Reconstructing the MSSM Lagrangian from LHC data

Michael Rauch

Rémi Lafaye, Tilman Plehn, Dirk Zerwas

What SFitter does

- Set of measurements
  - LHC measurements:
    kinematic edges, thresholds, masses, mass differences
    cross sections, branching ratios
  - ILC measurements
  - Indirect Constraints
    electro-weak: M_W, $\sin^2\theta_W$; $(g-2)_\mu$
    flavour: BR$(b \to s\gamma)$, BR$(B_s \to \mu^+\mu^-)$; dark matter: Ωh^2
  - or even ATLAS and CMS measurements separately
- Compare to theoretical predictions
  - Spectrum calculators: SoftSUSY, SuSPECT, ISASUSY
    [Allanach; Djouadi, Kneur, Moultaka; Baer, Paige, Protopopescu, Tata]
  - LHC cross sections: Prospino2 [Plehn et al.]
  - LC cross sections: MsmLib [Ganis]
  - Branching Ratios: SUSYHit (HDecay + SDecay) [Djouadi, Mühlleitner, Spira]
  - micrOMEGAs [Bélanger, Boudjema, Pukhov, Semenov]
  - g-2 [Stöckinger]

Parameter Scans

- MSSM parameter space is high-dimensional:
  - SM: 3+ parameters (m_t, α_s, α, $\ldots$)
  - mSUGRA: 5 parameters (m_0,$m_{1/2}, A_0, \tan(\beta), \mathrm{sgn}(\mu)$)
  - General MSSM: 105 parameters
- On loop-level observables depend on every parameter
  Simple inversion of the relations not possible
  $\Rightarrow$ Parameter scans
- Error estimates on parameters in the minimum

Find best points (best χ^2) using different fitting techniques:

- fixed Grid scan $\left(\begin{array}{c} + \text{ scans complete parameter space} \\ - \text{ many points needed } (\mathcal{O}(e^N)) \end{array}\right)$
- Gradient search (Minuit) $\left(\begin{array}{c} + \text{ Reasonably fast} \\ - \text{ Limited convergence, only best fit} \end{array}\right)$
- Weighted Markov Chains

Markov Chains

Markov Chain (MC):

- Sequence of points, chosen by an algorithm (Metropolis-Hastings), only depending on its direct predecessor
- Picks a set of "average" points according to a potential V (e.g. inverse log-likelihood, $1/\chi^2$)
- Point density resembles the value of V (i.e. more points in region with high V)
- Scans high dimensional parameter spaces efficiently [Baltz, Gondolo 2004]
- mSUGRA MC scans with current exp. limits
  [Allanach, Lester, Weber 2005-7; Roszkowski, Ruiz de Austra, Trotta 2006/7]

Weighted Markov Chains: Improved evaluation algorithm for binning: [Plehn, MR]

- Weight points with value of V: $\left(\frac{\text{number of points}}{\sum_{\text{points}} 1/V(\text{point})}\right)$ [based on Ferrenberg, Swendsen 1988]
- Maintain additional chain which stores points rejected because $V(\text{point}) = 0$

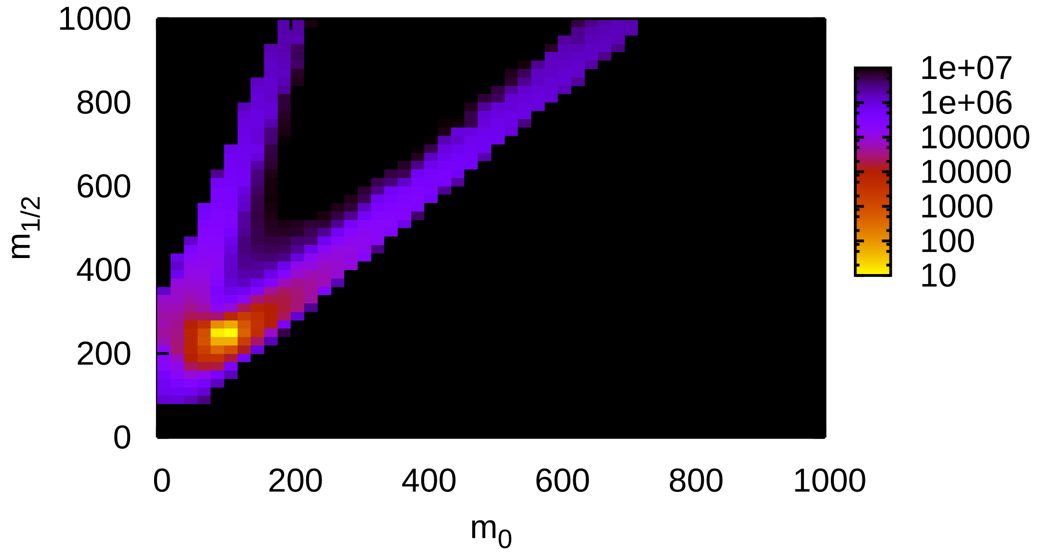
$+$ Fast scans of high-dimensional spaces $\mathcal{O}(N)$
$+$ Does not rely on shape of χ^2 (no derivatives used)
$+$ Can find secondary distinct solutions
$-$ Exact minimum not found $\Rightarrow$ Additional gradient fit
$-$ Bad choice of proposal function for next point leads to bad coverage of the space

mSUGRA as a Toy Model

mSUGRA with LHC measurements (SPS1a kinematic edges):
pick one set of ”measurements”, randomly smeared from the true values

Free parameters:
m_0, $m_{1/2}$, $\tan(\beta)$, A_0, $\mathrm{sgn}(\mu)$, m_t

SFitter output 1:
Fully-dimensional exclusive likelihood map
(colour:
minimum χ^2 over all unseen parameters)



SFitter output 2:
Ranked list of minima:

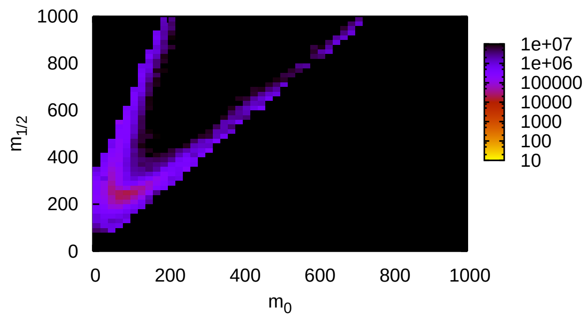
	χ^2	m_0	$m_{1/2}$	$\tan(\beta)$	A_0	μ	m_t
1)	1.32	100.4	251.2	12.7	-71.7	$+$	171.9
2)	7.18	106.3	243.6	14.3	-103.3	$-$	170.7
3)	13.9	103.5	258.2	12.2	848.4	$+$	174.4
4)	75.1	107.3	251.4	15.1	778.8	$-$	173.6

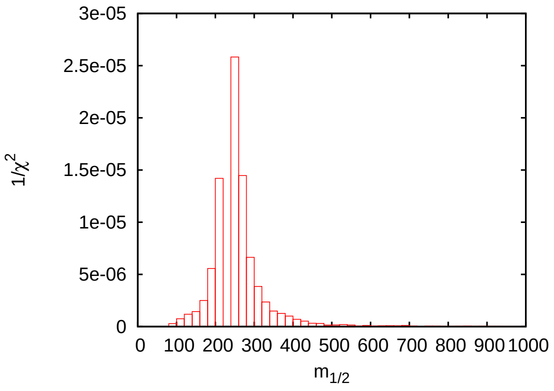
...

Bayesian or Frequentist?

SFitter provides full-dimensional log-likelihood map
$\longrightarrow$ "project" onto plotable 1- or 2-dimensional spaces

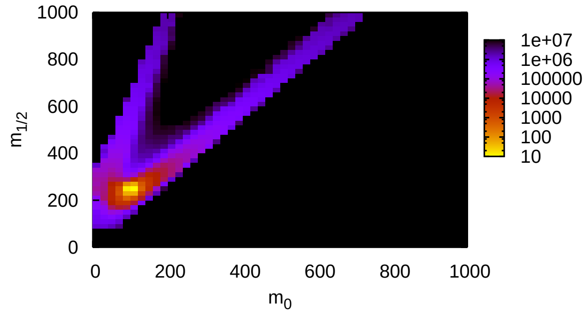
Bayesian:

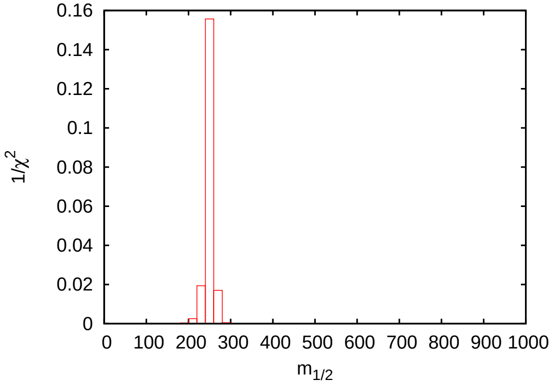




Marginalisation of χ^2 in all other directions

Frequentist:



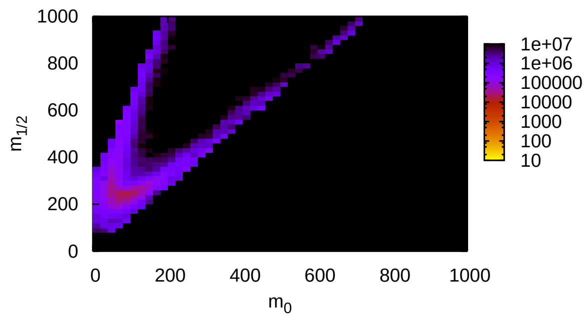


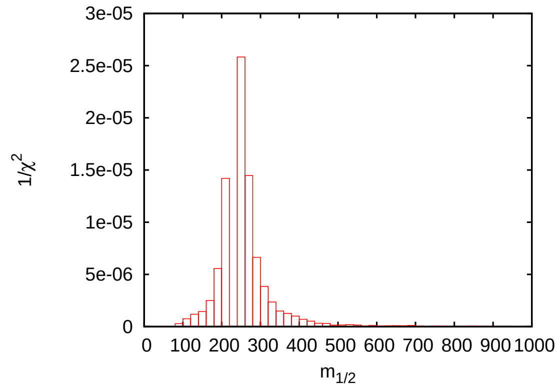
Profile likelihood: Value of bin is value of smallest χ^2 occuring in this bin

Bayesian or Frequentist?

SFitter provides full-dimensional log-likelihood map
$\longrightarrow$ "project" onto plotable 1- or 2-dimensional spaces

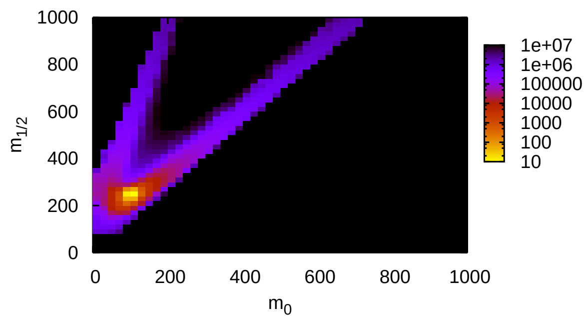
Bayesian:

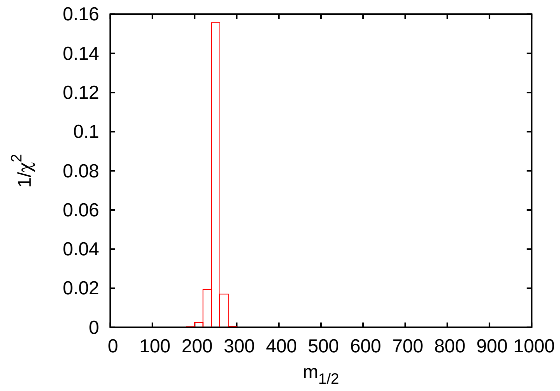




Marginalisation of χ^2 in all other directions

Frequentist:





Profile likelihood: Value of bin is value of smallest χ^2 occuring in this bin

Different methods answer different questions.
$\Rightarrow$ Bayesian and Frequentist!
Everybody can choose his/her favourite analysis . . .

Purely high-scale model

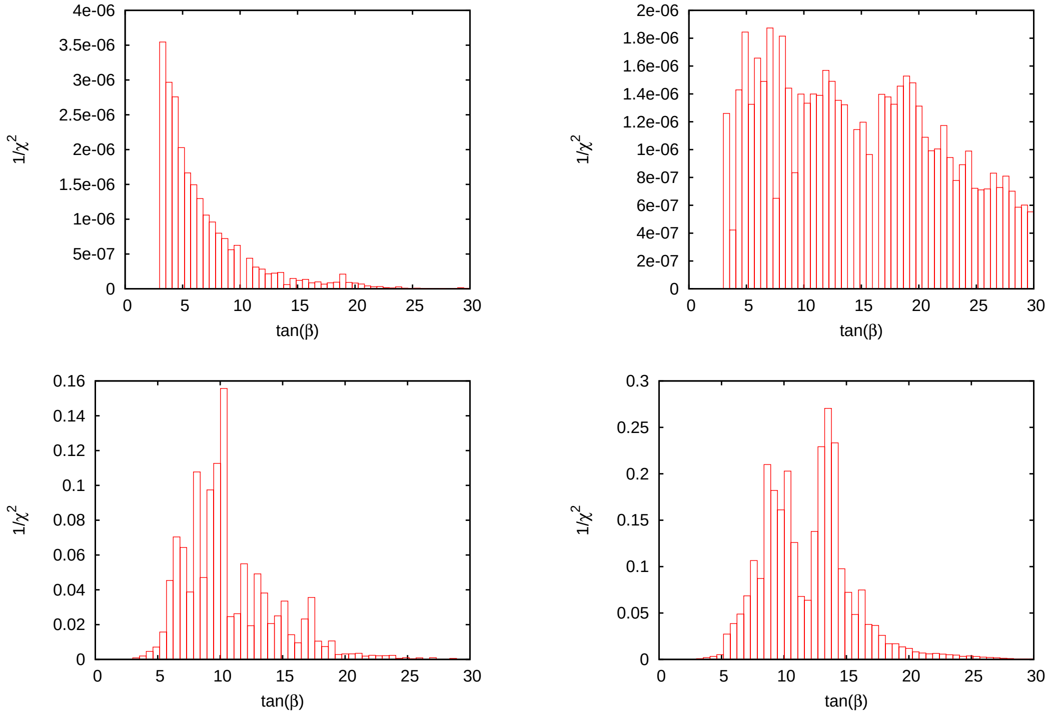
m_0, $m_{1/2}$, A_0 defined at the GUT-scale $\Leftrightarrow$ $\tan(\beta)$ defined at the weak scale

$\Rightarrow$ Replace $\tan(\beta)$ with high-scale quantity B

$\Rightarrow$ Flat prior in B yields prior $\propto \frac{1}{\tan(\beta)^2}$

SPS1a with LHC kinematic edges ($\tan(\beta)$ vs. $1/\chi^2$):

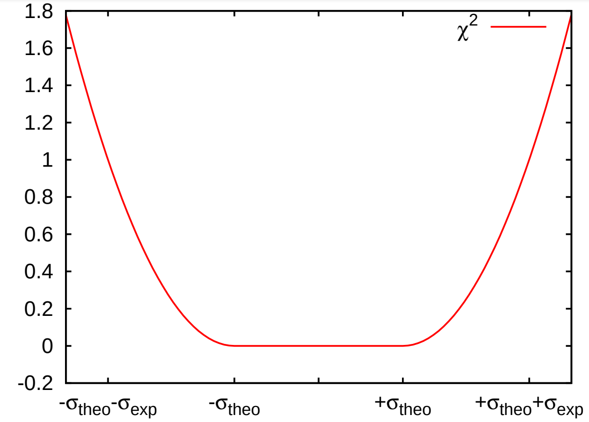
flat B prior | flat $\tan(\beta)$ prior



Bayesian:
Large influence of choice of prior
Choosing flat B prior strongly favours low values of $\tan(\beta)$.

Frequentist:
Two plots should be identical (no prior in χ^2 calculation)
Indirect influence via Markov Chain proposal function

Error determination



Treatment of errors:

- All experimental errors are Gaussian
$$\sigma^2_{\text{exp}} = \sigma^2_{\text{stat}} + \sigma^2_{\text{syst}(j)} + \sigma^2_{\text{syst}(l)}$$
- Systematic errors from jet ($\sigma_{\text{syst}(j)}$) and lepton energy scale ($\sigma_{\text{syst}(l)}$) assumed 99% correlated each
- Theory error added as box-shaped (RFit scheme [Hoecker, Lacker, Laplace, Lediberder])

$$\Rightarrow -2\log L \equiv \chi^2 = \sum_{\text{measurements}} \begin{cases} 0 & \text{for } |x_{\text{data}} - x_{\text{pred}}| < \sigma_{\text{theo}} \\ \left(\frac{|x_{\text{data}} - x_{\text{pred}}| - \sigma_{\text{theo}}}{\sigma_{\text{exp}}}\right)^2 & \text{for } |x_{\text{data}} - x_{\text{pred}}| \geq \sigma_{\text{theo}} \end{cases}$$

$\Rightarrow$ Parameter errors:

	SPS1a	$\Delta^{\text{theo}-\text{exp}}_{\text{flat}}$	$\Delta^{\text{theo}-\text{exp}}_{\text{zero}}$	$\Delta^{\text{theo}-\text{exp}}_{\text{gauss}}$	$\Delta^{\text{theo}-\text{exp}}_{\text{flat}}$
		LHC masses	LHC edges		
m_0	100	4.89	0.50	2.96	2.17
$m_{1/2}$	250	3.27	0.73	2.99	2.64
$\tan\beta$	10	2.73	0.65	3.36	2.45
A_0	-100	56.4	21.2	51.5	49.6
m_t	171.4	0.98	0.26	0.89	0.97

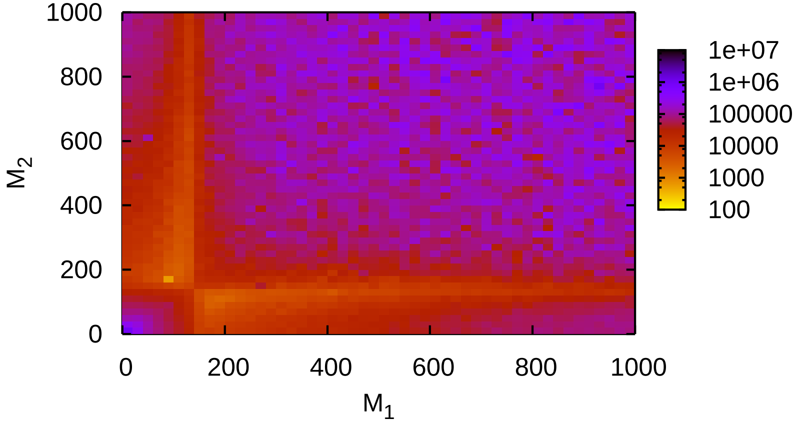
$\Rightarrow$ Use kinematic edges for parameter determination instead of masses

Weak-scale MSSM

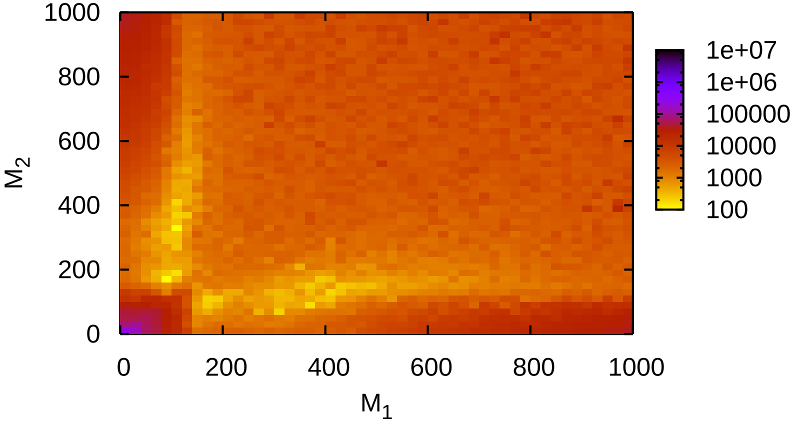
- No need to assume specific SUSY-breaking scenario
- Use of Markov Chains makes scanning the 19-dimensional parameter space feasible
- Lack of sensitivity on one parameter does not slow down the scan (no need to fix parameters)
- Underdetermined combinations of parameters can wash out correlations in plots
- Same SFitter output as before: Minima list and Likelihood map

MSSM using SPS1a spectrum and LHC kinematic edges:
(M_1, M_2, M_3, μ, $\tan(\beta)$, m_t sub-space)

Marginalisation

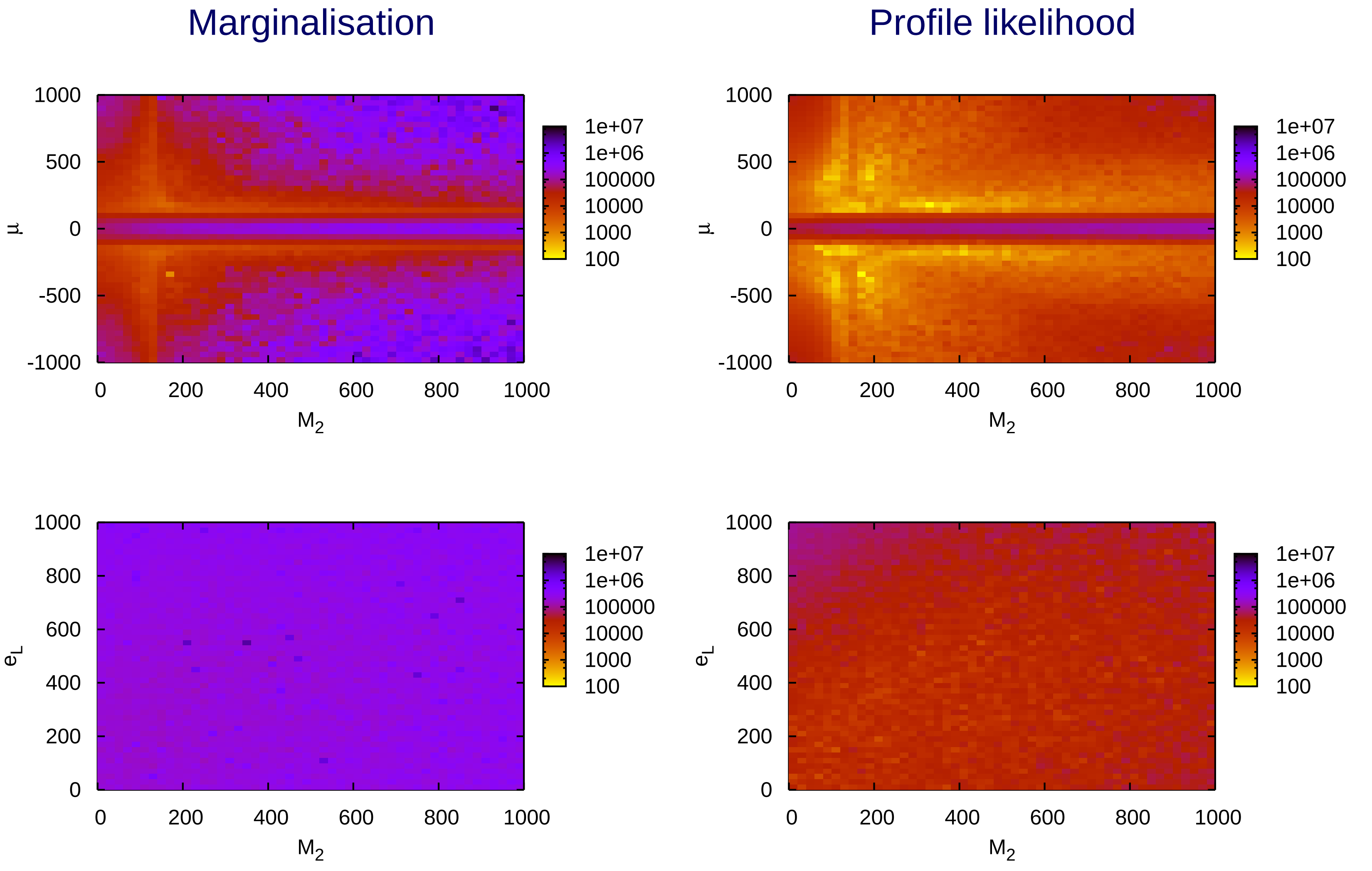


Profile likelihood

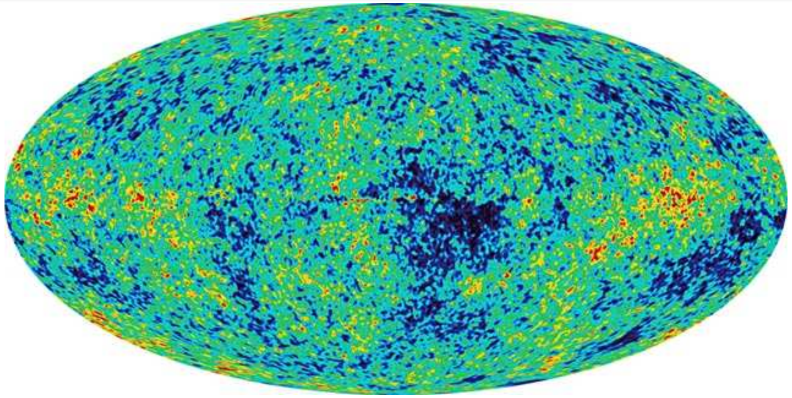


Weak-scale MSSM

MSSM using SPS1a spectrum and LHC kinematic edges:
(upper line: M_1, M_2, M_3, μ, $\tan(\beta)$, m_t sub-space
lower line: full parameter space)



Dark Matter



[NASA/WMAP Science Team]

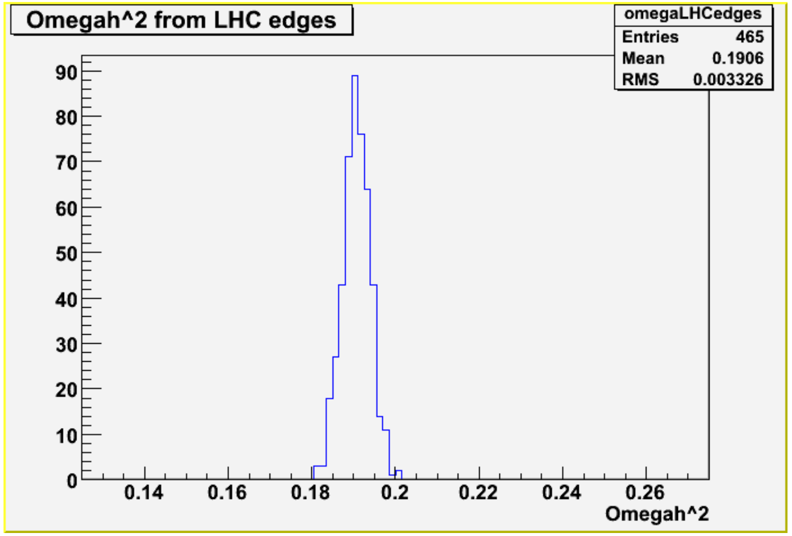
Content of the universe:

- 73% Dark energy
- 4% Ordinary matter
- 23% Dark matter

MSSM: χ_1^0 as LSP ideal candidate for cold dark matter (CDM): massive, weakly interacting

- SFitter: Determine Lagrangian parameters $\Rightarrow$ Spectrum and couplings
- e.g. micrOMEGAs: Calculate relic density $\Omega_{\text{CDM}} h^2 = n_{\text{LSP}} m_{\text{LSP}}$ [Bélanger et al.]
- $\Rightarrow$ Prediction of $\Omega_{\text{CDM}} h^2$
  LHC : $\Omega_{\text{CDM}} h^2 = 0.1906 \pm 0.0033$
  LHC+ILC: $\Omega_{\text{CDM}} h^2 = 0.1910 \pm 0.0003$
  (improvement by one order of magnitude)
- Compare with experiment
  (Measurement of the fluctuations of the
  cosmic microwave background):

WMAP: $\Omega_{\text{CDM}} h^2 = 0.1277 \pm 0.008$ [astro-ph/0603449]
Planck: $\Omega_{\text{CDM}} h^2 = \quad ? \pm 0.0016$

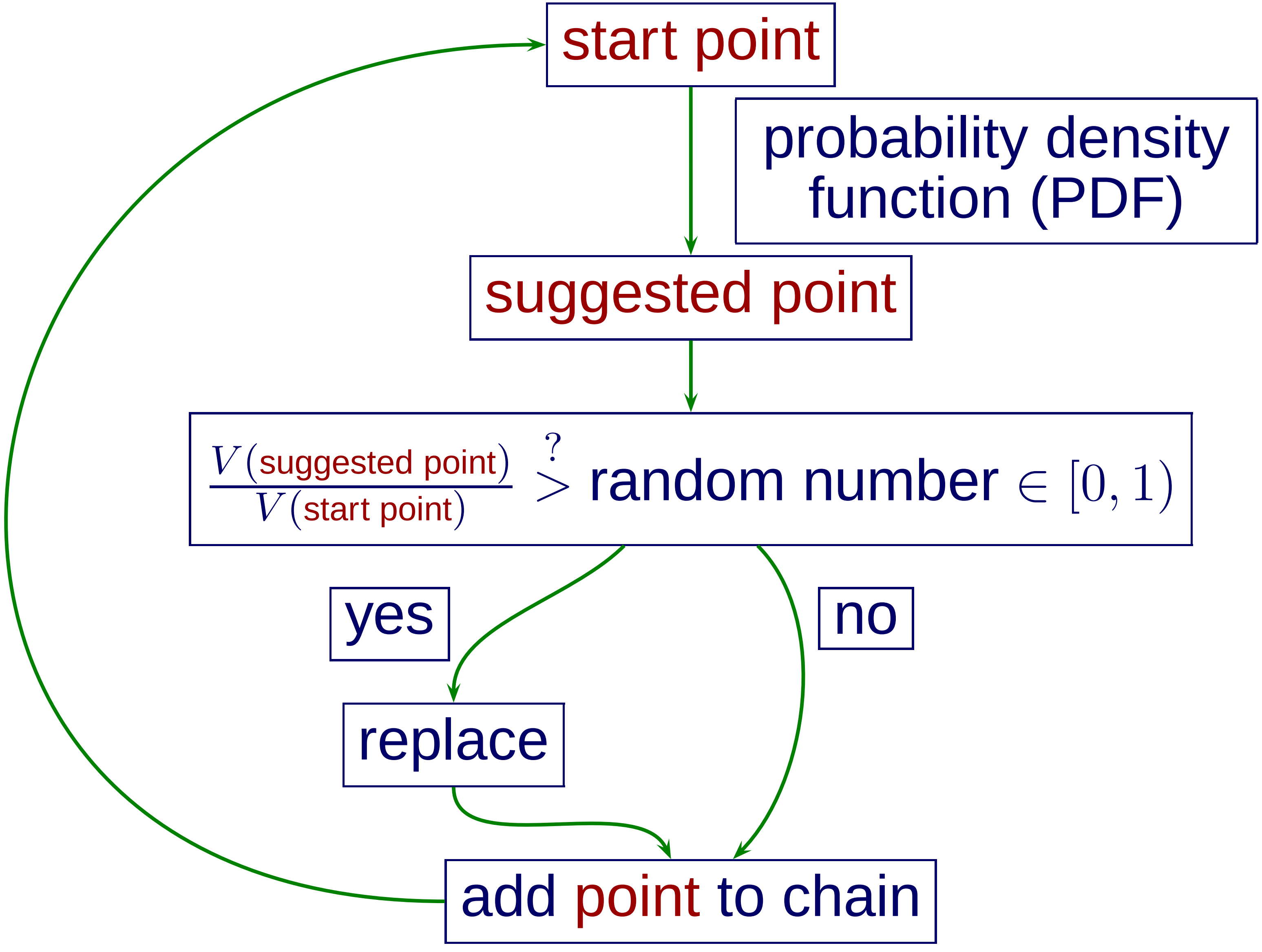


Summary & Outlook

- Parameter scans important to determine Lagrangian parameters from observables
- Problem of high-dimensional parameter spaces
- Markov Chains can do this effectively
- Improved algorithm developed
- Two types of output: Likelihood map and list of best points
- Both Bayesian and Frequentist from likelihood map
- Bayesian output significantly dependent on priors
- Tested with mSUGRA SPS1a:
  can reconstruct SPS1a from (simulated) LHC data
- Repeated procedure with weak-scale MSSM:
  reconstruction works as well
- SFitter (despite its name) not tied to SUSY
  $\longrightarrow$ extend to other models/problems

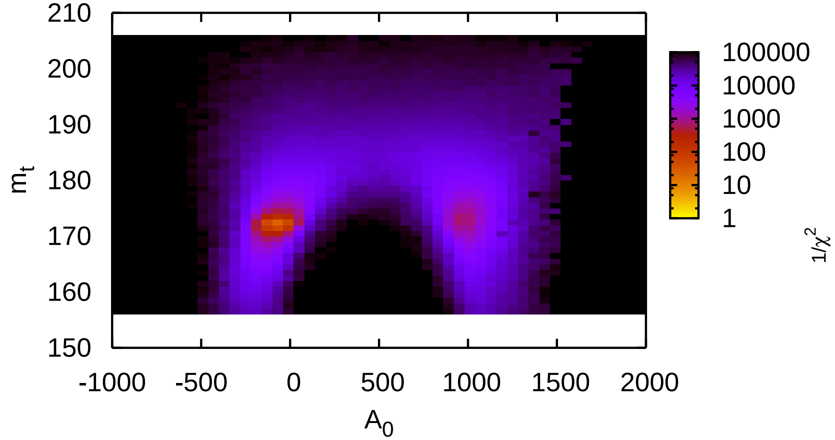
Backup Slides

Metropolis-Hastings Algorithm

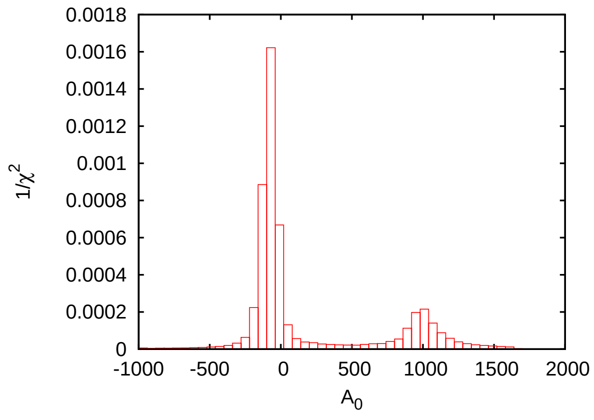


mSUGRA around Minima

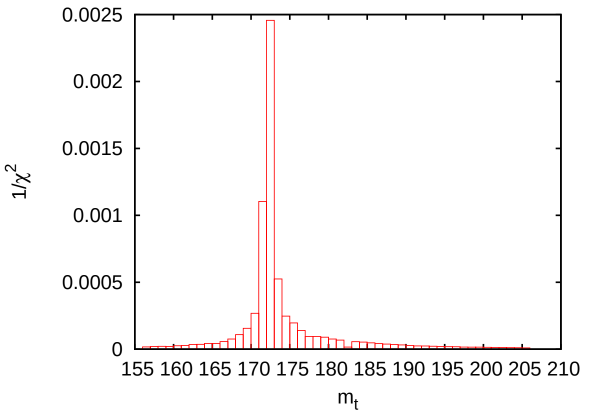
$\mathrm{sgn}(\mu) = +1$



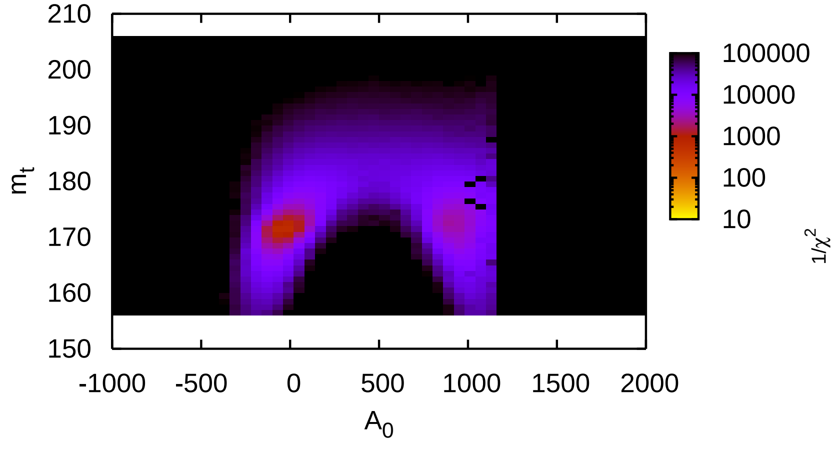
A_0



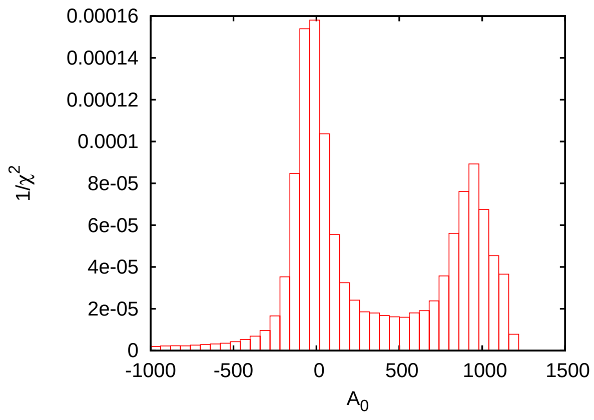
m_t



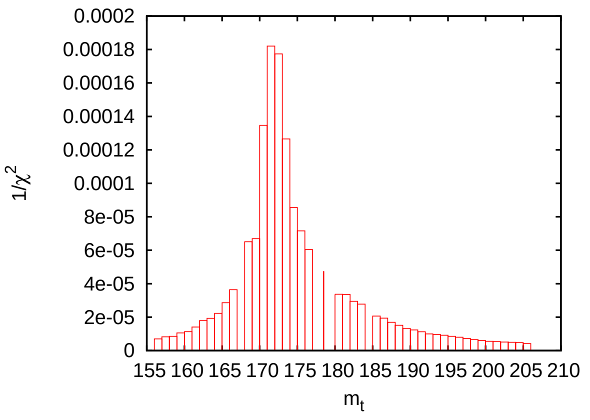
$\mathrm{sgn}(\mu) = -1$



A_0



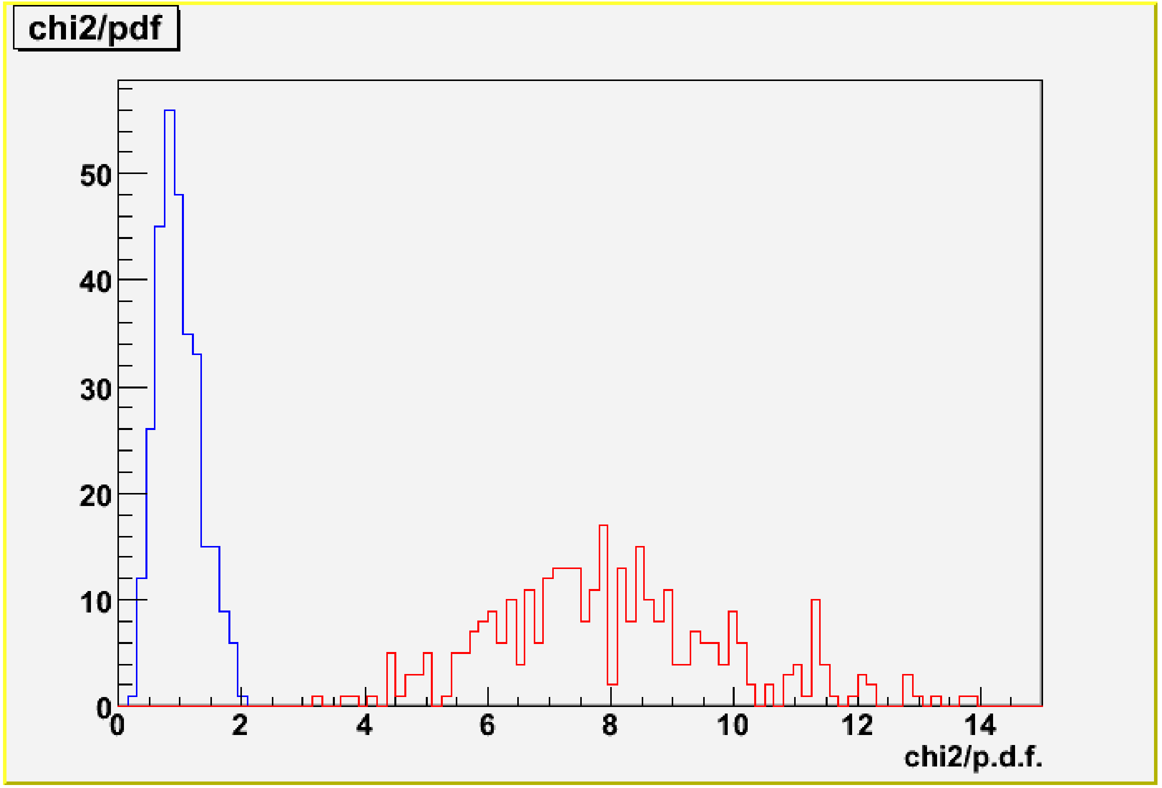
m_t



mSUGRA Minima

mSUGRA:
Correct solution vs. negative μ solution

Experimental results smeared by random number distributed as Gaussian around central value



[plot by D Zerwas 2006]

Experimental Input (edges)

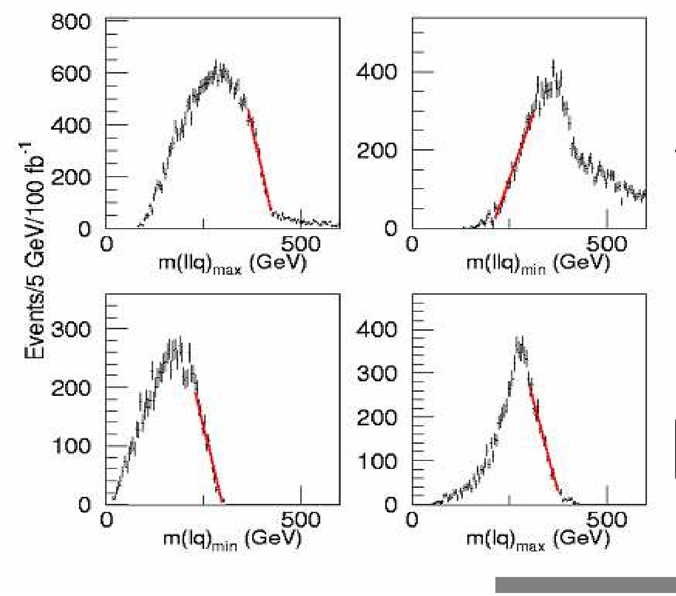
mSUGRA at LHC

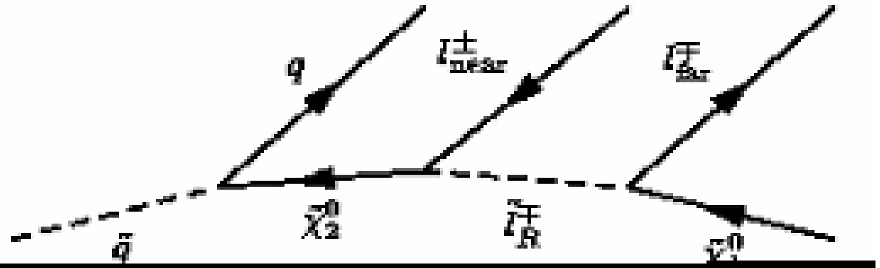
[Rémi Lafaye]

mSUGRA SPS1a as a benchmark point:

m_0=100 GeV, $m_{1/2}$=250 GeV, $\tan\beta$=10, A_0=-100 GeV, μ>0 and m_{top}=174.1 GeV

The LHC "experimental" data from cascade decays:





Variable	Value (GeV)	Errors Stat. (GeV)	Scale (GeV)	Total
$m_{\ell\ell}^{max}$	77.07	0.03	0.08	0.08
$m_{\ell\ell q}^{max}$	428.5	1.4	4.3	4.5
$m_{\ell q}^{low}$	300.3	0.9	3.0	3.1
$m_{\ell q}^{high}$	378.0	1.0	3.8	3.9
$m_{\ell\ell q}^{min}$	201.9	1.6	2.0	2.6
$m_{\ell\ell b}^{min}$	183.1	3.6	1.8	4.1
$m(\ell_L) - m(\tilde\chi_1^0)$	106.1	1.6	0.1	1.6
$m_{\ell\ell}^{max}(\tilde\chi_4^0)$	280.9	2.3	0.3	2.3
$m_{\tau\tau}^{max}$	80.6	5.0	0.8	5.1
$m(\tilde g) - 0.99 \times m(\tilde\chi_1^0)$	500.0	2.3	6.0	6.4
$m(\tilde q_R) - m(\tilde\chi_1^0)$	424.2	10.0	4.2	10.9
$m(\tilde g) - m(\tilde b_1)$	103.3	1.5	1.0	1.8

[redacted]

Theoretical errors:

- 3% for gluino and squark masses
- 1% for other sparticle masses

Experimental Input (edges)

(Obs)		= (meas) ± (exp) ± (theo)			
m_{h^0}	$=$	$109.53 \pm$	$0.25 \pm$	2.0	
m_t	$=$	$171.4 \pm$	$1.0 \pm$	0.0	
$\Delta m_{\tilde{\mu}_L,\chi_1^0}$	$=$	$106.26 \pm$	$1.6 \pm$	0.1	
$\Delta m_{\tilde{g},\chi_1^0}$	$=$	$509.96 \pm$	$2.3 \pm$	6.0	
$\Delta m_{\tilde{c}_R,\chi_1^0}$	$=$	$450.52 \pm$	$10.0 \pm$	4.2	
$\Delta m_{\tilde{g},\tilde{b}_1}$	$=$	$98.971 \pm$	$1.5 \pm$	1.0	
$\Delta m_{\tilde{g},\tilde{b}_2}$	$=$	$64.016 \pm$	$2.5 \pm$	0.7	
Edge(χ_2^0,$\tilde{\mu}_R$,χ_1^0)	$=$	$79.757 \pm$	$0.03 \pm$	0.08	$(m_{ll}^{\max})$
Edge($\tilde{c}_L$,χ_2^0,χ_1^0)	$=$	$446.44 \pm$	$1.4 \pm$	4.3	$(m_{llq}^{\max})$
Edge($\tilde{c}_L$,χ_2^0,$\tilde{\mu}_R$)	$=$	$316.51 \pm$	$0.9 \pm$	3.0	(m_{lq}^{low})
Edge($\tilde{c}_L$,χ_2^0,$\tilde{\mu}_R$,χ_1^0)	$=$	$392.8 \pm$	$1.0 \pm$	3.8	(m_{lq}^{high})
Edge(χ_4^0,$\tilde{\mu}_R$,χ_1^0)	$=$	$257.41 \pm$	$2.3 \pm$	0.3	$(m_{ll}^{\max}(\chi_4^0))$
Edge(χ_4^0,$\tilde{\tau}_L$,χ_1^0)	$=$	$82.993 \pm$	$5.0 \pm$	0.8	$(m_{\tau\tau}^{\max})$
Threshold($\tilde{c}_L$,χ_2^0,$\tilde{\mu}_R$,χ_1^0)	$=$	$211.95 \pm$	$1.6 \pm$	2.0	$(m_{llq}^{\min})$
Threshold($\tilde{b}_1$,χ_2^0,$\tilde{\mu}_R$,χ_1^0)	$=$	$211.95 \pm$	$1.6 \pm$	2.0	$(m_{llb}^{\min})$

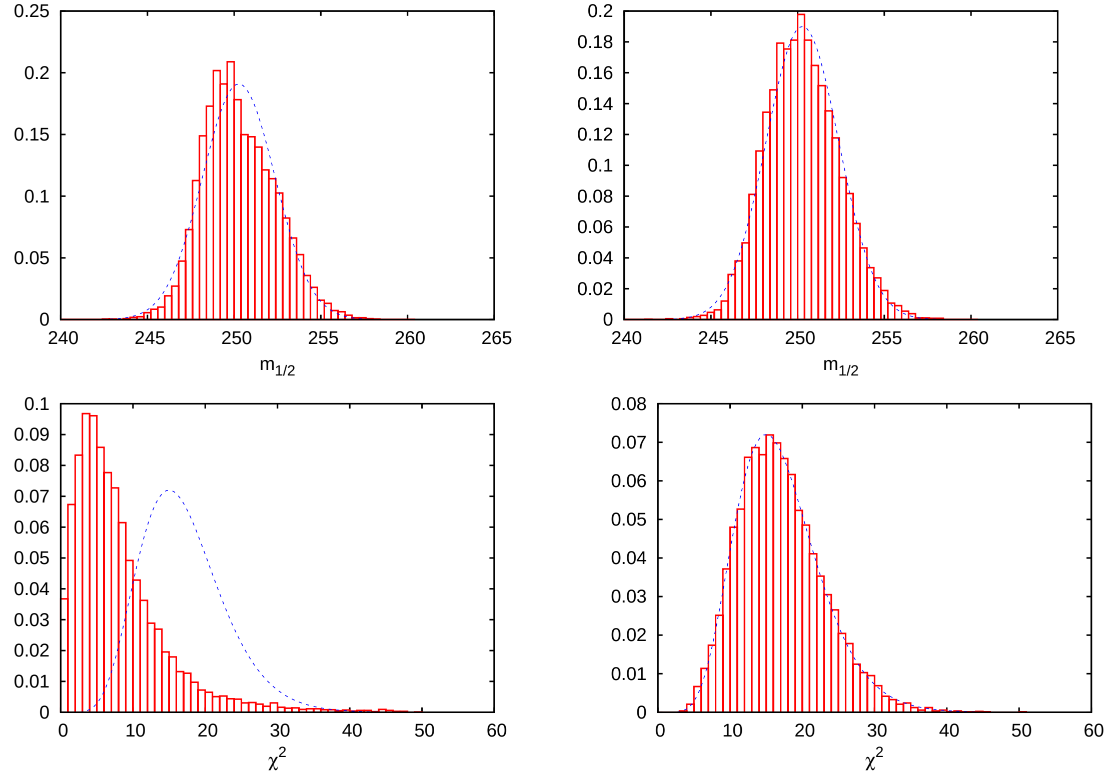
Error determination

Minuit output not usable for flat theory errors:

- Migrad function depends on parabolic approximation
- Cannot determine $\Delta\chi^2$ for Minos to yield 68% CL intervals

$\Rightarrow$ Need more general approach

- Perform $10,000$ toy experiments with measurements smeared around correct value
- Minimise each toy experiment
- Plot resulting distribution of parameter points and fit with Gaussian

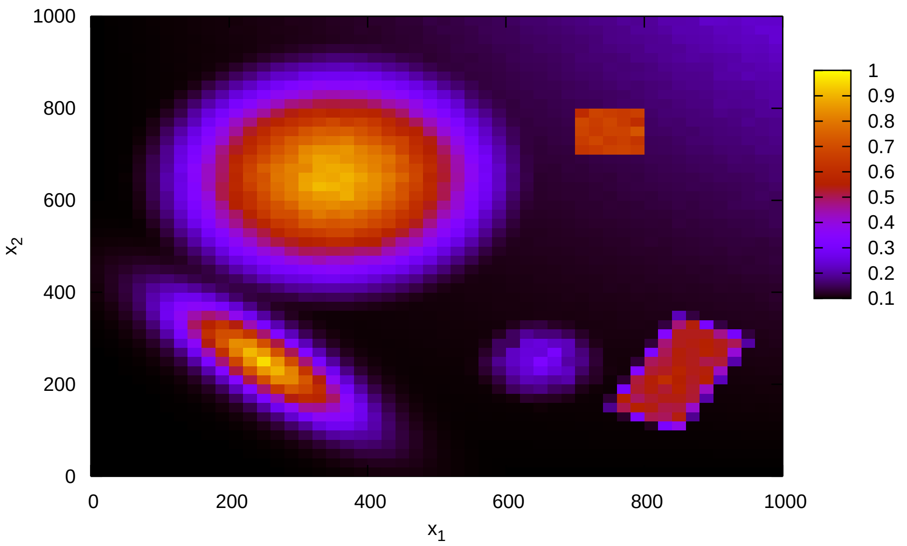


Flat theory errors   Gaussian theory errors

Example

Test function (5-dim):

- Small Hypersphere $r = 100$, $V_{\text{max}} = 75$ @ $(650, 250, 350, 350, 350)$
- Cuboid $d = (173, 120, 200, 200, 200)$, $V_{\text{max}} = 60$ @ $(850, 225, 650, 650, 650)$
- Cube $d = (100, 100, 300, 300, 300)$, $V_{\text{max}} = 25$ @ $(750, 750, 450, 450, 450)$
- Gaussian $\sigma = (50, 150, 150, 150, 150)$, $V_{\text{max}} = 16$ @ $(250, 250, 550, 550, 550)$
- Big Hypersphere $r = 300$, $V_{\text{max}} = 12$ @ $(350, 650, 650, 650, 650)$
- Background $V = 0.1 + 4 \cdot 10^{-30} \cdot x_1^2 x_2^2 x_3^2 x_4^2 x_5^2$




```
1. V=74.929@(655.00,253.72,347.83,348.57,349.59)
2. V=59.972@(850.04,224.99,650.00,649.99,654.56)
3. V=58.219@(849.97,225.01,587.08,650.01,650.02)
4. V=25.110@(750.00,749.99,450.00,450.01,450.01)
5. V=16.042@(245.45,253.44,552.51,542.58,544.75)
6. V=12.116@(350.70,650.40,650.36,650.40,650.38)
7. ...
```

Plot Details

- Parameters: $x_1, \ldots, x_5 \in [0, 1000]$
- Bins: 50×50
- PDF: Breit-Wigner ($\frac{1}{1+\Delta x_i^2/\sigma^2}$) with $\sigma = 100$
- Number of Markov chains: 9
- Number of points per chain: 10^7
- Number of function evaluations: $33,797,153$
- Acceptance ratio: 0.19
- Final r (measure of convergence): 1.815
- CPU time (3 GHz): 150 min