Experiment Simulation

CERN School of Computing 2006 Helsinki

Lecture 1



Francois' programme:



Aatos' programme:



Experiment Simulation

- Motivation for experiment simulation
- Principles of the Monte Carlo Method for experiment simulation
- GEANT4 tool kit implementation of the Monte Carlo Method for experiment simulation
- GEANT4 tool kit user interaction / customization / output
- Introduction to exercises

Slides/Handouts Organisation

- Handouts don't always show the same thing as the slides
 - Some slides have been changed
 - Some slides have been added
- The presented version of the slides can be downloaded from the CSC web

- Handouts don't always show the same thing as the slides
 - Some slides have been changed
 - Some slides have been added

Whenever this symbol appears on the slide:

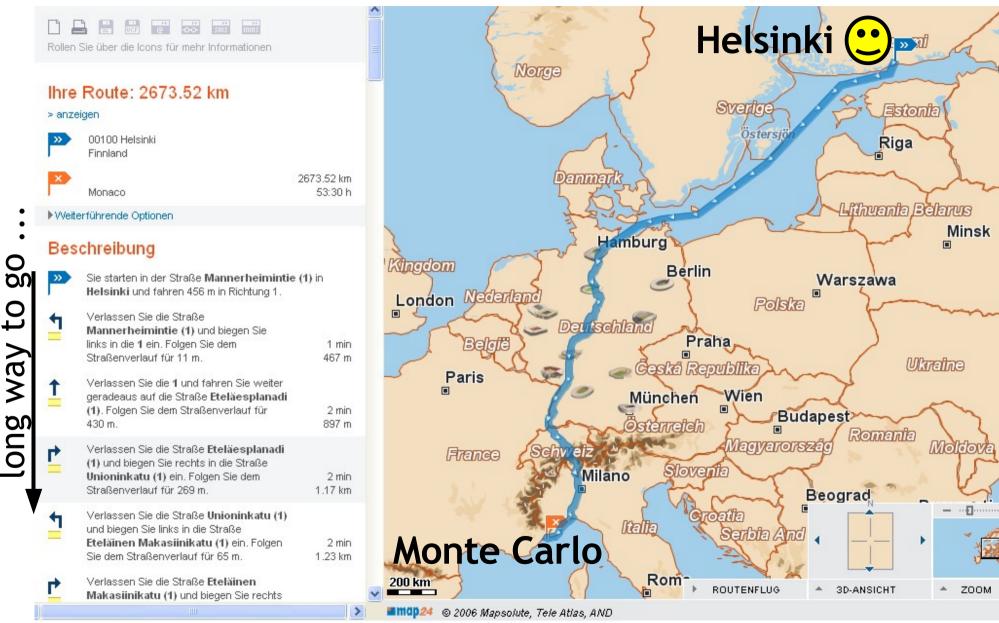


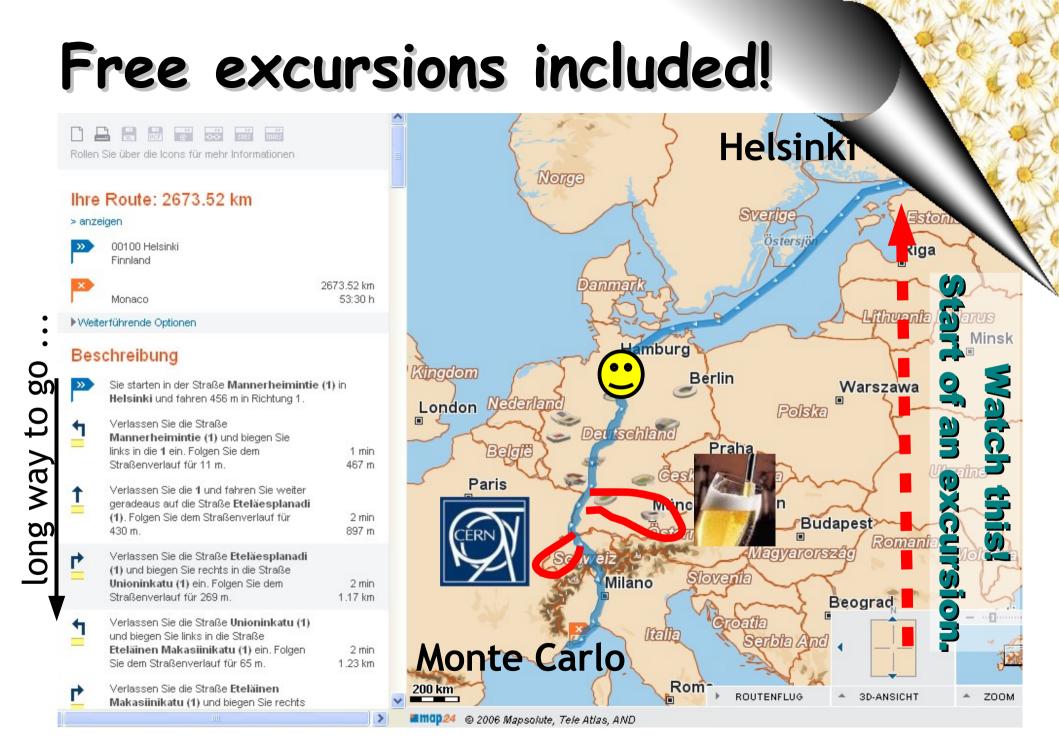
- Handouts don't always show the same thing as the slides
 - Some slides have been changed
 - Some slides have been added

Whenever this symbol appears on the slide:



Lecture Organisation



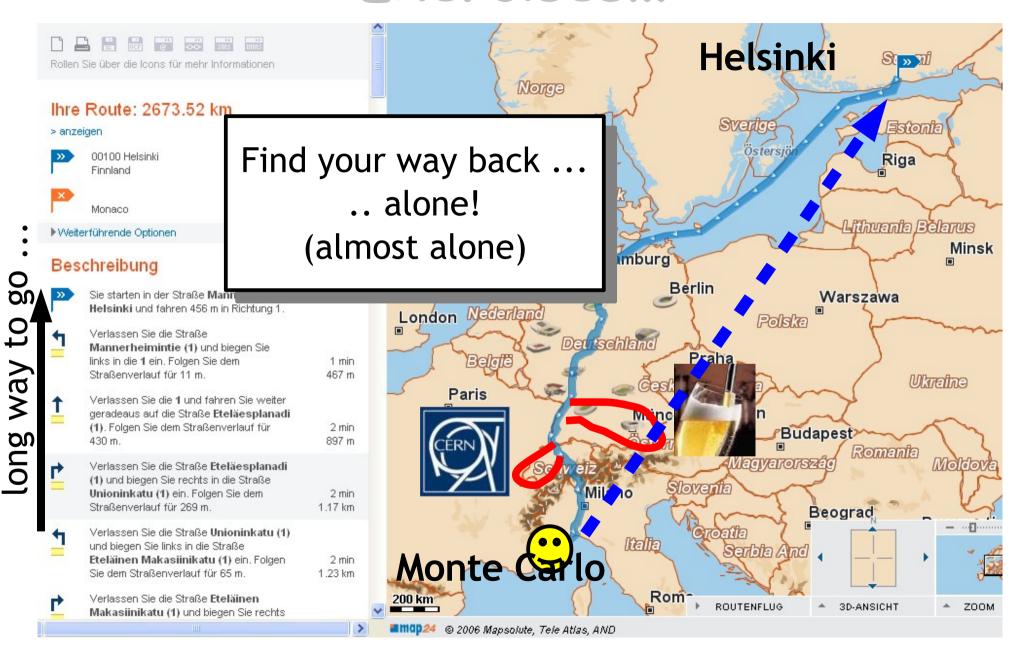


On an excursion ...

- ... we can relax!
- Excursions go into some details
 - related to the main track
 - but not required to follow the main track
- None of the examination questions are based on the material presented in excursions!

Marks the end of the excursion

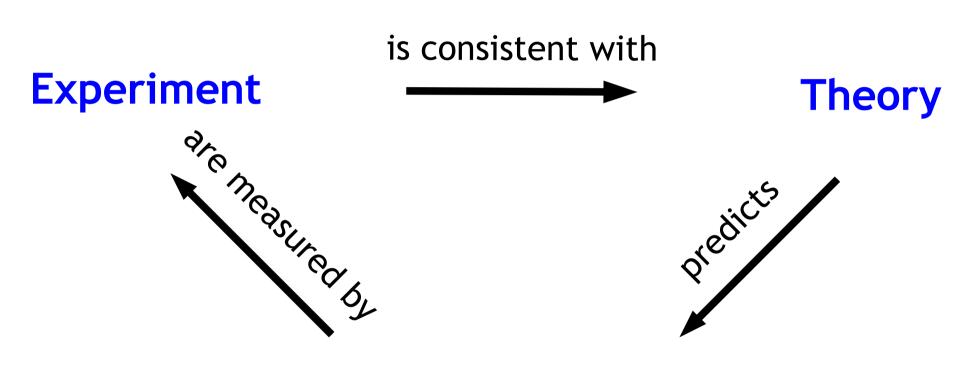
Exercises!!!



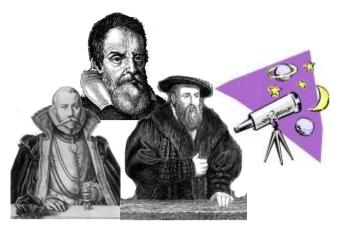
Why simulation?

- Some motivation:
 - from theory to experiment in high energy physics
 - Where does simulation come in?
 - Where does simulation come in?
- "Case study":
 - the Mickey Mouse theory and experiment
 - analogy to high energy physics
- Terminology and short explanation of terms
 - cross section, distributions

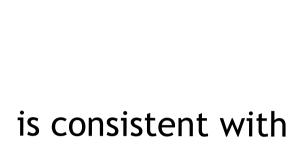
"Ideal" theory & experiment:



Observables



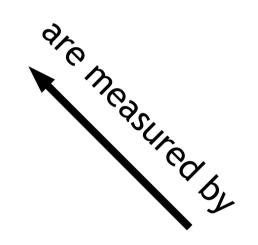
Experiment



 $F=\frac{d}{dt}(mv)$ $|F| = \frac{GM_AM_B}{|r_{AB}|^2}$



Theory

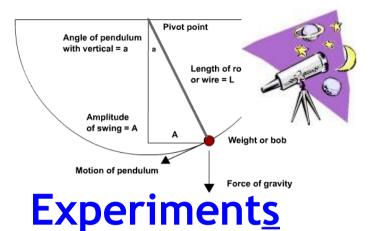




Observables



undetermined parameter!

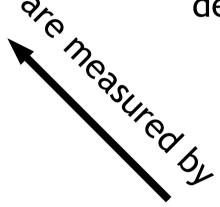


 $|F| = \frac{G M_A M_B}{|r_{AB}|^2}$

are consistent with

Theory

determine parameters





Observables



Quantumtheory & experiment

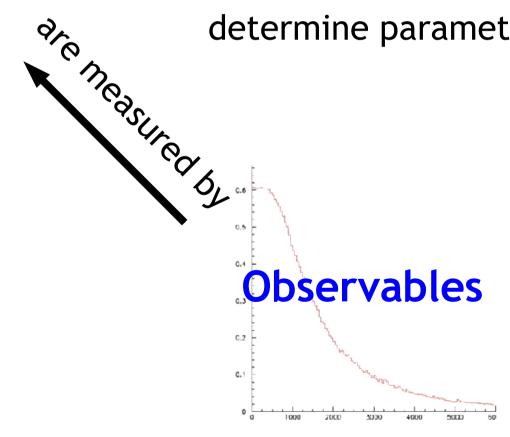
- (1) $H \mid \psi \rangle = i\hbar \frac{\partial}{\partial t} \mid \psi \rangle$
- (2) $\frac{P^2}{2m} + V = -\frac{\hbar^2}{2m}\Delta + V$
- (3)parameters
- (4) $(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0$

are consistent with



Theory

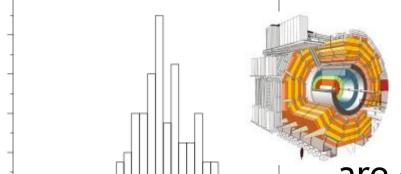
determine parameters



Experiments

microscopic probability distributions

macroscopic distributions



Experiments

 $(1) \quad H \mid \psi > = i\hbar \frac{\partial}{\partial t} \mid \psi >$

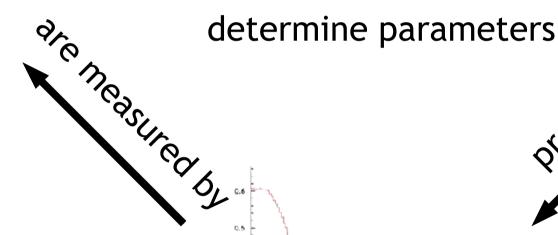
$$(2) \quad \frac{P^2}{2m} + V = -\frac{\hbar^2}{2m}\Delta + V$$

(3) parameters

$$(4) \quad \overline{(i\gamma^{\mu}\partial_{\mu} - m)\psi} = 0$$

Theory

are consistent with



Observables

predicts

microscopic probability distributions

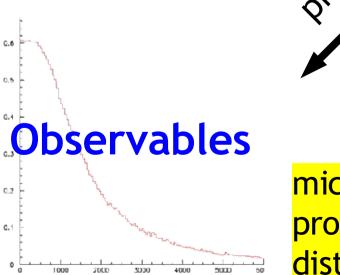
- We usually distinguish between three levels between theory and observables in HEP experiments:
 - fundamental interactions of not directly detectable particles predict distributions of detectable particles

 theory of interactions of detectable particles with atoms/molecules of the detector

- "digitization"

parameters

Theory



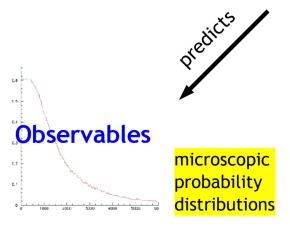
microscopic probability distributions

Level one:

- fundamental interactions of not directly detectable particles predict distributions of detectable particles
- we have one or more theories which needs to be checked thoroughly before we can accept the one or the other
- theories describe the isolated type(s) of interations which we still need to understand
- fundamental parameters unknown

Theory

parameters



Level one:

 fundamental interactions of not directly detectable particles predict distributions of detectable particles

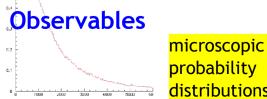
• Level two:

- theory of interaction of detectable particles with atoms/molecules of the detector
- we understand every type of interaction quite well
- these interactions happen in the detector and thus induce the measurement signals

- but, OH MY, there are SOOOOoooooo..... many interactions!!!!

- we have to understand this massive interaction-attack in order to understand our measurements! <mark>parameters</mark>

Theory



Level one:

 fundamental interactions of not directly detectable particles predict distributions of detectable particles

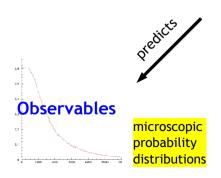
Level two:

 theory of interaction of detectable particles with atoms/molecules of the detector

• Level three:

- known as "digitization", detector response
- conversion of many microscopic particle interactions in the detector to a measurement signal (ADC counts,..)

 Theory
 - measured by specific electronics
 - depends on the type of detector
- NOT IN THIS LECTURES

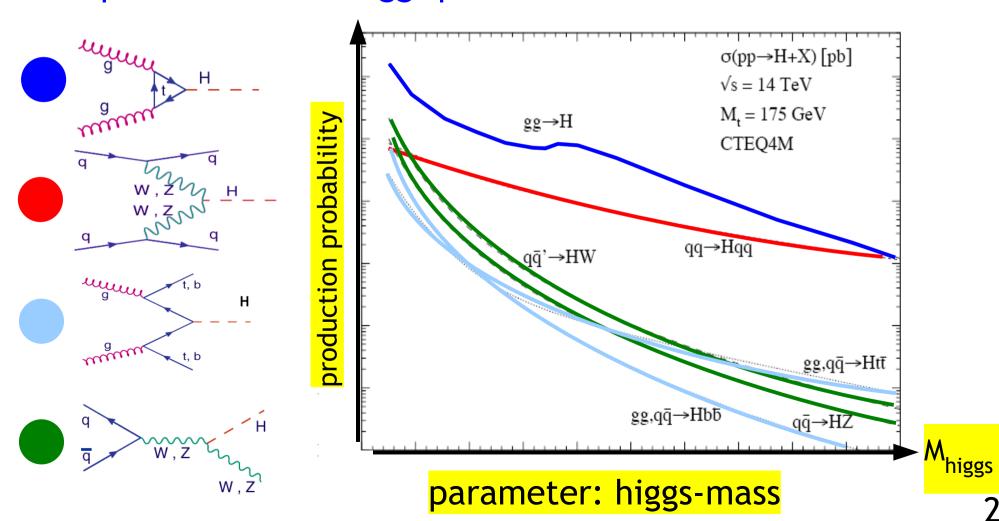


• Level one:

interactions of not directly detectable particles

• Example:

- quark, gluon interactions in the standard model, production of a higgs particle

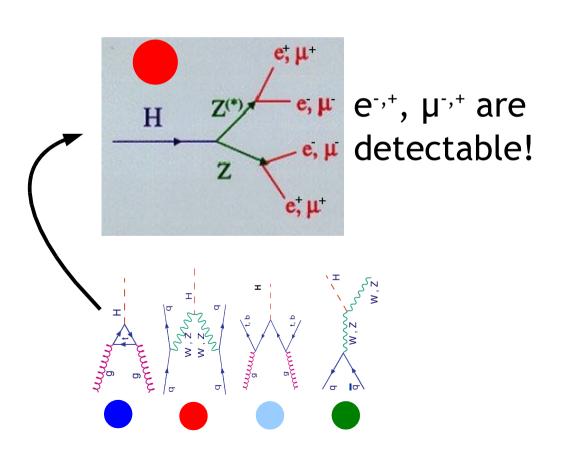


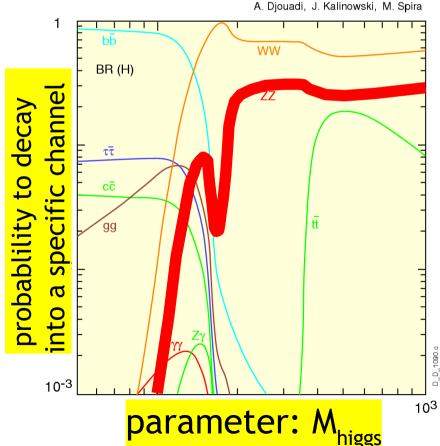
• Level one:

prediction of distributions of detectable particles

Example:

- higgs particle decays into different decay channels
- each channel has its own distribution of the kinematical properties of the end-products



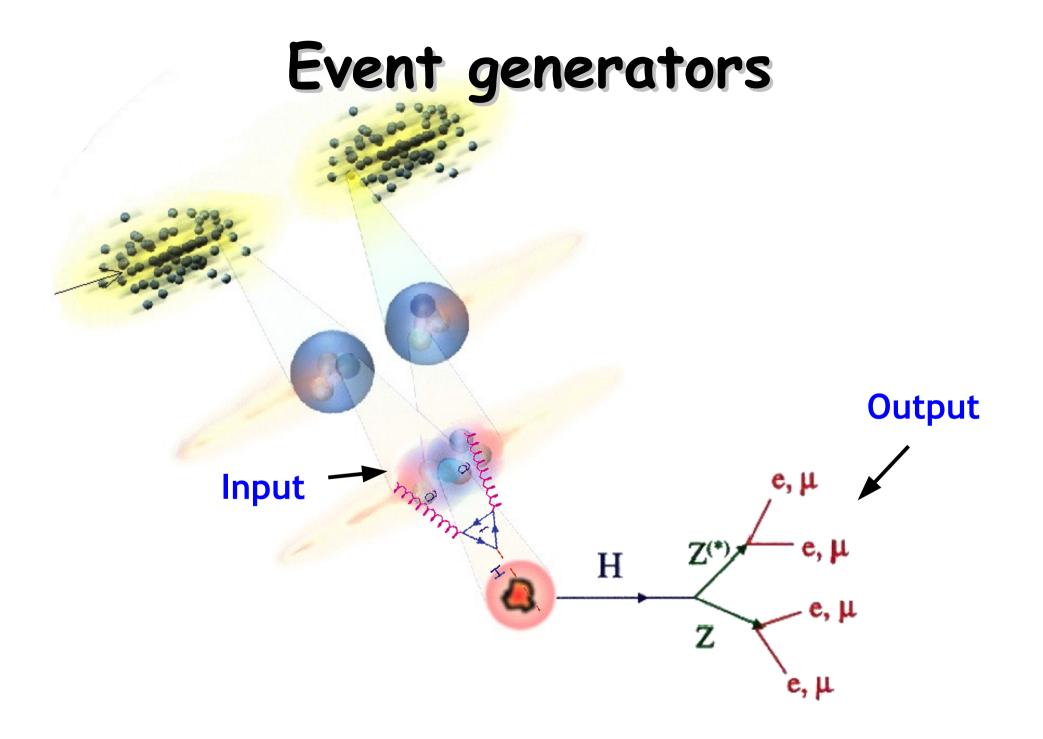


How to know what to look for?

- Standard Model
 - is one theory under test
 - there are others: sypersymmetry, ...
- Higgs particle production & decay
 - is one (quite improbable) process out of many other processes in proton-proton collision
 - all other processes have their own probability distributions
- Higss mass is only one of the undetermined parameters ...
- Its quite impossible to calculate all expected distributions in a deterministic, analytical manner
 - use statistical sampling methods: simulation!
 - generation of "stable particles" from simulated collisions

Event generators

- Event generators are simulation programs
 - to simulate the interaction of fundamental particles
 - up to the "stable" particles resulting from the interaction
- Many different packages available for HEP
 - incorporate one or many fundamental theories
 - sample interactions according to the process distributions based on the quantumtheories
- Adjustable parameters
 - allow you to scan the parameter ranges of yet undetermined parameters of the theories
- Good news: won't cover generators in these lectures!

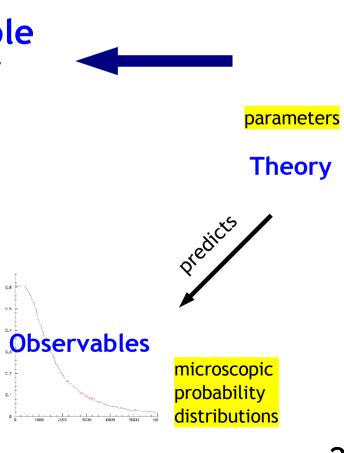


 Have three levels between theory and observables in high energy physics:

 fundamental interactions of not directly detectable particles predict distributions of detectable particles

 theory of interactions of detectable particles with atoms/molecules of the detector

- "dignization"



Rememberz

• Level two:

interactions of detectable particles

• Example:

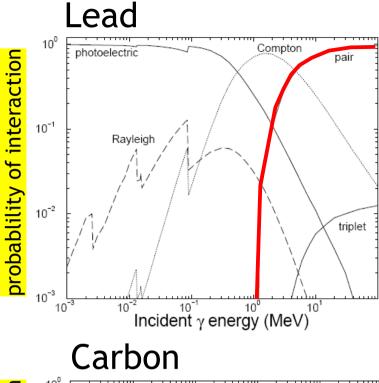
- pair creation

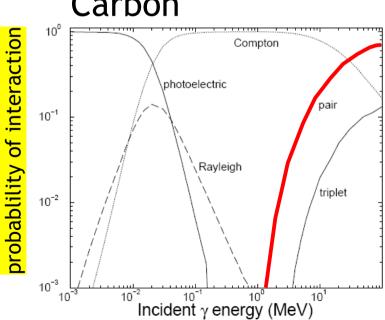
incident photon

atomic nucleus

well known,
no parameters!

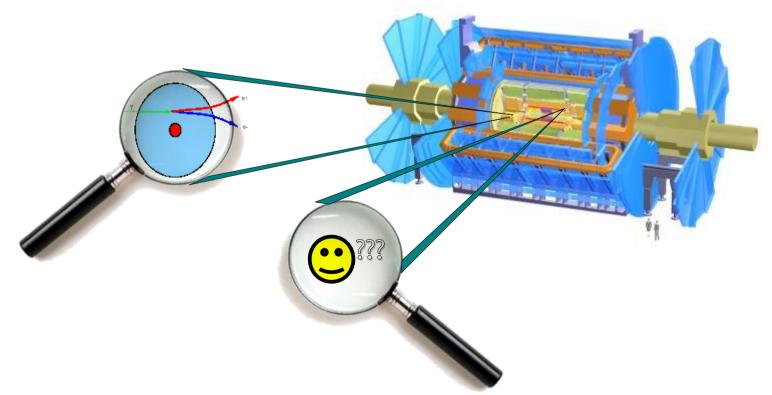
pair



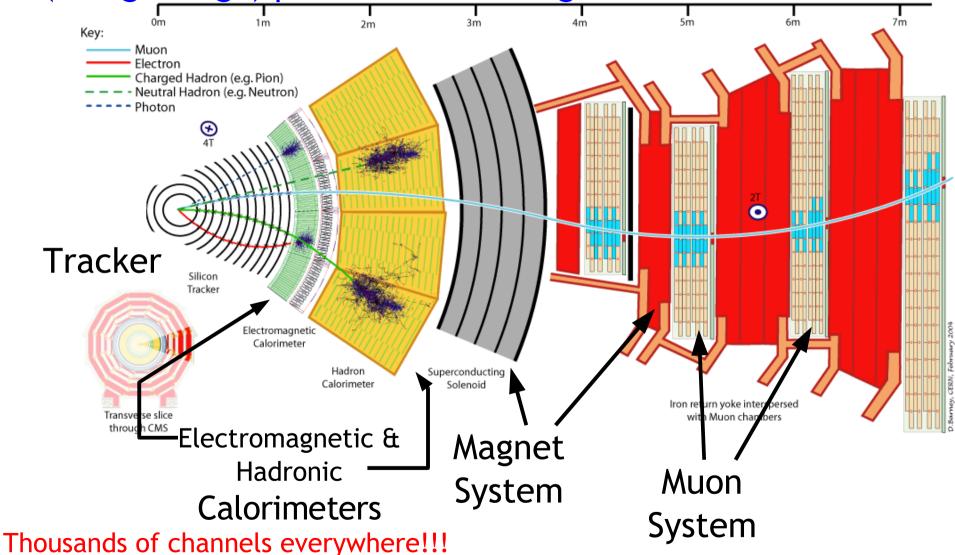


• Level two:

- the problem: how do we get the probability distributions of our particles <u>anywhere</u> in our detector?
- we need to know this distributions in order to understand what the detector has been measuring ...
- especially when there is more than the pair-creation process affecting our particles ...



Onion shell design exploiting the physics processes of ("long living") particles traversing bulk matter



Complex geometry

• Level two:

- the problem: how do we get the probability distributions of our particles <u>anywhere</u> in our detector?
- we need to know this distributions in order to understand what the detector has been measuring ...
- especially when there is more than the pair-creation process affecting our particles ...
- again: it's impossible to calculate these distributions in a analytical / deterministic way

- the solution:

- we simulate the fate of every single particle
- faithfully according to the known theories
- repeat the simulations often enough so that we get good estimates for the required distributions
- => EXPERIMENT SIMULATION, MONTE CARLO METHOD

The Mickey Mouse Case Study

- Simplified view on the topics covered so far
 - to "induce" some feeling for physics for computing students
 - in order to have a common base for the chapter yet to come ...

The Mickey Mouse <u>Material</u>

Somewhere in Disney World ...

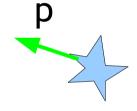
... where everything is 2D, of course:

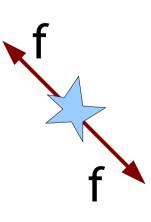
The M-material emits spontaneously

- at a time
- either one Pianissimo (p)
- or one Fortissimo (f)
- or two Fortissimos (f, f)

The physics of the p's is well understood: emission of p's occur at random times into random directions

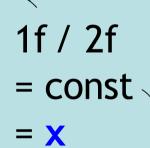






The Mickey Mouse Theory

- Mickey had thought very hard and has a theory: M-Theory
- The M-theory states that
 - f's are emitted into random directions
 - in case of two f's, they go always into opposite directions
 - the fraction of single to double f emission is constant (x), but unknown:
 a parameter in the M-theory
- How big is x, provided the theory is OK?
 - equivalent with: what is the production ratio of single f-events vs. double fevents?





The Mickey Mouse Theory

 Mickey had thought very hard and has a theory: M-Theory

The M-theory states that

- f's are emitted into random directions

Level one: fundamental interactions, undeterminded parameters, e.g. higgs mass

emission is constant (x), but unknown: a parameter in the M-theory

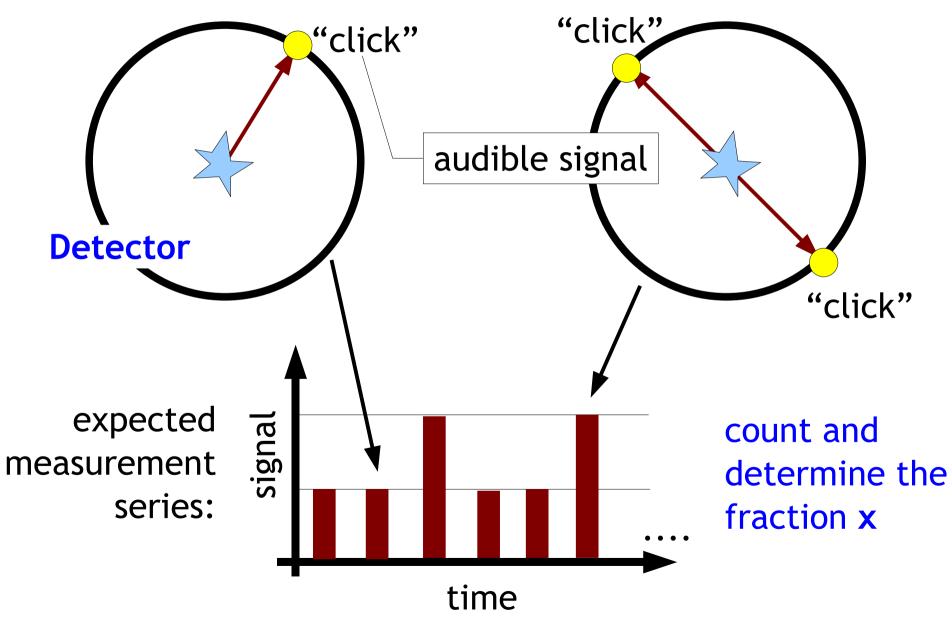
- How big is x, provided the theory is OK?
 - equivalent with: what is the production ratio of single f-events vs. double fevents?

1f / 2f = cons

= X



Mickey's "Gedankenexperiment"



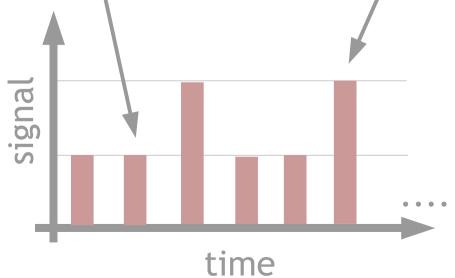
Mickey's "Gedankenexperiment" "click" "click"

Detector

Level two: interactions of particles with the detector

Click"

expected measurement series:



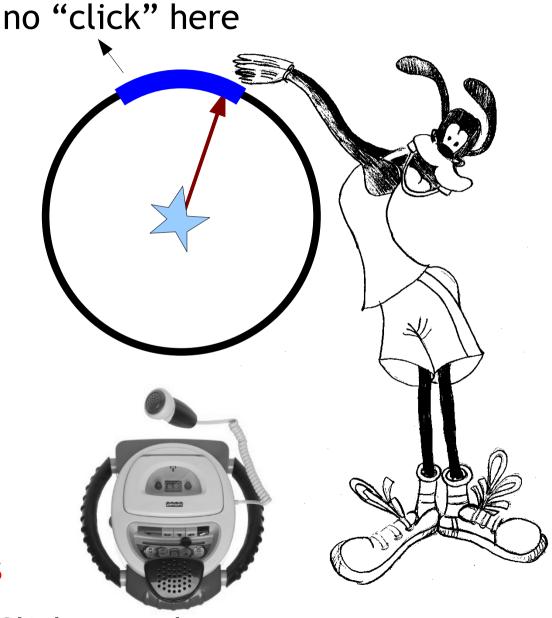
count and determine the fraction x

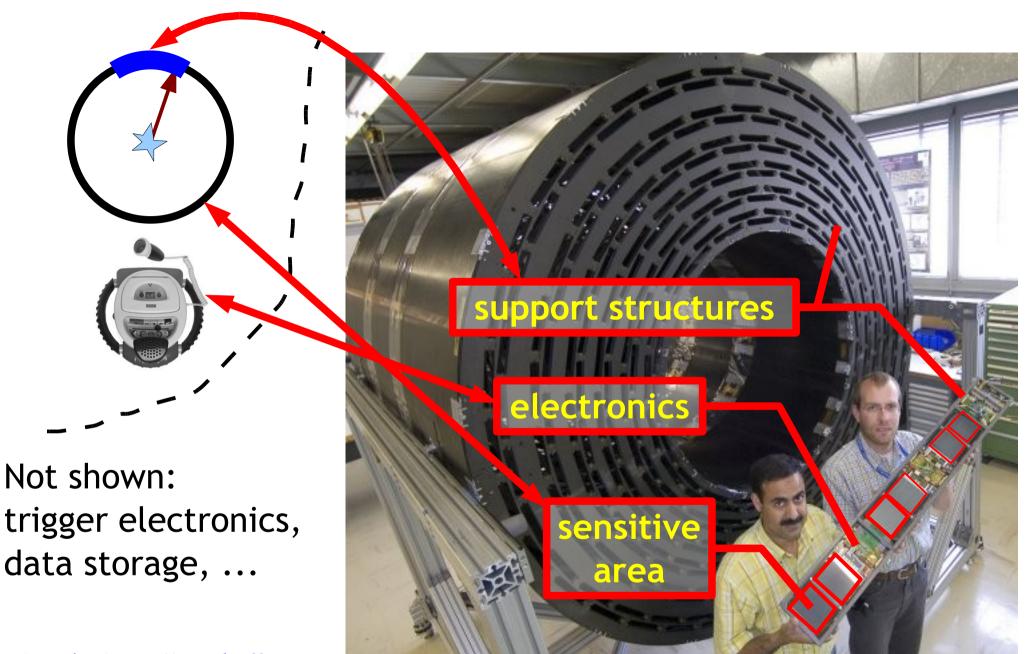
Mickey's experiment

Goofy builds the detector.

He has to use a cooling pipe taking 20% of the surface of the detector's sensitive area

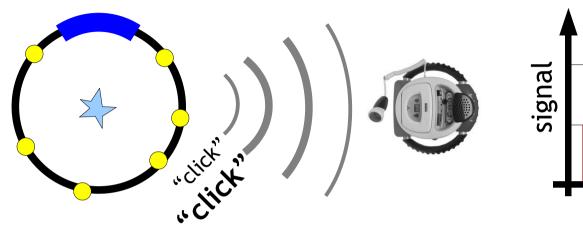
Technical Know How and How To for measuring observables

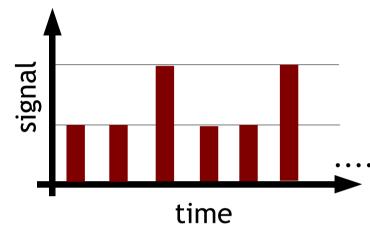


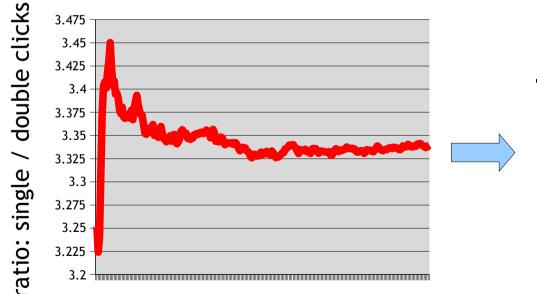


And: it's "only" a part of the whole detector!

Doing the experiment







single clicks

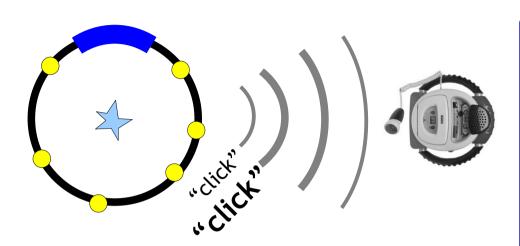
const. = x

double clicks

 $x \sim 3,33$

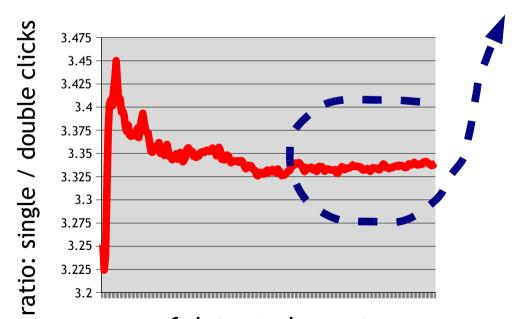
no. of detected events

Analysis:



Ratio between single and double clicks is constant.

=> consistent with
 Mickey's theory.



single clicks

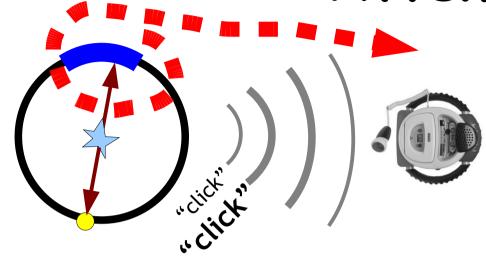
~ const. = x

double clicks

 $x \sim 3,33$

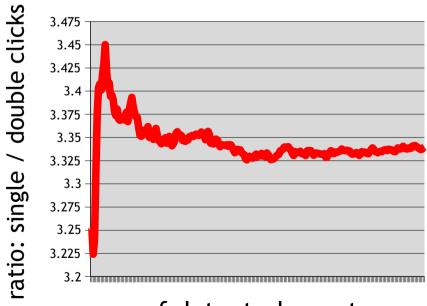
no. of detected events

Attention!!

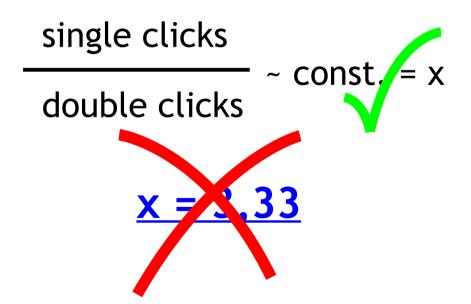


The measured ratio is biased!!

Count too many single click events!







Better analysis

- Our first excursion!!
- Take a closer look on
 - parameter in the M-Theory
 - indirect measurement of the parameter
- Analysis of the measurement
 - analytical method
 - simulation method

Analysis

p .. probability for a single f-event

1-p .. probability for a double f-event

x = p/(1-p).. ratio of single to double

- s.. fraction of sensitive area (4/5 in our case)
- 1-s .. fraction of the cooling pipe (20% = 1/5 in our case)

prob. of detecting a single click, in case of a single f-event:

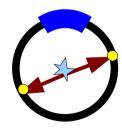
$$P_s = p \cdot s$$



prob. of detecting a double click, in case of a double f-event:

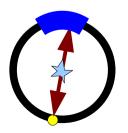
$$P_D = (1-p)\cdot[s - (1-s)]$$

= (1-p)\cdot(2s-1)



prob. of detecting a single click, in case of a double f-event:

$$P_F = (1-p)\cdot 2\cdot (1-s)$$



prob. of detecting a single click, in case of a <u>single</u> f-event:

$$P_s = p \cdot s$$



prob. of detecting a double click, in case of a <u>double</u> f-event:

$$P_D = (1-p)\cdot[s - (1-s)]$$

= (1-p)\cdot(2s-1)



prob. of detecting a single click, in case of a double f-event:

$$P_F = (1-p)\cdot 2\cdot (1-s)$$

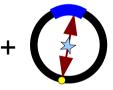


But what we measure is:

prob. of detecting a single click, in case of a <u>single or double</u> f-event:

$$\underline{P}_{S} = P_{S} + P_{F}$$





measured ratio single/double:

$$\underline{x} = \underline{P}_{S} / P_{D} = F(p), p = F^{-1}(\underline{x})$$

$$p = \frac{\underline{x} \cdot (2s-1) - 2 + 2s}{\underline{x} \cdot (2s-1) - 2 + 3s} \longrightarrow x = \frac{p}{1 - p} = 1/s \cdot [\underline{x} \cdot (2s-1) - 2 + 2s]$$

$$\times$$
 x ~ 3.33, x ~ 2., p ~ 0.66 ~ 2./3. for Mickey's setup

Simulation of Mickey's experiment

Analytical solution:

p.. probability for a single f-event

1-p .. probability for a double f-event

x .. ratio of measured single / double f-events

s .. fraction of sensitive detector

x .. ratio of single / double production rate, undetermined in the theory, i.e. p is undertermined.

$$x = \frac{p}{1 - p} = 1/s \cdot [\underline{x} \cdot (2s-1) - 2 + 2s]$$

Simulation

- p -> p_i in [0, 1/N, 2/N, ..., 1]
- for each p_i, generate M events: ~ p_i single f, (1-p_i) double f
- for each event, sample a random direction
- for each event, check, if the direction hits the cooling or not, and count the "clicks" accordingly
- thus, for each p_i an x_i has been simulated
- correlate p_i and \underline{x}_i to find the correction factor for the real experiment

Simulation of Mickey's experiment

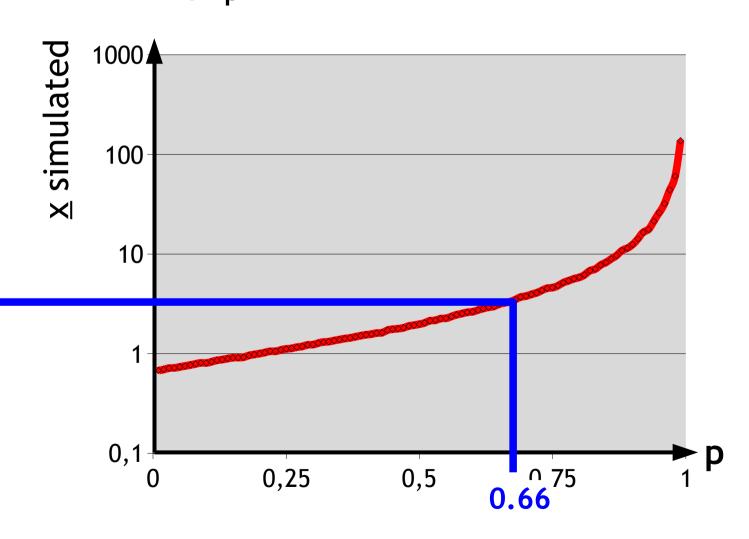
Analytical:

$$x = \frac{P}{1 - D} = 1/s \cdot [\underline{x} \cdot (2s-1) - 2 + 2s]$$

Simulation:

measured

 $x \sim 3.33$



- Already very simple setups are difficult to treat in an analytic / deterministic way!
 - rather complex expression for the correction factor!
- In HEP, theories, detectors, and analysis procedures are A LOT MORE complex!!
 - we need to apply other methods to understand the measurements of our detector in order to draw conclusions concerning the underlying physics
 - the Monte Carlo method is a required tool, whenever the analytical solution can't be given easily

Example: acceptance, efficiency

Acceptance a:

N ... number of events of a given type (e.g. single f emission)

N_d .. number of detected events of the same type

$$< N_d > = a < N >$$

Acceptance relates the avarage number of detected events of a given type with the avarage occurance of this type.

Detection Efficiency $\varepsilon(x)$:= probability of an event x to be detected if it has taken place

x.. physical variables (positions, momenta, ..) f(x).. distribution density of x (from physics-theory)

$$a = \int \epsilon(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

Mickey's case

prob. of detecting a single click, in case of a single f-event:

 $P_s = p \cdot s$

 $= p \cdot a$



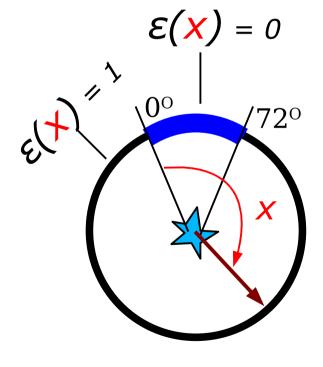
- p.. probability of emitting a single f-particle
- s.. fraction of sensitive detector area

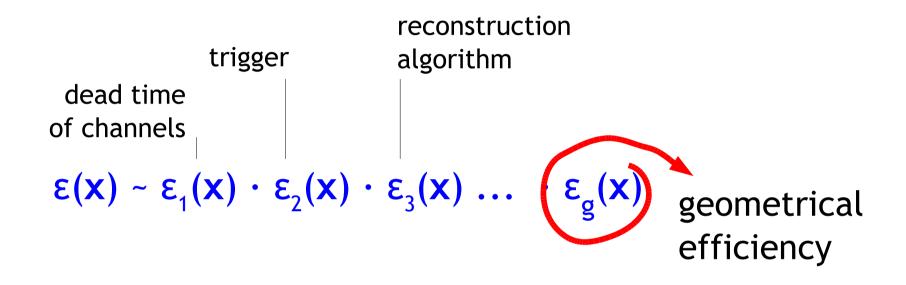
x.. angle of emission of an f-particle

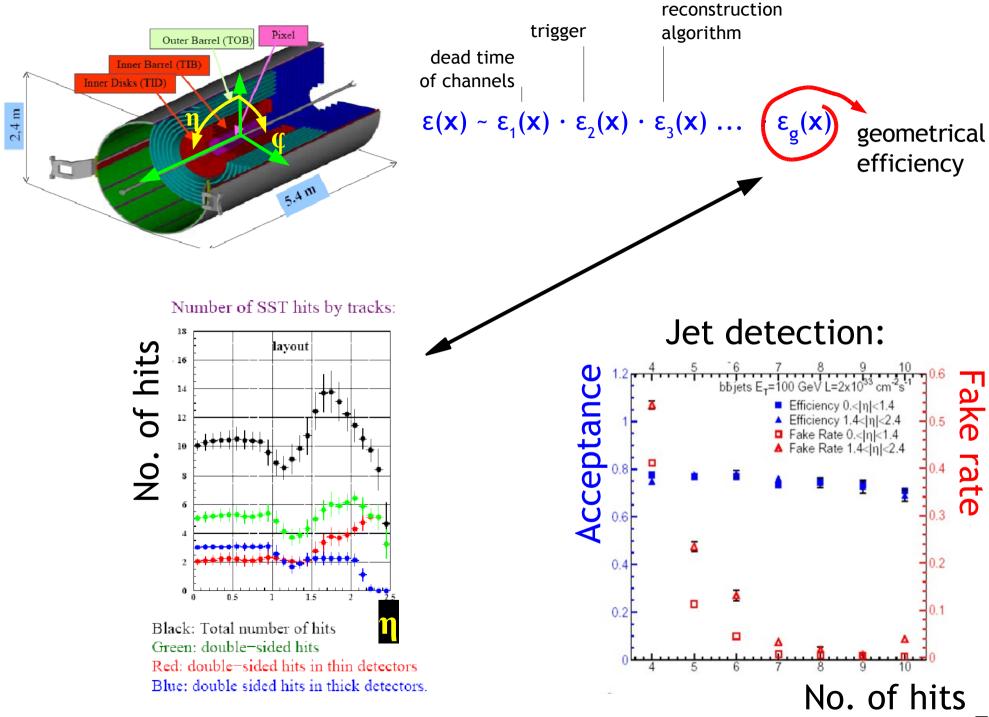
= $\mathbf{p} \cdot \int \epsilon(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$

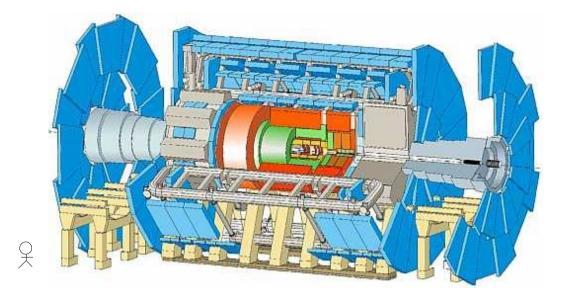
- f(x) .. distribution of x: all directions are equally probable: $f(x) = 1/360^{\circ}$
- $\varepsilon(x)$.. efficiency here only determined by the geometry:

$$\varepsilon(x) = \begin{cases} 0 \text{ for x in } [0^{\circ}, 72^{\circ}) \\ 1 \text{ for x in } [72^{\circ}, 360^{\circ}) \end{cases}$$









Generally, efficiency depends on many things:

- money
- geometry
- choice of sub-detectors (physical properties, deterioration, ..)
- event type
- trigger efficiencies
- reconstruction & analysis algorithms / SW & computing

How can we design & built a detector "efficiently" enough to measure what we want to measure?

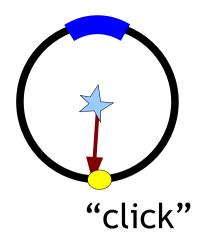
=> <u>simulation studies</u> contribute significantly to the taken decisions

There's more to understand ..

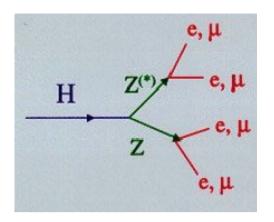
- Efficiency, acceptance described before cover
 - ony one aspect in an HEP experiment!!
 - only one type of event (e.g. emission of one f-particle)
 - don't tell you anything about other events
 - that you already understand very well
 - that you haven't thought of, yet ...
 - that you are not interested in measuring
 - that bias other measurements
- Some aspects, where **simulation** is extensively used to study and understand them
 - signal, background
 - noise, min. bias, ...

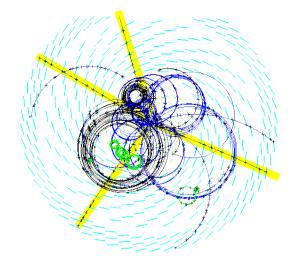
What we measure simultaneously

```
measured events
looking like events
                                                  from events of type E
         of type E
           <N<sub>unwanted</sub>> = <N<sub>min.bias</sub>> —— from other events
                           + <N<sub>noise</sub>> —— from somewhere else
  simultaneously
        detected
    other events
```

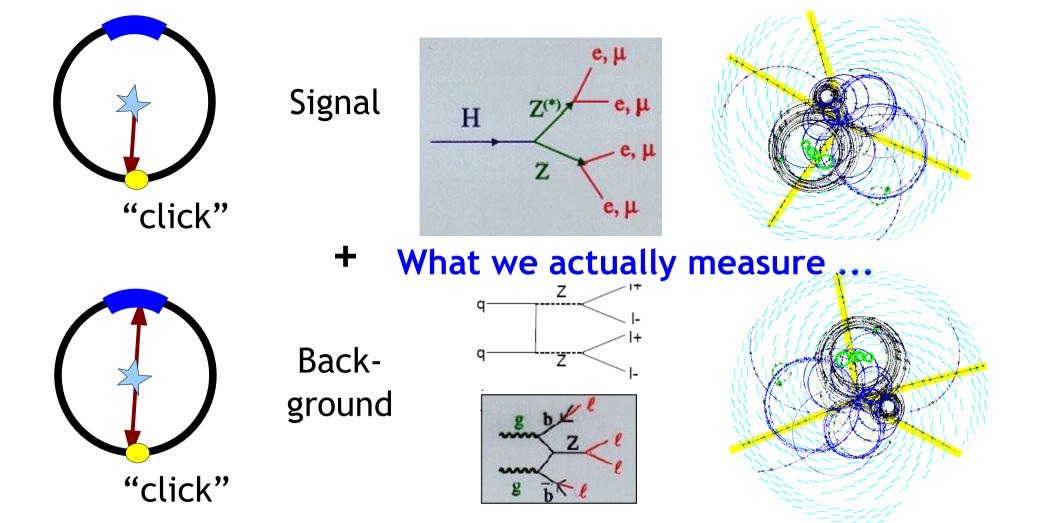


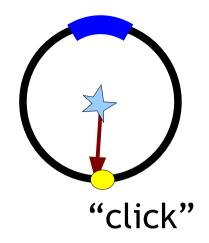
Signal



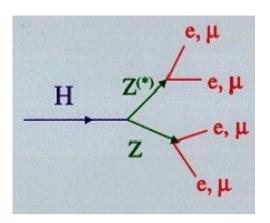


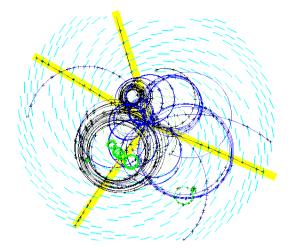
What we wish to measure!





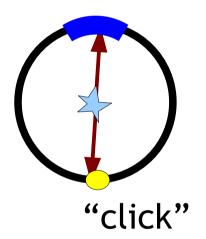
Signal



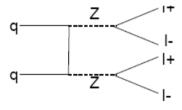


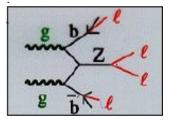
+

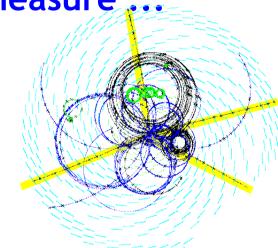
What we actually measure



Background



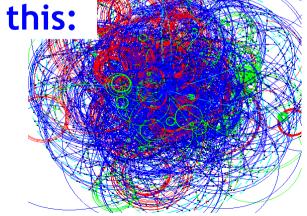




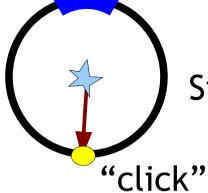
... if we manage to ignore this:

Mickey was lucky!

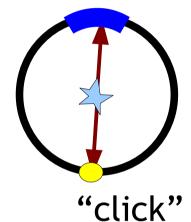
= minimal bias events and noise data



The four types of data

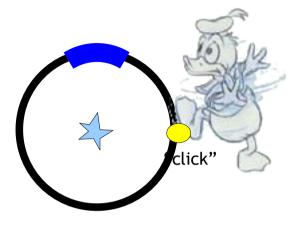


Signal

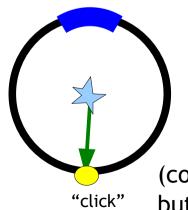


Background



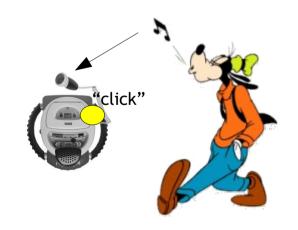


(comes from somewhere else)

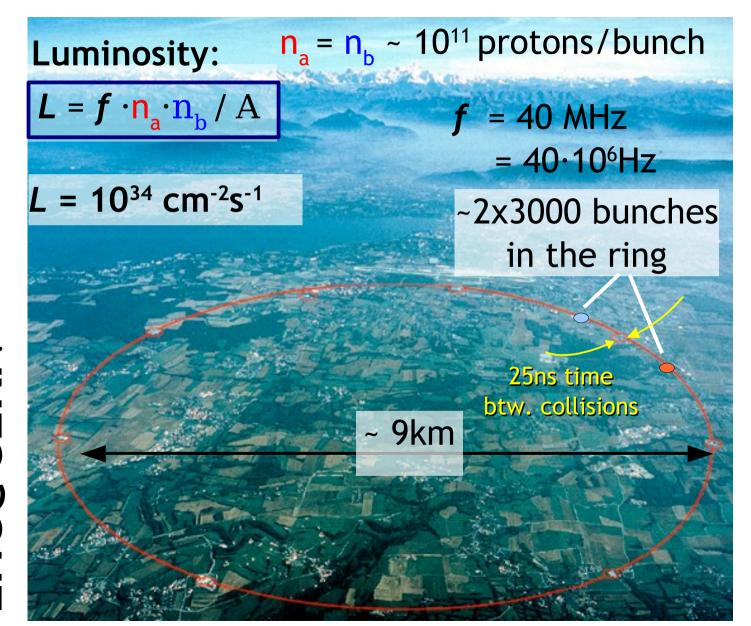


Minimal bias emission of a "Pianissimon"

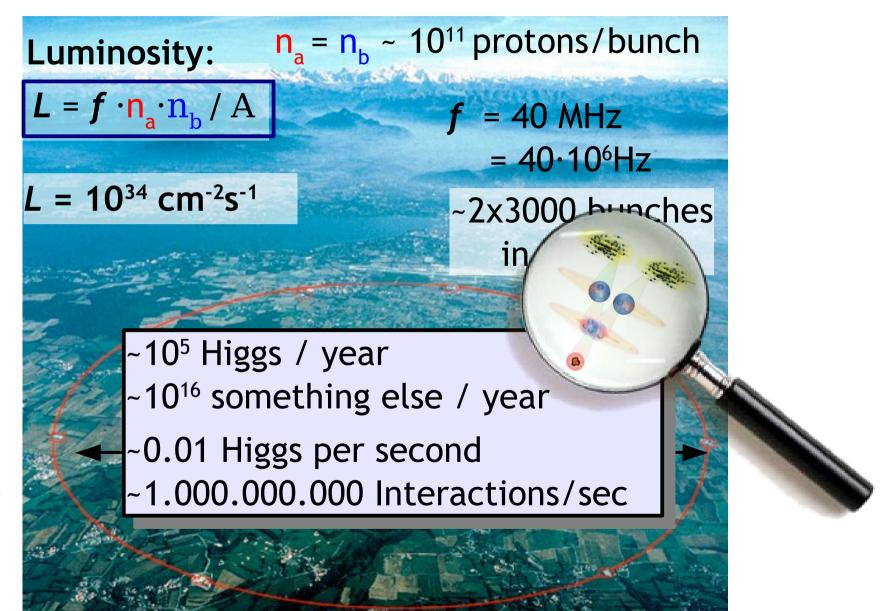
(comes from the interaction, but is not signal nor background)



Signal to noise, higgs case



Signal to noise, higgs case



LHC@CERN

Experiment = truffles pig

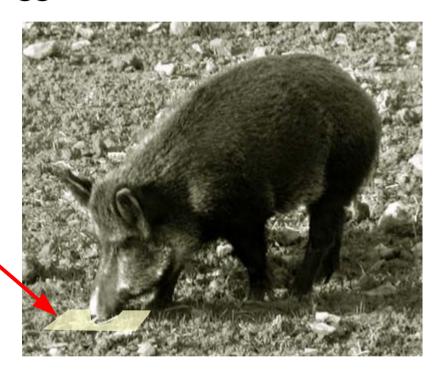
~10⁵ Higgs / year

→ 1 Higgs event in 10¹¹ events

~10¹⁶ something else / year

1 event .. $1dm^2 = 10 \times 10 \text{ cm}^2$

10¹¹ events .. 10^{11} dm² ~ $3 \cdot 10^{5}$ x $3 \cdot 10^{5}$ dm²



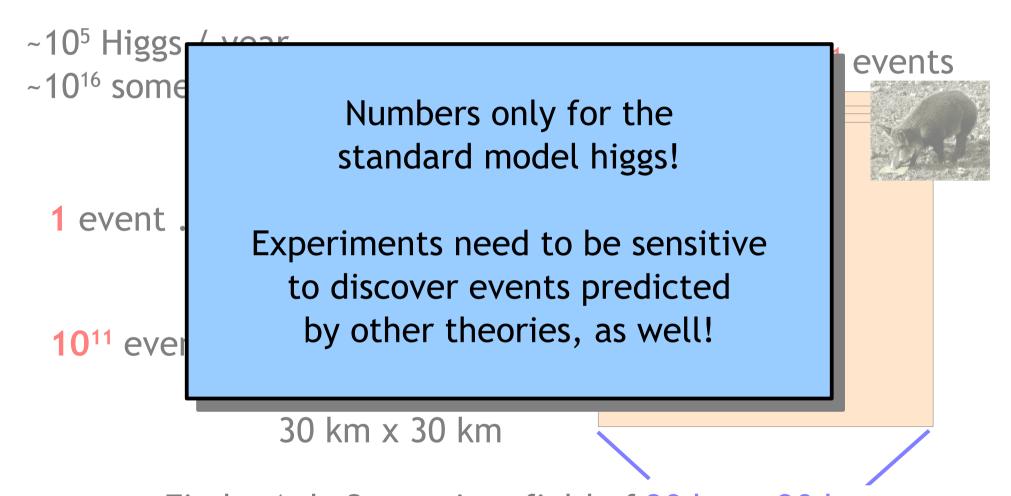
... just to visualize these numbers ...

Experiment = truffles pig

~10⁵ Higgs / year ►1 Higgs event in 10¹¹ events ~10¹⁶ something else / year 1 event .. $1dm^2 = 10 \times 10 \text{ cm}^2$ 10^{11} events .. 10^{11} dm² ~ $3.10^5 \text{ x } 3.10^5 \text{ dm}^2 =$ 30 km x 30 km

Find a 1 dm2 area in a field of 30 km x 30 km in not more than 1.5 min!!

Experiment = truffles pig



Find a 1 dm² area in a field of 30 km x 30 km in not more than 1.5 min!!

Let's summarize:

Theory

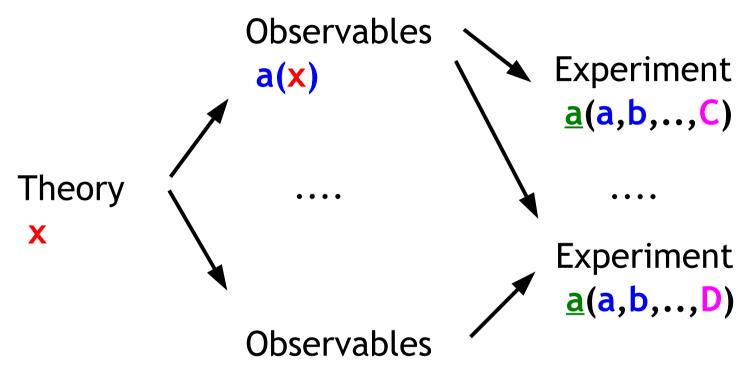
X

unknown
values for
parameters x
e.g. the mass of the Higgs

change in parameters, change in prediction of observables Observable **a(x)** Theory

Observable unknown b(x) values for parameters x e.g. the mass of the Higgs

change in parameters, change in prediction of observables

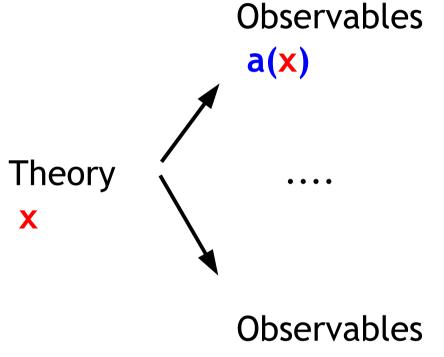


unknown b(x)
values for
parameters x
e.g. the mass of the Higgs

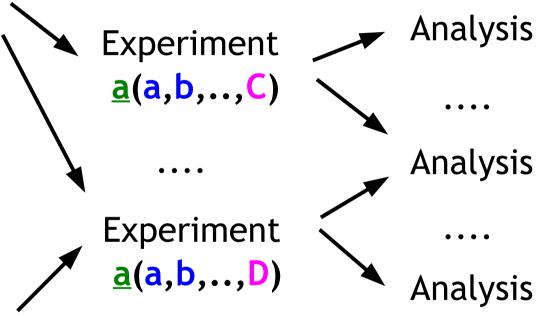
machine conditions C,D, efficiency, acceptance,.. measured observables <u>a</u>, <u>b</u>, ..

change in parameters, change in prediction of observables

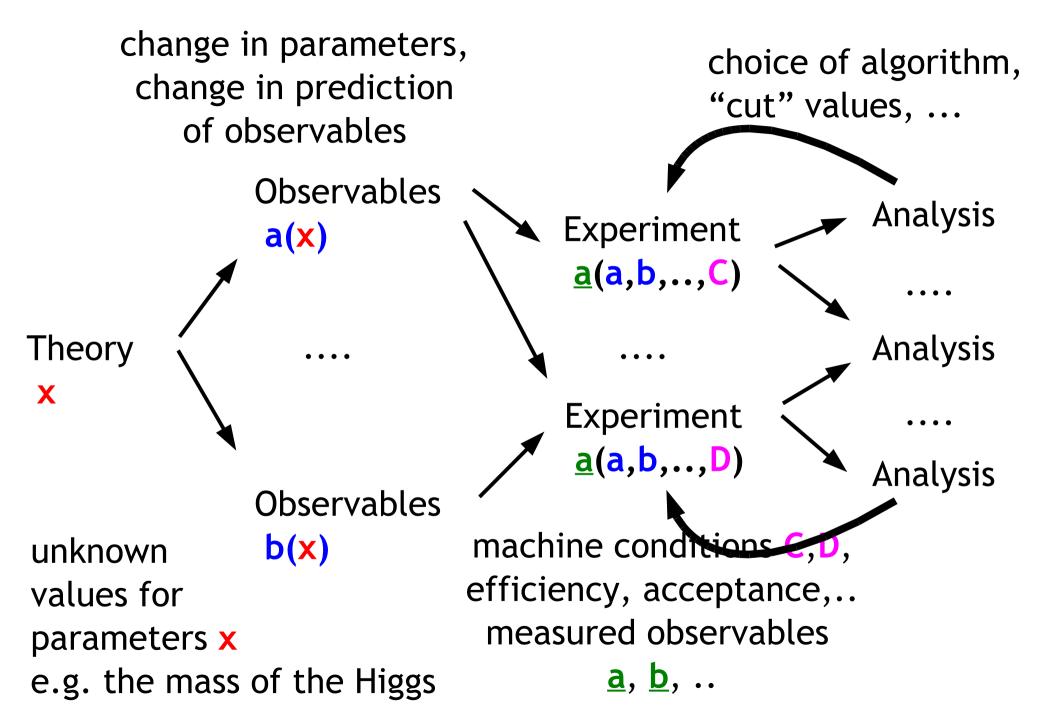
choice of algorithm, "cut" values, ...

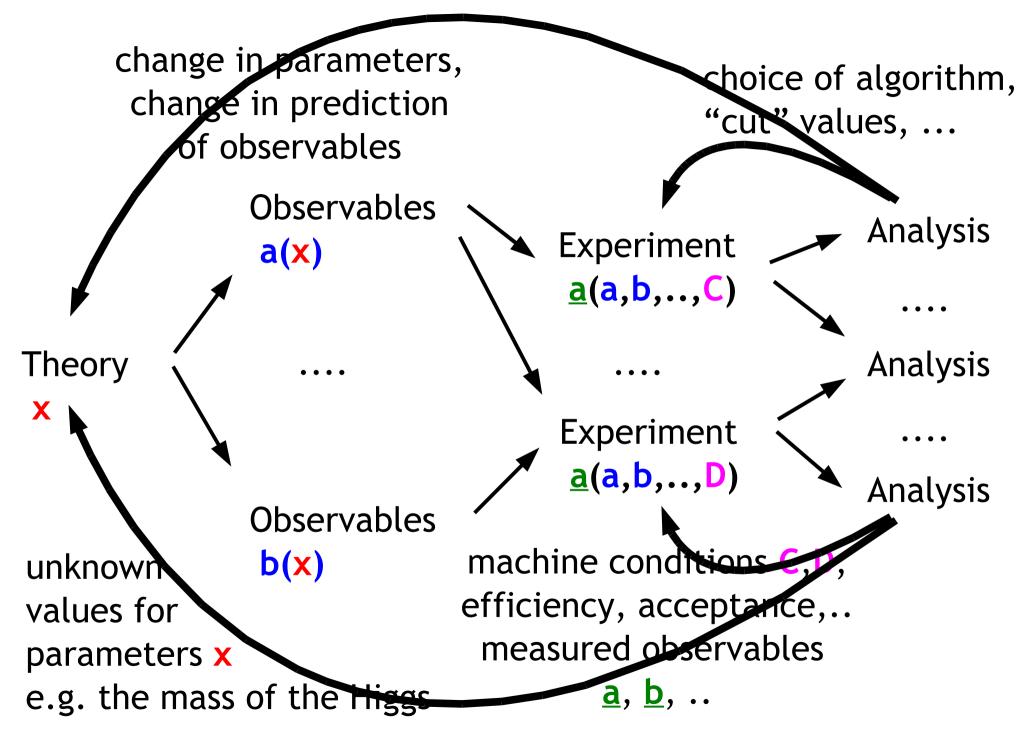


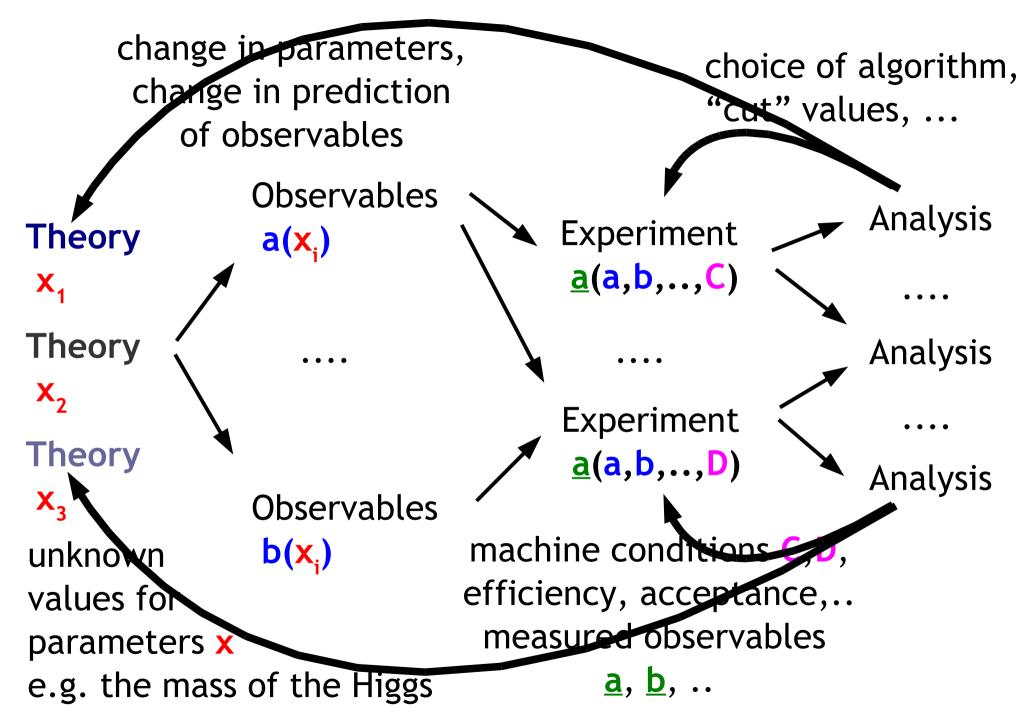
unknown b(x)
values for
parameters x
e.g. the mass of the Higgs



machine conditions C,D, efficiency, acceptance,.. measured observables $\underline{a},\underline{b},..$







change in parameters, change in prediction

choice of algorithm,

The need for simulation:

Because of the tremendous multiplicity of possible parameter constellations, it is impossible to design, operate, and "understand" today's HEP experiments without having corresponding simulation programs capable of "scanning" the relevant parameter ranges!

unkr

The

The

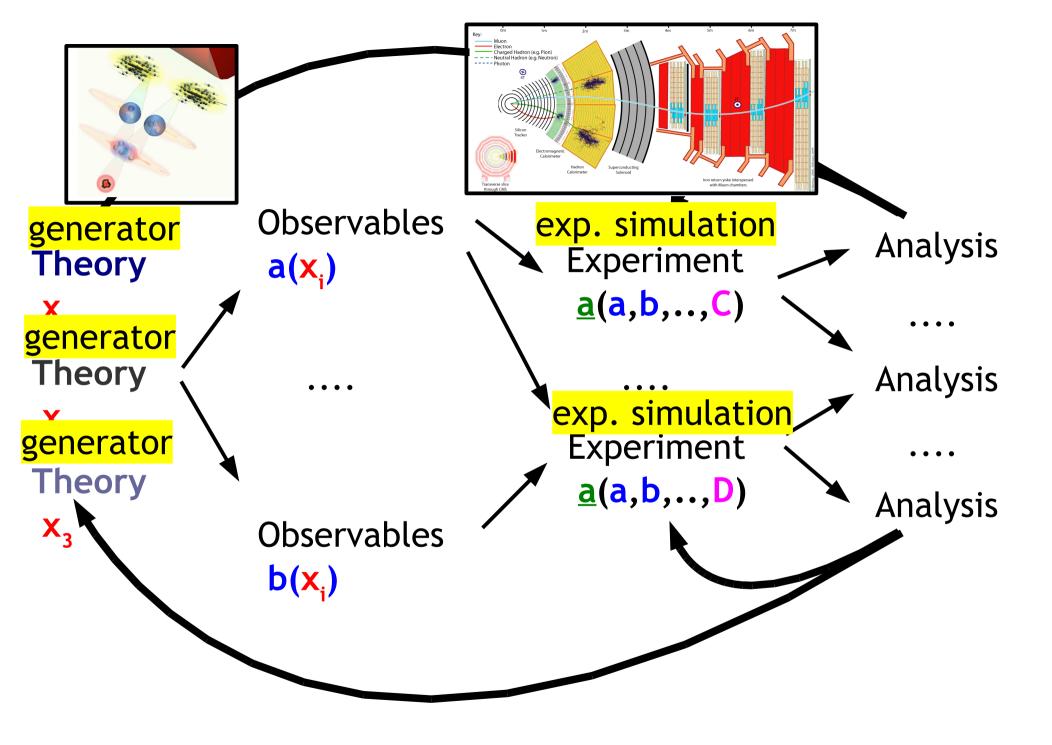
The

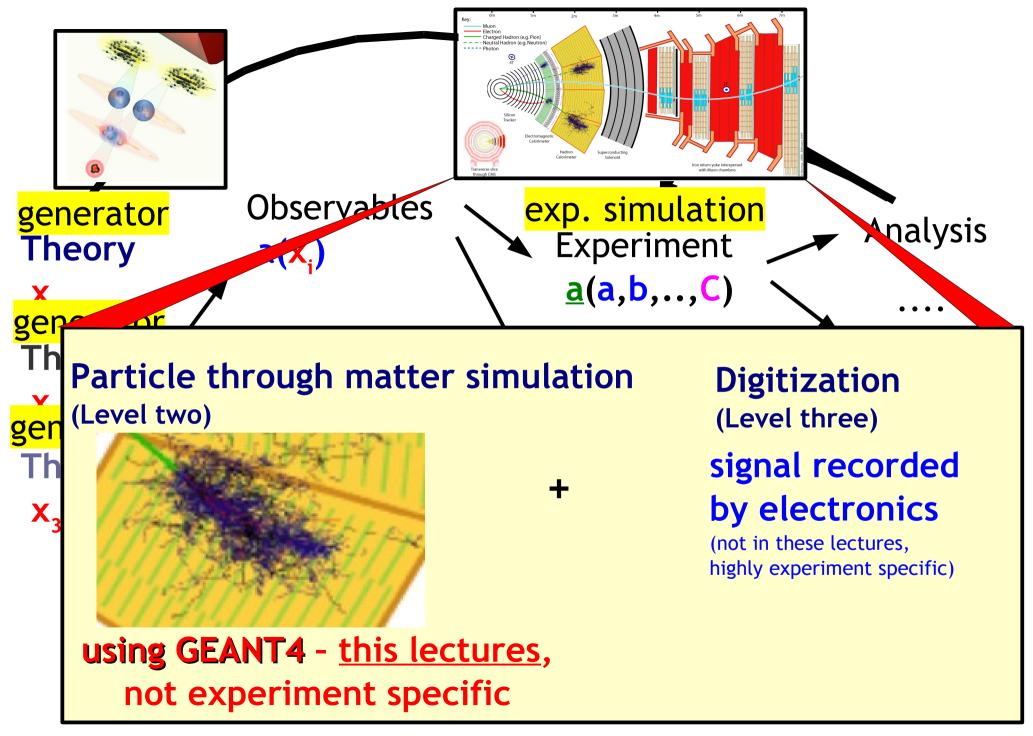
values for parameters x e.g. the mass of the Higgs

efficiency, acceptance,.. measured observables a, b, .. sis

sis

SÍS





GEANT4

"Geant4 is a toolkit for <u>simulating the passage of</u> <u>particles through matter</u>. It includes a complete range of functionality including <u>tracking</u>, <u>geometry</u>, <u>physics</u>

models and hits." [1]

[1] NIM A506 (2003), 250-303

GEANT comes from **GE**ometry **ANd** Tracking.

History of GEANT goes back to the 1970s (CERN)

Homepage: http://www.cern.ch/geant4

1994-1998: R&D phase, ~100 scientists from >10

experiments world wide;

First production release in 1998; Today's (2006) release: GEANT4.8.x => more than 10 years of work!!

Areas of application: <u>high energy physics</u>, medical applications, space science



Summmary ~1~

Need of simulation

- understanding
 - physics theories
 - experiments
 - analysis of measurements
- complexity of physics: generators
- complexity of detectors:
 - passage of particles through matter
 - response simulation ("digitization")

Simulation vs. analytical treatment

Mickey Mouse example

Types of data:

- signal and background
- minimum bias events
- noise