EURONU-WP6-12-53
FERMILAB-PUB-12-509-T
IDS-NF-036

Systematic uncertainties in long-baseline neutrino oscillations for large θ_{13}

September 27, 2012

Pilar Coloma[a], Patrick Huber[b],
Joachim Kopp[c], and Walter Winter[d]

[a,b] *Center for Neutrino Physics, Virginia Tech, Blacksburg, VA 24061, USA*
[c] *Fermilab, P.O. Box 500, Batavia, IL 60510-0500, USA and*
Max-Planck-Institut für Kernphysik, PO Box 103980, 69029 Heidelberg, Germany
[d] *Institut für Theoretische Physik und Astrophysik, Universität Würzburg,*
D-97074 Würzburg, Germany

Abstract

We study the physics potential of future long-baseline neutrino oscillation experiments at large θ_{13}, focusing especially on systematic uncertainties. We discuss superbeams, beta-beams, and neutrino factories, and for the first time compare these experiments on an equal footing with respect to systematic errors. We explicitly simulate near detectors for all experiments, we use the same implementation of systematic uncertainties for all experiments, and we fully correlate the uncertainties among detectors, oscillation channels, and beam polarizations as appropriate. As our primary performance indicator, we use the achievable precision in the measurement of the CP violating phase δ. We find that a neutrino factory is the only instrument that can measure δ with a precision similar to that of its quark sector counterpart. All neutrino beams operating at peak energies $\gtrsim 2$ GeV are quite robust with respect to systematic uncertainties, whereas especially beta-beams and T2HK suffer from large cross section uncertainties in the quasi-elastic regime, combined with their inability to measure the appearance signal cross sections at the near detector. A noteworthy exception is the combination of a $\gamma = 100$ beta-beam with an SPL-based superbeam, in which all relevant cross sections can be measured in a self-consistent way. This provides a performance, second only to the neutrino factory. For other superbeam experiments such as LBNO and the setups studied in the context of the LBNE reconfiguration effort, statistics turns out to be the bottleneck. In almost all cases, the near detector is not critical to control systematics since the combined fit of appearance and disappearance data already constrains the impact of systematics to be small provided that the three active flavor oscillation framework is valid.

[a]Email: pcoloma@vt.edu
[b]Email: pahuber@vt.edu
[c]Email: jkopp@fnal.gov
[d]Email: winter@physik.uni-wuerzburg.de

1 Introduction

The story of large θ_{13} has unfolded in fast succession from first hints in global fits [1–4], via direct indications from T2K [5], MINOS [6] and Double Chooz [7] to a discovery by Daya Bay [8], which was soon confirmed by RENO [9]. A recent global fit yields $\sin^2\theta_{13} = 0.023 \pm 0.0023$ [10] (see also Refs. [11,12], which find very similar values), where the error bars are entirely dominated by the reactor measurements. The precision of reactor experiments on θ_{13} will continue to improve and is not expected to be exceeded by beam experiments anytime soon (see, for instance, Ref. [13]).

The most important open questions in neutrino oscillations, within the context of three active flavors, are the determination of the neutrino mass hierarchy ($\mathrm{sgn}(\Delta m^2_{31})$) and the measurement of the CP violating phase δ. While there might be already some weak evidence for $\delta \sim \pi$ from global fits [11, 12], high confidence level CP violation (CPV) and mass hierarchy measurements cannot be performed with existing facilities, such as Daya Bay, RENO, Double Chooz, T2K, and NOνA in spite of the relatively large value of θ_{13} [14]. In the most aggressive scenario, *i.e.*, for upgraded proton drivers for both T2K and NOνA and mutually optimized neutrino-antineutrino running plans, CPV could be established at 3σ confidence level only for 25% of all values of δ. Therefore, a next generation of experiments is mandatory and a decision towards one of the proposed technologies—superbeam upgrades, a beta-beam or a neutrino factory—will soon be needed.

The determination of the mass hierarchy need not necessarily be performed in long-baseline experiments, given the relatively large value of θ_{13}. An independent determination of the mass hierarchy may be provided from the combination of T2K, NOνA and INO [15], from new proposals such as PINGU [16], from a reactor experiment with a relatively long baseline [17–19],[1] or from the combination of reactor and long baseline experiments with very high precision [21, 22]. Almost all of the long-baseline experiments studied in this work would allow for a high confidence level mass hierarchy discovery because of the sufficient length of the baselines and the chosen neutrino energies (see, for instance, Refs. [23–25]). Those setups with shorter baselines $\lesssim 500$ km, where it is not possible to determine the mass hierarchy from the long-baseline data alone, would have a very massive detector. In these cases, a large sample of atmospheric neutrino events will be available which in combination with the beam data allows for an extraction of the mass hierarchy [26–29]. Therefore, we will not focus on this observable in this study.

Regarding δ, the main focus in the literature so far has been on the question whether CPV can be detected, *i.e.*, whether the CP conserving cases $\delta = 0, \pi$ can be excluded. The discovery of leptonic CPV would support thermal leptogenesis [30], which could potentially lead to an explanation of the observed baryon–antibaryon asymmetry of the Universe—although a direct connection to the CP violating phases in the high energy theory can only be established in a model-dependent way. The CP asymmetry in vacuum is linearly proportional to $\sin\delta$, and great efforts have been made to optimize neutrino oscillation facilities for maximal sensitivity to this term. However, there are good reasons why $\cos\delta$ is also interesting. For example, if the neutrino mass matrix is determined by a symmetry to have the tri-bimaximal

[1] This may be rather challenging from the experimental point of view, see Ref. [20] for instance.

(bi-maximal) form, corrections originating from the charged lepton mass matrix may lead to the sum rule [31–33] $\theta_{12} \simeq 35^\circ\,(45^\circ) + \theta_{13}\,\cos\delta$. It is obvious that establishing such sum rules, which usually depend on $\cos\delta$, requires the measurement of the $\cos\delta$-term. An ideal long-baseline experiment would therefore have a relatively "flat" performance independent of δ and would be able to measure both terms with similar precision. The ultimate goal will be the measurement of δ with a precision comparable to the one achieved in the quark sector. In order to capture the whole parameter space in δ for fixed θ_{13} (or a relatively small range of θ_{13}), so-called "CP patterns" were proposed in Ref. [34, 35] to quantify the achievable precision as a function of the true δ. In Ref. [13] this dependence was studied in detail for different types of experiments, and the main factors that affect the achievable precision were identified. From the results presented in Ref. [13] it is clear that especially narrow band beams and setups with short baselines are typically optimized for the CPV measurement, *i.e.*, a good precision in the measurement of δ around the particular values 0 and π. On the other hand, more complicated (asymmetric) patterns arise in wide-band beams or in the presence of matter effects.

The key issue for long-baseline experiments at large θ_{13} is systematics. It is well known that especially signal normalization uncertainties affect neutrino oscillation measurements for large θ_{13}, see, *e.g.*, Refs. [23, 36, 37]. While in phenomenological studies near detectors are only in rare cases explicitly included or discussed, see *e.g.* Ref. [36] for T2HK and Ref. [38] for the neutrino factory, it is usually assumed that these can be described by an effective systematic error in the far detector in the range from 1% to 10%. The chosen values are "educated guesses" in the absence of explicit near detector simulations. This is unsatisfactory given the large impact of systematic uncertainties at large θ_{13}. Indeed, it is not even sufficient to use realistic numbers for the systematic errors, but it is equally important to implement them in an appropriate way, in particular taking into account correlations between the errors affecting different oscillation channels, different parts of the energy spectrum, etc. For instance, most conventional simulations assume that systematics are uncorrelated among all oscillation channels, but fully correlated among all energy bins and backgrounds. In the real world, cross sections are correlated among all channels measuring the same final flavor, fluxes among all channels in the same beam, *etc.* Furthermore, it is known that the matter density uncertainty affects the measurements for large θ_{13} for experiments with long baselines and high energies, see, *e.g.*, Refs. [39, 40].

In this study, we will explore the effect of systematic errors on the achievable precision in different experiments, and we will provide a detailed comparison between different setups under the same assumptions for the systematics. Our systematics treatment is an extension of the one used for multi-detector reactor experiment simulations [41–43]. In particular,

1. we use a detailed, physics-based and self-consistent systematics implementation including correlations, which is comparable for all experiments;

2. we explicitly simulate the near detectors, with comparable assumptions regarding statistics and geometry for all experiments;

3. we do not only choose particular values for the systematic errors, but we also study ranges which span the gamut from conservative to optimistic;

4. we use exactly the same assumptions for cross section and matter density uncertainties for all experiments. For the systematic uncertainties that depend on the particular type of neutrino beam (for instance, flux uncertainties or intrinsic beam backgrounds) we consider the same values for all experiments of the same type.

To facilitate the comparison between different facilities, we will use as a performance indicator the fraction of possible values of δ for which a certain precision can be obtained, similar to earlier figures showing the CPV performance. We thus treat the whole parameter space on equal footing, and our conclusions on the relative sensitivity of different experiments will not depend on any assumed "true" value of δ.[2]

For the experiment definitions and simulations, we have modified the AEDL language (Abstract Experiment Definition Language) of the GLoBES software [44, 45], which allows now for a flexible systematics implementation entirely in AEDL (without the need to write C code).[3]

The paper is organized as follows. In Sec. 2 the experimental setups are described, as well as the assumptions for the oscillation parameters, the treatment of systematics, the values chosen for the systematic errors and the definition of our performance indicator. In Sec. 3, we compare the results obtained with the new systematics implementation (using explicit near detector simulations and including correlations) to those obtained with the old implementation (using an effective description of the errors in the far detector). More details on the simulation of the various experiments can be found in appendix A, and the details on our statistical methods can be found in appendix B. We also illustrate for which experiments robust predictions can be made and for which ones more external information is needed. A comparison of the performance of all setups is presented in Sec. 4, where also the dependence on exposure is discussed. In Sec. 5 we identify for each experiment the relevant performance bottlenecks and provide guidance on which quantities should be optimized in each case. Finally, we summarize and conclude in Sec. 6.

2 Simulation techniques and systematics treatment

2.1 Experimental setups

Table 1 summarizes the main features of the setups studied in this work. We have chosen four representative *benchmark* setups for long baseline neutrino oscillation experiments:

Beta-beam: a high-γ ($\gamma = 350$) beta-beam [46, 47] has been chosen, since it provides a very good CPV discovery potential, even comparable to the one obtained at a neutrino factory, see *e.g.* Ref. [48]. The relatively long baseline ($L = 650$ km) is enough to

[2] It should be kept in mind, however, that the experiment which reaches the best overall precision in δ may not yield the best CPV discovery potential (and *vice versa*), since the latter depends on the achievable precision around the specific values $\delta = 0, \pi$.

[3] This is one of the key modifications of the software which is expected to be included in the GLoBES 4.0 release.

guarantee a measurement of the mass hierarchy given the large value of θ_{13} (see for instance Ref. [24]). The beam is aimed at a 500 kton Water Čerenkov (WC) detector. This setup will be referred to as BB350.

Neutrino factory: we consider a low energy version of the neutrino factory, with a parent muon energy of 10 GeV and a Magnetized Iron Neutrino Detector (MIND) detector placed at a baseline of 2 000 km [49]. This is the setup currently under consideration within the International Design Study for a Neutrino Factory (IDS-NF) [50]. It will be referred to as NF10 hereafter.

Off-axis conventional neutrino beam: here we follow the T2HK proposal given its high relevance in the literature and that a Letter of Intent (LoI) has already been submitted [28]. The experiment uses a WC detector with a fiducial mass of 560 kton, placed at a distance of 295 km from the source.

On-axis conventional neutrino beam: we study a setup with a relatively high-energy flux (taken from Ref. [51]) and with a 100 kt Liquid Argon (LAr) detector at 2 300 km from the source. This corresponds to one of the configurations under consideration within LAGUNA [52] and LAGUNA-LBNO [53]. We have checked that the Fermilab-to-DUSEL Long Baseline Neutrino Experiment (LBNE), with a 34 kt LAr detector at a baseline of 1 290 km [54], has a very similar performance. We will therefore refer to this setup WBB in the rest of this paper since the conclusions extracted from its performance would be generally applicable to both LBNO and LBNE.[4]

In addition to these setups, we will also considered four alternative setups with high relevance in the literature. In particular, we will discuss two out of the three options considered during the LBNE reconfiguration process: a new Fermilab-based beam line aimed at a 10 kt LAr surface detector placed at Homestake ($\mathsf{LBNE}_{\mathsf{mini}}$), and the existing NuMI beam with a new 30 kt LAr surface detector placed at the Ash River site ($\mathsf{NO}\nu\mathsf{A}^{+}$).[5] Moreover, we consider a lower energy version of the neutrino factory (NF5), with a muon energy of 5 GeV and a baseline of 1 300 km. (The detector technology in this case is still a MIND to make a direct comparison to the NF10 setup easier.) The fourth alternative setup discussed in this paper is combination of a low-γ ($\gamma = 100$) beta-beam with the SPL [27], labeled as BB+SPL. We use this setup to study whether a combination of different channels (in this case, CPT conjugates) can reduce the impact of systematic errors. More details about each setup are given in Appendix A.

Finally, sometimes we will compare the results for the setups listed in Table 1 to the results that would be obtained by 2020 from the combination of present facilities, that is, T2K, $\mathsf{NO}\nu\mathsf{A}$ and reactors. In order to do so, we assume that T2K and $\mathsf{NO}\nu\mathsf{A}$ will have run for 5 and 4 years per polarity by that date, respectively, and that the precision on the measurement

[4] We find a slightly worse performance for the LBNE setup, though.

[5] The third option considered within the LBNE reconfiguration process consists of a 15 kt LAr underground detector placed at the MINOS site in Soudan. However, we have checked that the performance of this setup is much inferior to that of all other setups discussed here, and therefore we do not consider it any further.

	Setup	$E_\nu^{\rm peak}$	L	OA	Detector	kt	MW	Decays/yr	$(t_\nu,t_{\bar\nu})$
Benchmark	BB350	1.2	650	–	WC	500	–	$1.1(2.8)\times10^{18}$	(5,5)
	NF10	5.0	2 000	–	MIND	100	–	7×10^{20}	(10,10)
	WBB	4.5	2 300	–	LAr	100	0.8	–	(5,5)
	T2HK	0.6	295	2.5°	WC	560	1.66	–	(1.5,3.5)
Alternative	BB100	0.3	130	–	WC	500	–	$1.1(2.8)\times10^{18}$	(5,5)
	+ SPL			–			4	–	(2,8)
	NF5	2.5	1 290	–	MIND	100	–	7×10^{20}	(10,10)
	$\mathsf{LBNE}_{\mathsf{mini}}$	4.0	1 290	–	LAr	10	0.7	–	(5,5)
	NOνA$^+$	2.0	810	0.8°	LAr	30	0.7	–	(5,5)
2020	T2K	0.6	295	2.5°	WC	22.5	0.75	–	(5,5)
	NOνA	2.0	810	0.8°	TASD	15	0.7	–	(4,4)

Table 1: Main features of the setups considered in this work. From left to right, the columns list the names of the setups, the approximate peak energy of the neutrino flux, the baseline, off-axis angle, detector technology, fiducial detector mass, beam power (for the conventional and superbeams) or useful parent decays per year (ion decays for the beta-beams and muon decays for the neutrino factories), and the running time in years for each polarity. Note, that the neutrino and antineutrino running is simultaneous for the neutrino factory setups NF10 and NF5 (the μ^+ and μ^- circulate in different directions within the ring). For beta-beams the number of useful ion decays is different for the two polarities, so we quote the number of useful ^{18}Ne (^{6}He) decays per year separately. For details on our simulations, see Appendix A.

of θ_{13} will be dominated by the systematic error reachable at Daya Bay. In order to simulate this combination we have followed Ref. [14].

It was noted in Ref. [13] that the achievable precision in δ around $\delta = \pm 90°$ is compromised in beta-beam setups because, unlike superbeams or neutrino factories, they cannot obtain a precise measurement for the atmospheric parameters through the ν_μ disappearance channels (see also Ref. [55]). Therefore, we have also combined the data obtained at BB350 with the disappearance data expected from T2K. Note that this is not necessary for BB+SPL since in this case the SPL beam would already provide a better measurement of the atmospheric parameters.

For all setups, we assume the near detector to be sufficiently far away from the neutrino production region to have the same geometric acceptance as the far detector. (A dedicated study would be needed to determine the appropriate distance for each setup.) This ensures that the ratio of near and far detector event rates is independent of energy. Apart from the baseline and mass, the near and far detectors are considered to be identical, *i.e.*, they share

the same energy resolutions, efficiencies,[6] energy binnings, *etc.*

2.2 Input values for the oscillation parameters

The following input values for the oscillation parameters, in agreement with the allowed ranges at 1σ from global fits [10–12], are used for all simulations in this paper:

$$\Delta m^2_{21} = 7.64 \times 10^{-5}\text{eV}^2, \quad \Delta m^2_{31} = 2.45 \times 10^{-3}\text{eV}^2;$$
$$\theta_{21} = 34.2^\circ, \quad \theta_{23} = 45^\circ, \quad \theta_{13} = 9.2^\circ.$$

Unless otherwise stated, a normal mass hierarchy is assumed. In our fits, we include Gaussian priors with a 1σ width of 3% for the solar parameters, 8% for θ_{23} and 4% for Δm^2_{31}. No external priors were used for θ_{13} and δ. We have checked that adding a prior on θ_{13} corresponding to the expected precision of the final Daya Bay measurement [8] does not affect our results, except for a very mild improvement in the measurement of δ around $\pm 90^\circ$ for some facilities where the intrinsic degeneracy is still present.

We compute the neutrino scattering cross sections in the target as a function of neutrino energy using GENIE version 2.6.0 [56]. We split the cross sections into Neutral Current (NC) and Charged Current (CC) contributions, and we further subdivide the latter into the three regimes Quasi-Elastic scattering (QE), RESonance production (RES), and Deep-Inelastic Scattering (DIS). The cross sections per nucleon vary by $\mathcal{O}(10\%)$ between targets with different proton-to-neutron ratios and nuclear masses, but for easier comparability among experiments, we use cross sections for ^{28}Si, which is an intermediate mass, isoscalar nucleus, throughout.

As a further simplification, no sign degeneracies have been considered in this analysis. This motivated by the fact that, thanks to the relatively large value of θ_{13}, almost all experiments presented in the comparison would most likely be able to measure the mass hierarchy, either from matter effects at long baselines (NF10, NF5, WBB, LBNE$_{\text{mini}}$, BB350) or from the analysis of atmospheric data at very massive WC detectors (T2HK, BB350, BB+SPL). The only exception to this could perhaps be NOνA$^+$, due to its relatively short baseline (735 km) and limited detector mass. Nevertheless, as already stated in Sec. 1, an independent determination of the mass hierarchy may be provided by other means anyway.

2.3 Systematic errors and their implementation

In this study, we treat systematic uncertainties for all experiments in the same framework. We implement beam flux uncertainties, fiducial mass uncertainties, cross section uncertainties, and the matter density uncertainty in the same way for all setups, whereas background uncertainties can only be the same within each class of experiments—superbeams, beta-beams, and neutrino factories. For each systematic error, we consider default, optimistic and conservative values, see Table 2. Note that the error estimates given in this Table

[6] The treatment of NC backgrounds takes into account possible differences, though, see Appendix A for details.

	SB			BB			NF		
Systematics	Opt.	Def.	Cons.	Opt.	Def.	Cons.	Opt.	Def.	Cons.
Fiducial volume ND	0.2%	0.5%	1%	0.2%	0.5%	1%	0.2%	0.5%	1%
Fiducial volume FD (incl. near-far extrap.)	1%	2.5%	5%	1%	2.5%	5%	1%	2.5%	5%
Flux error signal ν	5%	7.5%	10%	1%	2%	2.5%	0.1%	0.5%	1%
Flux error background ν	10%	15%	20%	correlated			correlated		
Flux error signal $\bar{\nu}$	10%	15%	20%	1%	2%	2.5%	0.1%	0.5%	1%
Flux error background $\bar{\nu}$	20%	30%	40%	correlated			correlated		
Background uncertainty	5%	7.5%	10%	5%	7.5%	10%	10%	15%	20%
Cross secs $\times$ eff. QE†	10%	15%	20%	10%	15%	20%	10%	15%	20%
Cross secs $\times$ eff. RES†	10%	15%	20%	10%	15%	20%	10%	15%	20%
Cross secs $\times$ eff. DIS†	5%	7.5%	10%	5%	7.5%	10%	5%	7.5%	10%
Effec. ratio ν_e/ν_μ QE*	3.5%	11%	–	3.5%	11%	–	–	–	–
Effec. ratio ν_e/ν_μ RES*	2.7%	5.4%	–	2.7%	5.4%	–	–	–	–
Effec. ratio ν_e/ν_μ DIS*	2.5%	5.1%	–	2.5%	5.1%	–	–	–	–
Matter density	1%	2%	5%	1%	2%	5%	1%	2%	5%

Table 2: The systematic errors considered in our analysis for superbeams (SB), beta-beams (BB), and neutrino factories (NF), respectively. Numerical values are shown for optimistic, default, and conservative assumptions. All numbers are based on external input and do not yet include any information from the near detector. Note that the background uncertainties listed here affect only detector-related backgrounds (NC events, charge or flavor misidentification), whereas the uncertainties related to the intrinsic beam background for superbeams are treated as flux errors.

†Despite showing only a single entry here, we use independent nuisance parameters for the ν_μ, ν_e, $\bar{\nu}_\mu$, and $\bar{\nu}_e$ cross sections.

*Despite showing only a single entry here, we use independent nuisance parameters for the ν_μ/ν_e, and $\bar{\nu}_\mu/\bar{\nu}_e$ cross section ratios. The uncertainty due to different detection efficiencies for the different lepton flavors has also been added in quadrature, and therefore only the error on the effective ratio is shown. Blank spaces indicate the cases when ν_e and ν_μ cross sections are allowed to vary in a completely independent way.

do *not* include the impact of the near detectors, which instead are explicitly simulated in our study. Note also that we do not claim that our default values will actually be exactly realized by any of the experiments, but the range from optimistic to conservative is likely to delimit the actual value.

In the following, we describe the types of the systematic errors and their physical underpinning. (More details are given in Appendix B.)

Flux errors. We assume the flux errors to be uncorrelated among different beam polarities, but fully correlated between the near and far detectors and between all oscillation channels in the same beam. For the superbeams, we include an additional flux error for each of the intrinsic ν_e, $\bar{\nu}_e$ and wrong-sign ν_μ components in the beam. We simulate $\nu_e + \bar{\nu}_e$ events in both near and far detectors, so that the former effectively measures the intrinsic beam

backgrounds and moreover provides some information on the ν_e, $\bar{\nu}_e$ cross sections. Note that the precision of these measurements depends on the size of the near detector because the event rate from the intrinsic beam contamination is much smaller than the one from the muon neutrinos in the beam. We choose the magnitude of the flux errors for superbeams based on what has been achieved in π-decay based neutrino beams like the MINOS experiment [57].[8] For the neutrino factory, the chosen flux errors correspond to the design values of the IDS-NF [58]. For the β-beams, similar flux errors should be attainable in principle because for both the neutrino factory and the beta beam, the parent decays have simple and well understood kinematics, and careful monitoring of the beam momentum distribution in the storage ring should be possible.

Cross section errors. We assume the external knowledge on cross sections to be universal for all experiments. We introduce separate systematic uncertainties for the QE, RES and DIS regimes, so that we have a set of three cross section errors for each of the four relevant neutrino species ν_μ, $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$. By treating the QE, RES and DIS cross sections as independent, we effectively introduce an uncertainty on the shape of the neutrino event spectrum.[9] In practice, the *ratios* of the $\overset{(-)}{\nu}_e$ and $\overset{(-)}{\nu}_\mu$ cross sections are known with greater precision than their absolute values, especially at high energy where lepton mass effects on the kinematics, effective nuclear form factors, *etc.* are less important [63]. Therefore, we altogether will have 12 nuisance parameters for the cross sections: three regimes (QE, RES and DIS) times two flavors (ν_e, ν_μ) times two polarities (ν, $\bar{\nu}$). In addition we will use six nuisance parameters for the QE, RES and DIS cross section ratios ν_e/ν_μ and $\bar{\nu}_e/\bar{\nu}_\mu$ whenever the final flavor cross section cannot be measured at the near detector (*i.e.*, for BB and SB). However, in the cases where the constraint on the cross section ratio is weaker than the corresponding constraint on the two cross sections, we *omit* the ratio altogether from the χ^2, and apply the cross section uncertainty directly to the different flavors (conservative cases in Table 2).

All cross section uncertainties are fully correlated among all channels measuring the same final neutrino flavor. For simplicity, we use the same error ranges for both neutrinos and antineutrinos as well as for electron and muon flavors (but the errors are still fully independent and uncorrelated between different flavors and/or polarities). As a result, the errors for ν_μ cross sections are generally slightly overestimated while the errors for $\bar{\nu}_e$ are underestimated. However, as we will illustrate later, only the cross section *ratios* have a sizable impact on the experimental sensitivity, and therefore this simplification does not affect our results.

In principle one might assume that, given lepton universality, the flavor ratios for cross sections should be entirely determined by kinematics and that, sufficiently far away from thresholds such as the electron or muon mass, the resulting differences between cross sections

[8] Even though our flux uncertainties may be optimistic compared to current estimates, their effect on the results is not very relevant, as it will be demonstrated in Sec. 5.

[9] Possible shape uncertainties for the cross sections within each regime are not considered for simplicity; see Refs. [59,60] for a related discussion. It also should be pointed out that nuclear effects do affect neutrino energy reconstruction [61,62] at low energies. Dedicated migration matrices are needed to study the impact of this effect, though.

for different flavor should be small. However, there is a multitude of effects that depend on the lepton mass and kinematics, in particular when the non-pointlike nature of the nuclear target is taken into account. Indeed, the cross sections depend on a set of form factors, all of which have, at least in principle, some dependence on the momentum transfer Q^2 and are thus are quite sensitive to the details of the lepton momentum distribution. For instance, for QE events a recent study [63] addresses these issues in detail and does indeed confirm differences as high as 10% for QE reactions at energies below 1 GeV. What is more, the uncertainties on these differences can be of the order of the difference itself. These uncertainties arise from possible uncertainties related to pseudo-scalar form factors and their Q^2-dependence.[10] Another source of uncertainty is the possible presence of second class currents. Overall, our assumptions on the uncertainty of the flavor ratio of the QE cross sections in this paper are based on Ref. [63]. There, it is shown that the uncertainty is strongly energy-dependent. We do not take this energy dependence into account, but our default value for the uncertainty (10%) corresponds to the regime where the QE cross section peaks (0.5 GeV), and we have chosen conservative (30%) and optimistic (2.5%) values taking into account that it varies over energy and it is difficult to assign a specific (theoretical) number. For the other two reaction channels, we have followed the discussion in Ref. [36], and our default values for the cross section ratios have been set at the 2% and 1% for the RES and DIS regimes, respectively.

Efficiency errors. For simplicity, efficiency errors are not treated as independent, but absorbed into the cross section uncertainties. Only the product of the cross section and efficiency for a given flavor and event type appears in observables, and thus, in our analysis, only the combined effect has been considered. Obviously, the physics governing the uncertainties on cross sections and efficiencies is completely different, but their effects can still be added in quadrature to obtain the error on the product of cross section and efficiency. This will be referred to as "effective cross section" from here on. We neglect the uncertainty on the *absolute* efficiencies because the uncertainty on the absolute cross section is typically much larger. Nevertheless, we do include a 5% uncertainty on the effective cross section ratios listed in Table 2 due to the different detection efficiencies for different lepton flavors. Such uncertainty is added in quadrature to the uncertainty on the cross section ratio. We do this independently for neutrinos and antineutrinos and for the QE, RES and DIS regimes. As for the optimistic and conservative cases, the corresponding errors have been taken to be a factor of two smaller and larger than the default value, respectively. Note that the particular value chosen here (5%), which in the following will be used for all experiments in order to treat them on equal footing, should be considered a rough estimate – in reality, the efficiency error can vary widely, depending for instance on the detector technology, reaction channel, *etc.* We assume that the near and far detectors are sufficiently similar for the efficiency errors to be correlated between them. A residual near–far difference can be at least partially absorbed into the fiducial mass errors.

[10] This Q^2-dependence is usually constrained by the Goldberger-Treimann relation, but experimental tests of the PCAC hypothesis, which is the foundation for Goldberger-Treimann relation, still have sizable uncertainties.

Near–far extrapolation errors. Using near detector data to reliably predict the neutrino spectrum at the far detector is quite challenging in practice, and the extrapolation is affected by a number of systematic uncertainties, for instance due to different geometric acceptance, different detector design, *etc.* Some of these uncertainties moreover have a strong energy dependence, which can, however, be reduced by carefully designing the experiment; for instance, the near detector should not be too close to the neutrino production volume to make its geometric acceptance as similar as possible to that of the far detector. Here, we assume for simplicity that the near–far extrapolation errors are energy and flavor independent, and can therefore be treated as an effective fiducial mass error, uncorrelated between the two detectors. We conservatively assume this error to be fairly large (2.5% in our default scenario, included in the fiducial volume error for the far detector), but it will be shown in Sec. 5 that it does not affect the experimental sensitivity significantly. The same assumptions regarding systematic uncertainties have again been made for all experiments in this case.

Background uncertainties. In addition to the uncertainties on the intrinsic beam backgrounds discussed above, we also include uncertainties on detector-related backgrounds such as NC events and charge or flavor misidentification. These errors are typically uncorrelated among channels and detectors. It is well known, though, that for large θ_{13} the overall impact of background uncertainties is small.

Matter density uncertainty. For relatively short baselines that traverse only the Earth's mantle, it is a good approximation to assume the matter density to be constant along the baseline. We compute the constant effective density separately for each experiment based on the density profile along the baseline derived from the Preliminary Reference Earth Model (PREM) [64]. The mantle density can be determined from seismic wave measurements with an uncertainty of typically about 5% or below. For specific trajectories, values of about 2% have been achieved [65], so we use this number as our default. In the conservative case, we use 5%, whereas in the optimistic scenario we assume an uncertainty of only 1%, which may be achievable with a dedicated geophysical campaign.

2.4 Performance indicators

One of the challenges in comparing future experiments based on their achievable precision in δ, which we will call $\Delta\delta$, is that this precision strongly depends on δ itself. In Refs. [13,34,35], it has therefore been proposed to show the performance as a function of the true δ, as illustrated in the left panel of Fig. 1 for two illustrative experiments. While these two experiments exhibit the same qualitative dependence on δ, the sensitivity of the experiment corresponding to the solid curve varies less with δ. This can be due to a number of reasons, like the baseline or the width of the energy spectrum (for a detailed discussion see Ref. [13]), but it is obvious that both experiments are optimized for their sensitivity to CPV since the error is smallest for $\delta = 0$ and π. Depending on the chosen value of the CP phase one finds that either the experiment represented by the dashed line or the one corresponding to the solid line is more precise. If matter effects are strong, or if the numbers of neutrino

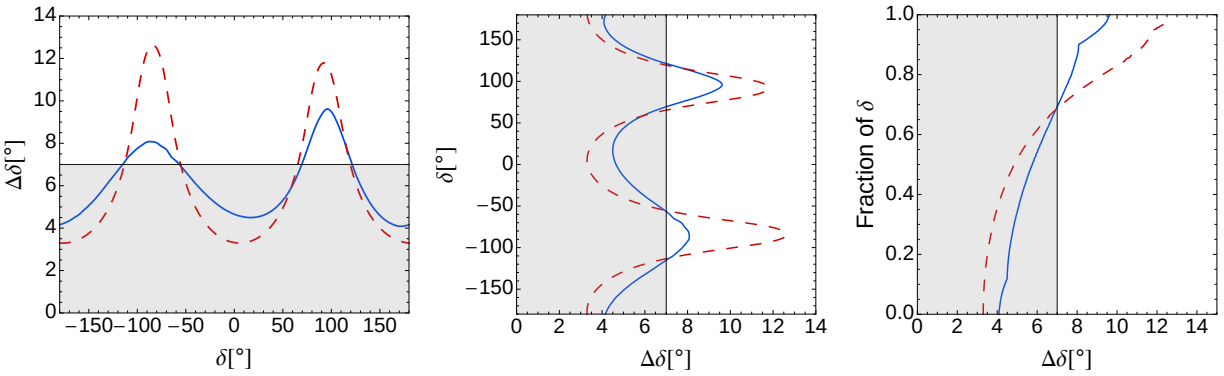

Figure 1: Left panel: Error on δ (1σ) as a function of the true δ for two sample experiment (solid and dashed curves). Middle panel: Left plot mirrored at the diagonal. Right panel: Fraction of true δ as a function of the precision of $\Delta\delta$ for these two experiments. The vertical axis shows for what fraction of the δ values a certain precision can be obtained. The lines show where the curves intersect.

and antineutrino events is very different, the pattern can be shifted in δ, complicating the situation further [13].

The new paradigm in neutrino oscillations for large θ_{13} is precision, and therefore different experiments should be compared based on how well they perform on this task within the remaining parameter space of interest—which is mainly the unknown phase δ. Since it is difficult to compare in a quantitative way the different experiments using a plot like the left hand panel of Fig. 1, we mirror the plot at the diagonal, as shown in the middle panel, and stack the values of true δ for which a certain precision can be obtained. The result of this procedure is the "fraction of δ" (sometimes also called "CP fraction") for which $\Delta\delta$ is smaller than a given number, as illustrated in the right panel of Fig. 1. The same approach was previously used to quantify the discovery potentials for CPV, mass hierarchy, and θ_{13} by showing the fraction of true values of δ for which a certain observable could be measured as a function of the true $\sin^2 2\theta_{13}$. With the fraction of δ on the vertical axis in Fig. 1 (right), the comparison no longer depends on relative phase shifts between the two curves from the left panel, it quantifies the performance assuming that all values of δ are equally likely and equally important. The disadvantage is that one cannot read off from the plot at which value of δ the performance of an experiment peaks. Note, however, that the discussed setups are typically optimized for CPV, which means that the optimal performance is in most cases achieved close to $\delta = 0$ or π [13].

A somewhat more subtle technical point, elaborated on in Refs. [34, 35], is the fact that the absolute performance can depend in a nontrivial way on the chosen confidence level. For instance, $\Delta\chi^2$ may behave in a highly non-Gaussian way far from the best fit point, in particular if the mass hierarchy degeneracy cannot be resolved. Here we assume that we are in the Gaussian limit, and sign degeneracies have not been considered. However, we note that the fraction of δ values for which a certain sensitivity is reached is also useful in the non-Gaussian case.

In the following sections, we will illustrate our results using plots similar to Fig. 1, right panel, as well as the standard plots showing the CPV discovery potential.

3 Impact of systematics

To begin our analysis of systematic uncertainties, we compare in Fig. 2 the results obtained with our new implementation of systematic errors to the ones obtained with the simpler, but widely used implementation, in which systematic errors are not correlated between different oscillation channels and near detectors are not explicitly simulated. Instead, their effect is parameterized in terms of effective uncertainties in the far detector, which are typically chosen between zero (statistical errors only) and $\mathcal{O}(1-10)\%$. In our new implementation, the near detectors are explicitly simulated, the systematic errors are correlated in a physical way, and they systematic are varied between the optimistic and conservative values from Table 2. Here we focus on the benchmark setups from Table 1, which give an idea on the ultimate performance that each type of beam could reach.

The upper row of plots in Fig. 2 shows the fraction of possible values of δ for which a certain precision $\Delta\delta$ can be achieved (at 1σ, for $\sin^2 2\theta_{13} = 0.1$), while the lower row shows the fraction of possible values of δ for which CPV can be established at 3σ, as a function of $\sin^2 2\theta_{13}$. The light gray bands show how the performance of each experiment would vary between the statistics-only (no systematics) limit and the case where each signal (background) channel in the far detector is assigned an uncorrelated 5% (10%) error. The matter density uncertainty is also included in this case: the right/lower edges of the gray bands has been computed assuming a 5% matter density uncertainty, whereas for the left/upper edges the matter density has been kept fixed. The colored (dark) bands in Fig. 2 show the results obtained with the new implementation of systematic uncertainties, with the width of the bands illustrating the difference between the optimistic and conservative scenarios from Table 2.

For the wide band beams operating at high enough energies (NF10, WBB), the old and new implementations yield very similar results. In fact, the bands obtained for these beams with the new implementation are roughly included as a subset in the ones obtained with the old method, which means that sharper predictions can be made now. We find that the predicted performance of NF10 changes only mildly, and that our results agree well with those presented in Ref. [66]. For BB350 and T2HK, on the other hand, there is an overall offset between the old and new systematics treatments, and the default values (solid curves) are not even within the old predicted ranges. In fact, it seems that the old treatment may have been too optimistic. As we shall demonstrate later, the main reason for this offset is that the ratio between ν_e and ν_μ cross sections is needed as an external input. For instance, the T2HK beam consists mainly of ν_μ, but the δ measurements relies on the detection of ν_e, for which the cross sections are difficult to measure in the near detector. The situation is precisely the opposite for BB350, for which ν_e are produced and ν_μ are observed at the far detector. Note that both experiments operate at relatively low energies, where QE and RES scattering dominate. These types of interaction have larger uncertainties than high-energy DIS scattering, hence the large difference between the widths of the light gray and dark

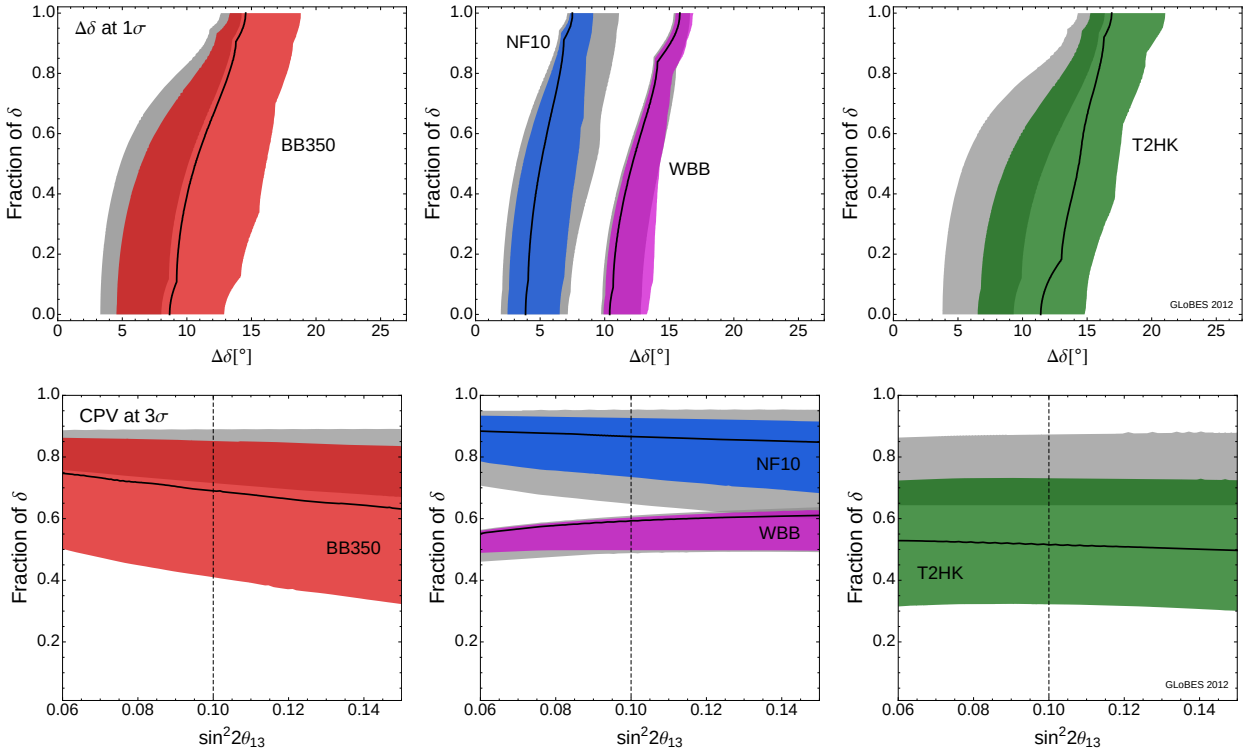

Figure 2: Comparison of experimental sensitivities predicted with the old systematics treatment using effective uncertainties in the far detector (light gray shadings) and the new treatment which includes near and far detectors and properly accounts for correlations of uncertainties between different channels, detectors, *etc.* (dark colored shadings). The upper row shows the fraction of δ values for which δ can be measured with a given precision (at 1σ, for $\sin^2 2\theta_{13} = 0.1$), the lower row shows the fraction of δ for which a 3σ discovery of CPV is possible as a function of the true $\sin^2 2\theta_{13}$. Different experiments are shown in different panels, as indicated in the legends. For the old systematics implementation, the effective errors are varied between no systematics and 5% normalization error for the signal (10% for the background), whereas for the new implementation the ranges between the optimistic and conservative cases from Table 2 are considered. The default values are shown as black curves. A true normal hierarchy has been assumed, and no sign degeneracies have been considered. The vertical dotted lines in the lower panels correspond to $\sin^2 2\theta_{13} = 0.1$, which is the true value chosen for θ_{13} in the upper panels.

colored bands in Fig. 2 for T2HK and WBB. Note also that our assumption of independent errors for the different cross section regimes introduces an effective shape error, this being especially relevant for BB350.

The widths of the bands in Fig. 2 can also be interpreted in terms of the robustness of the predictions. For NF10, the predictions are robust with respect to systematics since all signal channels depend on the detection of ν_μ or $\bar{\nu}_\mu$, the cross sections can be measured in the near and far detector(s). Another interesting result is perhaps that WBB outperforms T2HK at its default performance and is much more robust against systematic errors. This is a result of the relatively high neutrino energies (mostly in DIS regime) for WBB, in combination with the wide beam spectrum and the long baseline.

The relative position of the black lines within the new systematics bands in Fig. 2 shows how close the default scenario is to the optimistic performance. For example, for WBB, the default performance already approaches the optimistic limit, which means that further improvements in systematics will not lead to a major increase in sensitivity. Instead, as we will show in Secs. 4 and 5, an effort towards increasing statistics at the far detector would be more useful than a further reduction in systematic errors. For T2HK, on the other hand, the default curve lies in the middle of the band even though it has been simulated with exactly the same values of the systematic errors. This means that systematics are important, and an improvement will clearly help. We will discuss in Sec. 5 what is the relative importance of systematics and statistics for each of the setups under consideration.

Comparing the no systematics limit (statistics only) with the optimistic systematics in the new implementation (the two uppermost/leftmost limits for each setup), one can also read off how close the optimistic implementation is to the statistics only limit. In almost all cases, the optimistic choice is already close to the statistics limit, with the exception of T2HK. There, even in the optimistic case, the sensitivity to δ is limited by the uncertainty on the QE cross section ratio.

4 Performance comparison

The nominal performance of all setups listed in Table 1 is compared in Fig. 3, using the default values for the systematic errors according to Table 2. In the left panel, the fraction of δ values for which a given precision in δ can be achieved is shown. Considering only the benchmark setups BB350, NF10, WBB, and T2HK, it can be seen that the neutrino factory outperforms all other options by a factor of two. It is the only experiment which can achieve a precision comparable to the one obtained in the quark sector, where the CP phase is determined to be $\gamma = 70.4^{+4.3}_{-4.4}{}^\circ$ [67], depicted by the vertical gray band in the left panel. We also show in this figure, in addition to the setups listed in Table 1, the results that would be obtained by the year 2020 from the combination of T2K, NOνA and reactors.

We would like to point out the remarkable performance of BB+SPL, which outperforms any of the other superbeam and beta-beam options. As we will discuss in Sec. 5, the reason for this is the reduction in systematic errors related to the cross sections. For the other alternative setups, which can be regarded as smaller versions of the respective original

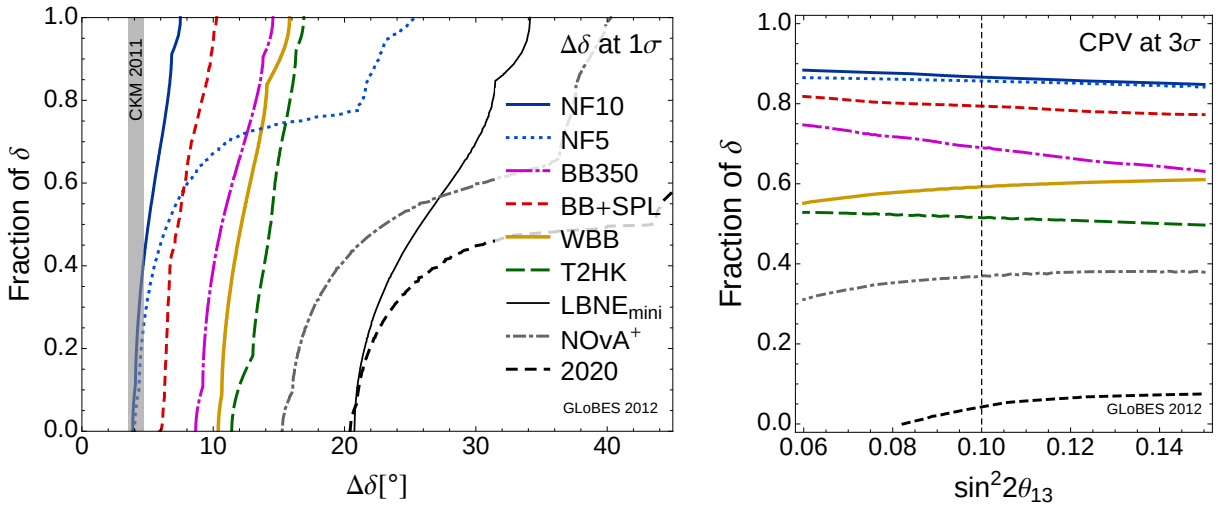

Figure 3: Comparison between the different setups from Table 1 for the default systematic errors listed in Table 2 (including near detectors). We have also included in the comparison the results that would be obtained by 2020 through the combination of T2K, NOνA and reactors. Left panel: Fraction of δ as a function of the precision at 1σ for $\sin^2 2\theta_{13} = 0.1$. Right panel: Fraction of δ for which CPV can be established at 3σ as a function of $\sin^2 2\theta_{13}$ in the currently allowed range. A true normal hierarchy has been assumed, and no sign degeneracies have been accounted for. In the right panel, $\mathsf{LBNE}_{\mathsf{mini}}$ is not shown because it does not reach 3σ sensitivity to CPV. The vertical dotted line in the right panel corresponds to $\sin^2 2\theta_{13} = 0.1$, which is the true value chosen for θ_{13} in the left panel. In the left panel, the vertical gray band depicts the current precision for the CPV phase in the quark sector, taken from Ref. [67].

proposals, the precision varies strongly as a function of the true δ. This is due to the fact that intrinsic degeneracies appear around $\delta = \pm 90^\circ$ superimposed to the true solutions, effectively worsening the observable precision on δ (see also Ref. [13]). In the particular case of the NF5 this is due to the coarse energy binning that we have used for our simulations, which is fixed by the available migration matrices.[11]

It is interesting to compare the precision on δ in Fig. 3, left panel, with the CPV discovery potential (right panel), which mainly corresponds to the precision at specific values of δ (0 and π). Again, neutrino factories emerge as the optimal setups, being able to observe CPV at 3σ for $\sim 85\%$ of the values of δ. Compared to the left panel, NF10 and NF5 perform very similar, which is expected from optimization studies [49]. However, this optimization does not include the full parameter space, which is much better covered by NF10 (left panel).[12] The performance of almost all setups is reduced for larger $\sin^2 2\theta_{13}$. The exceptions are NOνA$^+$ and WBB, which benefit from the increased statistics for larger values of θ_{13}. Note that the 3σ CPV discovery potential of T2HK is comparable to that of NOνA$^+$, whereas T2HK is clearly a better precision instrument (left panel). As far as the comparison between LBNE$_{\text{mini}}$ and NOνA$^+$ is concerned, both perform similar in the left panel with NOνA$^+$ exhibiting a stronger dependence on δ, as explained above. For CPV discovery (right panel), however, LBNE$_{\text{mini}}$ cannot compete at all because it does not reach 3σ for any value of δ. Finally, it should be noticed that the combination of T2K, NOνA and reactors (the 2020 line) would indeed be able to observe CPV at 3σ for some values of δ, although the fraction of δ values for which this is possible remains well below the 10% (see also Ref. [14]).

The comparison between different setups does not only depend on systematic uncertainties but also on exposure. We therefore show in Fig. 4 the exposure dependence of the precision on δ for all setups in Table 1. Here exposure ($\mathcal{E}$) is defined as the beam intensity (protons on target or useful parent decays) $\times$ running time $\times$ detector mass, and only the relative exposure compared to the nominal values $\mathcal{E}_0$ from Table 1 is shown. The left panel of Fig. 4 shows results for the benchmark setups, while the right panel displays results for the alternative setups. The bands reflect the variation of the performance between the optimistic and conservative choices for the systematics uncertainties (see Table 2).

Fig. 4 reveals several interesting features. First, for some experiments the gradient at the nominal exposure is significantly larger than for others. In particular, WBB, LBNE$_{\text{mini}}$, and NOνA$^+$ operate in the statistics limited regime ($\Delta\delta \propto 1/\sqrt{\mathcal{E}}$), where the systematics contribution is small. This makes the exposure the most relevant performance bottleneck for them (see also Sec. 5). Comparing LBNE$_{\text{mini}}$ with NOνA$^+$, NOνA$^+$ clearly exhibits a larger dependence on systematics. This dependence increases with exposure. In most other cases, when optimistic values are chosen for the systematics (lower edges of the bands) the scaling with exposure seems to be dominated by statistics ($\Delta\delta \propto 1/\sqrt{\mathcal{E}}$), while for conservative values (upper edges) the setups start to be more dominated by systematics and the curves are less steep. In these cases, the difference between optimistic and conservative systematics increases significantly with exposure. An interesting exception is BB+SPL,

[11] We have checked that, if the bin size is reduced by a factor of two, the dependence on δ is largely reduced since the intrinsic degeneracies are better resolved.

[12] In fact, the meeting point of the two curves in the left panel roughly corresponds to the values of δ which are relevant for a good CPV discovery potential and therefore the CPV sensitivity is the same.

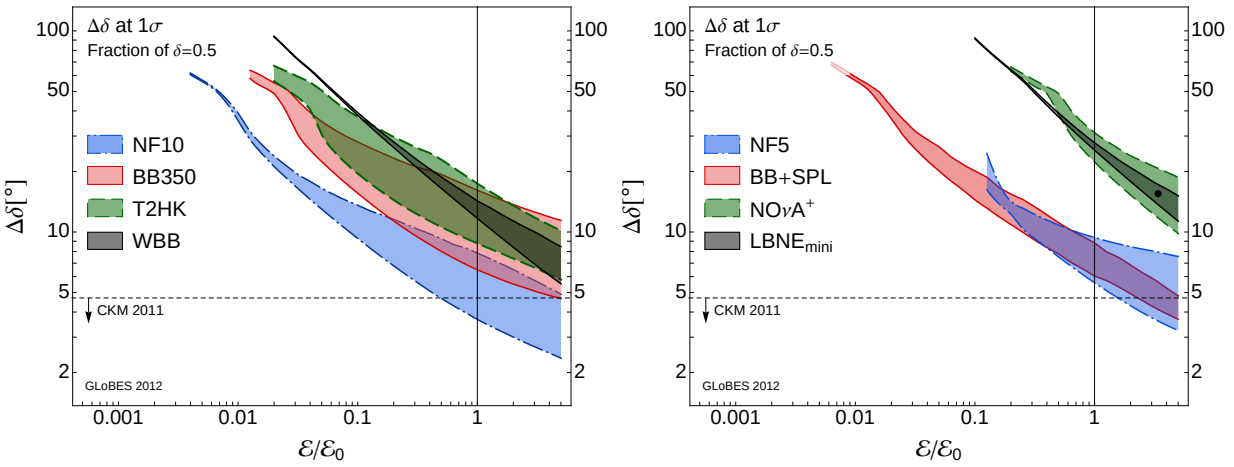

Figure 4: Error on δ (at 1σ, for $\sin^2 2\theta_{13} = 0.1$) as a function of exposure, where the bands reflect the variation in the results due to different assumptions for the systematic errors between the optimistic (lower edges) and conservative (upper edges) values in Table 2. In the left panel, the results for the benchmark setups from Table 1 are shown, while the right panel shows the results for the alternative setups. The nominal exposure $\mathcal{E}_0$, to which the exposure $\mathcal{E}$ on the horizontal axis is normalized, is the one given in Table 1. Here near detectors are included, and the results are shown for the median values of δ (fraction of δ is 50%). The current precision on the CP phase in the CKM matrix $V_{\rm CKM}$ is also indicated. In the right panel, the black dot indicates the luminosity for the original LBNE configuration (34 kt LAr detector [54]).

for which systematics are equally important regardless of the exposure. In this case, the dominant systematics (cross section ratios) are reduced by the combination of the two beams, so that even under conservative assumptions for the systematic uncertainties, the performance of BB+SPL is still dominated by statistics.

5 Performance bottlenecks and the role of the near detectors

Here we discuss the most important limiting factors for the performance of each individual experiment, *i.e.*, the key factors to be watched in the design and optimization of the experiment. As we will demonstrate, there is typically one dominant performance bottleneck, which is however different from experiment to experiment. We study the impact of:

1. systematic errors, including possible correlations,
2. exposure,
3. the near detector.

In order to identify the key systematic errors, we start by taking all of them at their default values and then switching off each group of systematic errors (flux errors, cross section

uncertainties, *etc.*) independently. This method reveals the systematic uncertainties that have the greatest impact, and which uncertainties are irrelevant for the measurement of δ.[13]

The impact of switching off groups of systematic uncertainties for the different experiments is shown in Fig. 5 and Fig. 6 for the benchmark and alternative setups, respectively, and for the default values of the systematics listed in Table 2. In both figures, the upper colored bars show how much the precision in δ would improve for each experiment if a given systematic error is switched off. (Only those groups of systematic errors that actually have a sizeable impact for each facility are shown.) For each experiment, the precision that would be reached in the statistics-only limit is also shown ("all off"). As mentioned above, neither do the different bars typically add up to the "all off" one, nor does the dominant systematics alone account for it. The reason are correlations among systematic uncertainties and between systematics and oscillation parameters, *i.e.*, the difference between the "all off" case and the other bars can be interpreted as the importance of these correlations. The impact of doubling the exposure (see also Fig. 4), as well as the performance loss which each experiment would suffer if no near detector was available, are also shown for each experiment.

Note that the edges of the bars shown in Figs. 5 and 6 correspond to the medians of the corresponding $\Delta\delta$ distributions, *i.e.*, for 50% of all δ values the precision will be better than the $\Delta\delta$ value shown in the figures, for the other 50% it will be worse. The δ values corresponding to the left and right edges of any given bar need not be the same since the median may change from one edge of the bar to the other. Small differences in the results would appear if instead we chose a fixed value of δ for all bars corresponding to the same experiment, or if we sampled the $\Delta\delta$ distributions not at their median, but at a fraction of δ other than 50%. Nevertheless we expect our general conclusions to remain unchanged.

Let us first discuss the impact that different systematic uncertainties have on the benchmark setups (Fig. 5). For NF10, the most important systematic uncertainty is the one on the matter density, as has been established earlier [39]. Improving the flux error or the understanding of the DIS cross sections marginally helps, with the relative importance of these two depending slightly on the value of δ. For BB350, the sensitivity is limited mostly by the errors on the QE and RES effective cross section ratios (the event rate is substantial in both regimes). Correlations among these turn out to be important because they lead to an effective shape error. For WBB, the impact of systematics is generally small, while the sensitivity is mainly limited by the exposure (see also Fig. 4). It should be noted here that WBB has been simulated with a LAr detector, which is the least studied among the detector technologies considered here. For instance, it is the only detector for which no tabulated detector reponse functions ("migration matrices") from detailed Monte Carlo simulations are available for the signal reconstruction up to now. For T2HK, the impact of systematic uncertainties (in particular on the QE cross section ratio and the intrinsic beam background) is generally large. Exposure is also important, but it is not the dominant limitation.

An interesting question is how much the near detectors actually help in the precision measurement of δ. We therefore show in Fig. 5 how the predicted δ precision changes when the

[13] We have also checked that the inverse procedure, *i.e.*, starting from the statistics only limit and switching on each group of systematic errors independently, leads to similar conclusions. However, this second procedure is less intuitive and therefore we will not show its results in the following.

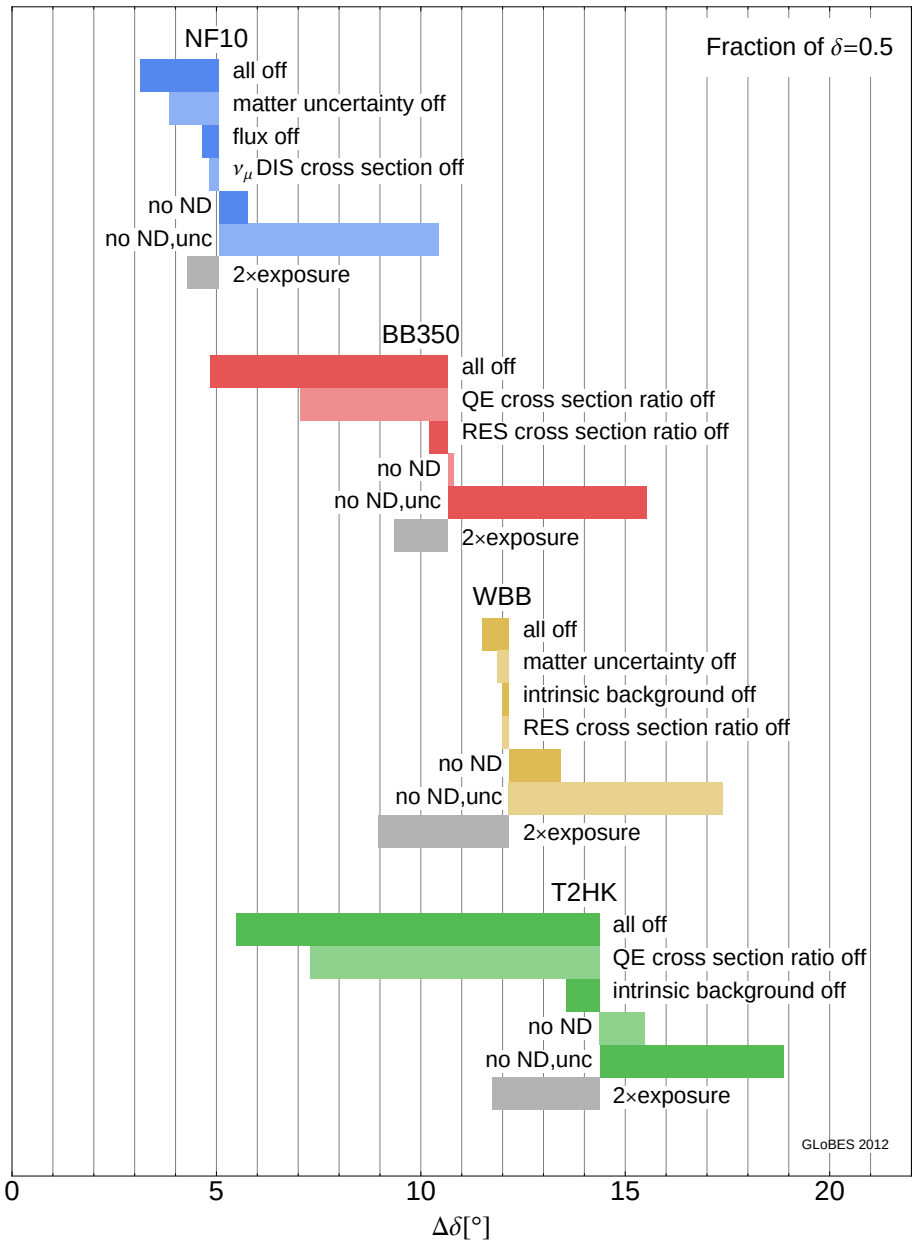

Figure 5: Dependence of the achievable precision in δ (at 1σ, for $\sin^2 2\theta_{13} = 0.1$) for the benchmark setups in Table 1 on systematic uncertainties, exposure, and near detectors. The bars show the improvement in the precision of δ compared to the default scenario if the dominant systematic errors are switched off separately. Here "all off" refers to the statistics-only limit, "matter uncertainty off" to no matter density uncertainty, "flux off" to no flux errors, "DIS ν_μ cross section off" to no DIS effective cross section errors for neutrinos and antineutrinos, "cross section ratio off" to fully correlated effective cross section errors for ν_e and ν_μ, and for $\bar{\nu}_e$ and $\bar{\nu}_\mu$, and "intrinsic background off" to no uncertainty on the intrinsic beam backgrounds. The effect of doubling the exposure is also shown, as well as two sets of results without a near detector: for "no ND" systematic uncertainties are still correlated between oscillation channels at the far detector, while for "no ND, unc", also correlations between appearance and disappearance channels are not included. The $\Delta\delta$ values shown here correspond to the median value of δ (*i.e.*, for 50% of δ values, the precision would be better, for the other 50% it would be worse).

near detector is not included in the analysis ("no ND"). A somewhat surprising result is that for none of the setups considered here omitting the near detector affects the achievable precision by more than 1–2 degrees. Also, in none of the cases the ND is the most critical factor. The main reason is that, even without a near detector, most systematic errors are correlated among the different oscillation channels, so that the nuisance parameters are constrained by the requirement of self-consistency among the different far detector channels (in particular appearance and disappearance). Note that this self-consistency requirement relies entirely on the validity of the three active flavor oscillation framework. Thus, in the absence of a near detector it is doubtful that meaningful bounds on physics beyond the Standard Model (such as sterile neutrinos or non-standard interactions [68, 69]) could be obtained.

In order to illustrate the importance of correlations between appearance and disappearance data, we also show in Fig. 5 results for the case where the near detector is omitted and in addition correlations between appearance and disappearance channels in the far detector are not included, *i.e.*, the appearance and disappearance data sets are assigned independent systematic errors ("no ND, unc"). In this case, as expected, the near detector plays a crucial role for all setups. The difference between these bars and the bars labeled "no ND" shows explicitly the importance of correlations and how disappearance data can be used to constrain systematic errors in the appearance sample in a very efficient way. The effect of the disappearance channel is particularly relevant for the BB350 setup, for instance. In this case, since ν_e disappearance is small, the far detector is particularly useful in constraining systematic effects related to flux uncertainties and ν_e cross section measurements. Therefore, the near detector does not provide any additional information and can be removed from the analysis with practically no impact on the precision. In addition, for BB350 (T2HK) the near detectors do not provide a measurement of the $\overset{(-)}{\nu}_\mu$ ($\overset{(-)}{\nu}_e$) cross sections needed for the far detector appearance measurement, but only of the flavor-conjugate cross section. (In T2HK, the near detector is in principle sensitive also to the $\overset{(-)}{\nu}_e$ cross sections due to the the intrinsic beam backgrounds, but statistics in these channels is too small to allow for a precise cross section measurement.) The experiment which benefits most from the near detector is WBB, where having the near detector is in principle more important than improving any of the systematic errors. The reason for this is that this setup is statistically limited, and is therefore not able to constrain both the nuisance parameters and the atmospheric oscillation parameters independently in the analysis. Thus, increasing the nominal exposure would be much more critical in this case.

Fig. 6 shows how the performance of the alternative setups in Table 1 depends on systematics, exposure, and the near detectors. We notice that for NF5, the relative impact of the systematic errors (including the matter density uncertainty) is smaller whereas exposure is somewhat more important than for NF10. In addition, NF5 is the only experiment for which the near detector is more important than systematics or exposure. For $\mathsf{LBNE}_{\mathsf{mini}}$ and $\mathsf{NO}\nu\mathsf{A}^+$, the near detector has also a larger effect than the systematic uncertainties, but the main limitation for these statistics-dominated experiments is exposure. In fact, an increase in statistics may also render the near detector unnecessary because, similar to WBB discussed above, $\mathsf{LBNE}_{\mathsf{mini}}$ and $\mathsf{NO}\nu\mathsf{A}^+$ need the near detector mainly because they are unable to constrain both the systematic nuisance parameters and the atmospheric oscillation pa-

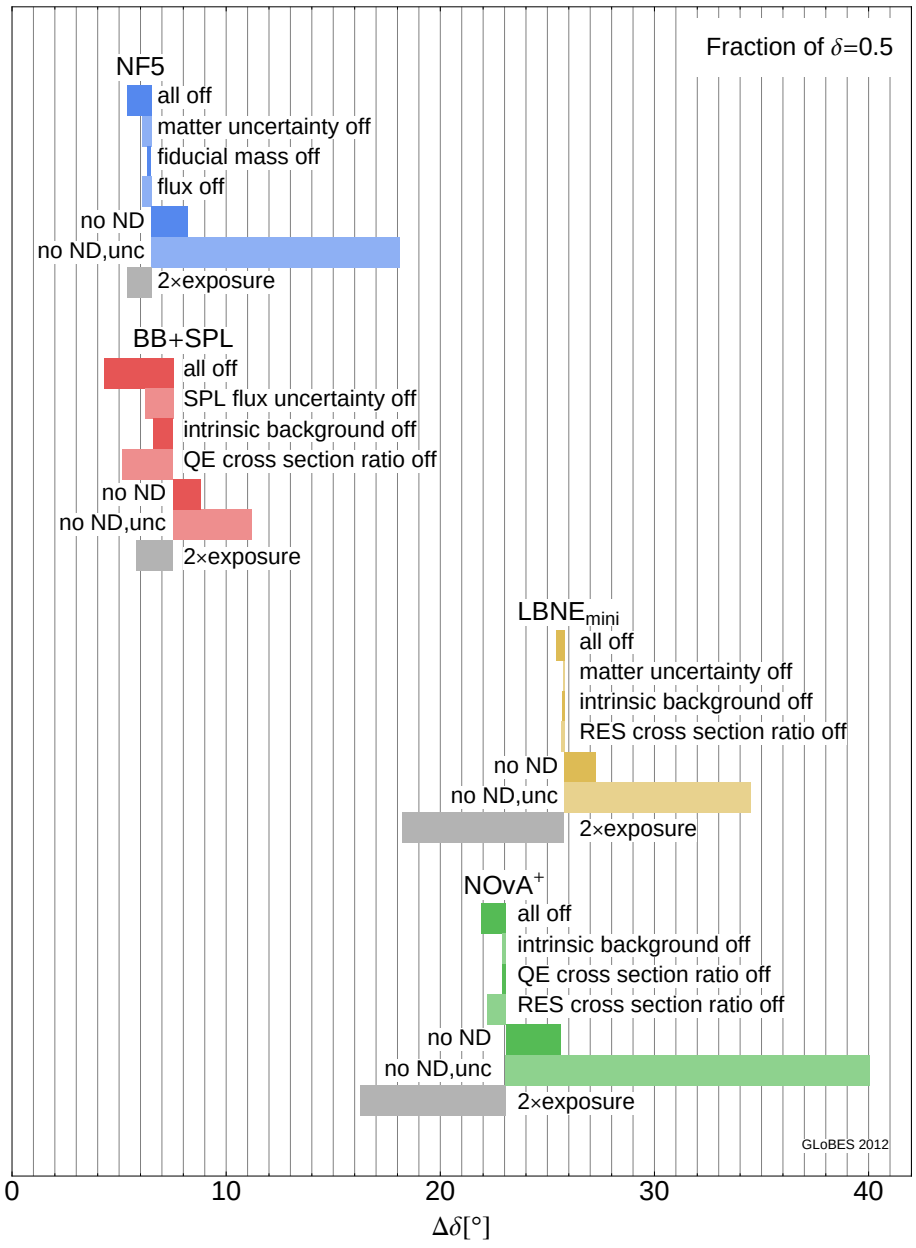

Figure 6: Dependence of the performance of the alternative setups in Table 1 on systematic uncertainties, exposure, and near detectors. The meaning of the labels and abbreviations is the same as in Fig. 5.

rameters independently using disappearance data. Thus, in the optimization of experiments of this type, the benefits of a near detector and of increased statistics have to be carefully weighed against each other.

An interesting question one could ask at this point is whether there are feasible ways of reducing systematic uncertainties, especially the ones on the QE cross sections. An interesting example for this is BB+SPL. In this setup, both ν_e and ν_μ cross sections can be measured precisely in the same detector, which reduces the impact of systematics and increases the absolute performance. This can be seen from the reduced length of the "all off" bar for BB+SPL compared to the BB350 or T2HK cases, which all operate in the low energy regime. Some further improvement would be achieved if the SPL flux uncertainty was reduced, though. Note that BB+SPL could in principle even compete with NF5 if the exposure could be significantly increased, the cross section ratios could be better constrained, or the SPL flux could be better understood. This shows how a combination of facilities can be of great help in reducing the impact of systematics on their performance. A similar effect would be obtained if an independent measurement of the ν_e/ν_μ cross section ratio was performed for both neutrinos and antineutrinos. The proposed low energy muon storage ring experiment ν-STORM [70] would be ideal for this measurement.

An additional method to reduce the impact of systematics could be a facility optimized for the second oscillation peak, see for instance Ref. [37], where an SPL-based experiment with a detector at 650 km instead of 130 km is proposed. This would be useful to increase the CPV discovery potential of the facility as well as to reduce the impact of systematic errors. Note that in Ref. [37], correlations between systematic uncertainties were not taken into account and near detectors were not simulated explicitly. We have checked that the conclusions still hold in the case where full correlations are taken into account. Indeed, the setup proposed in [37] exhibits the least dependence on systematic errors among all the experiments compared here.

6 Summary and conclusions

Systematic uncertainties in neutrino oscillation experiments are especially important for large θ_{13}. Hence, a dedicated comparison with a careful treatment of these uncertainties is needed to determine the optimal next-generation experiment, given the large value of θ_{13}. Also, the degree to which this optimization depends on the assumptions regarding systematic uncertainties should be carefully assessed.

In this study, we have analyzed and compared superbeams, beta-beams, and neutrino factories on an equal footing, paying special attention to systematic uncertainties. In particular, a realistic implementation of systematic uncertainties in the simulations used to predict the sensitivity of future experiments depends not only on individual numbers for certain systematic errors, but also on how these errors are correlated among different detectors, oscillation channels, *etc.* In most previous studies, only few types of systematic uncertainties were considered, and the respective error margins were chosen in order to account in an effective way for the real error menu. In particular, near detectors were typically not simulated explicitly, and correlations were neglected. In this paper, instead, we have used

explicit near and far detector simulations with comparable assumptions and an improved systematics implementation which takes into account all possible correlations. Moreover, to allow for a simple and fair comparison of different facilities, we have used identical assumptions on external input (in particular cross sections) wherever possible. Besides our default set of systematic errors, we also consider more conservative and a more optimistic scenarios (see Table 2), which should encompass the performance of a real experiment.

Table 1 summarizes the setups studied in this work. Since we expect that the mass hierarchy can de determined by all of the discussed setups (with the possible exception of NOνA$^+$), we have used the 3σ discovery potential for leptonic CP violation (CPV) and the achievable precision at 1σ in the measurement of δ ($\Delta\delta$) as our main performance indicators. While the first indicator depends on the performance of the experiment around the specific values $\delta = 0, \pi$, the second one treats all values of δ as equally important. Since the dependence of the experimental sensitivity on δ is in general complicated, we present our results in terms of the fraction of δ values for which a certain precision (or better) is achieved, see Fig. 1 for illustration.

We have compared our new systematics implementation with the previous effective treatment and have found good agreement except for T2HK and BB350, for which the near–far extrapolation depends strongly on the poorly known ratio of ν_e and ν_μ QE cross sections. Therefore, the performance of these experiments strongly depends on the systematics assumptions, and it is difficult to make self-consistent predictions. We have also discussed the impact of the true value of δ on the measurement. While the performance is relatively uniform for most experiments, especially the precision attainable from NOνA$^+$ and NF5 depends on δ. NF5, for instance, was clearly optimized for CPV, whereas a precise measurement independent of δ would require higher muon energies and longer baselines, as realized in the NF10 setup. Interestingly, T2HK does not exhibit such a strong dependence on δ, in spite of the narrow beam spectrum.

For each experiment under consideration here, we have also identified the main limitations to a further increase in sensitivity. We have considered systematic uncertainties, exposure, and the impact of the near detector as possible bottlenecks. The results can be summarized as follows:

Superbeams can be divided into two classes: low and high energy. For the low-energy experiment T2HK, the fact that the ν_e cross sections needed for appearance measurements cannot be easily obtained from the near detector, is clearly the most important limitation. This is especially relevant since it operates in the QE regime, where cross section uncertainties are large and it is very difficult to relate the measured $\overset{(-)}{\nu}_\mu$ cross sections to the $\overset{(-)}{\nu}_e$ cross sections needed for the appearance measurement. Although the intrinsic beam backgrounds were included in our near detector simulations, we could not identify a simple way of measuring the $\overset{(-)}{\nu}_e$ cross sections directly with the required precision. Uncertainties in the the intrinsic beam background and the limited exposure are also limiting factors for T2HK, and we find that the availability of a near detector is of some importance. For WBB, LBNE$_{\text{mini}}$, and NOνA$^+$, which operate at higher energies, systematic errors can be controlled to the level needed, which in turn implies that, from the systematics point of view, very robust predictions can be made

for these experiments. The critical issue for them is instead exposure. For instance, for LBNE$_{\text{mini}}$, investing in the far detector mass may be more important than constructing a near detector. We conclude that for superbeams, the impact of systematic uncertainties depends mainly on the beam energy, especially because cross section uncertainties are much smaller in the high energy DIS region than in the low-energy QE region. The separation into narrow-band beams (T2HK, NOνA$^+$) and wide-band beams (WBB, LBNE$_{\text{mini}}$), on the other hand, has turned out not to be the primary issue.

Beta-beams using ^{6}He and ^{18}Ne for the neutrino production also suffer from the fact that the ratio between the $\overset{(-)}{\nu}_e$ and $\overset{(-)}{\nu}_\mu$ QE cross sections is needed as an external input. If a low-γ beta-beam is however combined with an SPL-based superbeam (BB+SPL), the performance is much better and much more robust than the one of a high-γ BB350. In fact, BB+SPL is the only experiment that could compete with a neutrino factory. The reason is that BB+SPL uses both the $\overset{(-)}{\nu}_e \to \overset{(-)}{\nu}_\mu$ and $\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e$ channels, so that both $\overset{(-)}{\nu}_e$ and $\overset{(-)}{\nu}_\mu$ cross sections can be measured.

Neutrino factories achieve the best absolute precision (comparable to that in the quark sector) and are very robust with respect to systematic errors. This is due to two main factors: firstly, the energy of the beam lies in the DIS regime where cross section uncertainties are small; secondly, this is the only experiment where the final flavor cross section can be determined in a self-consistent way from the disappearance data. Depending on baseline and muon energy, the most relevant factors affecting their performance are the matter density uncertainty (for setups with longer baselines and high E_μ, such as NF10) or exposure and near detector (for setups with shorter baselines and low E_μ, such as NF5).

We also find, remarkably, that near detectors have a relatively small impact and help to improve the precision in δ by only about 1–2° or less. The reason is that most systematic uncertainties are correlated between appearance and disappearance channels and can therefore be constrained by the far detector alone, provided that statistics in the disappearance channel is good enough to break correlations between systematic effects and the atmospheric oscillation parameters. The near detector turns out to be practically useless at beta-beam facilities: since the ν_e disappearance channel does not depend very much on the atmospheric parameters, the ν_e data from the far detector are even more useful for constraining flux and cross section uncertainties than in other experiments. It should be kept in mind, however, that near detectors will still be required to constrain effects of new physics in neutrino oscillations. In addition, if a combined analysis of appearance and disappearance data is not possible, a near detector proves to be critical in order to constrain cross section and flux uncertainties, as expected. Moreover, a well designed near detector facility is an excellent safeguard against "unknown unknowns".

The most attractive superbeam option, as far as the impact of systematic uncertainties is concerned, is a high-energy wide-band beam like LBNO or LBNE operating in the DIS regime. The LBNE$_{\text{mini}}$ experiment may be the first step towards such an experiment. However, both LBNO and LBNE have a limited discovery potential for CPV, and would suffer strongly from a reduction in statistics (notice that the WBB setup studied here had a LAr detector

with a fiducial mass of 100 kton). The ultimate precision can be reached with a neutrino factory, which is the only experiment with a precision competitive to the one achieved in the quark sector. The BB+SPL combination of a $\gamma = 100$ beta beam and a superbeam is a very interesting option that is very robust with respect to systematic errors and has a performance closer to neutrino factory than any other superbeam or beta-beam. Previous studies have underestimated the performance of BB+SPL because they did not take into account correlations between systematic uncertainties. Predictions for T2HK and BB350 heavily rely on external input on the flavor dependence of the QE cross sections. Here, an independent ν_e cross section measurement, for instance by a facility like ν-STORM, is necessary.

Acknowledgments

We would like to thank E. Fernandez-Martinez for illuminating discussions, and M. Mezzetto who contributed during the early stages of this work. We thank M. Vagins for providing the T2HK fluxes in machine readable format. WW would like to acknowledge support from DFG grants WI 2639/3-1 and WI 2639/4-1. This work has been supported by the U.S. Department of Energy under award number DE-SC0003915. PC, PH and WW would like to thank GGI Florence for hospitality during their stay within the "What's ν?" program. PC would also like to thank CERN and Fermilab for their hospitality during completion of this work. This work has been also supported by the EU FP7 projects EURONU (CE212372) and INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442). Fermilab is operated by Fermi Research Alliance under contract DE-AC02-07CH11359 with the United States Department of Energy.

A Simulation details for experimental setups

This appendix summarizes the technical details of our simulations for each of the setups included in our study.

The main parameters of all setups, *i.e.*, baselines, detector technology, fiducial masses, *etc.*, are summarized in Table 1, and the corresponding references are given in Sec. 2.1. The beam powers quoted in Table 1 are based on the cited references, but note that in some cases they were computed from an anticipated running time and a number of protons on target (pot). This requires an assumption on the experimental duty cycle, *i.e.*, the number of useful seconds per year: T2HK assumes 130 useful days per year (1.12×10^7 secs, approx.); NOνA$^+$ assumes 1.7×10^7 sec$\times$yr^{-1}; LBNE$_{\text{mini}}$ assumes 2×10^7 sec$\times$yr^{-1}; and BB+SPL and WBB assume 10^7 sec$\times$yr^{-1}. The maximum running time has been restricted to ten years for all experiments. This is usually split into neutrino and antineutrino running, except for neutrino factories where both muon polarities are circulating inside the decay ring at the same time.

The setups considered in this paper were defined as close as possible to the existing ones in the literature, *i.e.*, no further optimization with respect to beam or detector parameters (efficiencies, energy resolutions, *etc.*) was done. Such a study would be especially relevant

for the setups with LAr detectors, since their performance is still uncertain. We have simulated the LAr detector performance flat signal efficiencies and NC rejection efficiencies from Refs. [71, 72] for WBB and from Ref. [54] for $\mathsf{LBNE}_{\mathsf{mini}}$ and $\mathsf{NO}\nu\mathsf{A}^{+}$. In all cases, the same energy resolution as in Ref. [54] has been considered, and NC backgrounds have been migrated to low energies using matrices from the LBNE collaboration [54]. In the absence of any information regarding the performance of a LAr detector at the surface, we have it to be the same as for an underground LAr detector. We have used migration matrices (tabulated detector response functions) to simulate the detector performance for neutrino factories [73], beta-beams [47] and the SPL [27]. The signal reconstruction for T2HK has been simulated following Ref. [36], while the flux has been taken from Ref. [28].

The following sources of backgrounds have been considered in our analysis. For neutrino factory-based setups, only NC and charge mis-identification backgrounds have been included, since these are the only relevant backgrounds according to recent simulations of the MIND [66, 73]. For conventional beams and superbeams, we have followed the references mentioned in Sec. 2.1 and have included as main backgrounds NC events mis-identified as CC events and intrinsic contamination of the beam. Finally, only NC backgrounds have been considered for beta-beams. It is well-known that atmospheric neutrinos could also constitute a relevant source of background for this kind of experiments because suppression factors below 10^{-3} are difficult to achieve. This is especially true for the low-energy beta-beam at $\gamma = 100$. We have nevertheless omitted the atmospheric background in our simulations since Ref. [74] showed that it is only relevant for small θ_{13}, while for large values of θ_{13}, suppression factors as large as 10^{-2} can be tolerated.

For high energy ν_μ beams, there is an additional background coming from the production of τ leptons in the far detector. All facilities considered in this paper are located around the first atmospheric oscillation maximum where $\nu_\mu \to \nu_\tau$ oscillations are strong. For high energy beams such as the NF10 or WBB, the ν_τ energy is sufficient to produce a considerable amount of τ leptons at the detector. The leptonic τ decays $\tau \to e\nu\nu$ and $\tau \to \mu\nu\nu$, which have a branching ratio of 17% each, can lead to events that mimic ν_μ or ν_e charged current interactions. This phenomenon, known as the τ-contamination, has been studied in the context of the high-energy (25 GeV) neutrino factory [75–77], for non-magnetized detectors [78], and for wide band beams [53]. It was shown in Ref. [49] that τ contamination does not constitute a problem for the golden channel at a neutrino factory, and Ref. [75] shows that in the disappearance channel, it would only be problematic for precision measurements in the atmospheric sector.[14]

Contamination by $\tau \to e\nu\nu$ decays could in principle also affect high energy conventional neutrino beams. In the case of a typical superbeam such as WBB, with its flux peaking around 4-5 GeV, the majority of the fake events would be reconstructed with an energy below 1.5–2 GeV and would therefore affect mainly the second oscillation maximum. This could affect the measurement of δ. However, it was shown in Ref. [79] that for a wide-band beam with a detector located at the first oscillation maximum, the second maximum is of little importance anyway because it is strongly affected by NC backgrounds. Note also that ν_τ cross sections have large uncertainties, which could contribute significantly to the overall

[14] If the impact in the atmospheric sector is very large, it may have an indirect effect in the achievable precision in δ.

	Setups	ν app	$\bar{\nu}$ app	ν dis	$\bar{\nu}$ dis
Benchmark	NF10	44880/35	8701/61	159532/19	209577/21
	BB350	2447/378	2262/330	93775/–	106750/–
	T2HK	4754/2106	2006/2290	33788/544	168685/5502
	WBB	1830/248	147/148	5526/763	1884/515
Alternative	NF5	11022/4	2916/11	18337/2	32891/2
	BB100	1203/96	1048/81	65926/–	44776/–
	SPL	10455/1546	4453/1695	214524/9	93039/4
	$\mathsf{LBNE}_{\mathsf{mini}}$	389/162	63/102	3330/533	941/1419
	$\mathsf{NO}\nu\mathsf{A}^{+}$	752/590	155/386	7335/1255	3179/2397

Table 3: Total number of signal/background events expected at the far detectors of the experiments considered here. Numbers have been obtained assuming $\theta_{12} = 32^\circ$, $\theta_{23} = 45^\circ$, $\theta_{13} = 9^\circ$, $\delta = 0$, $\Delta m^2_{21} = 7 \times 10^{-5}$ eV^2 and $\Delta m^2_{31} = 3 \times 10^{-3}$ (normal hierarchy). Disappearance channels at beta-beams have been assumed to be background-free.

systematic error budget. Kinematic cuts on the momentum distributions of the visible final state particles, which might help to reduce the ν_τ contamination in superbeam experiments, are currently under investigation [53].

From these arguments it is clear that a dedicated study is required to address the actual impact of τ contamination on precision measurements, both at neutrino factories and at conventional neutrino beams. However, since this is beyond of the scope of the present paper, we have not included τ-related backgrounds for any of the setups studied in this work.

Table 3 shows the total number of events expected at the far detectors of the setups under study in this paper. In each case, the size of the near detector has been chosen such that the results are dominated by the statistics at the far detector. For this purpose, we require at least 10 times more disappearance events at the near detector compared to those obtained at the far detector. In the case of very long baseline setups with high density detectors (for instance, NF10 or WBB) 25 tons would be enough to fulfill this requirement. However, for setups with short baselines and very massive far detectors (T2HK, BB350, BB+SPL) the near detector mass had to be increased to 50, 100 and even 1 000 tons.

As mentioned in Sec. 2.1, we assume that the near detector is located sufficiently far away from the neutrino source for the neutrino spectra at the near and far sites to be similar. In the context of a neutrino factory, Ref. [38] demonstrates that a detector at a distance of 1 km from the end of a 600 m long decay straight can be simulated with an effective baseline (relative to an imaginary pointlike source) of 1.27 km. In this case, geometry effects arising from the straight section of the decay ring and the detector extension are expected

to be small. We have therefore adopted this value for the neutrino factory-based setups. Similar values have been considered for superbeam setups since the decay pipe for pions would be comparable in size to the storage ring of a neutrino factory, or even smaller. In the case of beta-beams on the other hand, the storage ring would need straight sections around 2 500 m in length to keep its livetime (*i.e.*, the useful fraction of the decay ring, $\ell = \frac{L_s}{L_t}$, where L_s and L_t are the lenght of the straight sections and the total lenght of the decay ring, respectively) around 35% (see, for instance, Ref. [80]). Therefore, we take the effective near detector baseline to be 2 km.

Finally, we have assumed the near and far detectors to be identical regarding signal and background rejection efficiencies as well as bin sizes and energy resolution. However, this may not apply in general, and some differences may arise as a consequence of a different detector technology or design. In particular, a different detector technology may imply a different background rejection efficiency. In all cases under study, the most relevant source of background are NC events misidentified as CC events. Therefore, we have (conservatively) assumed that the systematic uncertainties on this background are uncorrelated between the two detectors. However, for the sake of simplicity we have assumed the rejection efficiencies for this background to be the same for the two detectors in all cases. The rest of systematic uncertainties have been taken to be fully correlated between near and far detectors (with the exception of those affecting fiducial masses, obviously).

B Details on systematics implementation

The standard implementation of systematic uncertainties in GLoBES [43–45], based on the pull method [39,81], has been extended for this work to allow for an easier and more realistic treatment. For each GLoBES rule r,[15] we define a Poissonian χ^2 according to

$$\chi^2_r = \sum_i 2\left(T_{r,i}(\vec{\Theta},\vec{\xi}) - O_{r,i} + O_{r,i} \ln \frac{O_{r,i}}{T_{r,i}(\vec{\Theta},\vec{\xi}))} \right) . \tag{1}$$

Here $T_{r,i}(\vec{\Theta},\vec{\xi})$ is the predicted number of events in the i-th energy bin for this rule, for a particular set of oscillation parameters $\vec{\Theta}$ and systematic biases ("nuisance parameters") $\vec{\xi}$. $O_{r,i}$ is the "observed" event rate, *i.e.* the rate corresponding to the set of assumed "true" oscillation parameters. Both $T_{r,i}$ and $O_{r,i}$ receive contributions from different oscillation channels c:

$$T_{r,i}(\vec{\Theta},\vec{\xi}) = \sum_c \left(1 + a_{r,c}(\vec{\xi})\right) S_{r,c,i}(\vec{\Theta}) . \tag{2}$$

[15] A rule corresponds to a realistic data set including typically several experimentally indistinguishable signal and background components ("channels"). For example, in a superbeam using a WC detector, rules could correspond to electron-like and muon-like events. The channels contributing to each of these would would be actual ν_e and ν_μ interactions on the one hand, and background from beam contamination, neutral current interactions, flavor mis-identification, *etc.*, on the other hand.

Here, $S_{r,c,i}(\vec{\Theta})$ is the event rate that channel c contributes to rule r. The rule- and channel-dependent auxiliary parameters $a_{r,c}$ are given by

$$a_{r,c} \equiv \sum_k w_{r,c,k}\, \xi_k \,, \tag{3}$$

where the coefficients $w_{r,c,k}$ (which can be either zero or one) specify if a particular nuisance parameter ξ_k affects the contribution from channel c to rule r ($w_{r,c,k} = 1$) or not ($w_{r,c,k} = 0$). Thus the $w_{r,c,k}$ determine how systematic uncertainties are correlated among the rules and channels, and thus among detectors, beam polarities, flavors, signal and background, *etc.* For instance, in the case of a fiducial mass error, one would choose $w_{r,c,k} = 1$ for all rules describing data from a particular detector, and $w_{r,c,k} = 0$ for all other rules. In the fit, the nuisance parameters ξ_k are minimized over along with the oscillation parameters, and so-called *pull* terms are added to the χ^2 to ensure that their magnitude cannot get much larger than the systematic uncertainties σ_k they are implementing. The total χ^2 is

$$\chi^2 = \sum_r \chi^2_r + \sum_k \left(\frac{\xi_k}{\sigma_k}\right)^2 . \tag{4}$$

The important new feature compared to the standard treatment of systematic uncertainties in GLoBES is that the nuisance parameters ξ_k are no longer associated with a particular rule, but are defined globally and can in principle affect any rule or channel.

References

[1] G. L. Fogli, E. Lisi, A. Marrone, A. Palazzo, and A. M. Rotunno, *Hints of $\theta_{13} > 0$ from global neutrino data analysis*, *Phys. Rev. Lett.* **101** (2008) 141801, [arXiv:0806.2649].

[2] T. Schwetz, M. Tortola, and J. W. F. Valle, *Three-flavour neutrino oscillation update*, *New J. Phys.* **10** (2008) 113011, [arXiv:0808.2016].

[3] M. C. Gonzalez-Garcia, M. Maltoni, and J. Salvado, *Updated global fit to three neutrino mixing: status of the hints of theta13 ¿ 0*, *JHEP* **04** (2010) 056, [arXiv:1001.4524].

[4] T. Schwetz, M. Tortola, and J. Valle, *Global neutrino data and recent reactor fluxes: status of three-flavour oscillation parameters*, *New J.Phys.* **13** (2011) 063004, [arXiv:1103.0734].

[5] **T2K** Collaboration, K. Abe *et. al.*, *Indication of Electron Neutrino Appearance from an Accelerator-produced Off-axis Muon Neutrino Beam*, *Phys.Rev.Lett.* **107** (2011) 041801, [arXiv:1106.2822].

[6] **MINOS** Collaboration, P. Adamson *et. al.*, *Improved search for muon-neutrino to electron-neutrino oscillations in MINOS*, *Phys.Rev.Lett.* **107** (2011) 181802, [arXiv:1108.0015].

[7] **DOUBLE-CHOOZ** Collaboration, Y. Abe *et. al.*, *Indication for the disappearance of reactor electron antineutrinos in the Double Chooz experiment*, *Phys.Rev.Lett.* **108** (2012) 131801, [arXiv:1112.6353].

[8] **DAYA-BAY** Collaboration, F. An *et. al.*, *Observation of electron-antineutrino disappearance at Daya Bay*, *Phys.Rev.Lett.* **108** (2012) 171803, [arXiv:1203.1669].

[9] **RENO collaboration** Collaboration, J. Ahn *et. al.*, *Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment*, *Phys.Rev.Lett.* **108** (2012) 191802, [arXiv:1204.0626].

[10] M. Gonzalez-Garcia, M. Maltoni, J. Salvado, and T. Schwetz, *Global fit to three neutrino mixing: critical look at present precision*, arXiv:1209.3023.

[11] G. Fogli, E. Lisi, A. Marrone, D. Montanino, A. Palazzo, *et. al.*, *Global analysis of neutrino masses, mixings and phases: entering the era of leptonic CP violation searches*, arXiv:1205.5254.

[12] D. Forero, M. Tortola, and J. Valle, *Global status of neutrino oscillation parameters after recent reactor measurements*, arXiv:1205.4018.

[13] P. Coloma, A. Donini, E. Fernandez-Martinez, and P. Hernandez, *Precision on leptonic mixing parameters at future neutrino oscillation experiments*, arXiv:1203.5651.

[14] P. Huber, M. Lindner, T. Schwetz, and W. Winter, *First hint for CP violation in neutrino oscillations from upcoming superbeam and reactor experiments*, *JHEP* **0911** (2009) 044, [arXiv:0907.1896].

[15] M. Blennow and T. Schwetz, *Identifying the Neutrino mass Ordering with INO and NOvA*, arXiv:1203.3388.

[16] E. K. Akhmedov, S. Razzaque, and A. Y. Smirnov, *Mass hierarchy, 2-3 mixing and CP-phase with Huge Atmospheric Neutrino Detectors*, arXiv:1205.7071.

[17] S. T. Petcov and M. Piai, *The lma msw solution of the solar neutrino problem, inverted neutrino mass hierarchy and reactor neutrino experiments*, *Phys. Lett.* **B533** (2002) 94–106, [hep-ph/0112074].

[18] S. Choubey, S. Petcov, and M. Piai, *Precision neutrino oscillation physics with an intermediate baseline reactor neutrino experiment*, *Phys.Rev.* **D68** (2003) 113006, [hep-ph/0306017].

[19] L. Zhan, Y. Wang, J. Cao, and L. Wen, *Experimental Requirements to Determine the Neutrino Mass Hierarchy Using Reactor Neutrinos*, *Phys.Rev.* **D79** (2009) 073007, [arXiv:0901.2976].

[20] X. Qian, D. Dwyer, R. McKeown, P. Vogel, W. Wang, *et. al.*, *Mass Hierarchy Resolution in Reactor Anti-neutrino Experiments: Parameter Degeneracies and Detector Energy Response*, arXiv:1208.1551.

[21] H. Nunokawa, S. J. Parke, and R. Zukanovich Funchal, *Another possible way to determine the neutrino mass hierarchy*, *Phys.Rev.* **D72** (2005) 013009, [hep-ph/0503283].

[22] A. de Gouvea, J. Jenkins, and B. Kayser, *Neutrino mass hierarchy, vacuum oscillations, and vanishing u—e3—*, *Phys. Rev.* **D71** (2005) 113009, [hep-ph/0503079].

[23] V. Barger, P. Huber, D. Marfatia, and W. Winter, *Which long-baseline neutrino experiments are preferable?*, *Phys.Rev.* **D76** (2007) 053005, [hep-ph/0703029].

[24] W. Winter, *Minimal Neutrino Beta Beam for Large θ_{13}*, *Phys. Rev.* **D78** (2008) 037101, [arXiv:0804.4000].

[25] J. Tang and W. Winter, *Neutrino factory in stages: Low energy, high energy, off-axis*, *Phys. Rev.* **D81** (2010) 033005, [arXiv:0911.5052].

[26] P. Huber, M. Maltoni, and T. Schwetz, *Resolving parameter degeneracies in long-baseline experiments by atmospheric neutrino data*, *Phys.Rev.* **D71** (2005) 053006, [hep-ph/0501037].

[27] J.-E. Campagne, M. Maltoni, M. Mezzetto, and T. Schwetz, *Physics potential of the CERN-MEMPHYS neutrino oscillation project*, *JHEP* **0704** (2007) 003, [hep-ph/0603172].

[28] K. Abe, T. Abe, H. Aihara, Y. Fukuda, Y. Hayato, *et. al.*, *Letter of Intent: The Hyper-Kamiokande Experiment — Detector Design and Physics Potential —*, arXiv:1109.3262.

[29] V. Barger, R. Gandhi, P. Ghoshal, S. Goswami, D. Marfatia, *et. al.*, *Neutrino mass hierarchy and octant determination with atmospheric neutrinos*, arXiv:1203.6012.

[30] M. Fukugita and T. Yanagida, *Baryogenesis Without Grand Unification*, *Phys.Lett.* **B174** (1986) 45.

[31] S. King, *Predicting neutrino parameters from SO(3) family symmetry and quark-lepton unification*, *JHEP* **0508** (2005) 105, [hep-ph/0506297].

[32] I. Masina, *A Maximal atmospheric mixing from a maximal CP violating phase*, *Phys.Lett.* **B633** (2006) 134–140, [hep-ph/0508031].

[33] S. Antusch and S. F. King, *Charged lepton corrections to neutrino mixing angles and CP phases revisited*, *Phys.Lett.* **B631** (2005) 42–47, [hep-ph/0508044].

[34] W. Winter, *Understanding cp phase-dependent measurements at neutrino superbeams in terms of bi-rate graphs*, *Phys. Rev.* **D70** (2004) 033006, [hep-ph/0310307].

[35] P. Huber, M. Lindner, and W. Winter, *From parameter space constraints to the precision determination of the leptonic Dirac CP phase*, *JHEP* **0505** (2005) 020, [hep-ph/0412199].

[36] P. Huber, M. Mezzetto, and T. Schwetz, *On the impact of systematical uncertainties for the CP violation measurement in superbeam experiments*, *JHEP* **0803** (2008) 021, [arXiv:0711.2950].

[37] P. Coloma and E. Fernandez-Martinez, *Optimization of neutrino oscillation facilities for large θ_{13}*, *JHEP* **1204** (2012) 089, [arXiv:1110.4583].

[38] J. Tang and W. Winter, *Physics with near detectors at a neutrino factory*, *Phys.Rev.* **D80** (2009) 053001, [arXiv:0903.3039].

[39] P. Huber, M. Lindner, and W. Winter, *Superbeams versus neutrino factories*, *Nucl. Phys.* **B645** (2002) 3–48, [hep-ph/0204352].

[40] T. Ohlsson and W. Winter, *The role of matter density uncertainties in the analysis of future neutrino factory experiments*, *Phys. Rev.* **D68** (2003) 073007, [hep-ph/0307178].

[41] P. Huber, M. Lindner, T. Schwetz, and W. Winter, *Reactor neutrino experiments compared to superbeams*, *Nucl.Phys.* **B665** (2003) 487–519, [hep-ph/0303232].

[42] P. Huber, M. Lindner, M. Rolinec, T. Schwetz, and W. Winter, *Prospects of accelerator and reactor neutrino oscillation experiments for the coming ten years*, *Phys. Rev.* **D70** (2004) 073014, [hep-ph/0403068].

[43] P. Huber, J. Kopp, M. Lindner, M. Rolinec, and W. Winter, *From Double Chooz to Triple Chooz: Neutrino physics at the Chooz reactor complex*, *JHEP* **05** (2006) 072, [hep-ph/0601266].

[44] P. Huber, M. Lindner, and W. Winter, *Simulation of long-baseline neutrino oscillation experiments with GLoBES*, *Comput. Phys. Commun.* **167** (2005) 195, [hep-ph/0407333]. http://www.mpi-hd.mpg.de/lin/globes/.

[45] P. Huber, J. Kopp, M. Lindner, M. Rolinec, and W. Winter, *New features in the simulation of neutrino oscillation experiments with GLoBES 3.0*, *Comput. Phys. Commun.* **177** (2007) 432–438, [hep-ph/0701187].

[46] J. Burguet-Castell, D. Casper, J. Gomez-Cadenas, P. Hernandez, and F. Sanchez, *Neutrino oscillation physics with a higher gamma beta beam*, *Nucl.Phys.* **B695** (2004) 217–240, [hep-ph/0312068].

[47] J. Burguet-Castell, D. Casper, E. Couce, J. Gomez-Cadenas, and P. Hernandez, *Optimal beta-beam at the CERN-SPS*, *Nucl.Phys.* **B725** (2005) 306–326, [hep-ph/0503021].

[48] J. Bernabeu, M. Blennow, P. Coloma, A. Donini, C. Espinoza, *et. al.*, *EURONU WP6 2009 yearly report: Update of the physics potential of Nufact, superbeams and betabeams*, arXiv:1005.3146.

[49] S. K. Agarwalla, P. Huber, J. Tang, and W. Winter, *Optimization of the Neutrino Factory, revisited*, *JHEP* **1101** (2011) 120, [arXiv:1012.1872].

[50] "International design study of the neutrino factory." http://www.ids-nf.org.

[51] A. Longhin, *Optimization of neutrino fluxes for European super-beams*, *PoS* **ICHEP2010** (2010) 325.

[52] **LAGUNA** Collaboration, D. Angus *et. al.*, *The LAGUNA design study - towards giant liquid based underground detectors for neutrino physics and astrophysics and proton decay searches*, arXiv:1001.0077.

[53] **LAGUNA-LBNO** Collaboration, A. Rubbia *et. al.*, *Expression of interest for a very long baseline neutrino oscillation experiment (LBNO)*, Tech. Rep. CERN-SPSC-2012-021, SPSC-EOI-007, CERN, 2012.

[54] **LBNE** Collaboration, T. Akiri *et. al.*, *The 2010 Interim Report of the Long-Baseline Neutrino Experiment Collaboration Physics Working Groups*, arXiv:1110.6249.

[55] A. Donini, D. Meloni, and S. Rigolin, *The Impact of solar and atmospheric parameter uncertainties on the measurement of theta(13) and delta*, *Eur.Phys.J.* **C45** (2006) 73–95, [hep-ph/0506100].

[56] C. Andreopoulos *et. al.*, *The GENIE Neutrino Monte Carlo Generator*, *Nucl. Instrum. Meth.* **A614** (2010) 87–104, [arXiv:0905.2517].

[57] **MINOS** Collaboration, P. Adamson *et. al.*, *Neutrino and Antineutrino Inclusive Charged-current Cross Section Measurements with the MINOS Near Detector*, *Phys.Rev.* **D81** (2010) 072002, [arXiv:0910.2201].

[58] **IDS-NF** Collaboration, S. Choubey *et. al.*, *International Design Study for the Neutrino Factory, Interim Design Report*, arXiv:1112.2853.

[59] E. Fernandez-Martinez and D. Meloni, *Importance of nuclear effects in the measurement of neutrino oscillation parameters*, *Phys.Lett.* **B697** (2011) 477–481, [arXiv:1010.2329].

[60] D. Meloni, M. Martini, and M. Martini, *Revisiting the T2K data using different models for the neutrino-nucleus cross sections*, *Phys.Lett.* **B716** (2012) 186–192, [arXiv:1203.3335].

[61] M. Martini, M. Ericson, and G. Chanfray, *Neutrino energy reconstruction problems and neutrino oscillations*, *Phys.Rev.* **D85** (2012) 093012, [arXiv:1202.4745].

[62] O. Lalakulich and U. Mosel, *Energy reconstruction in quasielastic scattering in the MiniBooNE and T2K experiments*, arXiv:1208.3678.

[63] M. Day and K. S. McFarland, *Differences in Quasi-Elastic Cross-Sections of Muon and Electron Neutrinos*, arXiv:1206.6745.

[64] A. M. Dziewonski and D. L. Anderson, *Preliminary reference Earth model*, *Phys. Earth Planet. Interiors* **25** (1981) 297–356.

[65] E. Kozlovskaya, J. Peltoniemi, and J. Sarkamo, *The density distribution in the Earth along the CERN-Pyhasalmi baseline and its effect on neutrino oscillations*, hep-ph/0305042.

[66] R. Bayes, A. Laing, F. Soler, A. C. Villanueva, J. G. Cadenas, *et. al.*, *The Golden Channel at a Neutrino Factory revisited: improved sensitivities from a Magnetised Iron Neutrino Detector*, arXiv:1208.2735.

[67] A. Lenz, U. Nierste, J. Charles, S. Descotes-Genon, H. Lacker, *et. al.*, *New Physics in B-Bbar mixing in the light of recent LHCb data*, arXiv:1203.0238.

[68] J. Kopp, M. Lindner, T. Ota, and J. Sato, *Non-standard neutrino interactions in reactor and superbeam experiments*, *Phys. Rev.* **D77** (2008) 013007, [0708.0152].

[69] S. Antusch, M. Blennow, E. Fernandez-Martinez, and J. Lopez-Pavon, *Probing non-unitary mixing and CP-violation at a Neutrino Factory*, *Phys.Rev.* **D80** (2009) 033002, [arXiv:0903.3986].

[70] **nuSTORM** Collaboration, P. Kyberd *et. al.*, *nuSTORM - Neutrinos from STORed Muons: Letter of Intent to the Fermilab Physics Advisory Committee*, arXiv:1206.0294.

[71] P. Coloma, E. Fernandez-Martinez, and L. Labarga, *Physics reach of CERN-based SuperBeam neutrino oscillation experiments*, arXiv:1206.0475.

[72] P. Coloma, T. Li, and S. Pascoli, *A comparative study of long-baseline superbeams within LAGUNA for large θ_{13}*, arXiv:1206.4038.

[73] A. B. Laing, *Optimisation of detectors for the golden channel at a neutrino factory*, .

[74] E. Fernandez-Martinez, *The gamma = 100 beta-Beam revisited*, *Nucl.Phys.* **B833** (2010) 96–107, [arXiv:0912.3804].

[75] D. Indumathi and N. Sinha, *Effect of tau neutrino contribution to muon signals at neutrino factories*, *Phys.Rev.* **D80** (2009) 113012, [0910.2020].

[76] A. Donini, J. J. Gomez Cadenas, and D. Meloni, *The τ-contamination of the golden muon sample at the Neutrino Factory*, arXiv:1005.2275.

[77] R. Dutta, D. Indumathi, and N. Sinha, *Tau contamination in the platinum channel at neutrino factories*, *Phys.Rev.* **D85** (2012) 013003, [arXiv:1103.5578].

[78] J. Tang and W. Winter, *Requirements for a New Detector at the South Pole Receiving an Accelerator Neutrino Beam*, *JHEP* **1202** (2012) 028, [arXiv:1110.5908].

[79] P. Huber and J. Kopp, *Two experiments for the price of one? – The role of the second oscillation maximum in long baseline neutrino experiments*, *JHEP* **1103** (2011) 013, [arXiv:1010.3706].

[80] S. Choubey, P. Coloma, A. Donini, and E. Fernandez-Martinez, *Optimized Two-Baseline Beta-Beam Experiment*, *JHEP* **0912** (2009) 020, [arXiv:0907.2379].

[81] G. L. Fogli, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, *Getting the most from the statistical analysis of solar neutrino oscillations*, *Phys. Rev.* **D66** (2002) 053010, [hep-ph/0206162].