3.4) applies, so both the distribution of the MLE and the shape of the likelihood surface for a given realised MLE are known gaussian. In fact, because the observed information converges to the Fisher information in the large-sample limit, both share the same covariance.

This suggests the following strategy: generate an ensemble of candidate MLE points by sampling from the known Gaussian surface in the extended space. The corresponding physical $\Delta\chi^2$ can then be recovered using Eq. (4.4), with the covariance matrix defining the $\Delta\chi^2_{\text{Ext}}$ surface taken to be the same one used to generate the MLE. To obtain this matrix and understand the scaling required to correct for Poisson statistics, only a few simulated data fits are needed.

As a simple proof of concept, we demonstrate this procedure assuming true Normal Ordering and using the X_S–X_C linear model. We fix the mass ordering and assume precise knowledge of θ_{23} in order to isolate the effect of the non-linear dependence on $\delta_{\mathcal{CP}}$ in generating non-flat critical values. This represents the simplest generalisation of the archetypal one-dimensional fit with a boundary. The steps of the procedure are illustrated schematically in Fig. B.2.



(a) 1σ, 90%, 2σ critical values.

(b) 1σ, 90%, 2σ, 3σ critical values.

Figure B.3.: Comparison of critical values obtained with the linear model (blue) and the official FC method (red).

The critical values obtained under different true points are shown in Fig. B.3 and compared with the official FC values. Our values are higher than those produced by the standard F-C method used by T2K, but this disagreement is expected since the irregularities of the θ_{23} and MO degeneracies are neglected. For example, $\delta_{\mathcal{CP}} = \pi/2$ is not a genuine boundary because the model can accommodate the predictions beyond the true NO maximum by allowing the best fit to move to the opposite ordering (this is why the FC values have no significant bump at $\delta_{\mathcal{CP}} = \pi/2$).

Nevertheless, the relative shape appears approximately correct, which given the simplicity of this implementation and how much is being neglected, is quite encouraging. This does not apply to the 3σ however, this is expected given we know the likelihood surface of the extended space can deviate from Gaussianity at this significance. It is hoped that implementing a model that also accounts for these other irregularities, will reduce these discrepancies.

The key point, however, is the efficiency of this procedure. Each critical value required less than 20 minutes to generate on a single modest CPU, compared with the hundreds of hours typically needed for the conventional FC method. Even if it turns out to be necessary to use the full F-C procedure to account for all systematic and marginalisation effects, this much faster calculation could be used to interpolate between a smaller number of expensive-to-compute points on the critical surface, reducing the amount of computation required to map out the critical surface with sufficient accuracy.

B.4. Increasing exposure

It is well established that these critical values vary with exposure [44]. An interesting application of the above framework is to visualise how these thresholds evolve as statistics increase. To demonstrate this, we apply the procedure described above while progressively decreasing the Gaussian width, which serves as a proxy for increased exposure. For this, a symmetric Gaussian (equal sensitivity in both the X_S and X_C directions) was used so that the true point is unimportant. The red and black concentric circles of Fig. B.4b show the size of the 90%CL in the (X_S, X_C) space (smaller circles correspond to higher statistics), while the corresponding 1σ critical values are marked by the red and black points in Fig. B.4a. The reason the 90% CL are show in Fig. B.4b is because global changes in probability mass affect all critical values, so it is useful to see how the vast majority of the density lies in relation to the physical circle.

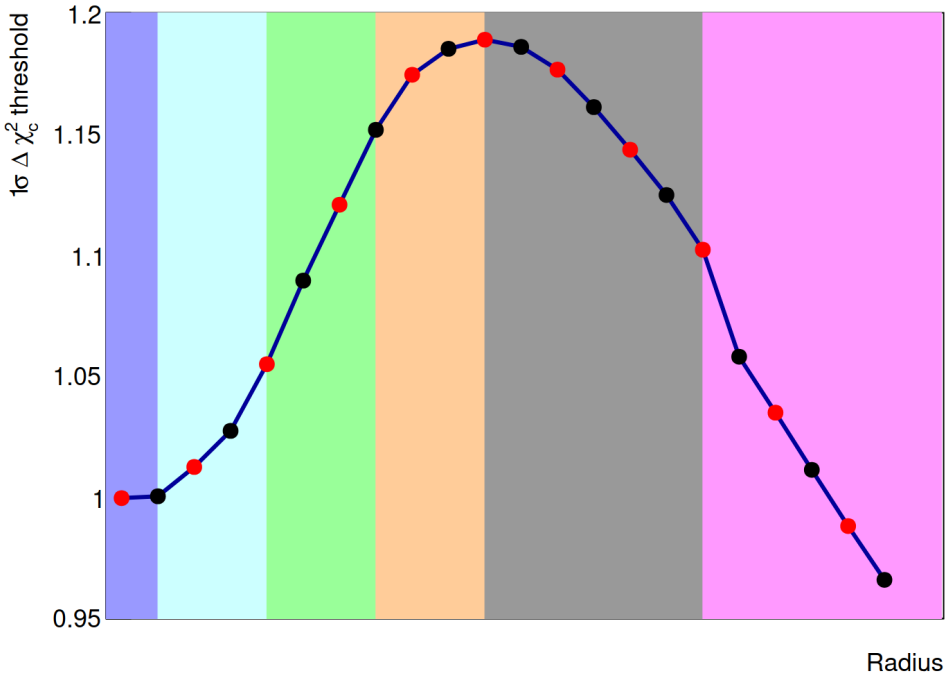
B.4.1. Remarks

What we are essentially doing is examining the distribution of the distance marked 'C' in Fig. B.2 (this is why a symmetric sensitivity was used as it avoids needing to mentally account for differences in X_S–X_C sensitivities).

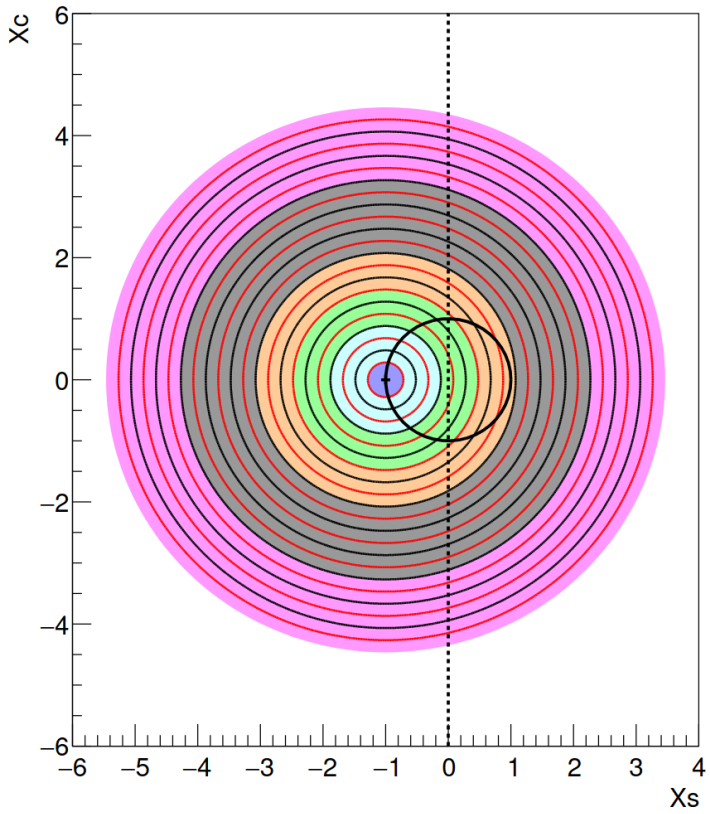
We note that not only do the critical values change with increasing Gaussian width, but the rate of this change itself evolves. In fact, the behaviour appears to pass through several identifiable stages, which we have marked in different colours in Fig. B.4a. Below we attempt to attach an interpretation to each, going from smallest width to largest (or left to right on Fig. B.4a):

- Blue: In this region the physical subspace segments the Gaussian contour as a line. The constraint therefore behaves like a Gaussian, with $\Delta\chi^2_c \approx 1$ (i.e. we recover Wilks' theorem).
- Light blue: As the elliptical contour widens, a larger portion of the physical circle, and therefore its curvature, becomes viable. Consequently, the critical values increase and begin to deviate from Gaussian behaviour.
- Green: In this region, degeneracies begin to play a significant role, as even relatively mild fluctuations can migrate toward, and therefore produce best-fit points on, the far side of the physical subspace (the side opposite to the assumed true point). These realisations correspond to much larger values of $\Delta\chi^2$, and consequently the critical values rise sharply.
- Orange: As a larger fraction of the circle becomes visible, the rate of this increase starts to quench, and the previously large "C" distances become more typical.

- Grey: At this point, a significant number of throws extend outside the physical circle in all directions. As a result, the trend changes and the circle values begins to get smaller.
- Cyan: Eventually, a critical mass is reached where the Physical parameter space appears increasing point-like compared to the Gaussian contour. In this case all toy experiments have similar $\Delta\chi^2$ values regardless of the value of $\delta_{\mathcal{CP}}$ and this causes the critical value of $\delta_{\mathcal{CP}}$ to drop rapidly.



(a)



(b)

Figure B.4.: Critical values as a function of increasing Gaussian width. The red and black concentric contours in the lower panel correspond to the red and black critical values shown in the upper panel.

Appendix C.

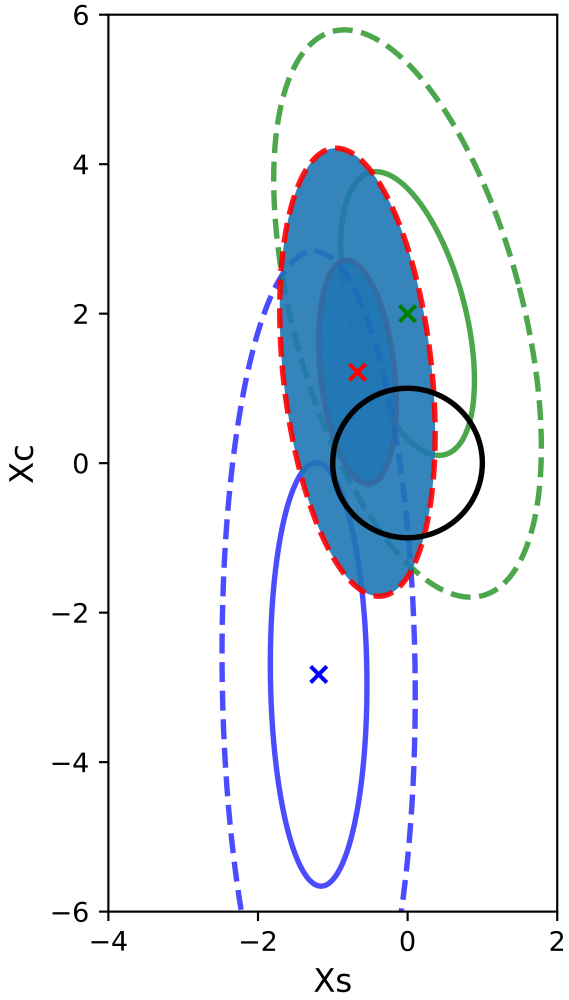
Combining experimental constraints

One interesting application of the linearisation concept is its potential to enable simple combinations of constraints across multiple experiments.

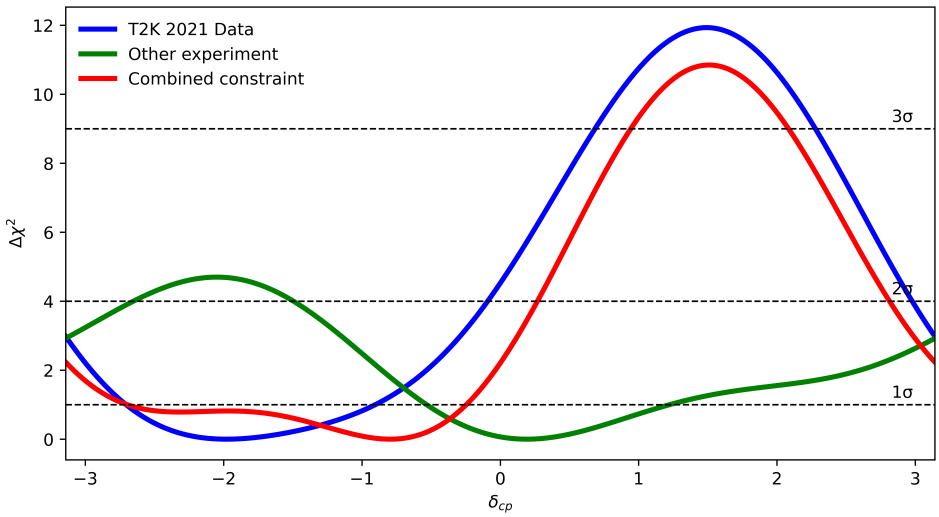
It is common for experiments with complementary measurements to produce joint fits. This helps to break degeneracies and improve sensitivity to parameters measured by both experiments. Most recently, T2K and NOvA (a US-based long-baseline neutrino experiment [97, 98]) have been preparing a joint analysis. A rigorous implementation of this type of analysis, however, requires a huge effort on the part of the participating collaborations. While this is of course the preferable approach, the time and person-power required often make it impractical.

In the absence of a full joint fit, linearised models provide a simple and efficient way to produce approximate joint constraints. The key idea is that sensitivities are Gaussian in the extended linear space, and combining Gaussian constraints is trivial since the combination of Gaussians is itself a Gaussian. Therefore, each individual experiment can perform their respective fits and the results can then be combined in that space. The physical joint fit is then recovered by projecting back into the physical subspace. This can be done very easily if the individual results can be well approximated by 2D Gaussians. But even if the full likelihood calculation is retained, the fact that interpretation of systematics is typically more straightforward in the extended space could also help when identifying important systematic uncertainties for the combination.

To sketch this idea we have performed this using the X_S–X_C linear model. The Gaussian dimension of the T2K-OA21 data fit has been combined this with the constraints of a hypothetical experiment with similar sensitivity but apparently contradictory conclusions about $\delta_{\mathcal{CP}}$. The results are shown in Fig .C.1. In this example the joint fit in $\delta_{\mathcal{CP}}$ seems paradoxical, as the combined interval is larger than the T2K constraint alone. However in the X_S–X_C space we can see why: The combined result technically has greater sensitivity, as evidenced by the smaller elliptical contour, but it exhibits larger intervals because its best-fit value lies closer to the physical subspace.



(a) Blue contour: Gaussian constraint from the T2K 2021 data fit. Green contour: constraint from a hypothetical experiment. Red: contour from combining both.



(b) Physical $\delta_{\mathcal{CP}}$ constraint obtained by evaluating the 2D likelihood surfaces on the unit circle.

Figure C.1.: Combined experimental constraints in X_S–X_C space and $\delta_{\mathcal{CP}}$.

Bibliography

[1] E. C. Drummond and W. A. Wooster, Proceedings of the Royal Society of London. Series A **1**17, 109 (1927).

[2] A. Franklin, *The Spectrum of β Decay: Continuous or Discrete? A Variety of Errors in Experimental Investigation* (Springer Netherlands, Dordrecht, 2009), pp. 211–235.

[3] D. Griffiths, *Introduction to elementary particles* (John Wiley & Sons, 2008).

[4] W. Pauli, Phys. Today **3**1N9, 27 (1978).

[5] C. W. Kim and A. Pevsner, *Neutrinos in physics and astrophysics* (Hardwood academic publishers, 1993).

[6] F. L. Wilson, American Journal of Physics **3**6, 1150 (1968).

[7] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Science **1**24, 103 (1956).

[8] R. Davis, D. S. Harmer, and K. C. Hoffman, Phys. Rev. Lett. **2**0, 1205 (1968).

[9] Kamiokande-II, K. S. Hirata *et al.*, Phys. Rev. Lett. **6**3, 16 (1989).

[10] A. I. Abazov *et al.*, Phys. Rev. Lett. **6**7, 3332 (1991).

[11] GALLEX, P. Anselmann *et al.*, Phys. Lett. B **2**85, 376 (1992).

[12] Kamiokande-II, K. S. Hirata *et al.*, Phys. Lett. B **2**05, 416 (1988).

[13] D. Casper *et al.*, Phys. Rev. Lett. **6**6, 2561 (1991).

[14] Super-Kamiokande, Y. Fukuda *et al.*, Phys. Rev. Lett. **8**1, 1562 (1998), arXiv: hep-ex/9807003.

[15] SNO, Q. R. Ahmad *et al.*, Phys. Rev. Lett. **8**9, 011301 (2002), nucl-ex/0204008.

[16] The Nobel Prize Committee, The Nobel Prize in Physics 2015 – Takaaki Kajita, Arthur B. McDonald, https://www.nobelprize.org/prizes/physics/2015/summary/, 2015, Awarded for the discovery of neutrino oscillations, which shows that neutrinos have mass.

[17] A. Ereditato, editor, *The State of the Art of Neutrino Physics: A Tutorial for Graduate Students and Young Researchers* (World Scientific, Singapore, 2018).

[18] Particle Data Group Collaboration, S. Navas *et al.*, Phys. Rev. D **1**10, 030001 (2024).

[19] F. Englert and R. Brout, Physical Review Letters **1**3, 321 (1964).

[20] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Physical Review Letters **1**3, 585 (1964).

[21] P. W. Higgs, Physical Review **1**45, 1156 (1966).

[22] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group, S. Schael *et al.*, Phys. Rept. **4**27, 257 (2006), arXiv: hep-ex/0509008.

[23] G. Voutsinas, E. Perez, M. Dam, and P. Janot, Phys. Lett. B **8**00, 135068 (2020), arXiv: 1908.01704.

[24] H. Nunokawa, S. J. Parke, and J. W. F. Valle, Prog. Part. Nucl. Phys. **6**0, 338 (2008), arXiv: 0710.0554.

[25] N. Cabibbo, Phys. Rev. Lett. **1**0, 531 (1963).

[26] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **4**9, 652 (1973).

[27] P. B. Denton, arXiv: 2501.08374 (2025).

[28] G. Altarelli and F. Feruglio, New J. Phys. **6**, 106 (2004), arXiv: hep-ph/0405048.

[29] CUORE, D. Q. Adams *et al.*, Nature **6**04, 53 (2022), arXiv: 2104.06906.

[30] G. Wang et al., Cupid: Cuore (cryogenic underground observatory for rare events) upgrade with particle identification, 2015, arXiv: 1504.03599.

[31] B. Pontecorvo, Sov. Phys. JETP **7**, 172 (1958).

[32] B. Pontecorvo, Sov. Phys. JETP **6**, 429 (1957).

[33] B. Pontecorvo, Zh. Eksp. Teor. Fiz. **5**3, 1717 (1967).

[34] Z. Maki, M. Nakagawa, and S. Sakata, Progress of Theoretical Physics **2**8, 870 (1962), https://academic.oup.com/ptp/article-pdf/28/5/870/5258750/28-5-870.pdf.

[35] B. Kayser, eConf **C**040802, L004 (2004), arXiv: hep-ph/0506165.

[36] C. Giunti and C. W. Kim, *Fundamentals of neutrino physics and astrophysics* (Oxford university press, 2007).

[37] F. J. P. Soler, Colin D. Froggatt, Franz Muheim, *Neutrinos in particle physics, astrophysics and cosmology* (CRC Press, 2008).

[38] P. B. Denton and R. Pestes, JHEP **0**5, 139 (2021), arXiv: 2006.09384.

[39] C. Jarlskog, Phys. Rev. Lett. **5**5, 1039 (1985).

[40] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Phys. Rev. Lett. **1**3, 138 (1964).

[41] L. Wolfenstein, Physical Review D **1**7, 2369 (1978).

[42] S. P. Mikheyev and A. Yu. Smirnov, Sov. J. Nucl. Phys. **4**2, 913 (1985), [Yad. Fiz.42,1441(1985); ,305(1986)].

[43] G. Barenboim, P. B. Denton, S. J. Parke, and C. A. Ternes, Phys. Lett. B **7**91, 351 (2019), arXiv: 1902.00517.

[44] T2K, K. Abe *et al.*, Phys. Rev. D **1**03, 112008 (2021), arXiv: 2101.03779.

[45] R. A. Wall, Prob3++, https://webhome.phy.duke.edu/~raw22/public/Prob3++/.

[46] M. Ghosh, P. Ghoshal, S. Goswami, N. Nath, and S. K. Raut, Phys. Rev. D **9**3, 013013 (2016), arXiv: 1504.06283.

[47] H. Minakata and S. Uchinami, JHEP **0**4, 111 (2010), arXiv: 1001.4219.

[48] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967).

[49] JUNO, F. An *et al.*, J. Phys. G **4**3, 030401 (2016), 1507.05613.

[50] F. Suekane, *Neutrino Oscillations: A Practical Guide to Basics and Applications*, Lecture Notes in Physics Vol. 898 (Springer, 2015).

[51] S. E. Kopp, Phys. Rept. **4**39, 101 (2007), arXiv: physics/0609129.

[52] M. Finkemeier, Radiative corrections to π_{l2} and k_{l2} decays, 1996, arXiv: hep-ph/9501286.

[53] A. Longhin, L. Ludovici, and F. Terranova, Eur. Phys. J. C **7**5, 155 (2015), arXiv: 1412.5987.

[54] NA61/SHINE, N. Abgrall *et al.*, Eur. Phys. J. C **7**6, 617 (2016), arXiv: 1603.06774.

[55] T2K, K. Abe *et al.*, Eur. Phys. J. C **8**3, 782 (2023), arXiv: 2303.03222.

[56] J. L. Hewett *et al.*, Fundamental physics at the intensity frontier, 2012, arXiv: 1205.2671.

[57] K. V. J. Mistry, *Neutrino Interactions* (Springer International Publishing, Cham, 2023), pp. 15–32.

[58] T2K, K. Abe *et al.*, Nucl. Instrum. Meth. A **6**59, 106 (2011), arXiv: 1106.1238.

[59] T2K, K. Abe *et al.*, Phys. Rev. D **8**7, 012001 (2013), 1211.0469, [Addendum: Phys.Rev.D 87, 019902 (2013)].

[60] C. Densham *et al.*, Design and Development of the T2K Pion Production Target, in *Particle Accelerator Conference (PAC 09)*, p. WE6RFP033, 2010.

[61] T2K, T. Sekiguchi *et al.*, Nucl. Instrum. Meth. A **7**89, 57 (2015), arXiv: 1502.01737.

[62] T2K, K. Abe *et al.*, Phys. Rev. D **9**6, 092006 (2017), 1707.01048, [Erratum: Phys.Rev.D 98, 019902 (2018)].

[63] S. E. Kopp, Phys. Rept. **4**39, 101 (2007), arXiv: physics/0609129.

[64] E. T. Atkin, *Neutrino Oscillations Analysis at the T2K experiment including studies of new uncertainties on interactions involving final state hadrons*, PhD thesis, Imperial College London, 2022.

[65] T2K, K. Suzuki, Nucl. Phys. B Proc. Suppl. **2**29-232, 453 (2012).

[66] T2K ND280 FGD, P. A. Amaudruz *et al.*, Nucl. Instrum. Meth. A **6**96, 1 (2012), arXiv: 1204.3666.

[67] T2K ND280 TPC, N. Abgrall *et al.*, Nucl. Instrum. Meth. A **6**37, 25 (2011), arXiv: 1012.0865.

[68] T2K, K. Abe *et al.*, JHEP **1**0, 114 (2020), arXiv: 2002.11986.

[69] ND280 Collaboration, T2K ND280 Detector Technical Design Report, 2006.

[70] T2K UK, D. Allan *et al.*, JINST **8**, P10019 (2013), arXiv: 1308.3445.

[71] S. Dolan *et al.*, Phys. Rev. D **1**05, 032010 (2022), arXiv: 2108.11779.

[72] K. Yasutome *et al.*, T2K Collaboration Report No. Technote 430, version 2.0.5, 2023 (unpublished).

[73] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*, 2 ed. (Springer, Berlin, 1994).

[74] Super-Kamiokande, Y. Fukuda *et al.*, Nucl. Instrum. Meth. A **5**01, 418 (2003).

[75] C. W. Walter, The Super-Kamiokande experiment, in *Physics and Astrophysics of Neutrinos*, pp. 19–43, World Scientific, 2008, arXiv: 0802.1041.

[76] A. Goldsack, *Reactor neutrinos in a gadolinium-loaded Super-Kamiokande*, DPhil thesis, University of Oxford, 2022.

[77] Super-Kamiokande, M. Jiang *et al.*, PTEP **2**019, 053F01 (2019), arXiv: 1901.03230.

[78] A. N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd english ed. ed. (Chelsea Publishing Company, New York, 1956), Originally published in 1933 in Russian.

[79] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).

[80] F. James, *Statistical Methods in Experimental Physics*, 2 ed. (World Scientific, 2006).

[81] Particle Data Group, M. Tanabashi *et al.*, Phys. Rev. D **9**8, 030001 (2018).

[82] L. Mellet, *From T2K to Hyper-Kamiokande : neutrino oscillation analysis and preparation of the time synchronization system*, Theses, Sorbonne Université, 2023.

[83] T. J. Loredo, Slides for Banff Discovery Workshop, https://www.astro.cornell.edu/people/timothy-loredo, 2010, Department of Astronomy, Cornell University.

[84] K. Cranmer, Practical Statistics for the LHC, 2014, arXiv: 1503.07622.

[85] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics Vol. 3 (Cambridge University Press, 2000).

[86] A. R. Brazzale and V. Mameli, Likelihood asymptotics in nonregular settings: A review with emphasis on the likelihood ratio, 2023, arXiv: 2206.15178.

[87] S. Algeri, J. Aalbers, K. Dundas Morå, and J. Conrad, Nature Reviews Physics **2**, 245 (2020), arXiv: 1911.10237.

[88] G. J. Feldman and R. D. Cousins, Physical Review D **57**, 3873 (1998), arXiv: physics/9711021.

[89] H. Chernoff, The Annals of Mathematical Statistics **25**, 573 (1954).

[90] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C **71**, 1554 (2011), arXiv: 1007.1727, [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[91] T2K Collaboration / Technical Note 396 Report No.,

[92] H. Li, *Support Vector Machine* (Springer Nature Singapore, Singapore, 2024), pp. 127–177.

[93] H. Yokomakura, K. Kimura, and A. Takamura, Phys. Lett. B **544**, 286 (2002), arXiv: hep-ph/0207174.

[94] Intensity Frontier Neutrino Working Group, A. de Gouvea *et al.*, Working Group Report: Neutrinos, in *Snowmass 2013: Snowmass on the Mississippi*, 2013, arXiv: 1310.4340.

[95] C. Bronner, Accelerator neutrino recent results from T2K, https://doi.org/10.5281/zenodo.6683821/, 2022.

[96] D. Barrow *et al.*, T2K Collaboration Report No. Technical Note 399, 2020 (unpublished), Internal technical note.

[97] F. N. A. Laboratory, Fermi National Accelerator Laboratory Report No. Design-2007-01, 2007 (unpublished).

[98] The NOvA Collaboration, S. Abubakar *et al.*, Phys. Rev. Lett. **136**, 011802 (2026).

List of figures

List of tables