CMS pMSSM workflow

Malte Mrowietz

04.11.2020

What’s going on in CMS with the pMSSM?

Scan model points from the pMSSM

- **Markov chain Monte Carlo** (MCMC)
- millions of points

Simulate events for a suitable **subset** with FastSim

- **Oversample** McMC in interesting regions
- ~ 500,000 points

Run CMS SUSY analyses over signal

Evaluate the results with simplified likelihood

- **Bayesian interpretation** possible due to MCMC sampling

Scanning with Markov chain Monte Carlo (McMC)

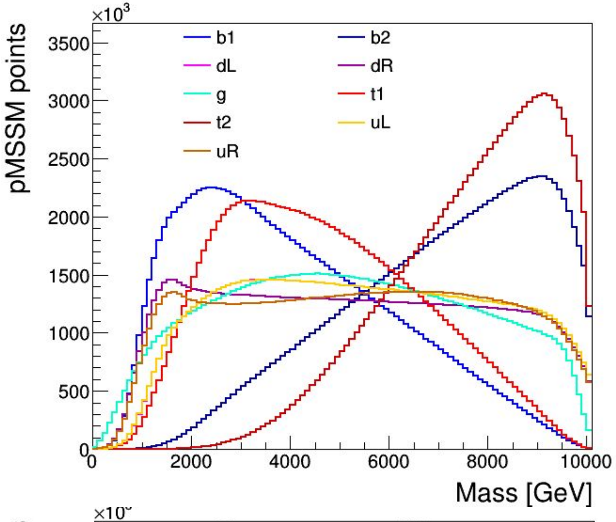
- Large parts of the (low-energy) pMSSM are constrained by all sorts of results (b-physics, LEP, Higgs mass etc)
- Want to sample the favored regions with higher frequency
- MCMC employs a likelihood to sample the space
  - Likelihood based on Higgs mass and low-energy observables
  - MCMC is <u>not</u> a minimization: it produces a posterior density
  - Efficient sampling of interesting pMSSM subspace, enables Bayesian interpretation

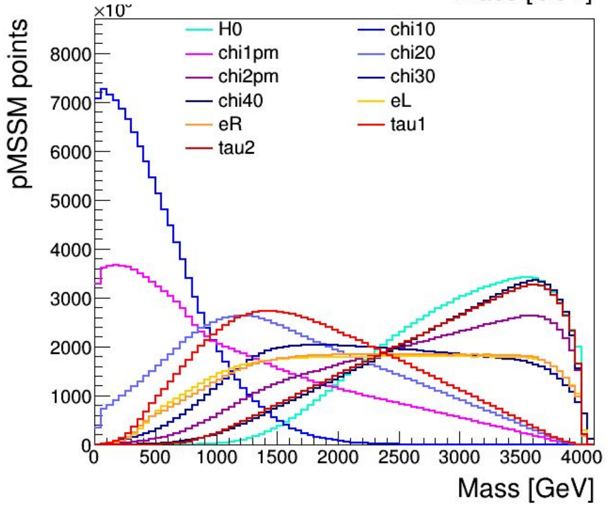
Considered ranges

Prior for MCMC: flat in linear pMSSM parameters

M_{q1}, M_{q3}, M_{u1}, M_{u3}, M_{d1}, M_{d3}	[0 , 10 TeV]
M_{l1}, M_{l3}, M_{r1}, M_{r3}	[0 , 4 TeV]
M_1, M_2, μ	[-4 TeV, 4 TeV]
M_3	[0 , 10 TeV]
A_t, A_b, A_l	[-7 TeV, 7 TeV]
m_{A0}	[0 , 4 TeV]
tan (β)	[2 , 60]

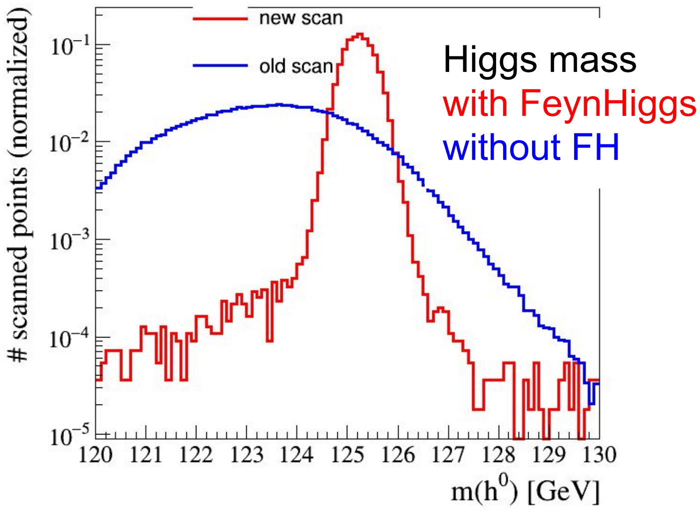
- Strong parameters up to 10 TeV:
  - Much of parameter space has no phenomenologically active strongly interacting particles
- Electroweak parameters up to 4 TeV:
  - No strong particles expected at LHC for m = 4 TeV, whole cascade phenomenology open
  - Long-lived phenomenology if μ small and 3 TeV < M_1, M_2 < 4 TeV
- tan (β) lower bound: non-perturbative at GUT scale

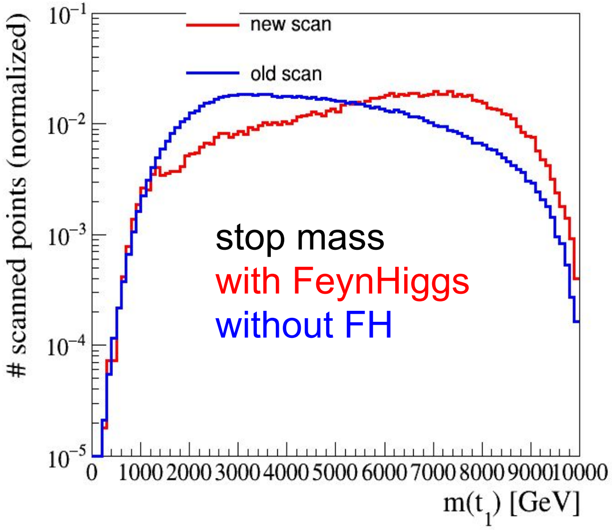




The McMC likelihood (part 1)

- Encoded prior knowledge into likelihood
- Avoided controversial results: if they turn out to be false, the scan could have a lingering bias
- public tools are available to implement low-energy constraints, e.g.:
  - Superiso (includes correlations among observables)
  - SPheno
  - New: FeynHiggs for Higgs mass likelihood





The McMC likelihood (part 2)

correlations not treated (SPheno)

- $BR(B^+ \rightarrow \tau \nu)$
- $BR(D_s \rightarrow \tau \nu)$
- $BR(D_s \rightarrow \mu \nu)$
- $\Delta(\rho)$

$L = \prod L_i$, i = flavour observables above

L_i: Gaussian distribution:

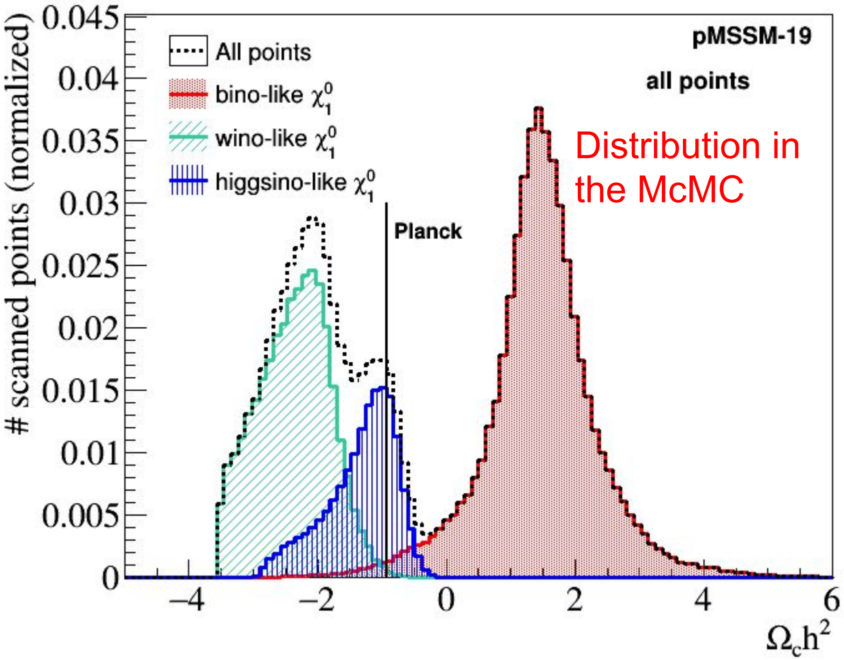
- centered on measurement
- width = measurement error
- evaluate at model prediction = likelihood

Superiso chi2 (correlations treated)

- $\Delta_0 (B \rightarrow K \gamma)$
- $BR(b \rightarrow s \gamma)$
- $BR(B_s \rightarrow \mu \mu)$
- $BR(B_d \rightarrow \mu \mu)$
- $BR(b \rightarrow s \mu \mu)$
- $BR(b \rightarrow s e e)$
- $BR(B0 \rightarrow K^{*0} \gamma)$

Picking a smaller, desirable subset of points

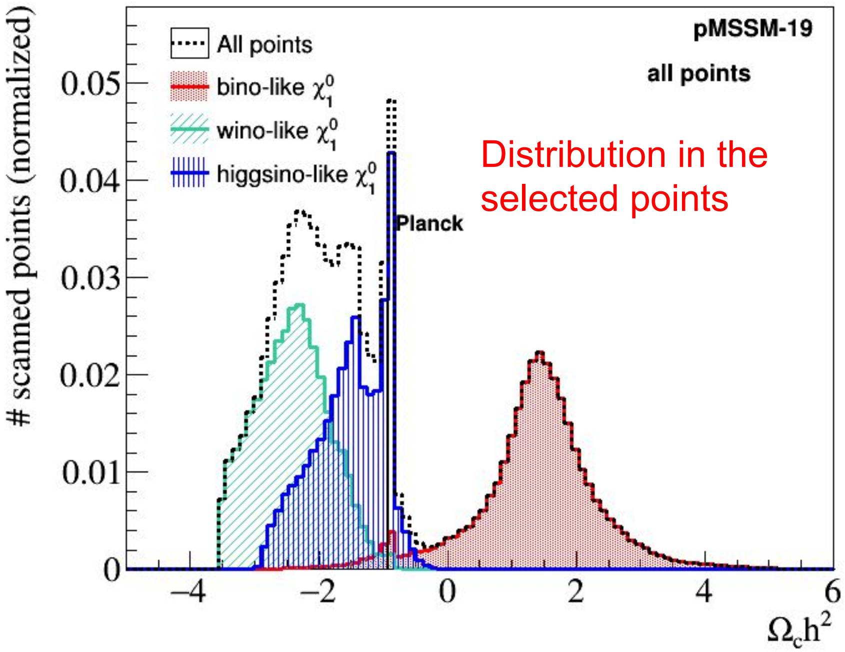
- Reduce point autocorrelation
- Over-sample regions we might want to zoom into:
  - by a factor of 3 if the point has ΔEW < 100 (as defined in *https://arxiv.org/pdf/1304.6732.pdf*)
  - by a factor of 10 if the point has m_{stop} < 1 TeV
  - by a factor of 5 if the point has m_{stop} < 1.5 TeV
  - by a factor of 20 if the point has a relic density compatible with the Planck measurement
  - undersample SModelS excluded points (using analyses that are not used later)
  - regions relevant for exciting but fluid results from the community, e.g., a_{μ},
- Weight all distributions by 1/(oversampling factor)



Distribution in the McMC

Picking a smaller, desirable subset of points

- Reduce point autocorrelation
- Over-sample regions we might want to zoom into:
  - by a factor of 3 if the point has ΔEW < 100 (as defined in *https://arxiv.org/pdf/1304.6732.pdf*)
  - by a factor of 10 if the point has m_{stop} < 1 TeV
  - by a factor of 5 if the point has m_{stop} < 1.5 TeV
  - by a factor of 20 if the point has a relic density compatible with the Planck measurement
  - undersample SModelS excluded points (using analyses that are not used later)
  - regions relevant for exciting but fluid results from the community, e.g., a_μ,
- Weight all distributions by 1/(oversampling factor)



Distribution in the selected points

Simulation and gen-level filter

No need to simulate events that we know won't pass a trigger

A gen-level filter might look something like this:

- Loose **OR** gen-level event filter (Filter efficiencies above 1%)
  - gen H_T > 140 GeV
  - gen muon with p_T > 15 GeV
  - gen electron with p_T > 15 GeV
  - gen photon with p_T > 70 GeV
  - gen tau with p_T > 30 GeV
  - leading photon p_T > 30 GeV + sub-leading photon p_T>18 GeV
  - two or more gen objects of type: electron p_T > 5, muon p_T > 2.5, photon p_T > 30
  - detector stable chargino that reaches the muon system

Conclusion

- Sample the pMSSM using Markov chain Monte Carlo
  - Good way to **incorporate prior (low-energy) results** into sampling
  - **Efficient sampling** of low-energy posterior density
  - **Bayesian interpretation** of scan is possible
  - Need to sample many more points than are actually analyzed
- Sample McMC posterior to obtain a subset of pMSSM models to analyze
  - **Increase scan resolution** by oversampling interesting regions (or undersampling less interesting ones)
- Bayesian interpretation of final posterior density
  - Solid interpretation framework
  - Much broader set of statements possible compared to frequentist interpretation
  - Need to be very careful about the conditions under which the conclusions are valid

Backup

Markov chain Monte Carlo Scan Workflow

1: Throw parameter point

Create parameter point
Value for each parameter are independently thrown from gaussian centered on value of previous point

2: Create physical spectrum

Use SPheno to create physical parameter spectrum (particle masses etc.).
Get Higgs mass from FeynHiggs

3: Calculate observables

Use SuperIso & SPheno to calculate flavour observables

4: MCMC decision

Calculate likelihood and
R=likelihood/likelihood$_{\text{previous point}}$

- Accept point if likelihood at least a good as for previous point
- Else, accept if R>u, where u=[0,1], flat

How long, and where to store

- If Bayesian interpretation desired: MCMC needs to converge. Try to apply convergence heuristics
- Points in same chain are autocorrelated, need to select a sufficiently sparse subset before continuing
- Convergence of the chain only matters for Bayesian interpretation
- How to store output:
  - Current version of the scan saved the compressed slha text files on disk
    - Produces large I/O for transferring files from worker node to storage
    - Very space inefficient
  - Better: store points directly in database format (ROOT, HDF5,etc.)

Further thoughts on the MCMC

- Consider MCMC step time (in CMS: 10-20s per point)
  - Do not compute quantities that are not necessary for the MCMC at this point
  - Main time consumer for us: disk I/O
  - It may be worth investing in an interface does not need to write to disk
- CMS uses private MCMC implementation, but: there are public libraries that implement MCMC:
  - Different choices of step function
  - Diagnostic tools
  - Potentially better optimized