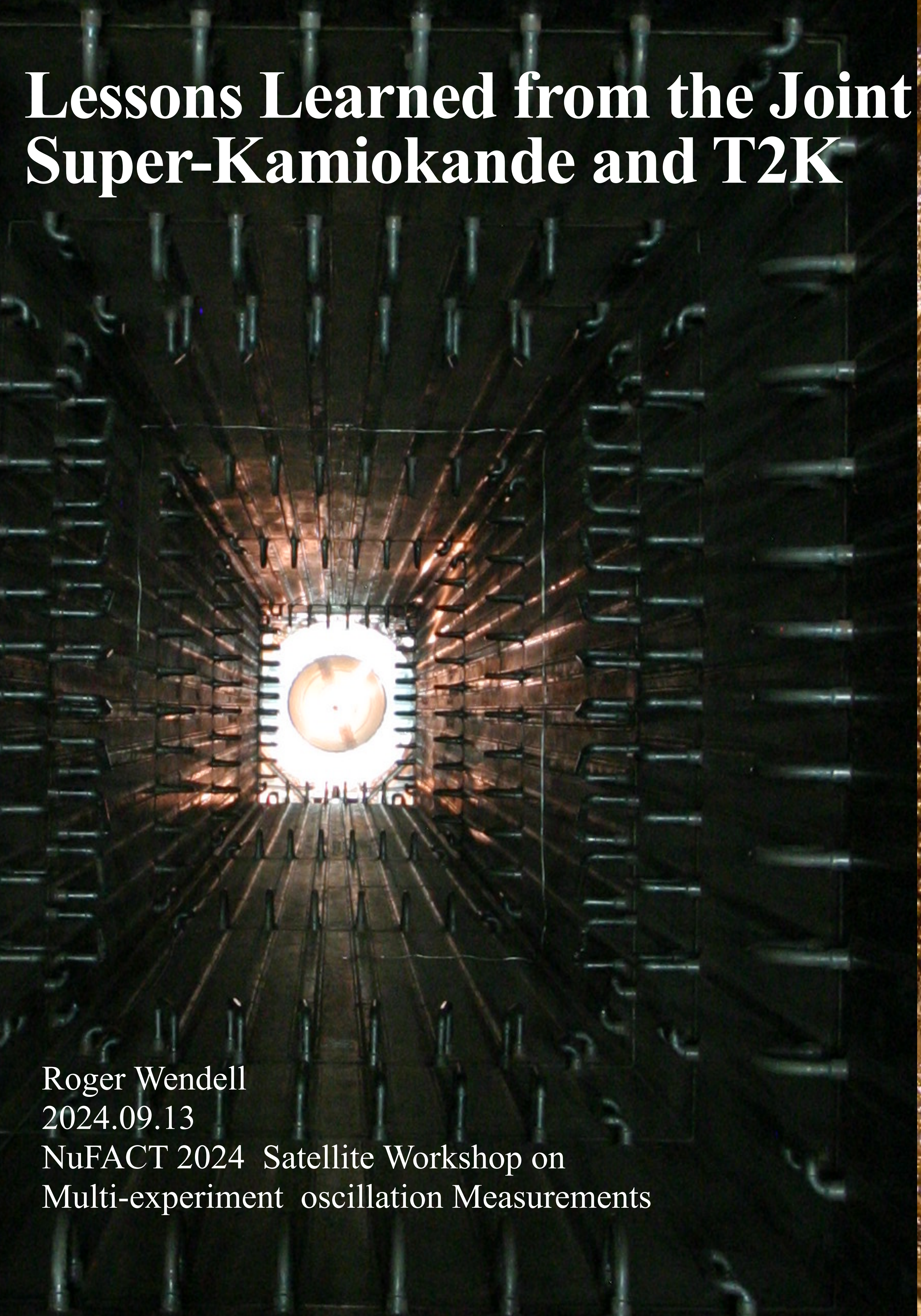
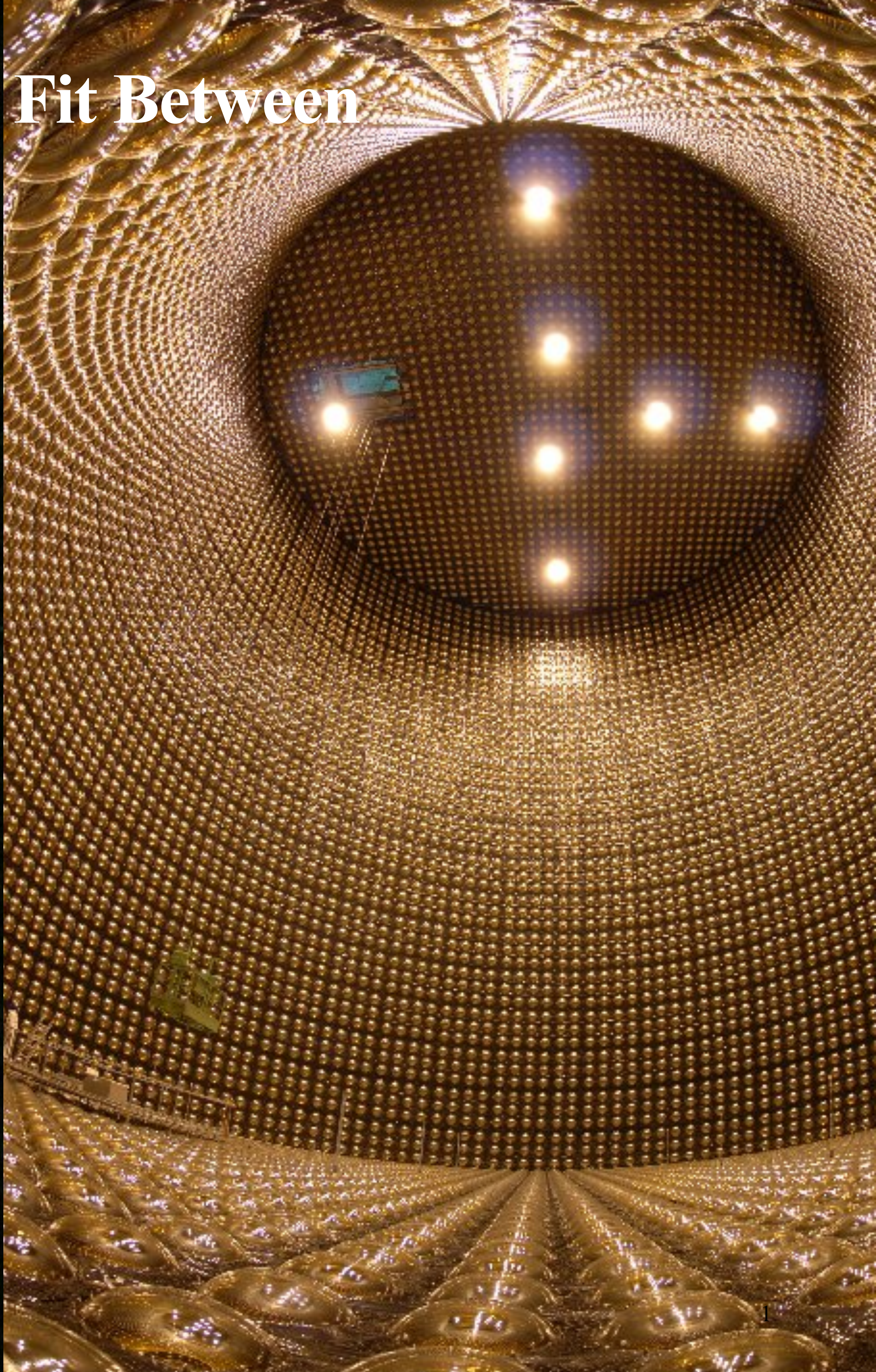
Lessons Learned from the Joint Fit Between Super-Kamiokande and T2K





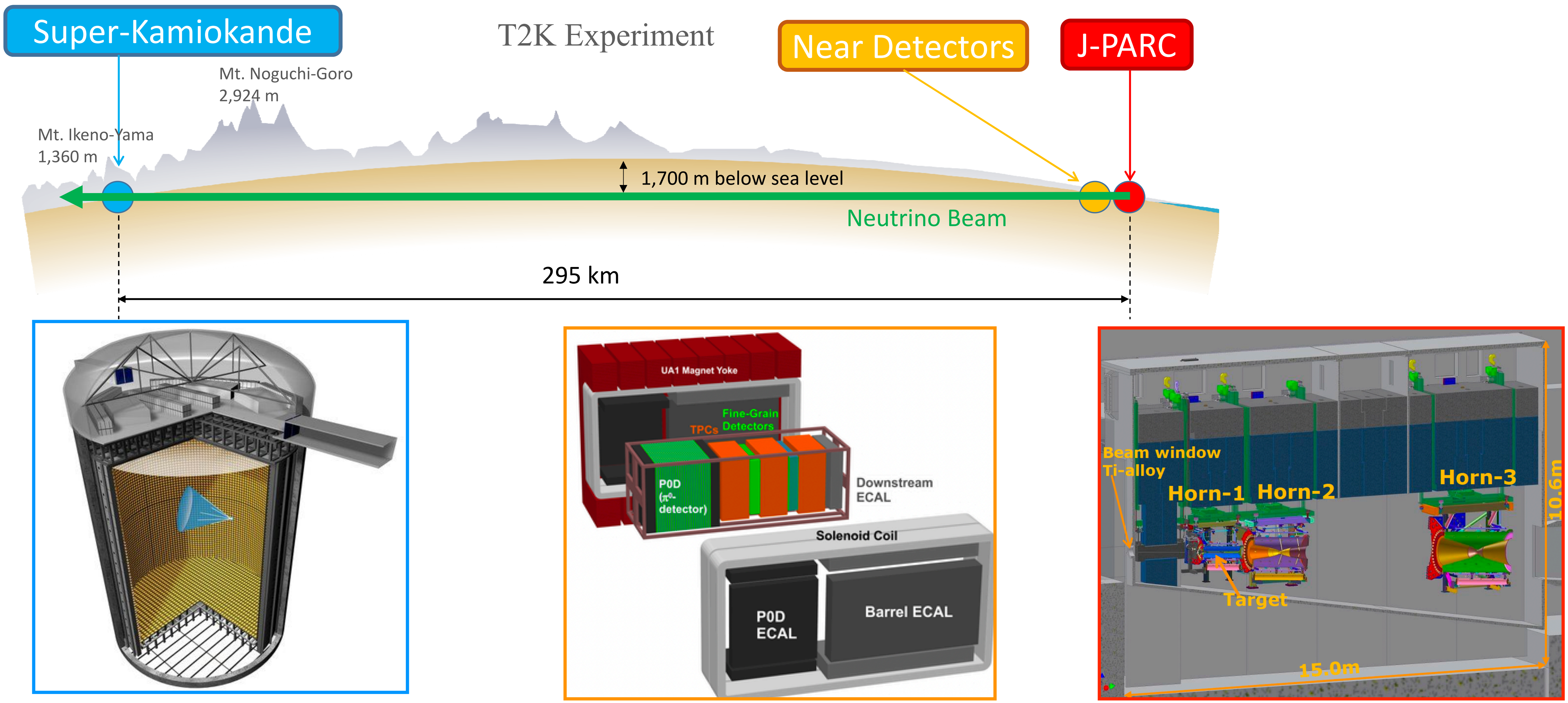
Roger Wendell
2024.09.13
NuFACT 2024 Satellite Workshop on
Multi-experiment oscillation Measurements

Warning

- Most of the non-technical aspects of the following slides are based entirely on R. Wendell's recollections and impressions of the path to a joint fit between Super-K and T2K
  - Many moving parts over many years
  - Some memories and impressions have faded
  - Some details have been omitted for brevity

- Despite the relative proximity between the Super-K and T2K experiments, getting to a robust analysis required a lot of technical and political work by many dedicated people

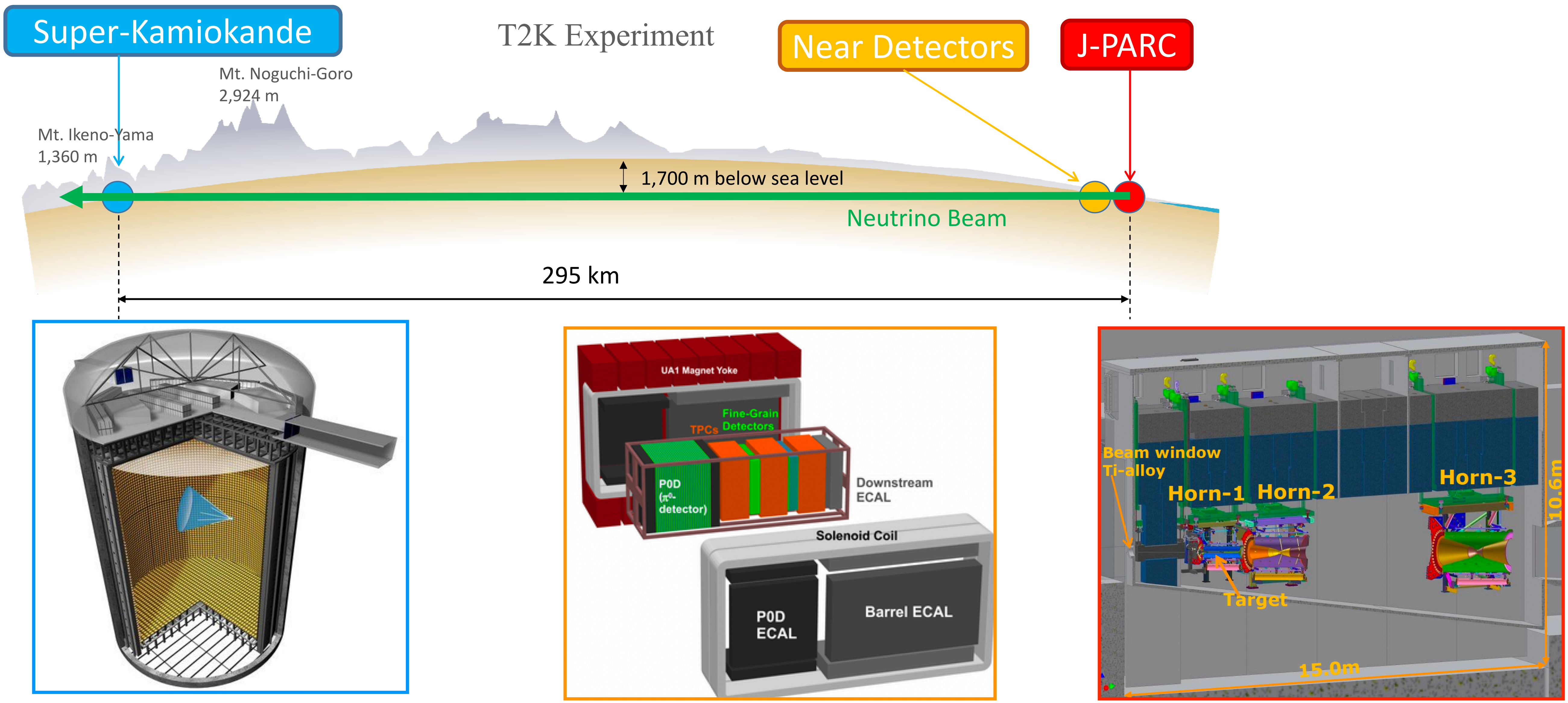
Background information

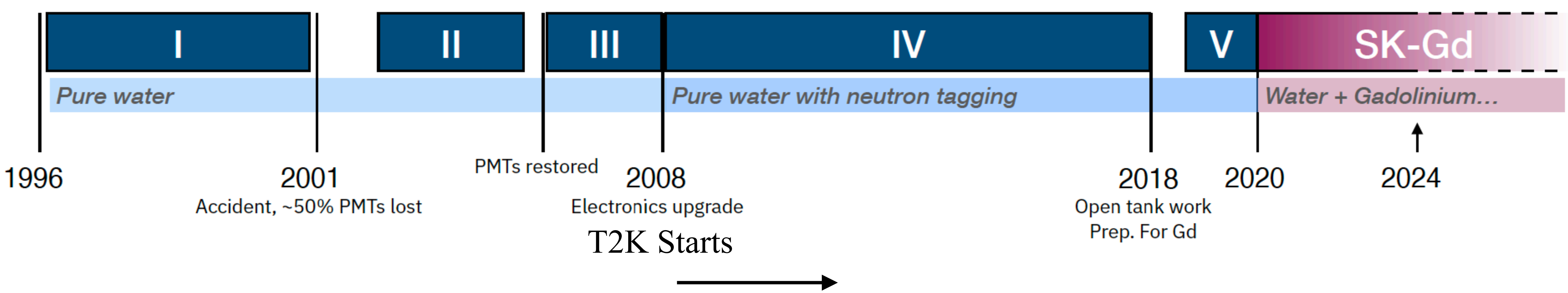


- Super-Kamiokande (SK) and T2K are separate experiments, with separate cultures and analysis tools
  - SK data taking started in 1996
  - T2K data taking started in 2009

- However, the SK detector is used as part of T2K
  - This makes a joint analysis of beam (T2K) and atmospheric (SK) data attractive
  - …but surprisingly challenging

Background information





Background information: Political Challenges

Background information: Political Challenges

- Proposals for a joint analysis started *as early as 2011*, maybe even earlier by various parties on the experiments
  - Some were not formal proposals (add χ^2…), but discussed among the management class

- Reasons to not do a joint fit heard throughout the years
  - Rejected: "T2K has just started taking data, needs to produce its own results first"
  - Opinion: Natural objection that is likely to confront any fit between two collaborations

  - Rejected: "Don't want to be bound to the *other* experiment while extending our analysis"
  - At that time:
    - T2K implementing **new** analysis techniques, **new** reconstruction algorithm
    - SK developing **new** event selections (neutrons) with "tried and true" tools
  - Opinion: A common framework (MC, reconstruction, error model) is important for extracting the most from a joint fit. Hard to divert resources when collaborations are growing separately

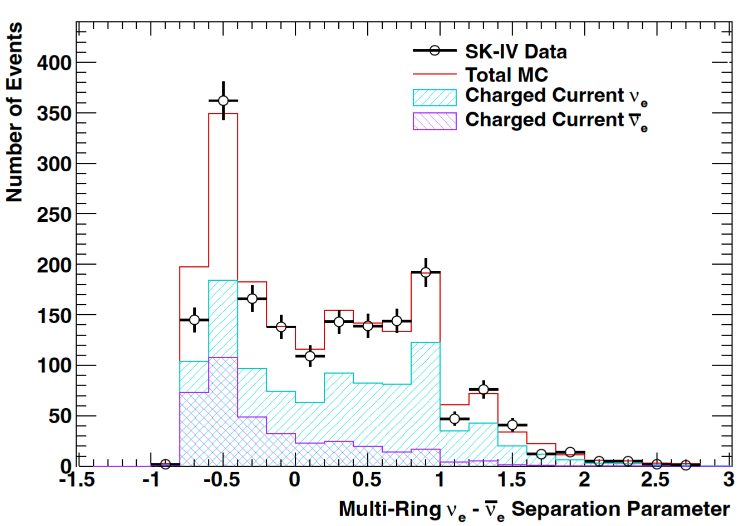
  - Rejected: "Is there really any benefit?" , "Is it even feasible"?
  - Opinion: Intuitively many understand a joint fit is "better" but demonstrating it realistically is important to convince people.
  - Opinion: Despite a lot of overlap among collaborators many people are mostly focused on either "T2K" or "Super-K" and do not follow the other experiment's analysis closely

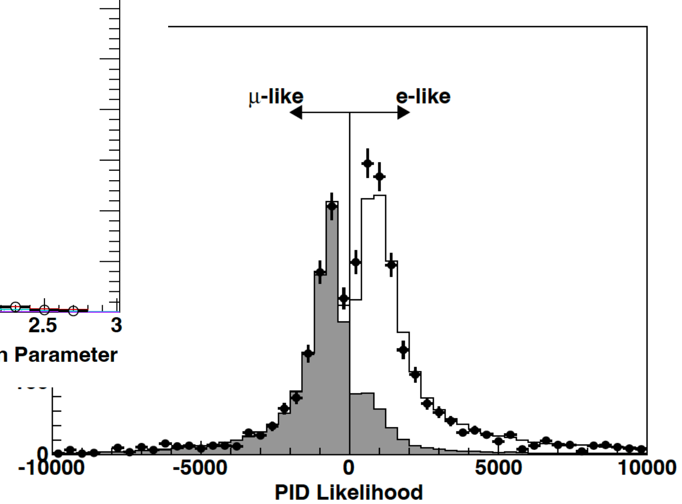
SK+T2K Analysis: A Staged Approach

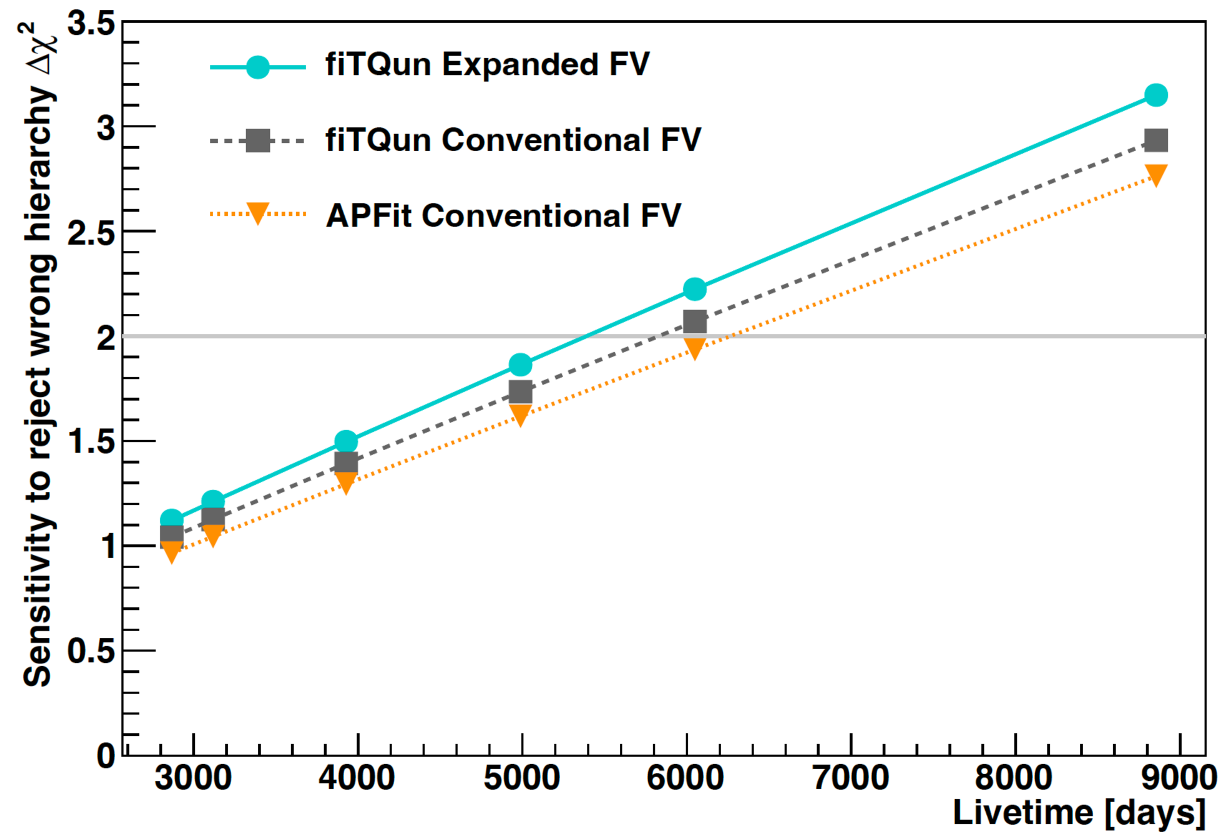
- Overcame many of the political difficulties with a staged approach
  - Opinion: Starting work towards a real joint fit early was essential to building trust and for demonstrating the need and benefit of the analysis

- Demonstrate capabilities of new reconstruction algorithm on atmospheric neutrino analysis with a dedicated analysis on the Super-K side
  - Starting in **2016**, *built a new SK analysis* with "T2K analysis tool" (fiTQun)
  - Implemented some changes and improvements to the algorithm on the way

True Number of Rings	fiTQun Reconstruction 1R	2R	$\geq$ 3R	APFit Reconstruction 1R	2R	$\geq$ 3R
True 1R	95.0%	4.64%	0.41%	95.9%	3.85%	0.29%
True 2R	27.8%	66.7%	5.56%	42.5%	52.8%	4.63%
True $\geq$3R	7.04%	25.5%	67.5%	20.2%	33.0%	46.8%





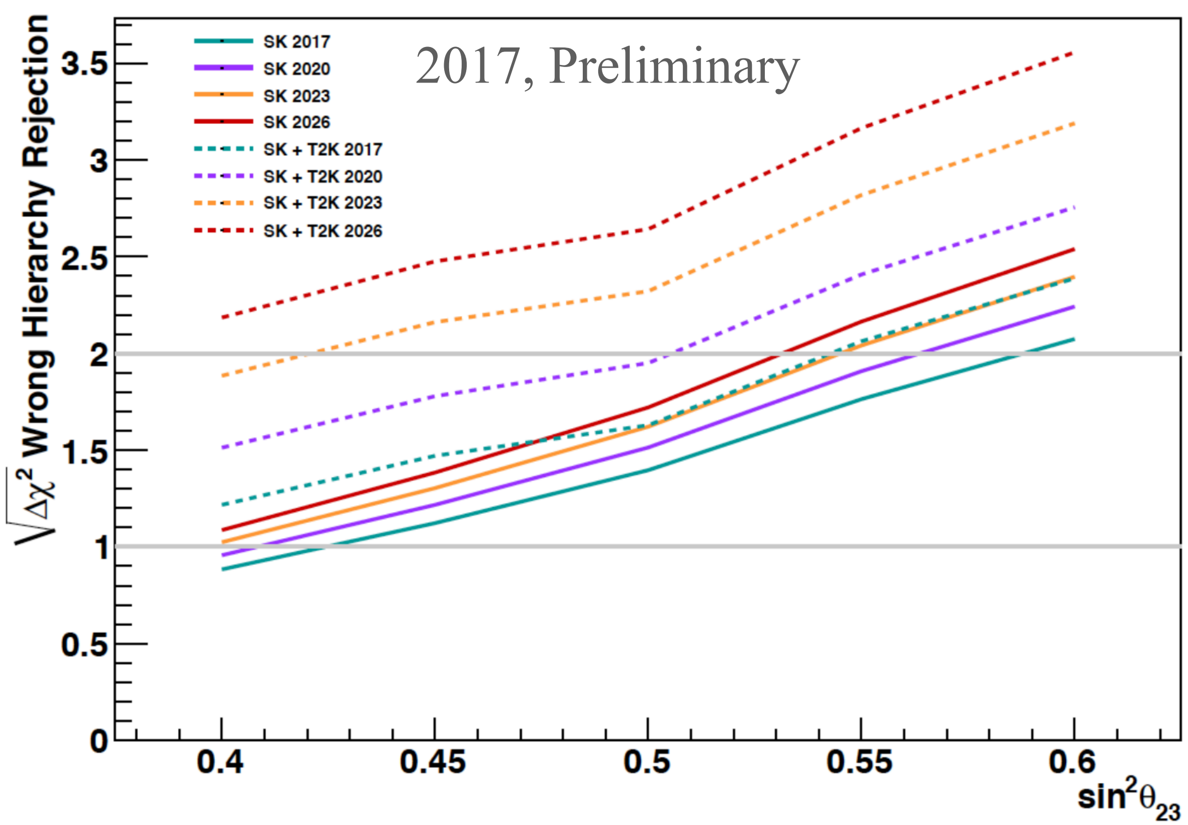


SK+T2K Analysis: A Staged Approach

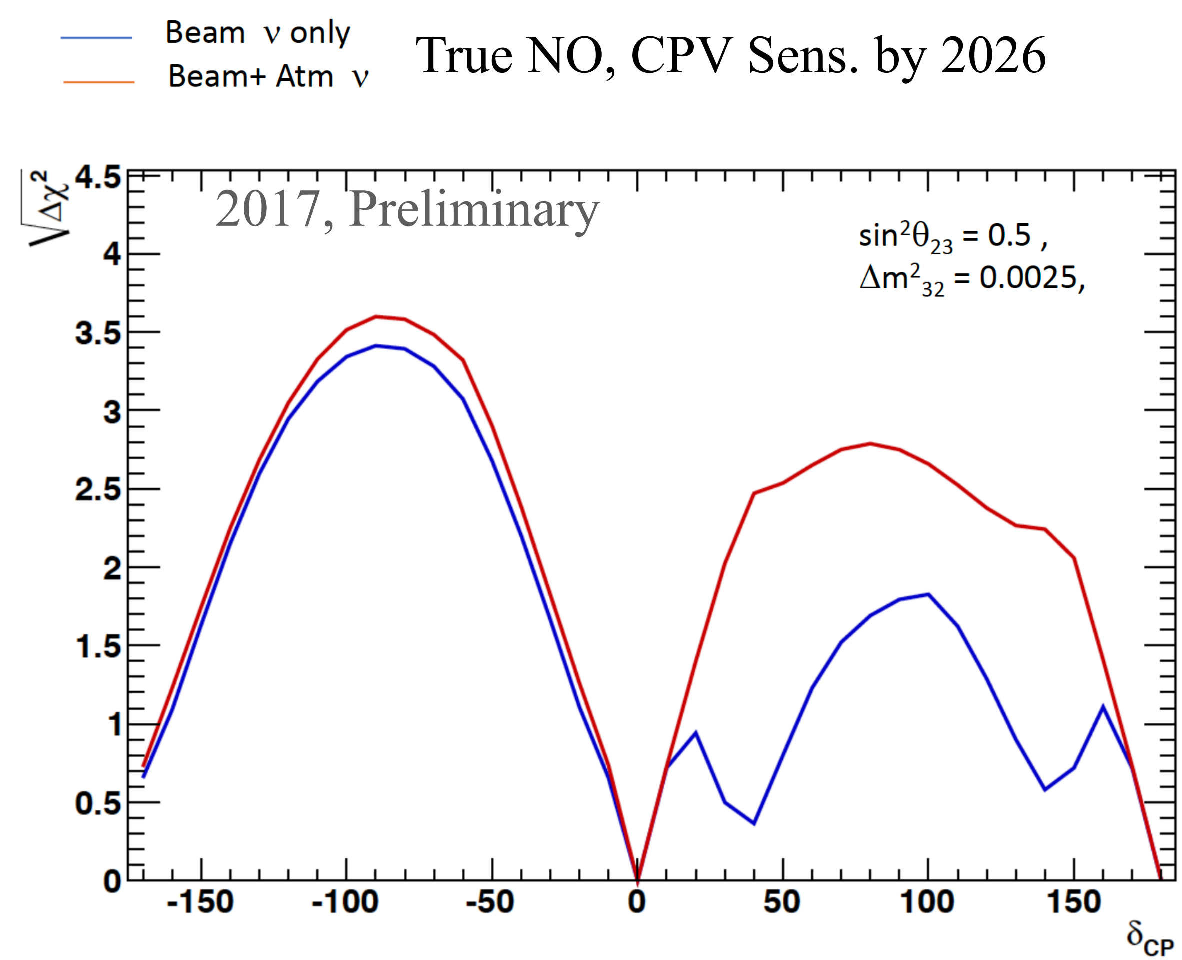
- Overcame many of the political difficulties with a staged approach
  - Opinion: Starting work towards a real joint fit early was essential to building trust and for demonstrating the need and benefit of the analysis

- Demonstrate expected sensitivity benefit of a joint fit using that SK analysis, reweighted T2K beam flux (public) and simple assumptions about systematic errors and their correlations [Plots from **2017**]

True NO, MO Sens. by 2026



True NO, CPV Sens. by 2026



  - Opinion: Clearly shows benefit to *both* experiments using a consistent framework

SK+T2K Analysis: A Staged Approach

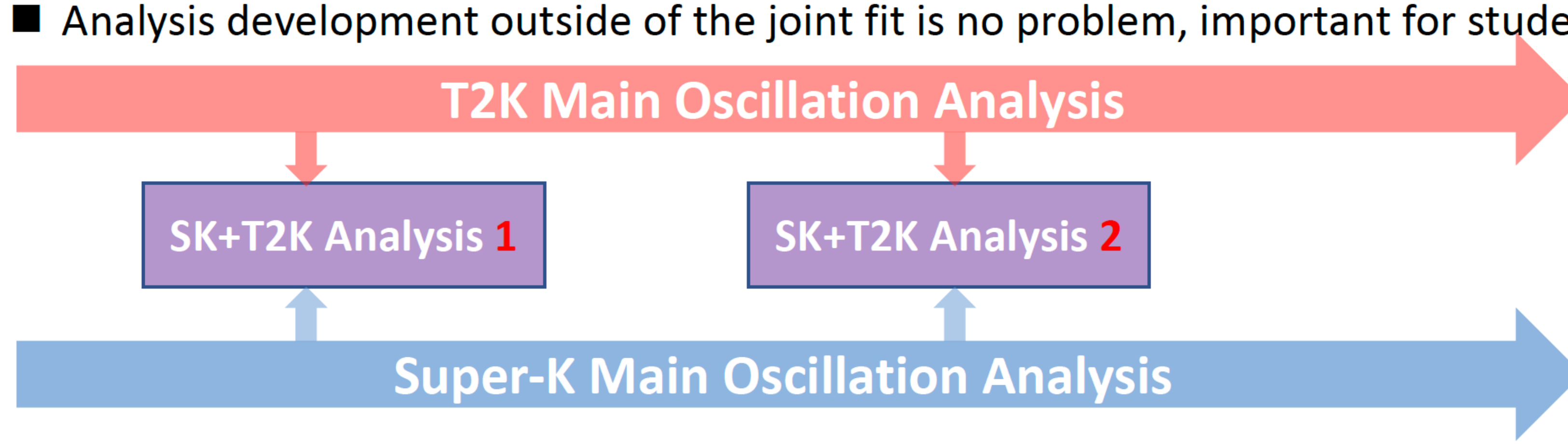
- Overcame many of the political difficulties with a staged approach
  - Opinion: Starting work towards a real joint fit early was essential to building trust and for demonstrating both the need and benefit of the analysis

- With these results in hand, begin private discussions with leadership from both experiments and collecting interested analyzers
  - Presentations at analysis and collaboration meetings with basic idea of a joint analysis

May 2018

Analysis Tenents

- Open Access For Analysis
  - T2K and Super-K analysis inputs (Data, MC, Systematic Errors+Covariances) are openly available, with consideration for blind analysis issues. Free choice of analysis method.
- Go With What Works
  - Use existing analysis samples from T2K and SK; there is no need to develop additional samples explicitly for a joint analysis
- Leave it to the Experts
  - Sample and bin definitions, detector calibration, (some) systematic errors should be left to the people that know them the best
- No restrictions on T2K only or SK only analysis development
  - Analysis development outside of the joint fit is no problem, important for students



SK+T2K Analysis: A Staged Approach

- Overcame many of the political difficulties with a staged approach
  - Opinion: Starting work towards a real joint fit early was essential to building trust and for demonstrating the need and benefit of the analysis

- Wrote MOU based on feedback from public and private discussions, circulated to collaboration leadership
  - 6~8 Months of iteration, finally signed in July 2019
  - <u>Despite preparations still ran into unforseen problems later in the analysis</u>

Article I
Governance

Article II
Working Group Participation

Article III
Data, Sharing, and Responsibilities

Article IV
Officialization and Publication of Results

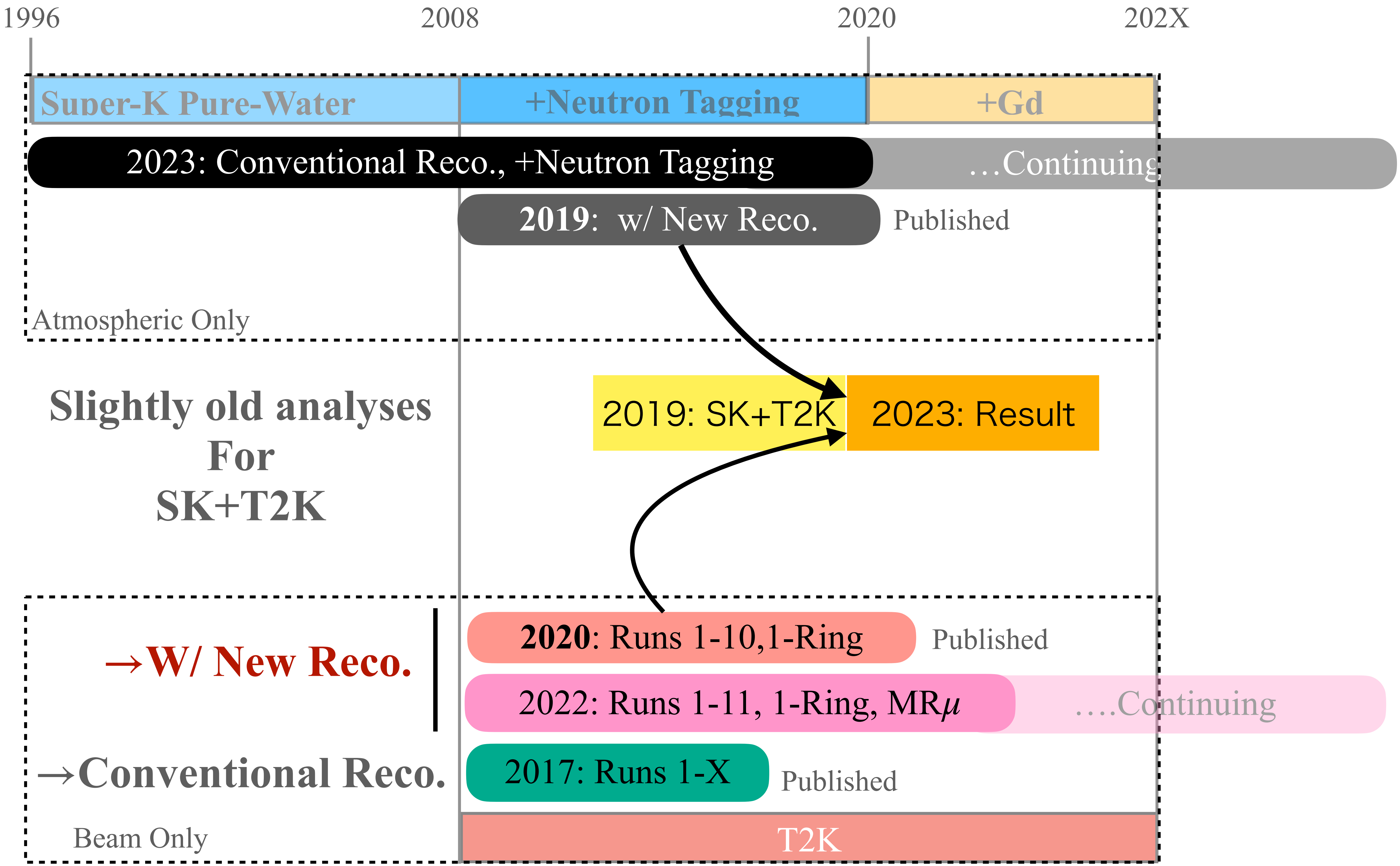
Article V
Scope

Article VII
Duration

- Written to address basic analysis tenets but with enough flexibility to allow modifications where necessary

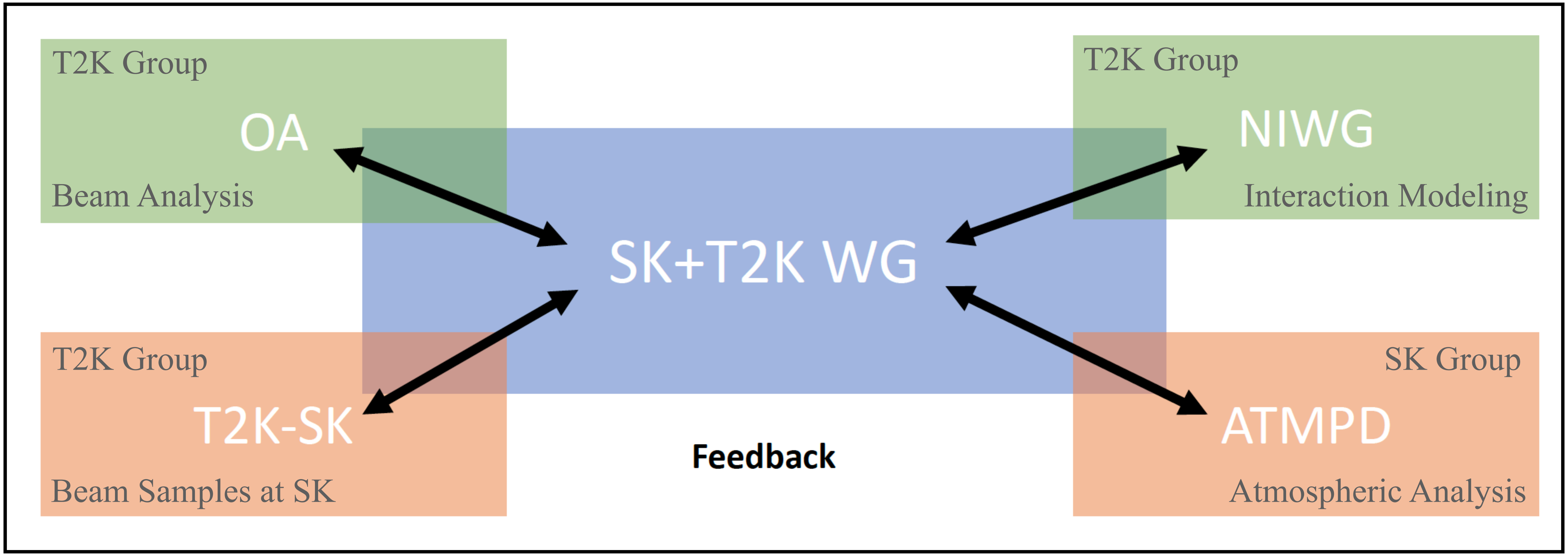
- Opinion: Takes a *long* time to establish political viability and complete paperwork to realize an analysis.
- *Hard* to write a document that balances
  - Respect for individual experiments's methods
  - The need to adapt those methods for a robust analysis

SK+T2K Analysis: Diagram of Staged Approach



- Formal joint fit activities started in 2019

SK+T2K Analysis: Working Group

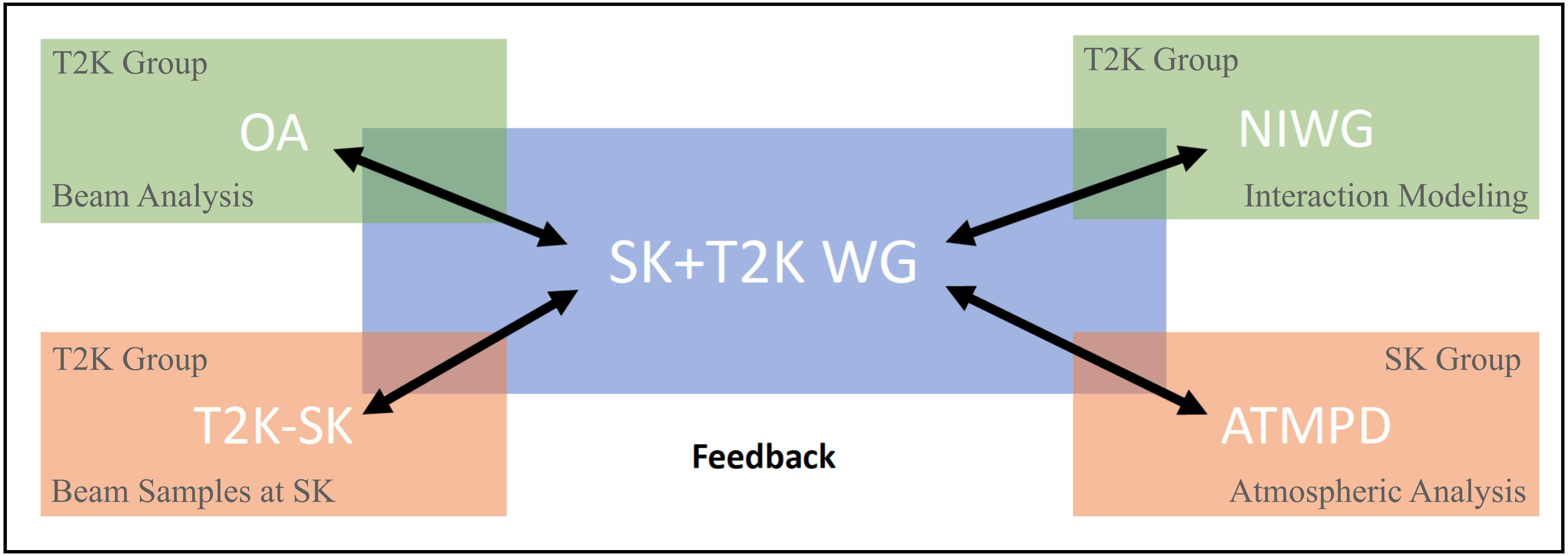


- Formed when MOU was signed
  - Conveners populated with experts from main analysis groups on both T2K and SK,
  - Feedback at all stages of the analysis, regular updates to both collaborations

- Freedom to select from *existing* data, MC, systematics, analysis tools from both experiments
- Freedom to *propose* changes to those, subject to approval by both collaborations

- Opinion: This structure was extremely useful when developing the analysis and made it straight-forward to implement (and adapt) models as well as to avoid making the "wrong move" at key junctions

SK+T2K Analysis: Working Group



- Lead convener and SK representative both well known and trusted by *both* collaborations
  - Both hold leadership positions in one or both of the experiments
  - Both have experience with analysis on both experiments

- Opinion: Somewhere in the WG structure it is important to have people that can interact seamlessly with the leadership on both experiments
- In the SK+T2K case important for :
  - Getting rapid inter-collaboration agreement on issues not covered by the MOU
    - Eg. How to choose speakers for talks, organize officialization procedures
  - Delivering analysis "inputs" to WG members (MC, bin definitions, etc.)
  - Setting up special meetings, etc.

Analysis Development

Analysis Development: Starting Point

- Chosen baseline analyses from each experiment

	T2K 2020	SK 2018
Samples	5 (All below ~1 GeV)	18 (7 below 1 GeV)
Generator	NEUT 5.4	NEUT 5.4
CCQE Model	Spectral Function (Benhar)	Local Fermi Gas (Nieves)
Reconstruction	fiTQun v4	fiTQun v6
Systematic Errors	Non-linear response, correlated	Linear response, uncorrelated
Prior Constraints	T2K ND280, external data	Some external data
Flux Tuning	NA61, other hadron production experiments	Muon flux used to tune hadron production

- Many differences between analyses (not all listed) despite using the same detector

- Goal for SK+T2K Analysis:
  - Unify interaction model, reconstruction, detector systematics

Analysis Development: Interaction Model

- SK historically uses LFG : better agreement with atmospheric data out-of-the-box, while T2K uses Spectral function (SF) for its agreement with ND data

- Kinematic overlap between T2K and low energy (< 1 GeV) necessitates a common interaction model for those samples
  - Several interaction uncertainties for SF are well-developed and well-constrained by T2K ND
    - No such interface for LFG
    - Error model on SK side is not as precise

- → Choose T2K Model for Low Energies
  - A new atmospheric MC based on NEUT with T2K model settings was generated and validatied for use in Joint Fit
  - Add new systematics important for atmospheric analysis
    - Normalization errors for NC π^0
    - CCQE differences between ν_e and ν_μ (δ_{CP} impact in ATM)
    - π^+ momentum spectrum for $CC1\pi$ interactions

- Question: Can T2K model and ND measurement be used above 1 GeV?
  - Not obvious as model developed for lower energies

Analysis Development: Interaction Model

	Low-energy sub-GeV atm + beam	High-energy multi-GeV atm
CCQE	T2K model with ND280 constraint, correlated in low-E/highE (except for high-Q^2)	
	high-Q^2 params w/ND280	high-Q^2 params w/o ND
	add ν_e/ν_μ ratio unc. (CRPA)	
2p2h	T2K model w/ND280	SK model (100% error) + T2K-style shape
Resonant	T2K model w/ND280 + new pion momentum dial + NC1π0 uncertainties	SK model for 3 dials common with T2K, use more recent larger T2K priors
DIS	T2K model w/ND280	SK model
ν_τ	SK model (25% norm on top of other syst) for other systematics checked that we have no numerically unstable values	
FSI	T2K model w/ND280	T2K model w/o ND280 should be mostly same as SK model
SI	T2K model, correlated in low-E/high-E only applied to FC and PC for atm, PN not applied to atm	

Analysis Development: Interaction Model

	Low-energy sub-GeV atm + beam	High-energy multi-GeV atm
CCQE	T2K model with ND280 constraint, correlated in low-E/highE (except for high-Q^2)	
	high-Q^2 params w/ND280	high-Q^2 params w/o ND
	add ν_e/ν_μ ratio unc. (CRPA)	

- Significant amount of ν_e in atmospheric flux, but not in T2K
  - T2K ND fit doesn't constrain uncertainty on ν_e/ν_μ cross section well, no energy dependence
  - Important interaction systematic for atmospheric constraint on δ_{cp}

- Can we use CCQE model at all energies?
  - T2K ND constraint only covers Q2 region where proton is below Cherenkov threshold
  - → Most multi-GeV events at SK do not have a visible proton ring so T2K model should be OK

- High-Q2 parameters are more important at multi-GeV energies, but no energy dependence in T2K model
  - Use three copies of T2K Q2 errors in multi-GeV samples, uncorrelated and unconstrained by T2K ND

Analysis Development: Interaction Model

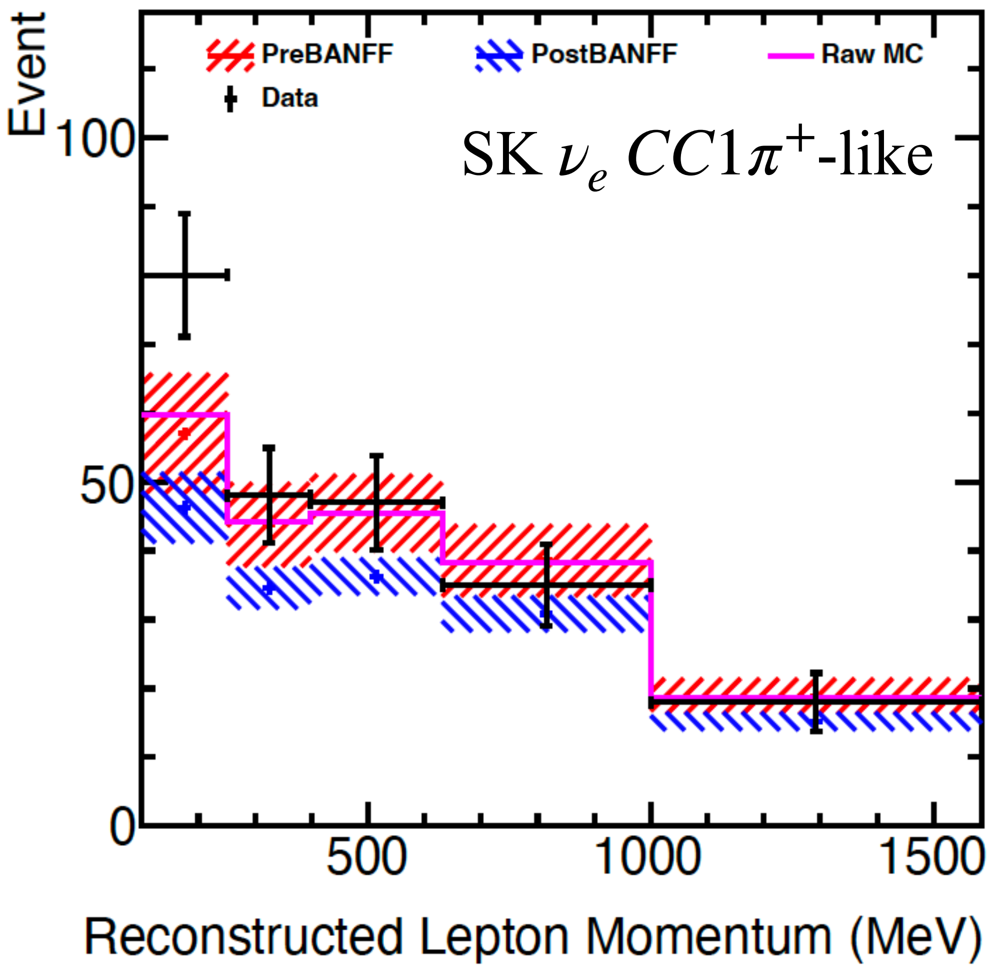
	Low-energy sub-GeV atm + beam	High-energy multi-GeV atm
Resonant	T2K model w/ND280 + new pion momentum dial + NC1π0 uncertainties	SK model for 3 dials common with T2K, use more recent larger T2K priors

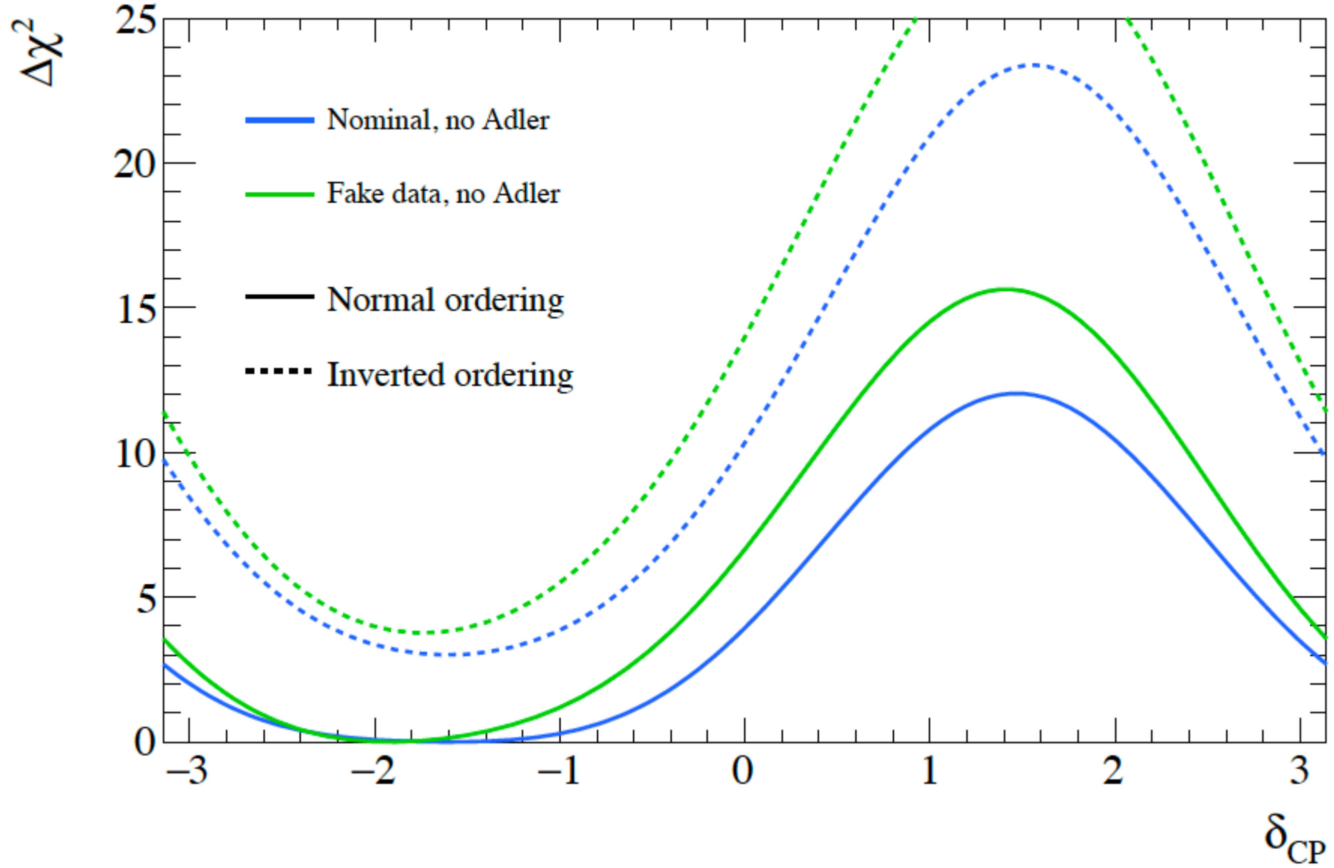
- Source of resonant events different at low (mostly through delta) and high energies, so separate uncertainties
  - Same model used at low-energy and high energy
  - Uncorrelated; consistent prior uncertainties, more conservative
- T2K NC $1\pi^0$ backgrounds are small so no explicit constraint on these, but SK uses an explicit π^0-like sample
  - Add normalization uncertainty for these events 30% , (100% for NC coherent)
- Additional pion uncertainties to address known data/MC discrepancy in samples enhanced in $CC1\pi$ interactions where the pion is below Cherenkov threshold (more next slides)

- Opinion: Despite very similar modeling, neither the SK nor T2K interaction model was sufficient for the joint fit. Careful consideration of many effects was necessary to make sure we did not over- or undercover systematics
  - Very important when dealing with powerful constraints from near detector!

Analysis Development: CC1pi+ Data Excess

- Downward-going (smaller oscillation effect) atmospheric neutrino data

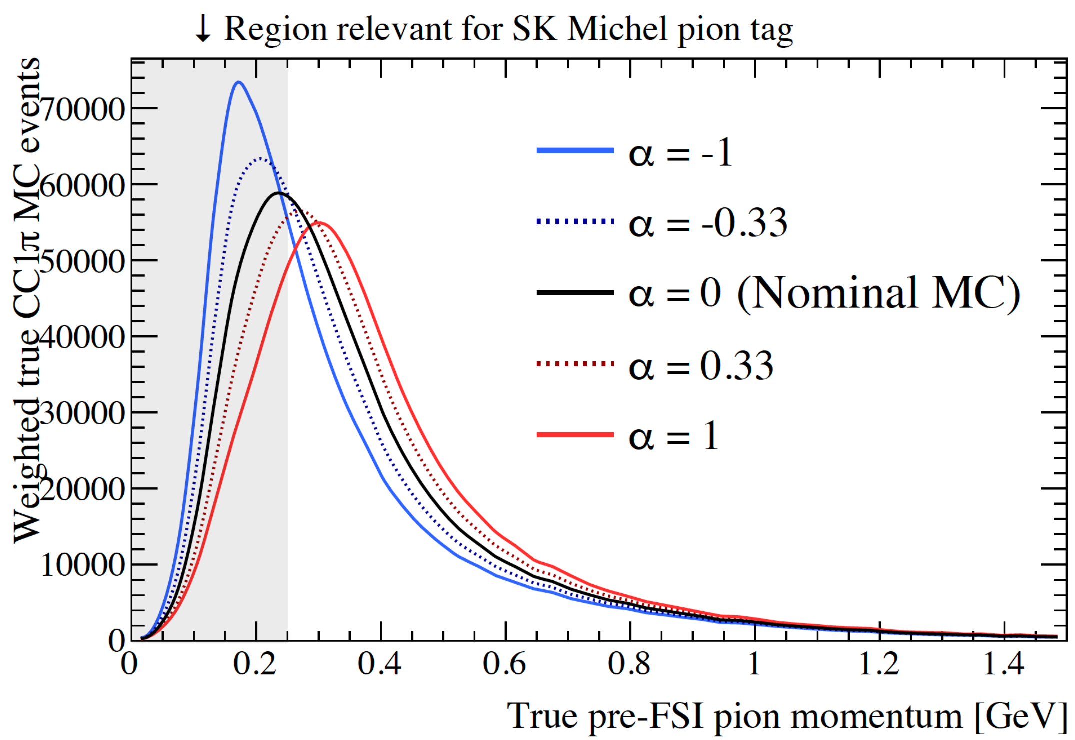




- Addition of ND constraint spoils data/MC agreement in pion-like sample, may bias δ_{cp} measurement in joint fit
  - For both T2K and SK, most of CP effect appears as a normalization change in these samples — large effect! green line in right plot

- Opinion: Not a problem in the SK only fit due to larger systematics and baseline cross section.
  - "Asimov" sensitivities are not enough to study such a problem
  - Cannot assume either experiment's analysis is sufficient "as is": combination is likely to uncover surprises!

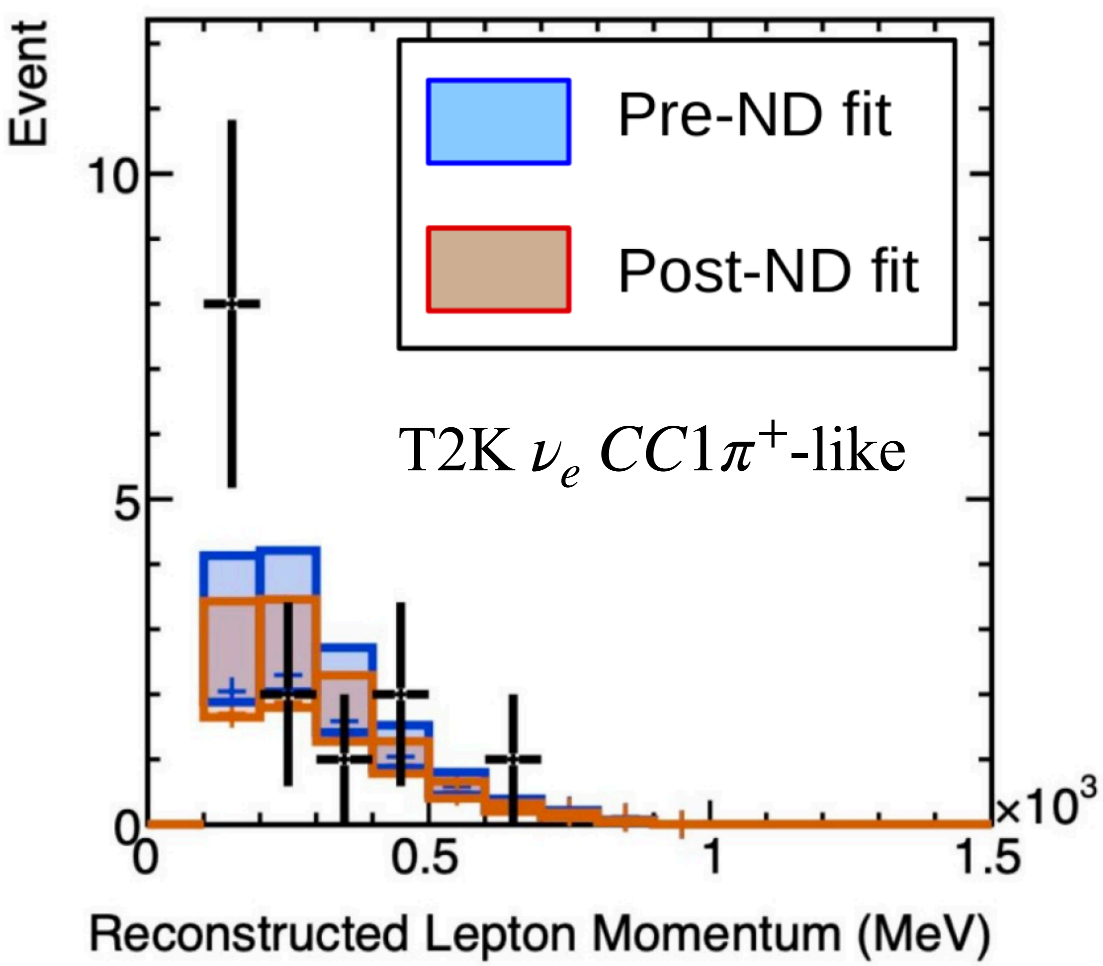
Analysis Development: CC1pi+ Data Excess Treatment



$$w(x \mid \alpha) = \frac{(1+\alpha x)^n}{\langle (1+\alpha x')^n \rangle_{x'}} = \frac{(1+\alpha x)^n}{\sum_{i=0}^{n} c_i \alpha^i}$$

- New parameter for ad-hoc reweighting in pion rest frame, without affecting lepton momentum

- New cut to reduce background neutron contamination in decay-electron samples (*not shown*) to both SK and T2K samples

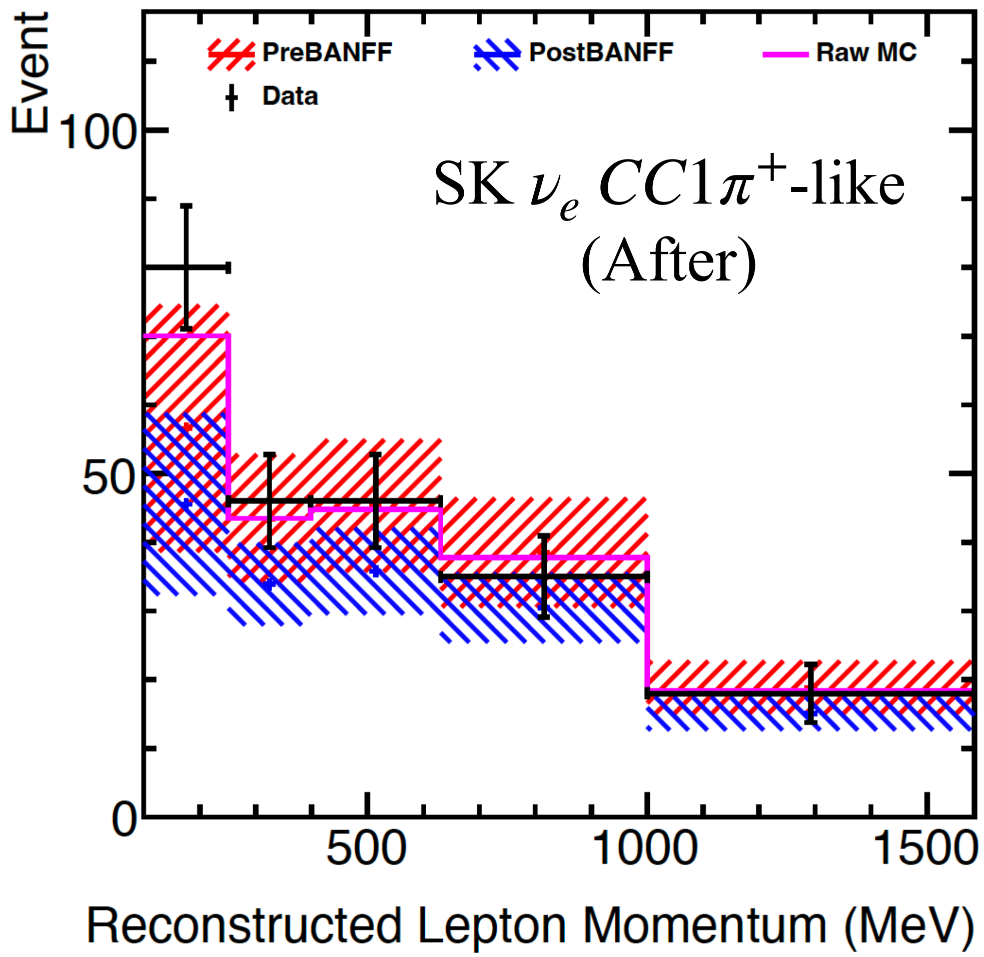


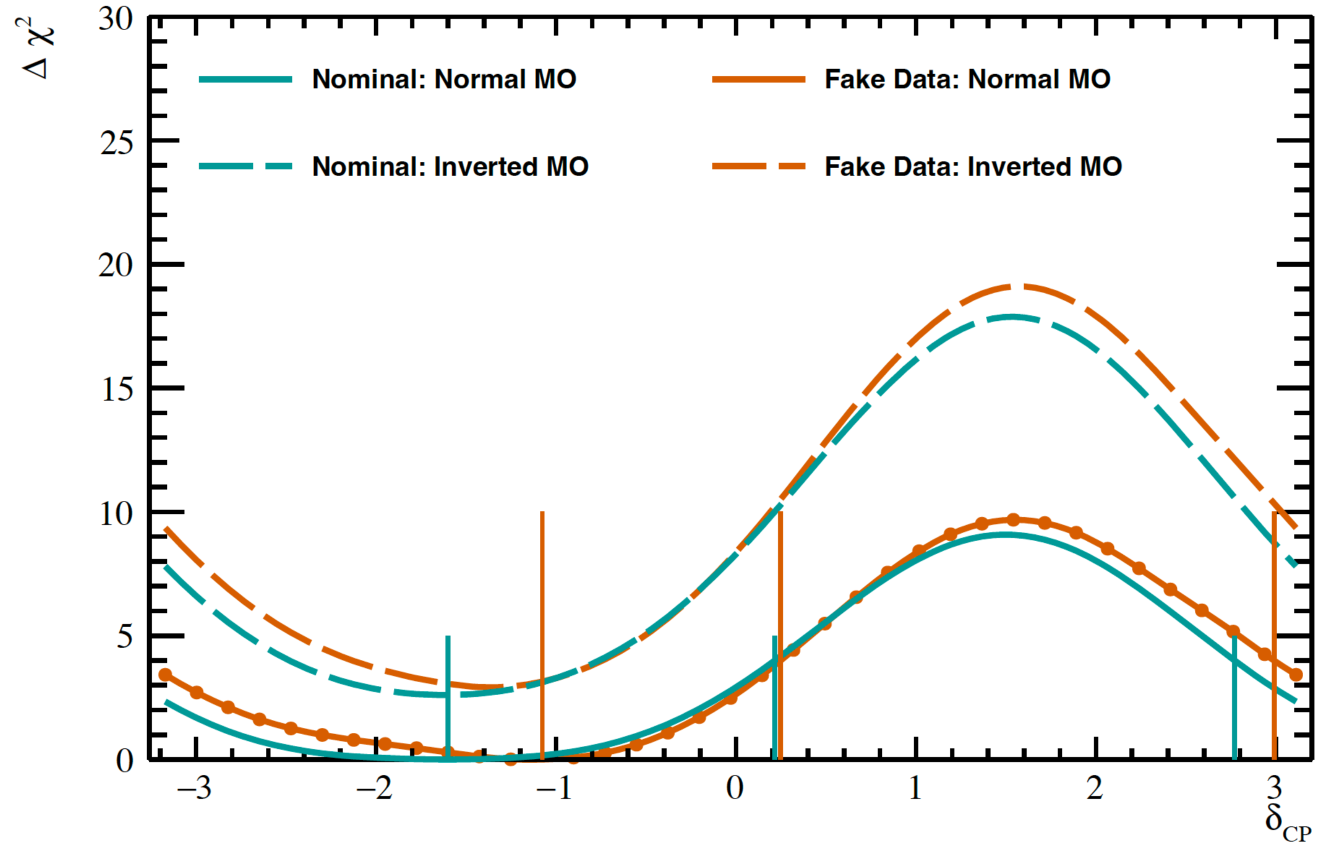
- Data excess is momentum dependent, strongest at lower energies; potentially mis-PID?

- Introduce migration systematic errors between samples at lowest momentum affecting *both* experiments

  - +20% on SK atmospheric sub-GeV ν_e 1 d.e. sample and -2.3% on sub-GeV ν_μ 1 d.e. sample both at $p_e < 251.2$ MeV/c
  - +20% on T2K FHC ν_e 1 d.e. sample and -2.3% on FHC ν_μ (only 1 d.e. events) sample both at $p_e < 251.2$ MeV/c

Analysis Development: CC1pi+ Data Excess





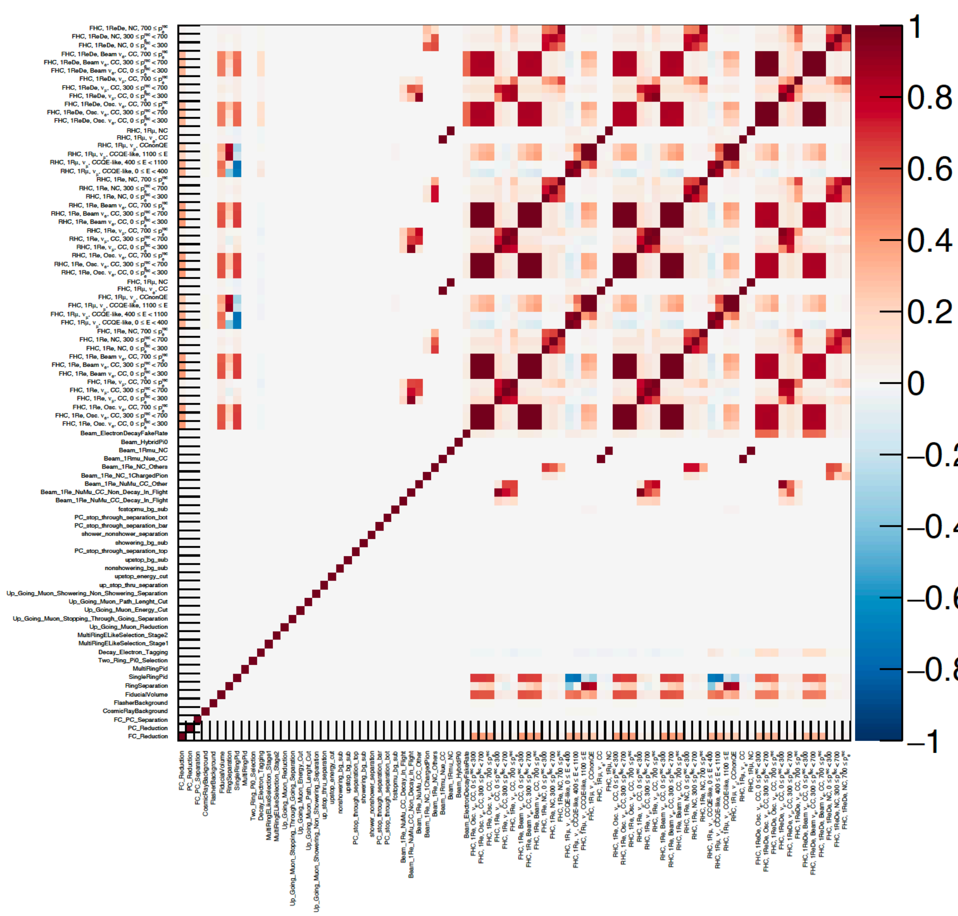
- Final treatment is sufficient to prevent conclusion-changing bias in the analysis
- The full nature of this data excess is still being investigated in the WG, T2K, and in SK

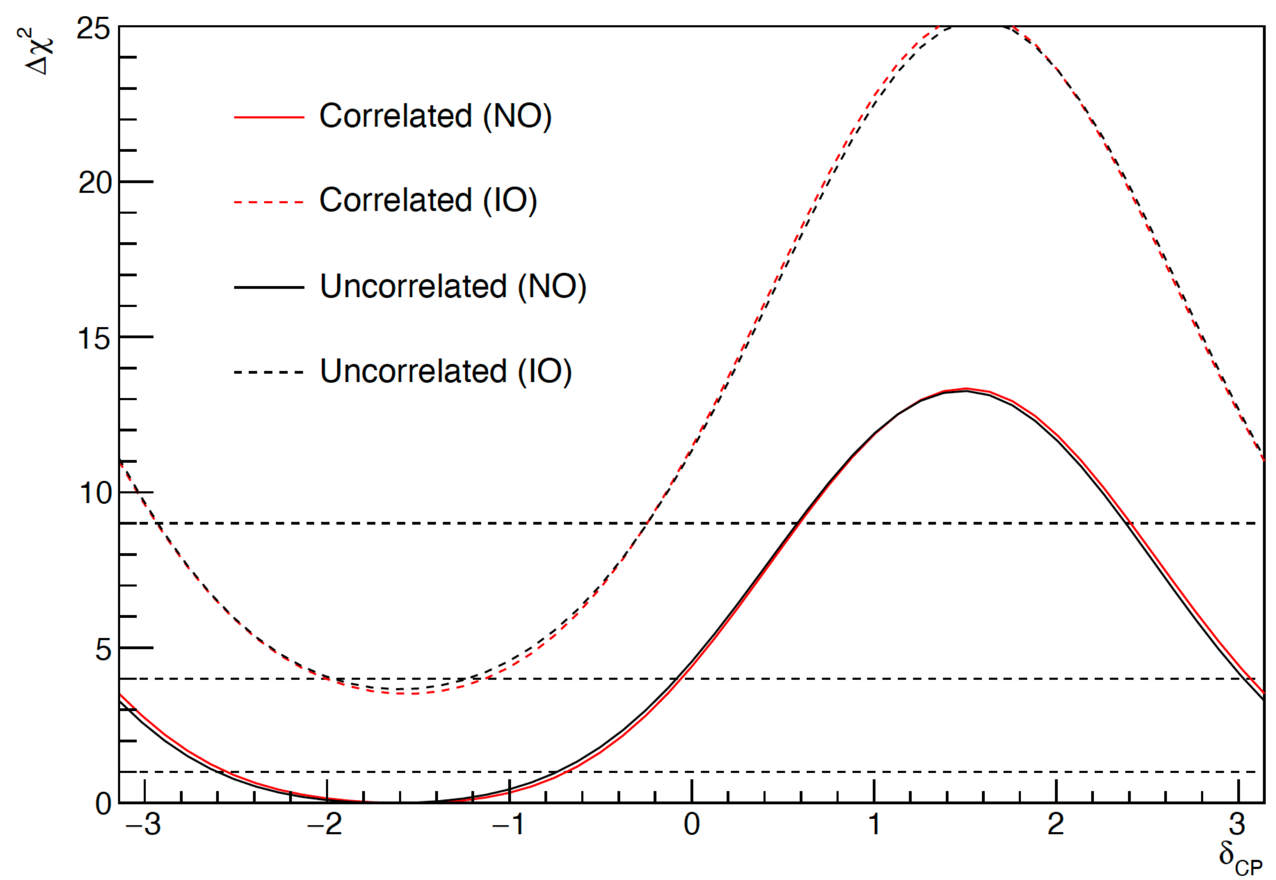
- Opinion: Very important to build flexibility into the Joint Fit framework (MOU) as well as to secure a cooperative spirit among participants
  - SK+T2K joint fit may have had a strongly biased result without the ability to make the above changes to the analysis

Analysis Development: Detector Systematics

- Correlating detector systematics seems "obvious" given that the events are recorded in the same detector
  - Technically challenging and time consuming…
  - After fully testing whether correlations would have significant effects on the analysis, found that most do not (bottom right)
    - Energy scale uncertainty is an exception

- Opinion: Difficult to know in advance exactly what parts of the analysis will be most important
  - Analysis may take more time than expected if one is really careful
  - Access to more human resources to check everything would have been helpful





Analysis Development: New Ways To Solve Old Problems

- 4 Analyses with different methods and choices for cross validation

	FA2	FA1	BA1	BA2
Style	Frequentist	Frequentist	Bayesian	Bayesian
Osc Prob.	Event-by-event	Binned	Binned	Event-by-event
Nuissance Par.	Profiled	Profiled	Marginalize	Marginalize
Fit	Fixed grid	Fixed grid with	MCMC	MCMC
Cores	CPUs	GPUs	CPUs	GPUs
Fast Osc.	Neighbor Ave.	Semi-Analytic	Semi-Analytic	Down-sampling
Prod Height	Neigbor Ave.	Binned	Binned	Binned

- Fast Osc: MC statistics insufficient to sample fast oscillation probability for atmospheric events with L/E > 2 km/GeV
  - → Oscillation probability needs to be averaged to create smooth Δm^2 contours without sacrificing sensitivity to other parameters

- Opinion: Working with many different fitters was important for debugging and for addressing technical challenges from new perspectives
  - New Fast Osc. Treatments developed; now being adopted at SK
  - Computing on fixed grids to save resources without compromising integrity
  - Development of fast GPU-based method for frequentist tests over a large phase space
  - …

Main Lessons

Major Lessons Learned

- It is important to prepare *long* in advance for a real joint fit between experiments
  - Political and bureaucratic problems likely
  - Some of these resolve with time (people get used to the idea) other require real work
  - If both experiments can pursue their own analyses with an eye for a future joint fit, a lot of time could be saved

- Being flexible is extremely important
  - SK+T2K benefitted greatly from willingness on both sides to adapt analyses for joint fit
  - SK moving to the "T2K interaction model" for low energies allowed for a more robust analysis in the end

- Sharing more information is generally better
  - SK+T2K analyzers had access to data, MC, systematic errors, technical documents, some of the relevant software, as well as many experts on both sides
  - Found problems that led to interesting and important developments that made the analysis more robust overall
  - Learned many details about each others analyses; definite cross pollination

- Actually doing the analysis is important
  - Good for advancing the field and addressing common problems
    - E.g. no interaction model is perfect but perhaps we can get there someday
  - A second analysis is now underway, with more interest than from the first round

END