INTERVAL ESTIMATION, AND THE PRACTICE OF FLAVOR PHYSICS

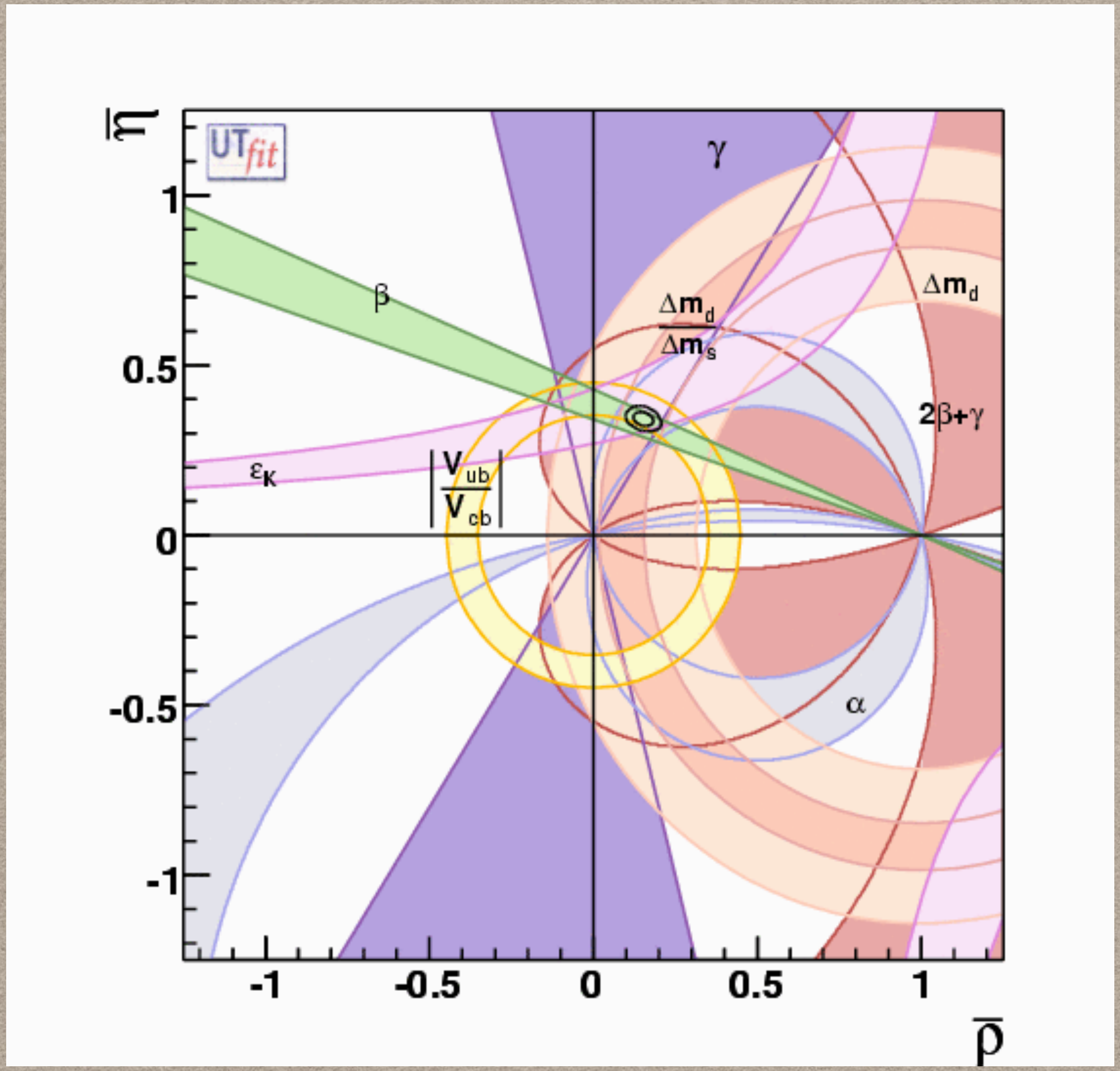
GIOVANNI PUNZI
UNIVERSITY & INFN-PISA

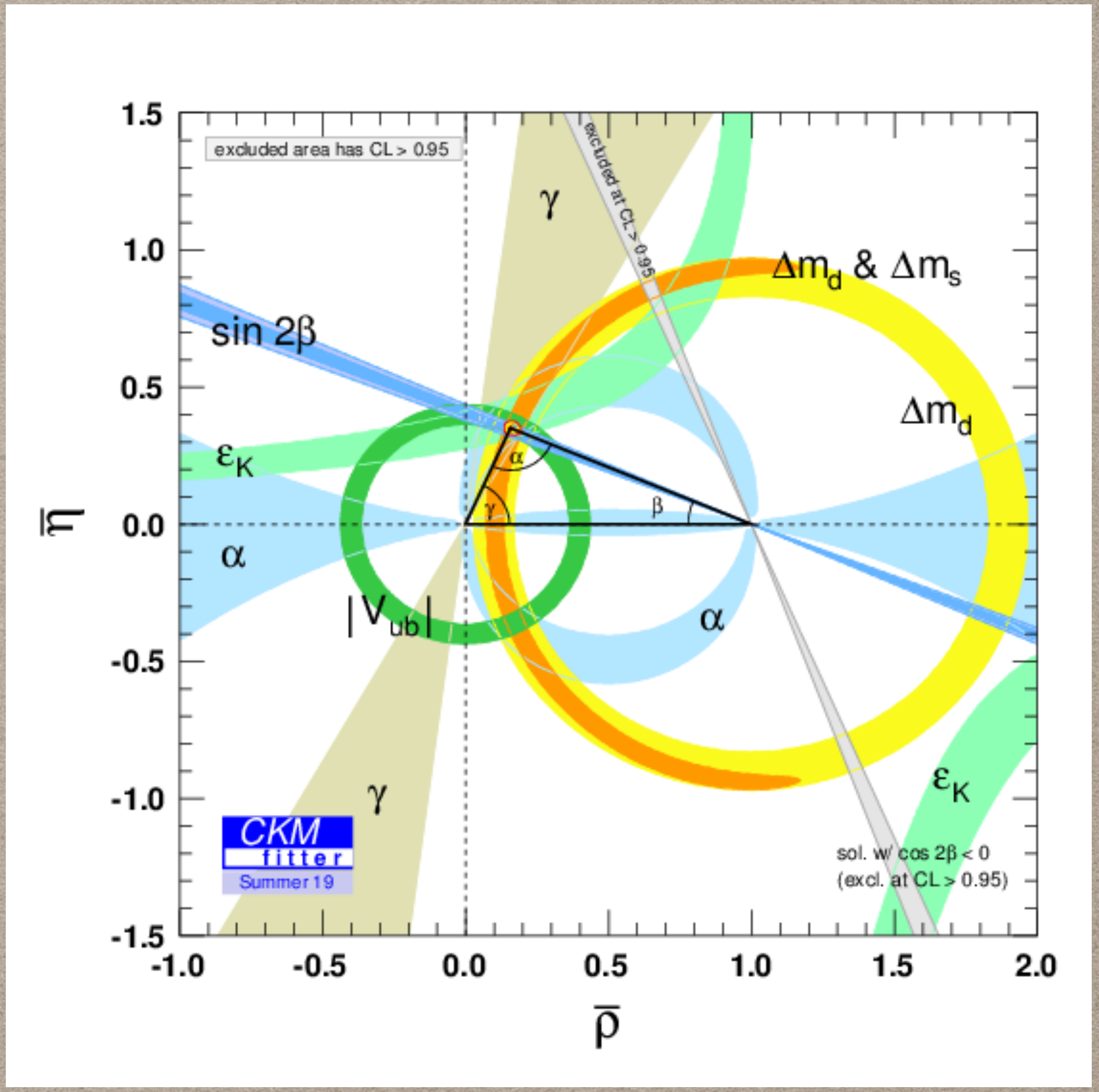
GIOVANNI.PUNZI@CERN.CH

OUTLINE

- With no pretense of completeness, I will discuss some practical issues in the field.
- We heard in the previous talk about alternative approaches to Interval Estimation
- In practice, in many cases more than one are used for the same measurement. Partly because of their different conceptual merits - but one reason is the practical need for some approximations - particularly on the frequentist side, that is what I focus on.
- One point is systematic treatment.
  - Bayesian need to "vary priors", in a non well-specified way; while frequentists need to make approximations. I will go in some details about this.
- The other point I will discuss is optimization of sensitivity.
  - Interval estimation is just the final step of the measurement process - an important ingredient for success is starting with careful experiment design.

POSSIBLY THE BEST KNOWN EXAMPLE

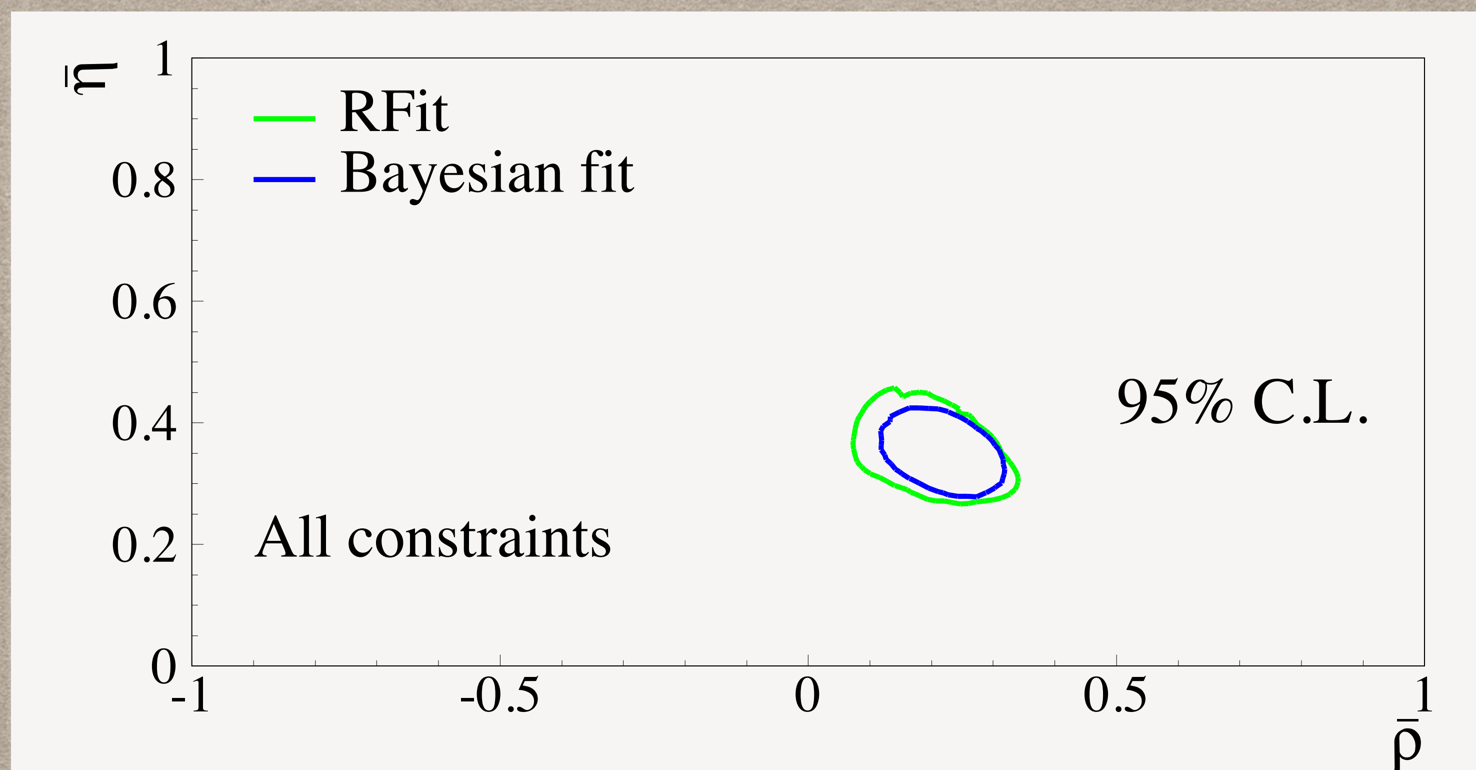




- CKM parameters (rho, eta) are a center point of Flavor physics
- Two groups worked for years publishing interval estimates from the whole of available data: CKMfit (frequentist), UTfit(Bayesian)

CKMFIT/UTFIT COMPARISON

- Comparison and discussion between these groups have being ongoing for long. Also analyzed in detail in a CERN workshop (2003)
- Summary conclusion: mostly similar when given the same likelihoods, difference is mainly in the systematic treatment. So it should be interesting.



different in the two approaches. As a consequence, the region defining the 95% (99%) confidence level for the UT parameters is wider by 30% (20%) in the frequentist as compared to the Bayesian approach. Further tests have shown that, if the same likelihoods are used for input quantities, the output results become almost identical. The main origin of the difference between the results in the Bayesian and the frequentist method is therefore the likelihood associated to the input quantities. But these differences will decrease progressively as the theoretical uncertainties will be reduced or related to experimental ones.

HOW FLAVOR-PHYSICS PRACTICE EVOLVED

(Disclaimer: what follows comes from a combination of INSPIRE citations, priv. comm. from statistics committees, and personal experience)

- Fully Bayesian methodology, just as in UTFit is still being used; sometimes with the help of new tools as Markov-Chain MC. However, there is more attention to the frequentist coverage side - often a frequentist method is also presented, or it is used as a technical tool to produce frequentist coverage in a more practical way.
- On the frequentist side, things are more varied:
  - Feldman-Cousins ordering is in wide use. The delta-chi^2 used by CKMfit is asymptotically equivalent.
  - CLs is a different frequentist approach in use, but less in Flavor than in High-PT (possibly due to its lower focus on rejection of H_0 ?)
  - Handling of nuisance parameters <u>still an important issue today.</u> Largely based on the same approach of CKMFit, with some attempts at improving over those approximation **- next slides**

Citations of FC paper

Belle	150
IceCube	113
CDF	70
CMS	68
H.E.S.S.	64
ANTARES	57
LHCb	52
BaBar	51
D0	51
T2K	39
MACRO	29
Pierre Auger	29
ATLAS	27
CLEO	27
MINOS	26
MEG	22
NOMAD	15
NOvA	15
OPERA	15
AMANDA	14

THE ISSUE WITH SYSTEMATICS

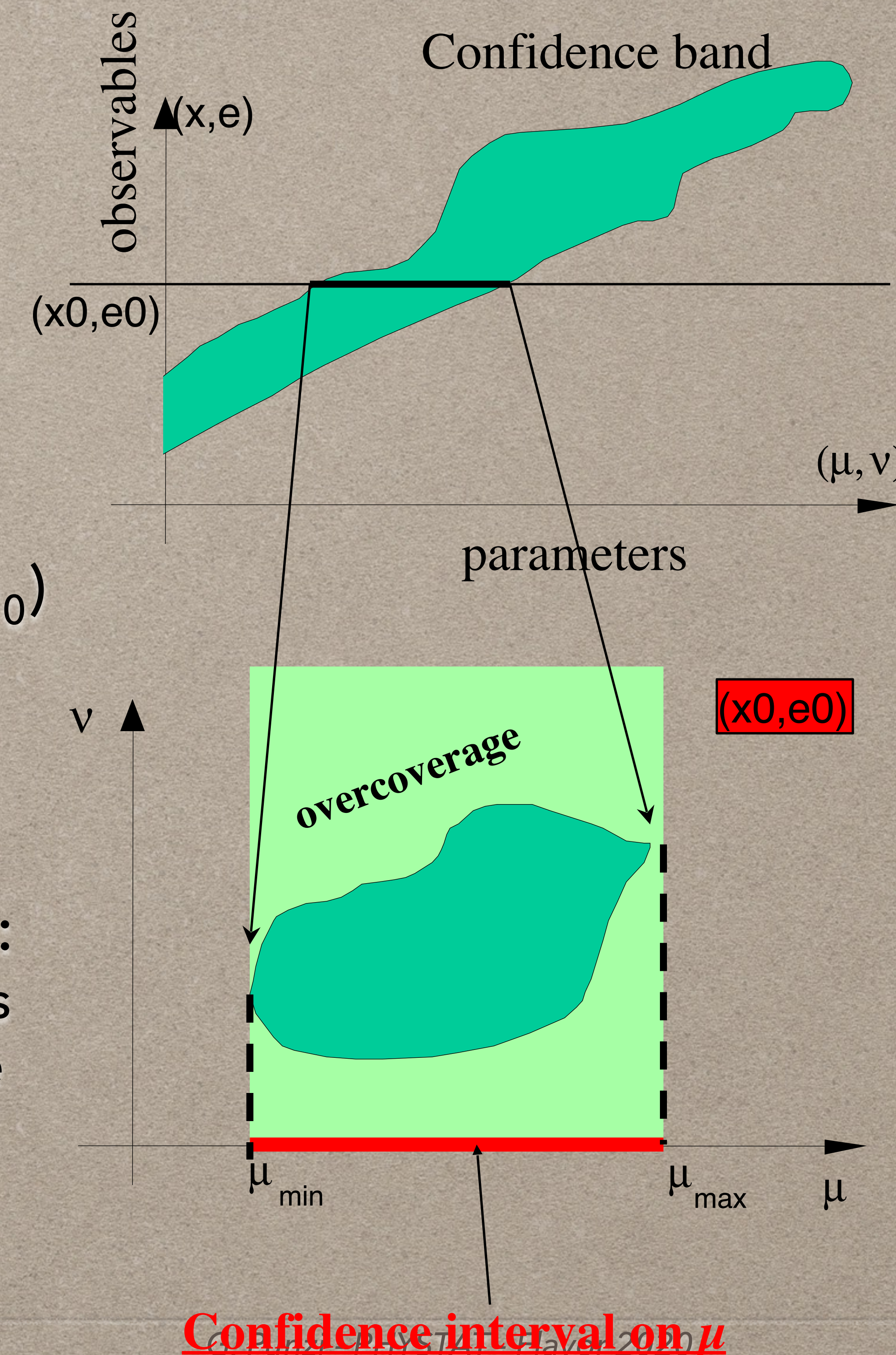
- Often the pdf p(x;μ) is actually a p(x;μ,v), where v is an unknown parameter I don't care about, but it influences my measurement (nuisance)
- I might also have some info of v from another measurement y: q(e;v).
  My problem is then: p(x,e; μ,v) = p(x;μ,v)*q(e;v), but I am only interested in μ
- In Bayesian approach, it is easy to get rid of v: evaluate the posterior, marginalized on v :

$$p(\mu \,|\, x, e) = \int p(\mu, \nu \,|\, x, e) d\nu \propto \int p(x, e \,|\, \mu, \nu) p(\mu) p(\nu) d\nu$$

- The only issue is the usual Bayesian question of choice of priors. This can also be non trivial, but will not discuss it further here [an example of surprising effects of choice of priors was shown at PhyStat05 by LeDiberder]
- I will look more closely at the frequentist case, where issues are of a more practical nature

NEYMAN CONSTRUCTION WITH NUISANCE

- The rigorous frequentist way to deal with systematic uncertainties is simple in principle:
  1. Build a confidence band, treating the nuisance parameter as any other parameter: p((x, e) ; (μ, ν))
  2. Get CR in (μ, ν) from measurement (x_0, e_0)
  3. <u>Project</u> onto μ space to get rid of information on ν
- There are however significant issues that have essentially prevented its practical use:
  - CPU - expensive, especially in large dimensions
  - Typically blows up interval/large over coverage
  - Sensitive to ordering algorithm
  - Limit for 0 uncertainty

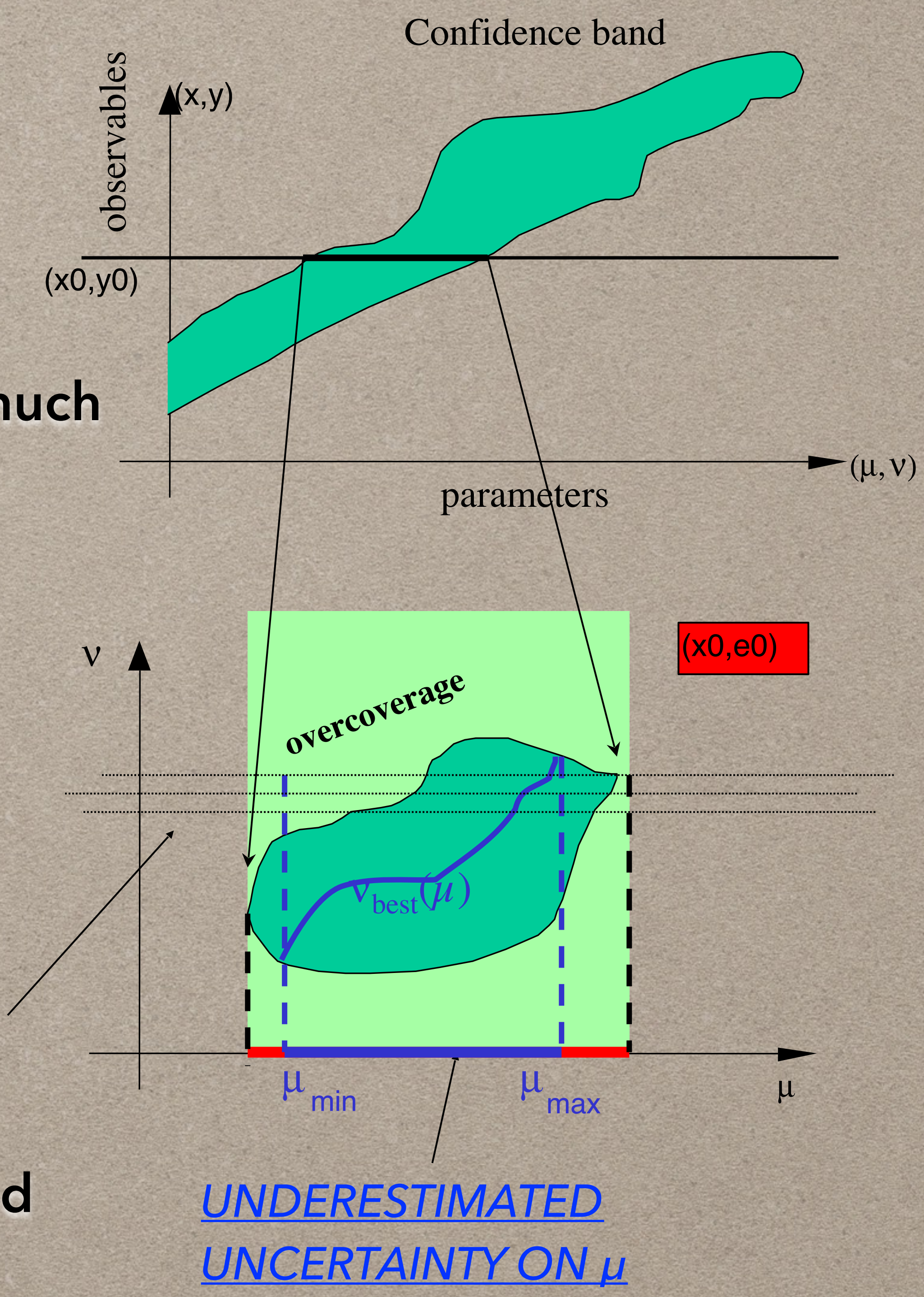


THE 'PLUG-IN'/'PROFILE' APPROACH

1. Define a new (profile) pdf:
   $p_{prof}(x;\mu) = p(x\ ;\ \mu, \nu_{best}(\mu))$
   where $\nu_{best}(\mu)$ maximizes $p(x_0\ ;\ \mu,\nu)$

2. Use $p_{prof}(x;\mu)$ to obtain Conf. Limits

Scanning limited to the μ space -> computationally much easier ! This is what CKMfit and most others do.



- Only checks coverage in a small subspace. Also, it depends on the observed value x_0
  -> "flip-flopping" fallacy, as defined by FC
  -> **undercoverage**, albeit usually modest

- Natural choice of ordering profile-likelihood ratio:
  $LR_{prof}(\mu) = p(x\ ;\ \mu, \nu'_{best}(\mu))/p(x\ ;\ \mu_{best}, \nu_{best}(\mu))$

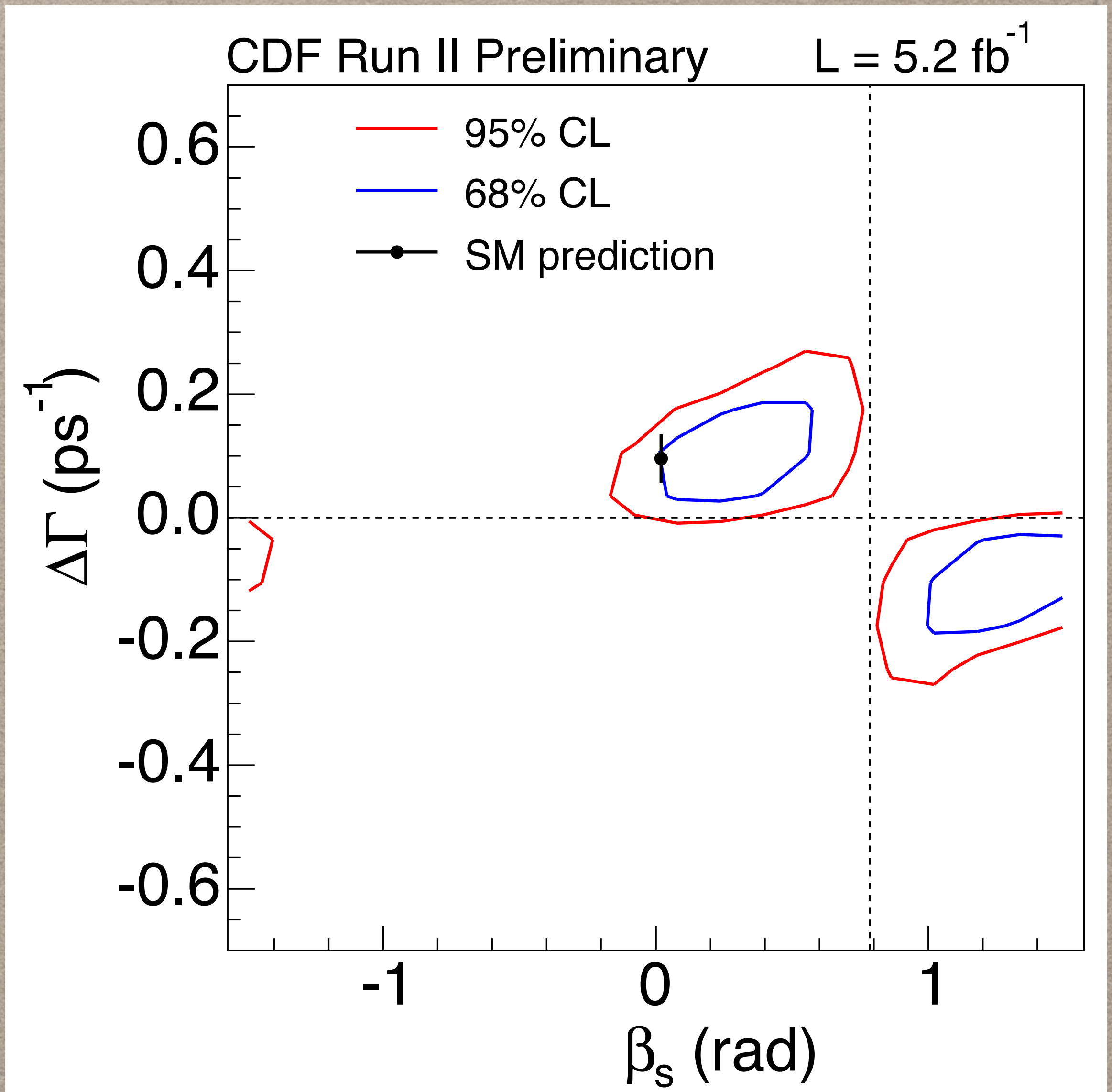
- **Profile method:** exploit the asymptotic chi2 distribution of LR allows cut $L_{prof}(\mu) > c$ with no need for MC. Sometimes the chi^2 is used directly.
  -> Very convenient, but further approximated

These two methods make the bulk of today's papers

GOING BEYOND: A REAL-LIFE EXAMPLE

- An attempt at moving past the usual approximations, in a full-fledged flavor physics measurement: CDF measurement of CP-violating phase β_s in $B_s \rightarrow J/\psi\phi$ [Phys. Rev. D 85, 072002 (2012)]

- 2-D relevant parameter space.
- Multiple solutions
- Highly non-gaussian/non linear contours
- 25 nuisance parameters



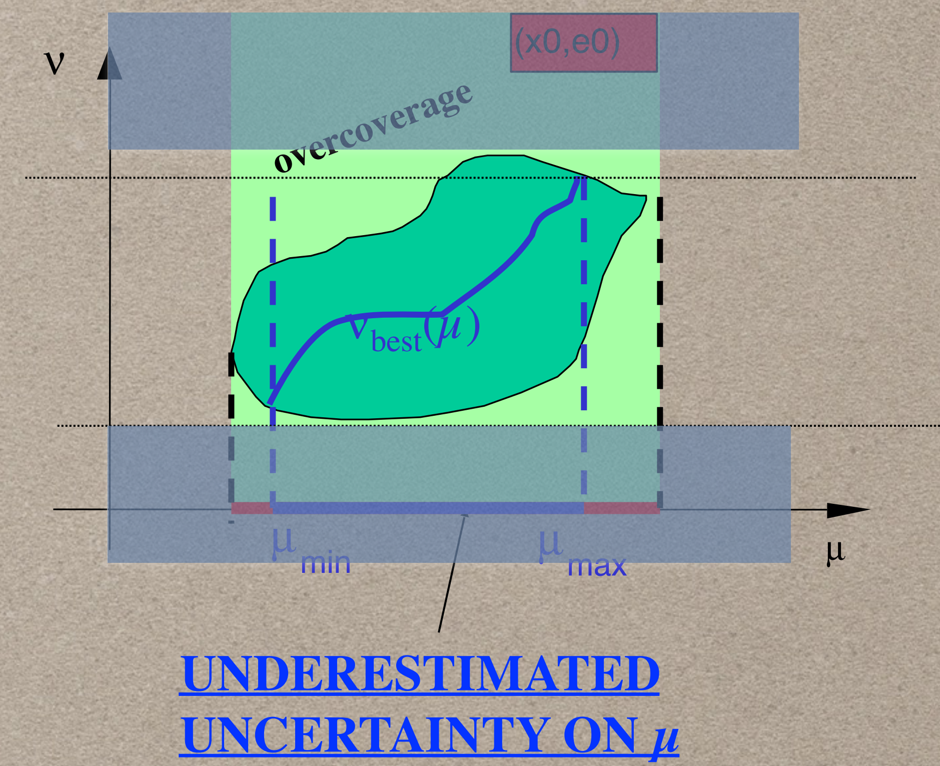
A HELPFUL LITTLE THEOREM

[R.Berger and D.Boos, JASA 89, 427 (1994) 1012]

concerns is defined as follows. Let C_β be a $1 - \beta$ confidence set for the nuisance parameter when the null hypothesis is true. Intuition suggests that we might be able to restrict the maximization to the set C_β. Indeed we show in section 2 that

$$p_\beta = \sup_{\theta \in C_\beta} p(\theta) + \beta \tag{2}$$

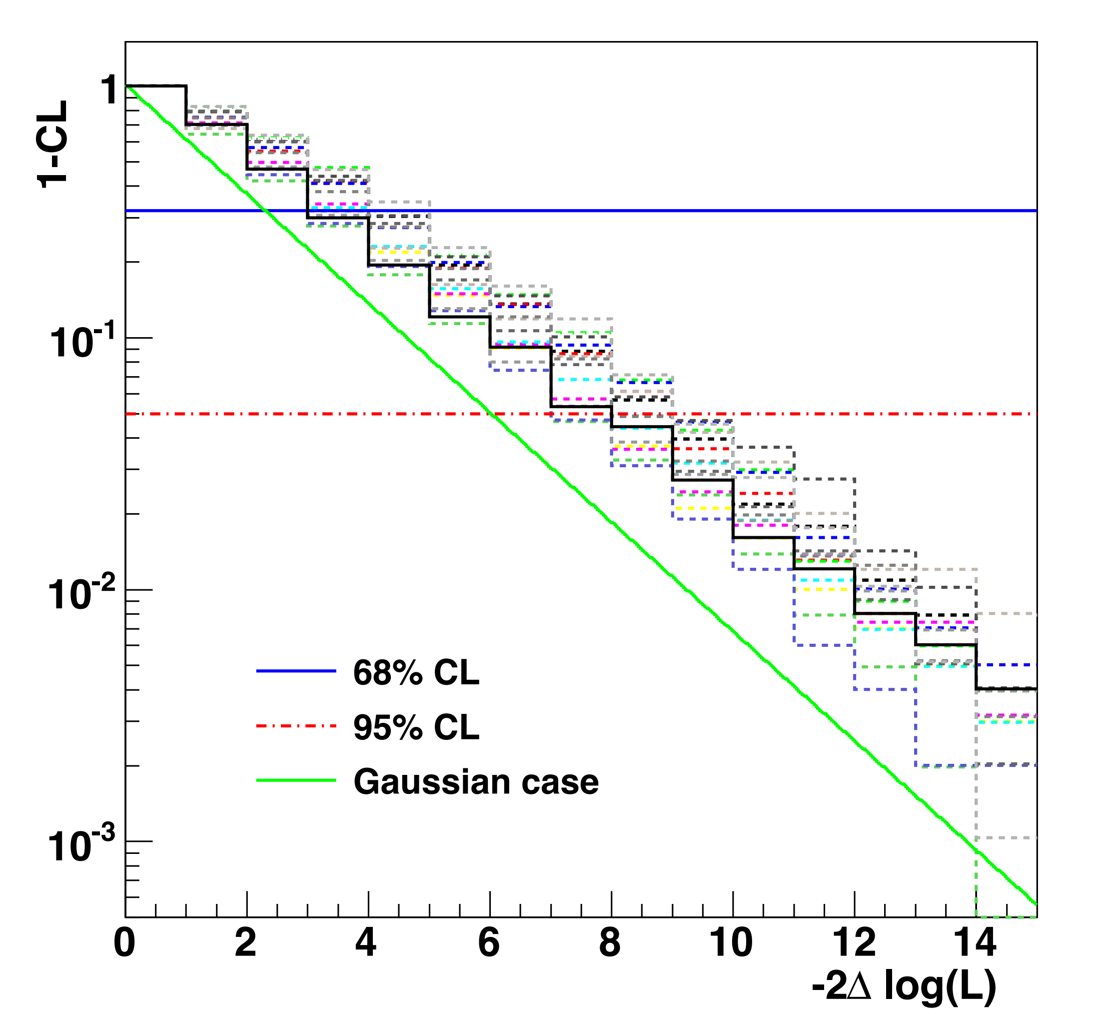
is an alternative valid p value. This p value may be preferred to p_{sup} on computational grounds (due to maximizing over bounded sets) and on statistical principles (restricting interest to likely regions of θ). The value of β and the confidence set C_β should of course be specified before looking at the data.

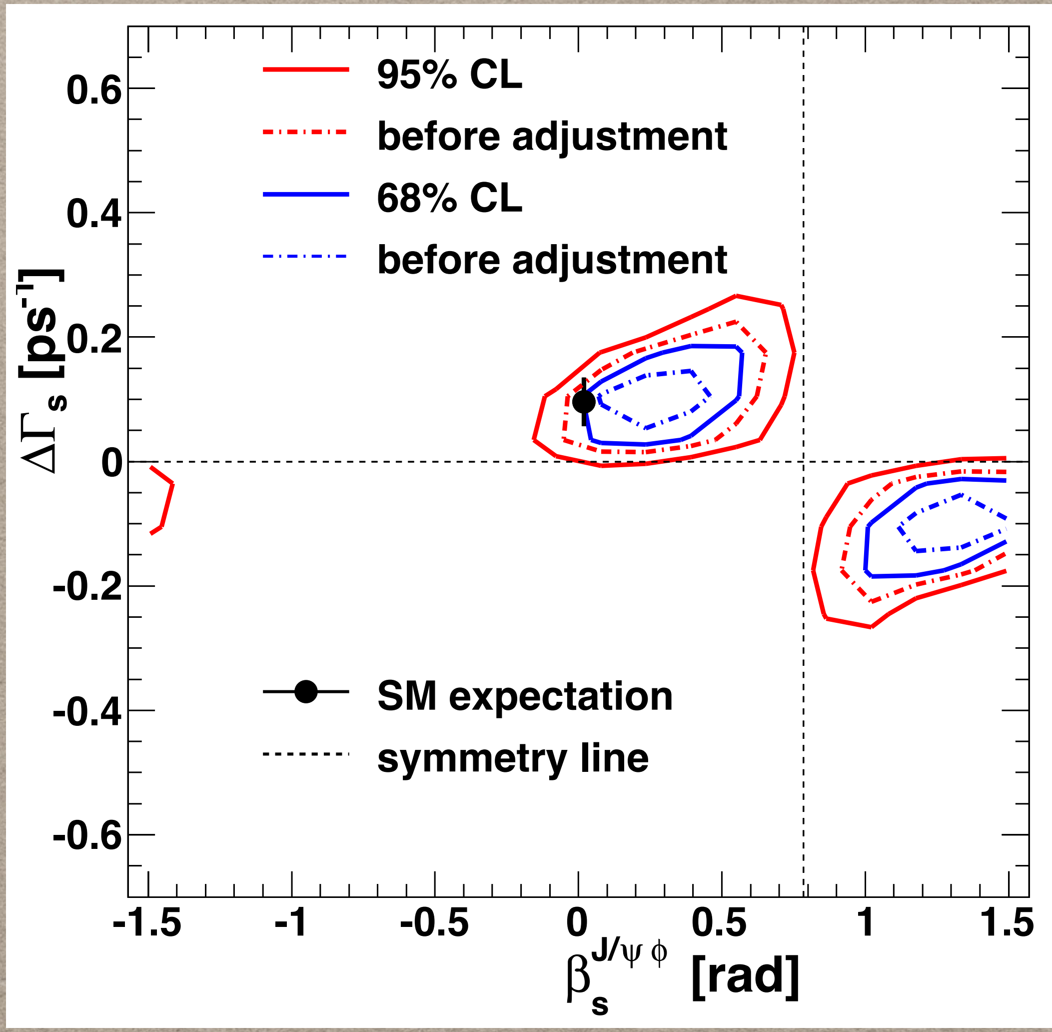


- One can limit the scan of nuisance parameters to a confidence region (1-β) for their values, provide one then corrects (1-CL) -> (1-CL)+β.
  - Example: set β=0.01 and derive limits at CL=96% to obtain valid limits at CL=95% accounting for nuisance paramenters
- Reduced scanning computational load, reduce overcoverage, limit variations

IMPROVED EVALUATION OF SYSTEMATICS

- Use profile-LR ordering, with MC simulation to get actual distribution
- Instead of plugging in a single v value, **sample a few points on boundary of 'box' defined by Berger-Boos**
  - Picked 5-σ for nuisance, to be prepared to exclude SM with high significance.
- Lesson learned: significant effects from both non-asymptotic of LR, and systematic uncertainties. NB: event yields in the thousands.

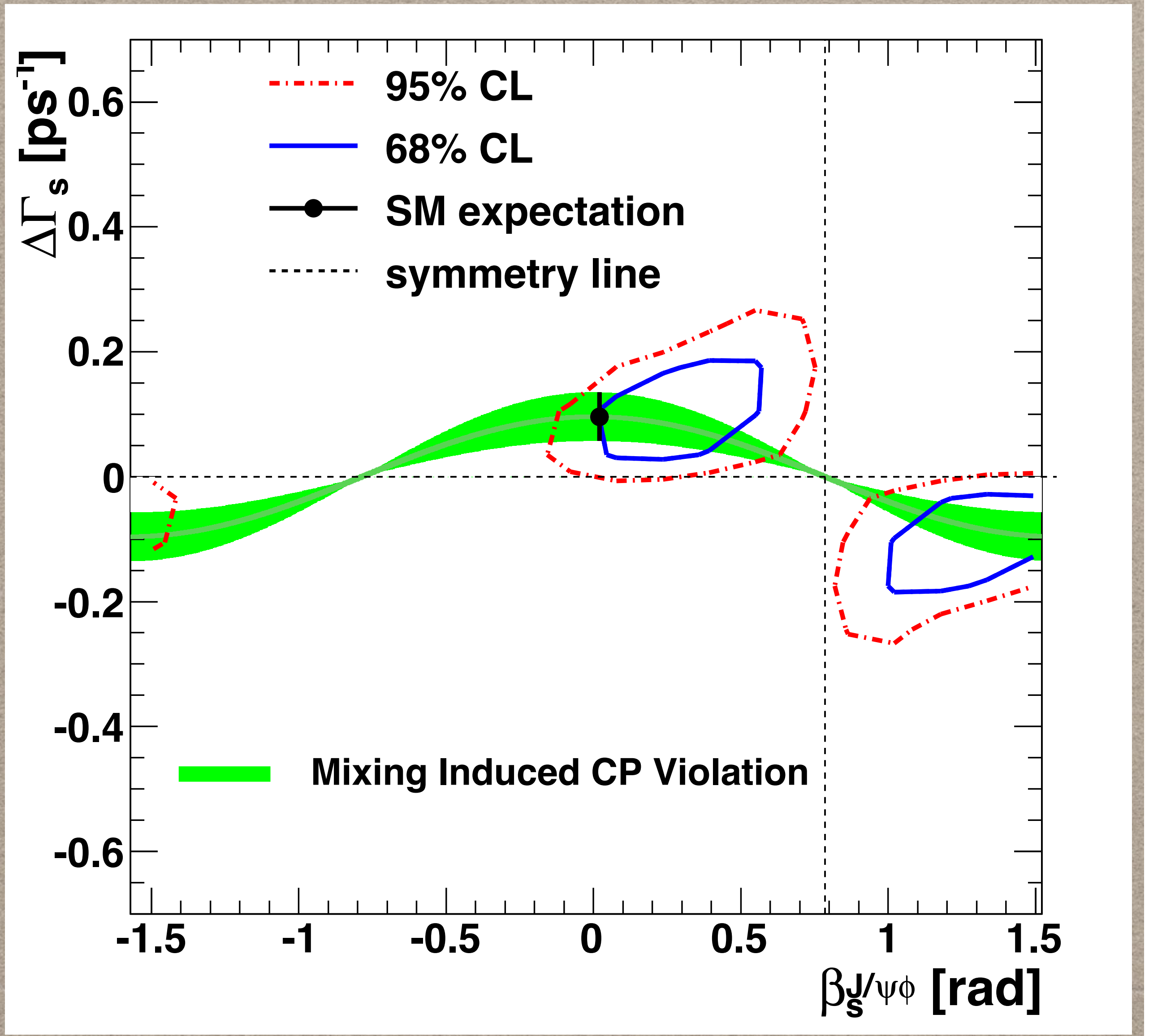




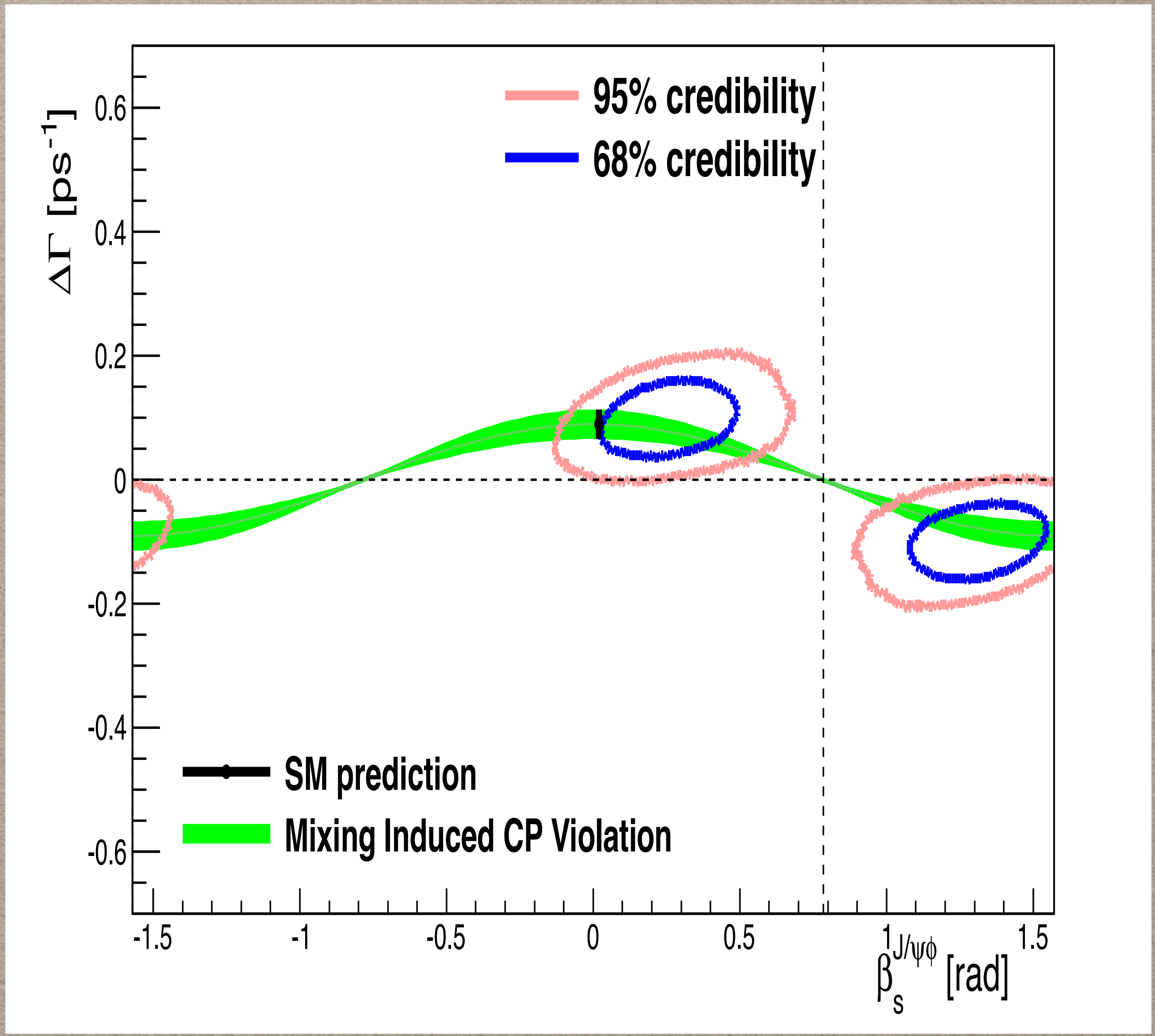
F-C VS BAYESIAN

- The same paper also reports a (fully) Bayesian analysis as 'crosscheck'. Includes prior variations (non pictured).
- Bayesian yields similar regions (a bit smaller, if you do not include prior variations)

Fully frequentist, FC w/ systematics (B-B clipped)



Bayesian w/systematics MCMC (no prior variations)



CAN ORDERING ALSO BE IMPROVED?

[ARXIV:0511202 (PHYSTAT05)]

- LR ordering is a good thing - but can't tell nuisance from physics parameters.
- Things can be improved by choosing an ad-hoc nuisance-aware ordering function:

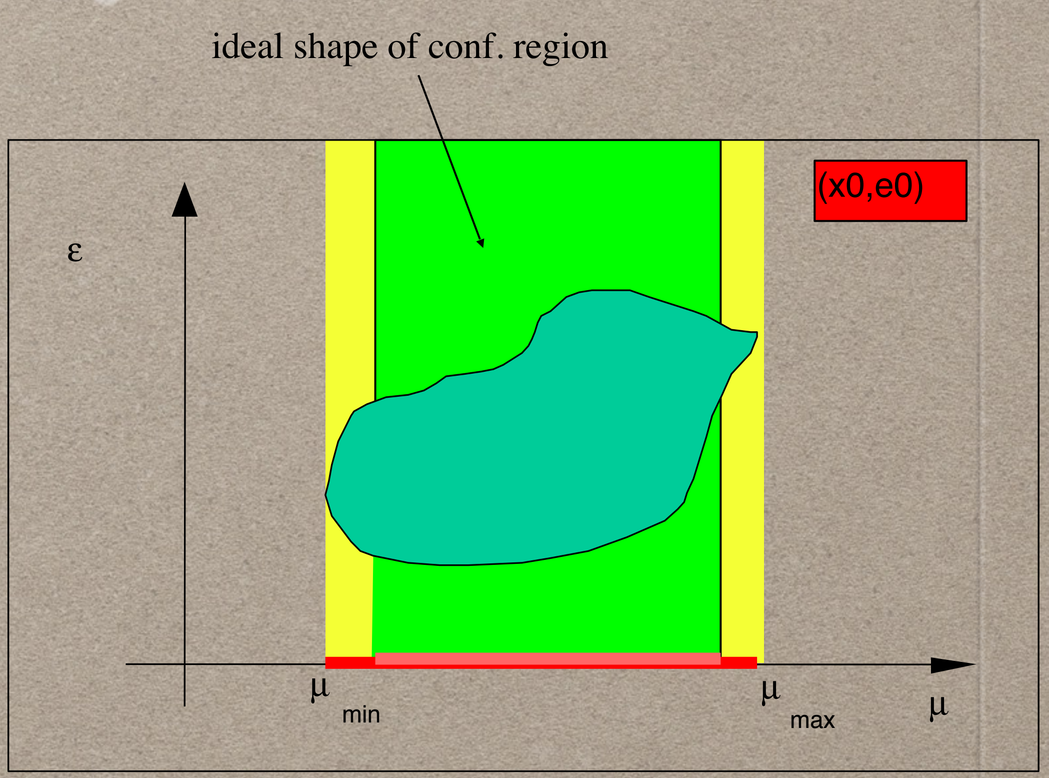
$$f(x,e;\mu) = \int_{f_0(x')<f_0(x)} p(x'|e;\mu,\hat{\nu}(e))dx'$$

where $f_0(x)$ *is the ordering function* in absence of systematics

- This particular ordering is ***independent of nuisance*** (facilitates computation) and ensures efficient use of the confidence band, minimizing "wasted coverage"
- Integration must still be done for several values of v (but the previous tricks still apply)
- If LR is used as $f_0(x)$, it is approximated by the profile-LR:

$$LR_{\text{prof}} = \frac{\sup_\nu p(x;\mu,\nu)}{\sup_\mu \sup_\nu p(x;\mu,\nu)}$$

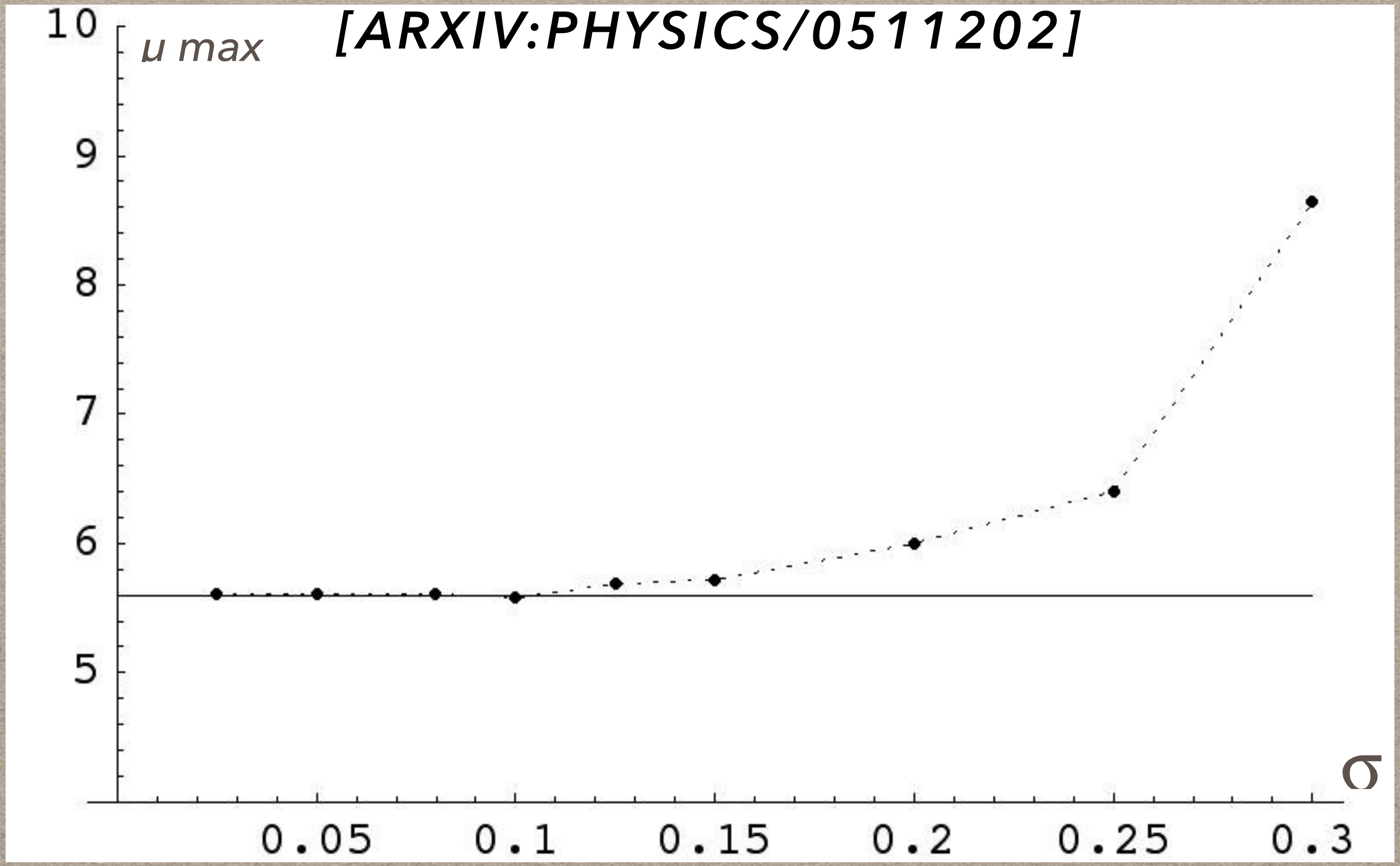
(note this is different from $LR = \dfrac{p(x;\mu,\nu)}{\sup_\mu \sup_\nu p(x;\mu,\nu)}$)



- This ordering has an additional good property (next slide)

THE 0-LIMIT ISSUE

- In most approaches, the confidence region does not approach the result obtained without systematics when σ(syst)→0 !
- Annoying, especially considering the limit is often *tighter* than in absence of systematics (I believe it was pointed out by G.Feldman at CL workshop@FNAL)
- This is prevented by the ordering shown in previous page
(fine print: in discrete cases may require some parameter tuning)



[Both these improvements have not been exploited much yet]

A SUMMARY ON FREQUENTIST SYSTEMATICS

	Ordering	Integration	Nuisance scan	Flip-flop	0-limit	Computation
MINOS,PROB,"Profile", TRolke	LR_{prof}	Approximate, assumes chi2 distrib.	$\nu = \nu_{best}(\mu, x_0)$	~N	NO	Easy
"Plugin", CKMFIT-RFit, Stat Sin 19, 301	LR_{prof}	Exact	$\nu = \nu best(\mu, x0)$ (CKMFIT/RFit assumes ranges)	Y	NO	Moderate
"Bayesian","Smeared","Hybrid"	any	Exact	Averaged over	N	OK	Easy
PRD 85, 072002 (sin $2beta_s$)	LR_{prof}	Exact	Exact (BB-clipped, boundary	N	NO	Moderate
CKMfit-Scan	LR_{prof}	Exact	Exact (numerical)	N	NO	Heavy
physics/0511202	*Special	Exact	Exact (projection)	N	OK	Moderate

SPEAKING OF POWER: OPTIMIZING YOUR ANALYSIS

- Another topic of interest is *optimization* of the measurement (selection/other user choices)
- Back to the "point-H_0 vs continuous-H_1" scenario: a recurring issue is the choice between "optimizing for limits" vs "optimizing for discovery".
  In flavor physics, excluding H_0 is not necessarily a remote, if lucky, possibility. H_0 may be a null BR for a quite reasonably existing rare process; of CPV in a channel where it has not been observed, but may quite be (and at times, is).
- The multi-D nature of many flavor physics measurements comes as an additional complication

EXAMPLE OF MULTI-D SENSITIVITIES

- Recent DUNE paper (Aug 28) [ArXiv: 2008.12769], exemplifies typical paradigms for quantifying sensitivity of a future experiment:
  - 'average' limits (assuming H_0)
  - 'expected' signal reach
  (# signal events > median limit)

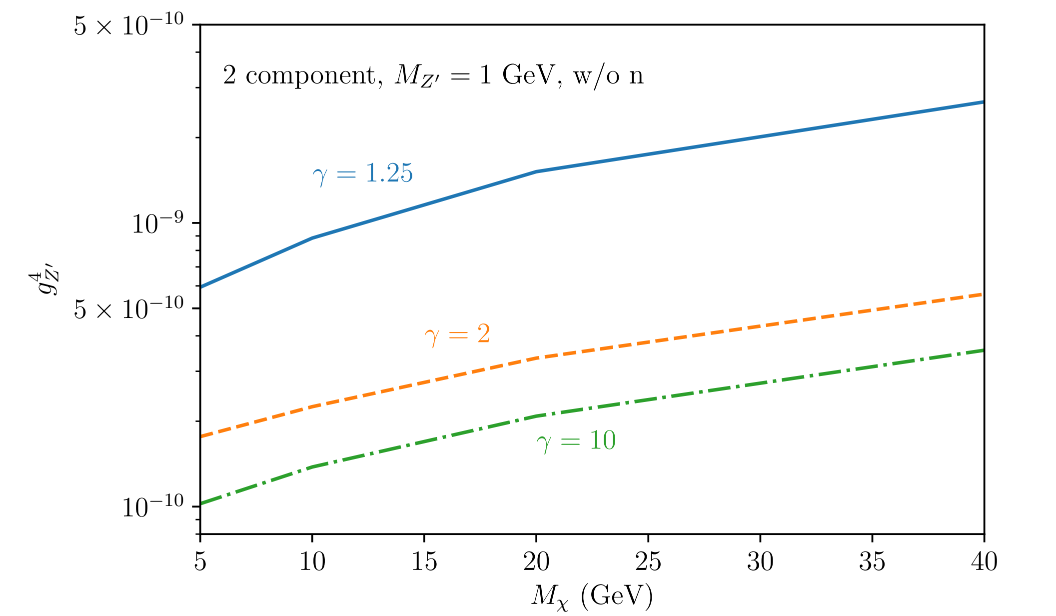


Fig. 26 Expected 5σ discovery reach with one year of DUNE livetime for one 10 kt module including neutrons in reconstruction (top) and excluding neutrons (bottom).

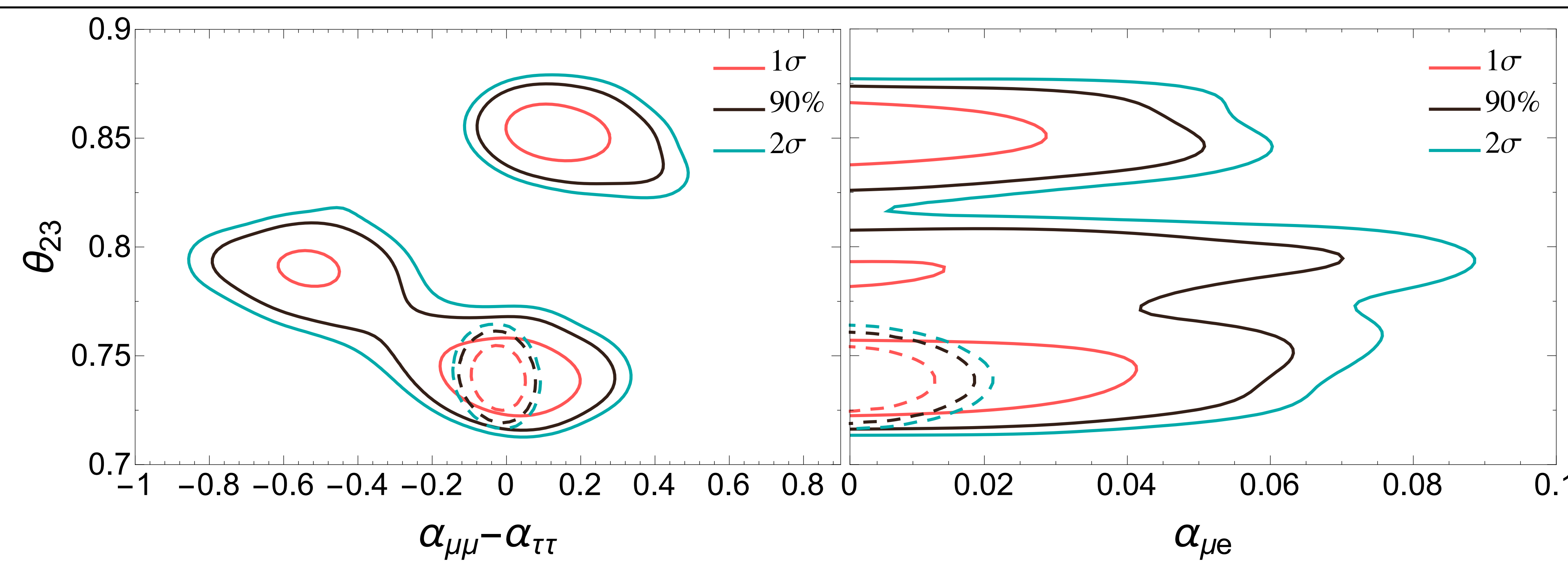


Fig. 6 Expected frequentist allowed regions at the 1σ, 90% and 2σ CL for DUNE. All new physics parameters are assumed to be zero so as to obtain the expected non-unitarity sensitivities. A value $\theta_{23} = 0.235\pi \approx 0.738$ rad is assumed. The solid lines

A DIFFERENT APPROACH: “SENSITIVITY"AS A REGION

[ARXIV:PHYSICS/0308063]

- Differs from usual notion of *sensitivity as a number*
- *Def:* The **sensitivity region** of a *search* is the **set**:

$$\mathcal{S} = \{\mu: 1 - \beta_\alpha(\mu) > CL\}$$

- Theorem: the following two facts hold simultaneously:

  1) If the true $\mu \in \mathcal{S}$, the probability of discovery is ***at least*** CL
  ("discovery" = excluding H_0@ signif. α)

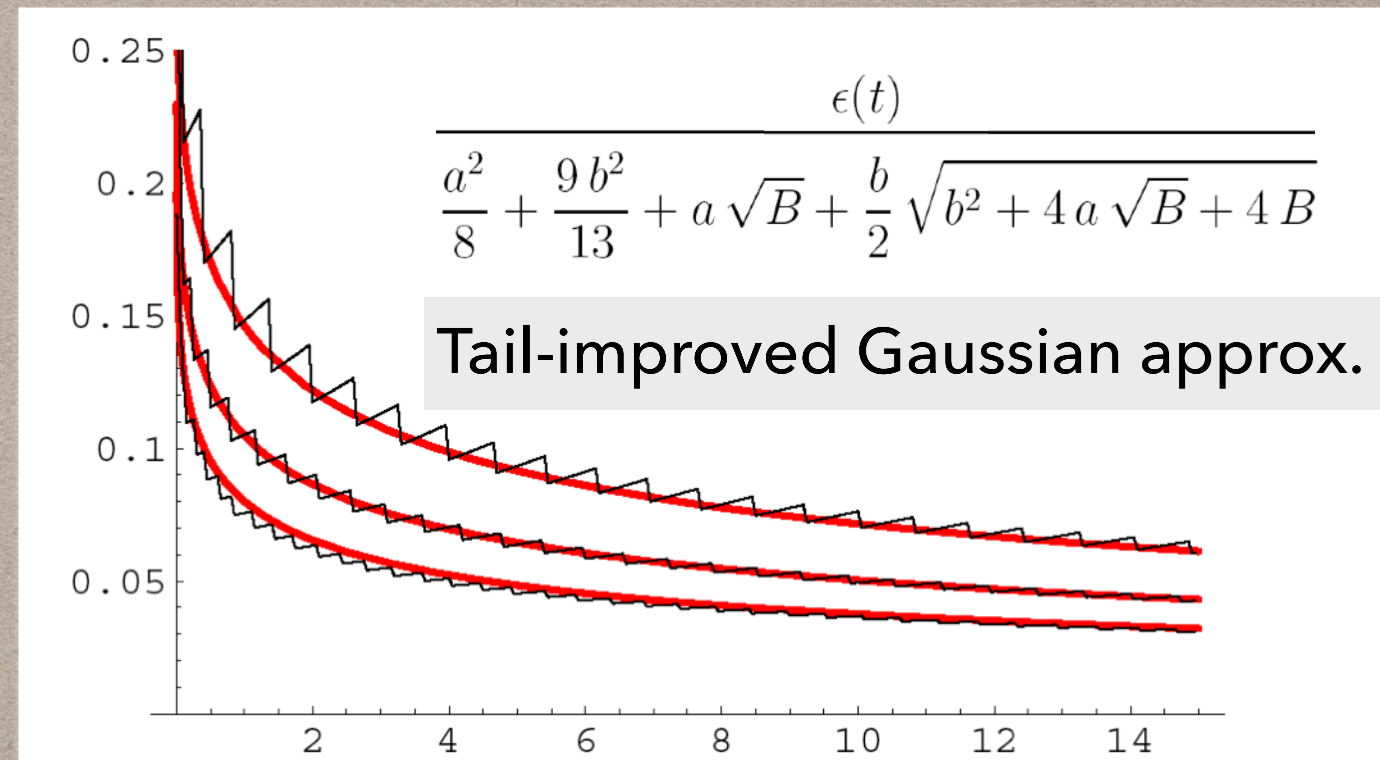
  2) In case H_0 is accepted instead, ***every*** $\mu \in \mathcal{S}$ will ***always*** be excluded @CL
  (independently of the true value of μ !)

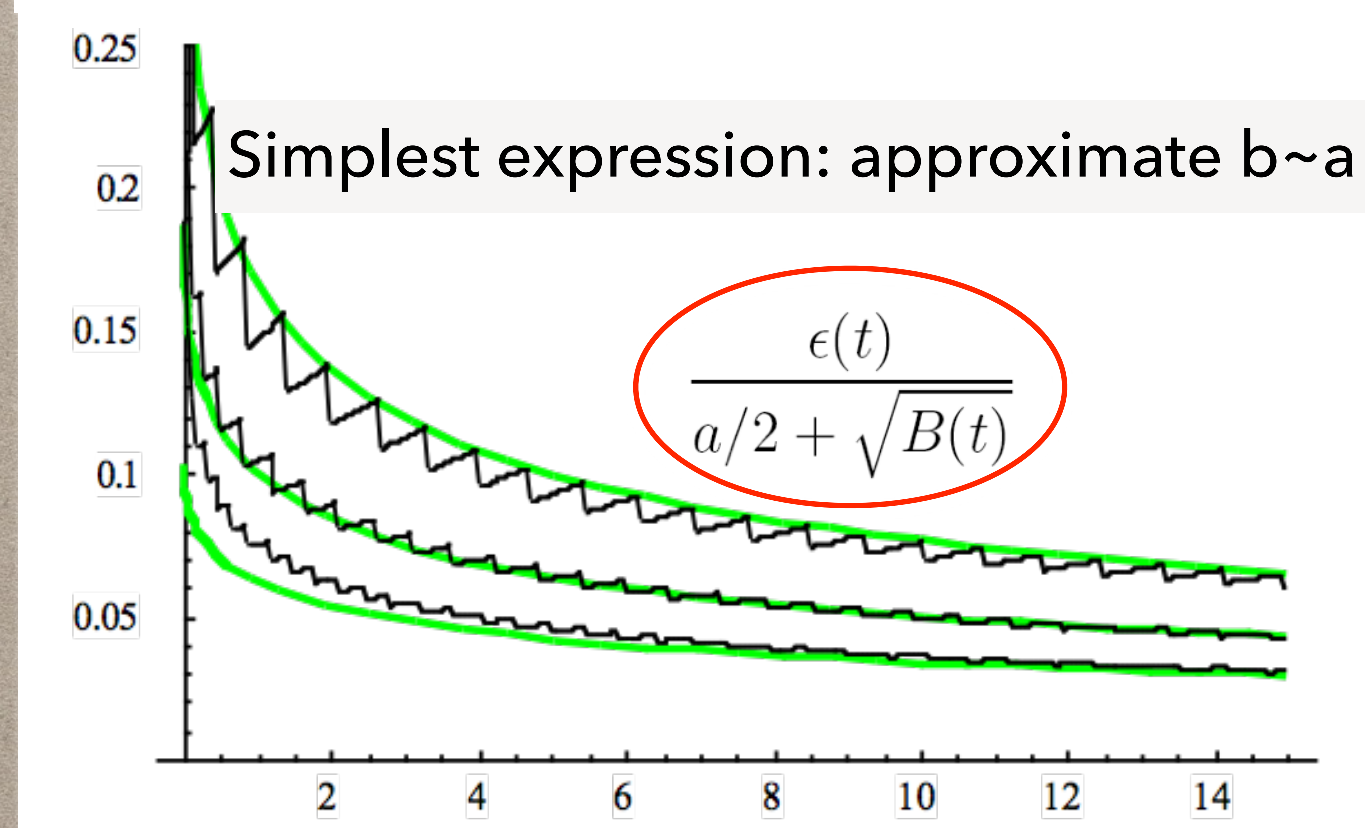
Optimization means to make S **as large as possibile** (“Unified” view of sensitivity). If it is not growing in all directions, it means there are physics choices - but this is good.

- NB: Independent of metrics and of expected signal. Independent of ordering for limits (fine print: acceptance region of the test should be excluded *before* any critical region is excluded. F-C usually works fine)

APPLICATION TO 'COUNTING EXPERIMENTS'

- For the Poisson+Background problem, the sensitive region is a half-line in the number of expected signal events $S(\mu)>S_{min}(B)$. Optimization then simply amount to minimizing $S_{min}(B)$
- This can be recast in terms of maximizing a function of the efficiency $\epsilon(t)$, <u>independent</u> of absolute cross section for signal
- Convenient approximate expressions can be written as functions of (a,b) = # of σ for (α,1-CL)





USAGE IN FLAVOUR PHYSICS

- Increasingly popular in HEP, particularly in Flavour
- Vast majority of existing papers simply maximize:

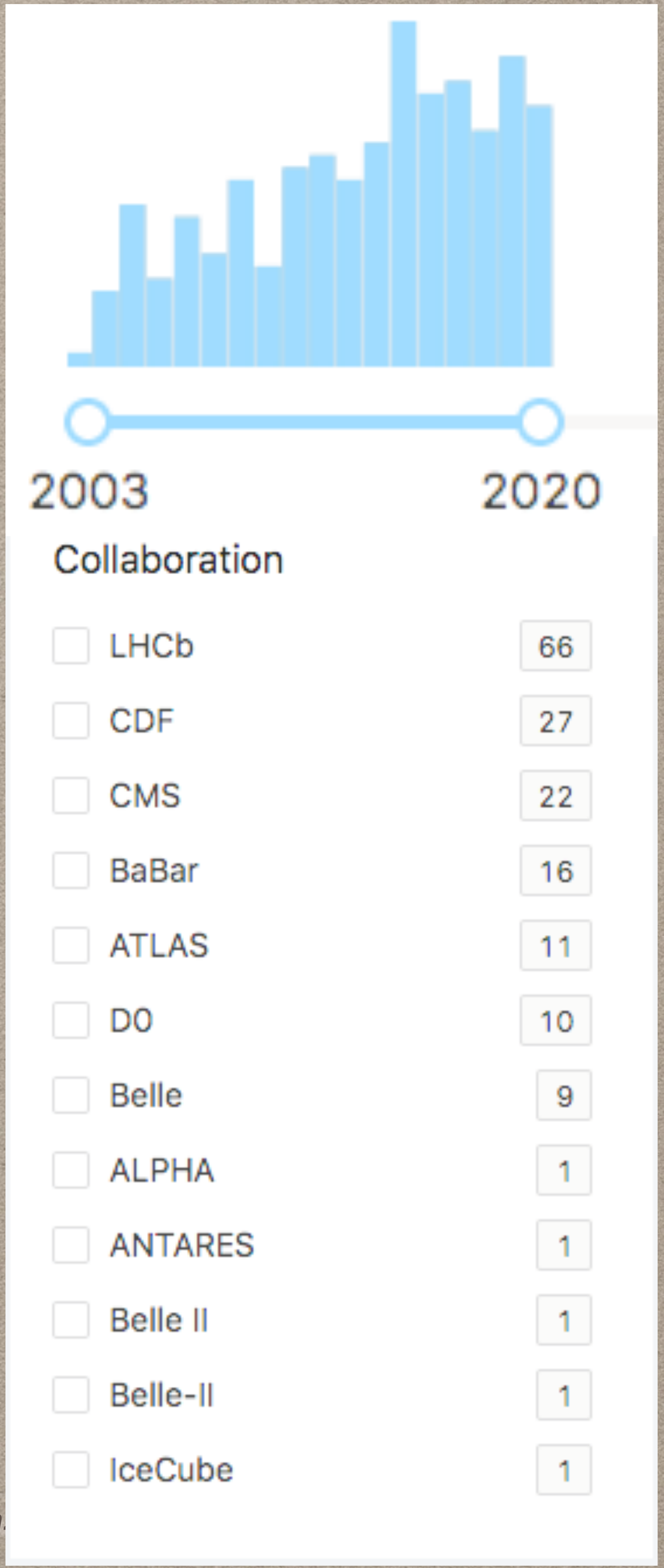
$$\frac{\epsilon}{a/2 + \sqrt{B}}$$

- This "out-of-the box" solution isn't bad - but was initially intended as a pedagogical example for the simplest possible case. There is still room to do better:

1. Can adapt to the actual likelihood fit -> more accurate optimization

   Not too difficult to explicitly solve the equation 1 - $\beta_\alpha(\mu)$ > CL in your specific case, and maximize the resulting region

2. Apply it to multi-D problems, not just counting experiments. The concept can be exploited also to make physics-driven choices. A bonus for Flavor physics.



G. Pun.

CONCLUSIONS

- There has been progress in interval estimation over time, and Flavor physics has benefitted. I like to think PHYSTAT helped.
- Analyses today more sophisticated and more conscious of issues. Often use several methods in parallel.
- There is still room for more progress. In particular, we are not yet making full use of the increased availability of computing power to get rid of old approximations that aren't necessary anymore.
- Computing power brought a revolution to deep learning and AI - there is no reason it should not do the same for Statistics.

BACKUP