The Logic of an EPP experiment

Go back to Rutherford and the logical steps of his experiment

Key elements in the Rutherford experiment – physical quantities

- **Energy of the collision** (driven by the kinetic energy of the α particles) the meaning of $\sqrt{s}$
- **Beam Intensity** (how many α particles /s)
- **Size and density of the target** (how many gold nuclei encountered by the α particles);
- **Deflection angle θ**
- **Probability/frequency of a given final state** (fraction of α particles scattered at an angle θ);
- **Detector efficiency** (are all scattered α particles detected?); includes acceptance (geometrical acceptance…).
- **Detector resolution** (how good θ angle is measured?)

PREPARATION, OBSERVABLES, INSTRUMENTAL EFFECTS

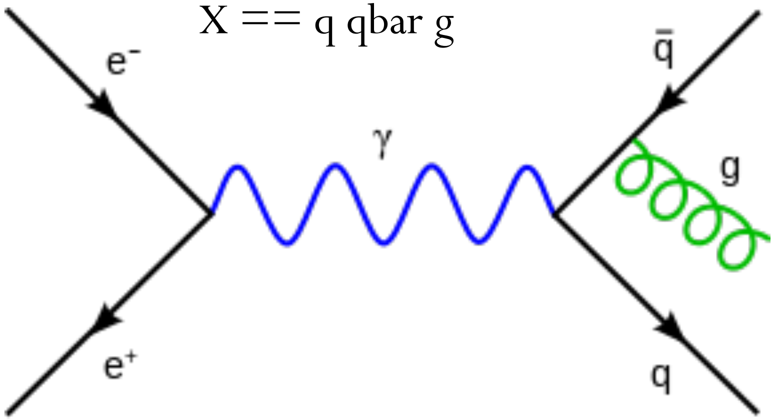
"Logic" of an EPP experiment - I

- Collision or decay: → **process to look at**
  - Initial state (proj. + target) OR (decaying particle);
  - Final state X = all particles produced
- Quadri-momentum conservation should always be at work
- In principle there is no need to measure ALL final state particles: a final state could be: → $\mu^+\mu^-$ + X ("inclusive" search)
- Possible final states:
  - a + b → a + b : elastic collision (e.g. pp→ pp)
  - a + b → X : inelastic collision (e.g. pp→ppπ^0)
- The experimentalist should set-up an experimental procedure to select the final state he/she searches. First of all he should be able **to count the number N_X of final states X**.

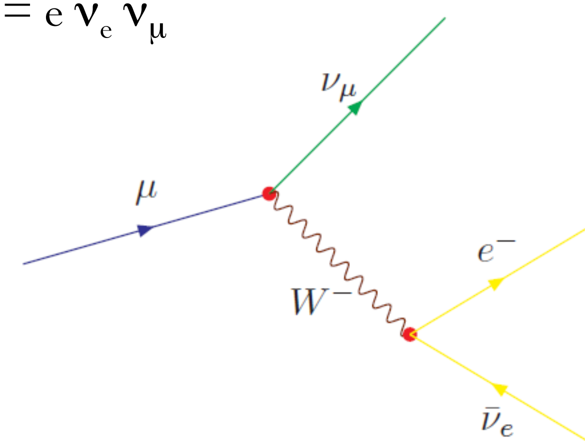
Why count ? – I

- Why count ?
- Because QFT based models allow to predict quantities (like ***cross-sections***, ***decay widths*** and ***branching ratios***, see later) that are proportional to "***how probable is***" a given final state.

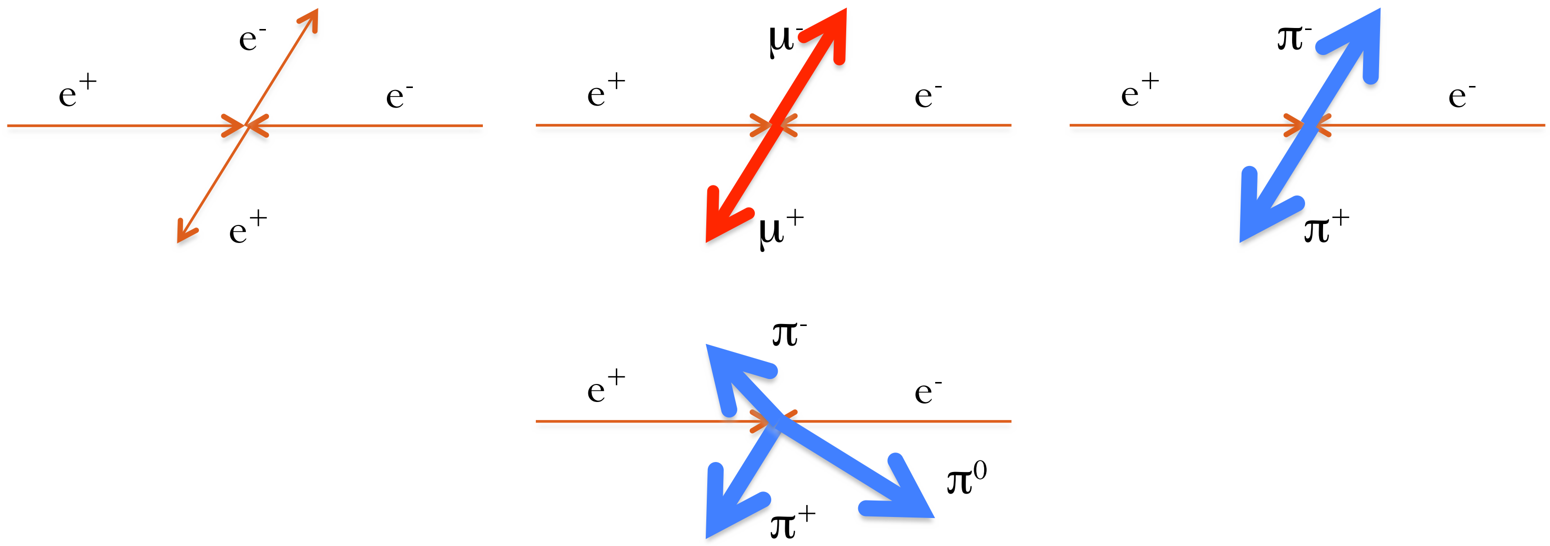
Example of collision:
X == q qbar g



Example of decay:
$X == e\ \nu_e\ \nu_\mu$



GENERAL COMMENT ON OBSERVABLES:
if masses and kinetic energies of each projectile and target are known
can the outcome of each collision be predicted?
NO !
only the probability of each possible outcome can be predicted …



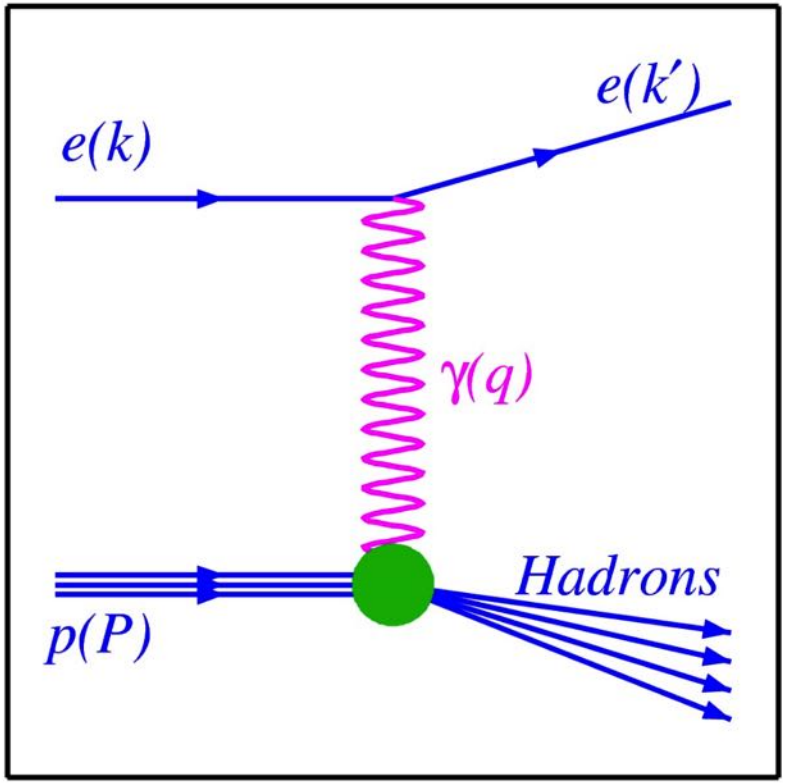
In every collision e^+e^- "***toss the dices***" and choose a possible final state
The theory allows us to evaluate the probability of the final states.
With the experiment one can only measure the frequency of the final states
and compare it to the predicted probability

Why count ? – II

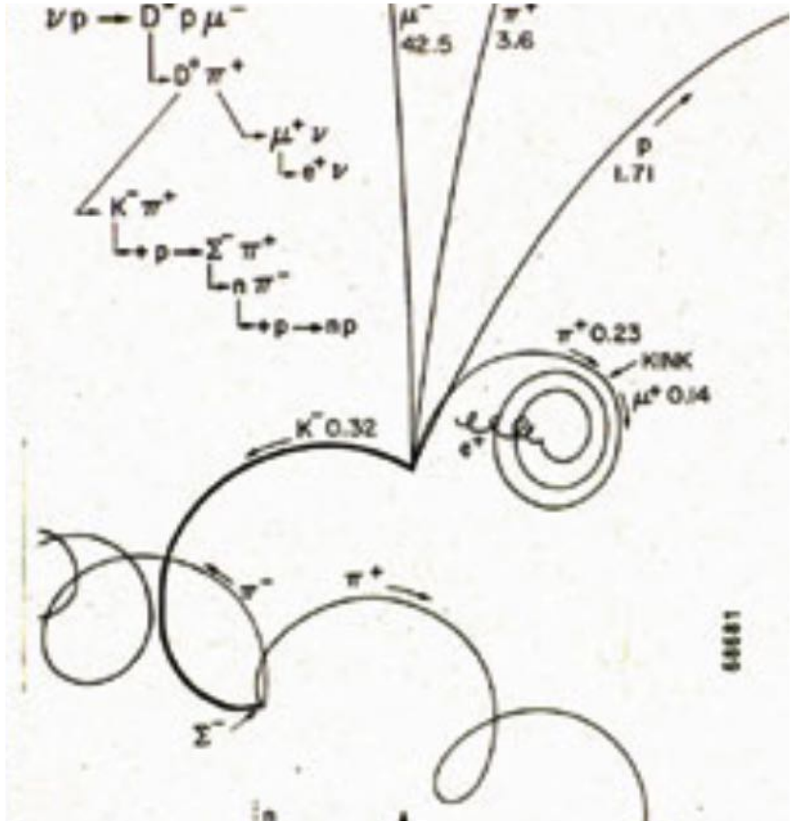
- Given a collision or a decaying particle you have several possibilities, several different final states.
- So: if I have produced N initial states (either a+b collisions or decaying particles), and out of them n times I observe the final state I am looking for, I can access this probability that should be $\approx n/N$
- Let's introduce the concept of **Event:**
  - The collection of all the particles of the final state from a single collision.
  - It is a collection of particles with their quadri-momenta.
  - Be careful not to overlap particles from different collisions.

Event: a “photo” of a collision/decay

Inclusive Event: measure the electron only



Exclusive Event: measure all particles to “close” the kinematics



Why random variables

- **Intrinsic quantum nature of the phenomena we are considering**
- **Instrumental effects**

- Example: the angular distribution in the Rutherford scattering → the variable is the deflection angle θ
  - ➔ from "physics" you expect f(θ): this is the PDF of the quantity θ
  - ➔ let's include the instrumental effects: θ = true; θ' = measured
    - → efficiency ε(θ)
    - → resolution R(θ-θ')
  - The measured "histogram" will be (see later)

$$g(\theta') = \int \varepsilon(\theta) R(\theta - \theta') f(\theta) d\theta$$

"Logic" of an EPP experiment - II

- An ***ideal detector*** allows to measure the quadri-momentum of each particle involved in the reaction.
  - Direction of flight;
  - Energy $\boldsymbol{E}$ and/or momentum modulus $|\boldsymbol{p}|$;
  - Which particle is (e.g. from independent measurements of $\boldsymbol{E}$ and $|\boldsymbol{p}|$, $\boldsymbol{m}^2=\boldsymbol{E}^2-|\boldsymbol{p}|^2$) ➔ Particle ID
- BUT for a ***real detector***:
  - Not all quadri-momenta are measured: some particles are out of acceptance, or only some quantities are accessible, there are unavoidable **inefficiencies**;
  - Measurements are affected by **resolution**
  - Sometimes the particle nature is "confused"

“Logic” of an EPP experiment - III

- Selection steps:
  1. **TRIGGER SELECTION**
     - Retain only “interesting events”: from bubble chambers to electronic detectors
     - ➔ “logic-electronic” eye: decides in a short time O(μs) if the event is interesting or not.
     - In some cases (e.g. pp), it is crucial since interactions are so probable…
     - LHC: every 25 ns is a bunch crossing giving rise to interactions: can I write 40 MHz on “tape” ? A tipical event has a size of 1 MB ➔ 40 TB/s. Is it conceivable ? And how many CPU will be needed to analyze these data ? At LHC from 40 MHz to 200 Hz ! Only one bunch crossing every 200000 !
     - “pre-scale” is an option
     - e^+e^-: the situation is less severe but a trigger is in any case necessary.

“Logic” of an EPP experiment - IV

2. **EVENT RECONSTRUCTION**: Once you have the final event sample, for each trigger you need to reconstruct at your best the kinematic variables.
3. **OFFLINE SELECTION**: choice of a set of discriminating variables on which apply one of the following:
   - cut-based selection
   - discriminating variables selection
   - multivariate classifier selection
4. **PHYSICS ANALYSIS**: analysis of the sample of ***CANDIDATES***

The selection strategy is a crucial part of the experimentalist work: defined and optimized using *simulated data samples*.

"Logic" of an EPP experiment - V

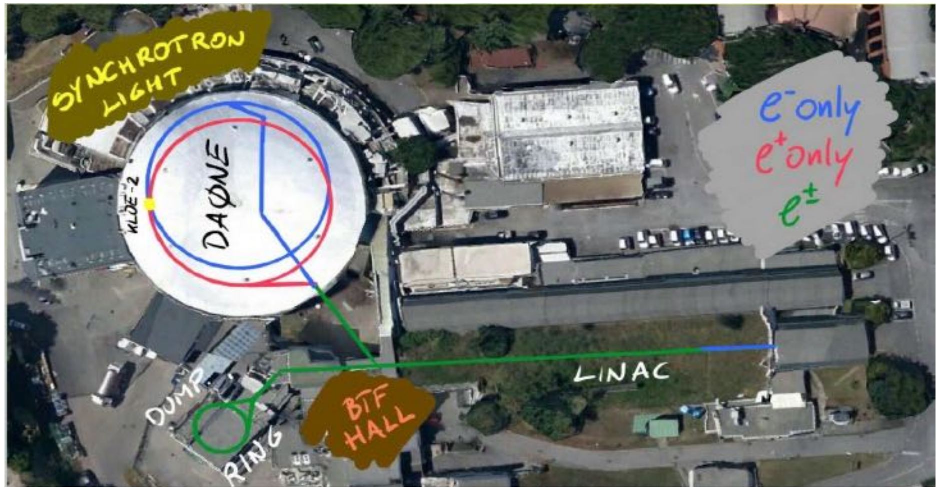
- Simulated samples of events: the Montecarlo.
  - "Physics" simulation: final state with correct kinematic distributions; also dynamics in some cases is relevant.
  - "Detector" simulation: the particles are traced through the detector, interactions, decays, are simulated.
  - "Digitization": based on the particle interactions with the detector, signals are simulated with the same features of the data.
- ➔ For every interesting final state MC samples with the same format of a data sample are built. These samples can be analyzed with the same program. In principle one could run on a sample without knowing if it is data or MC.
- To design a "selection" strategy for a given searched signal one needs: ***signal MC samples*** and ***background MC samples***.

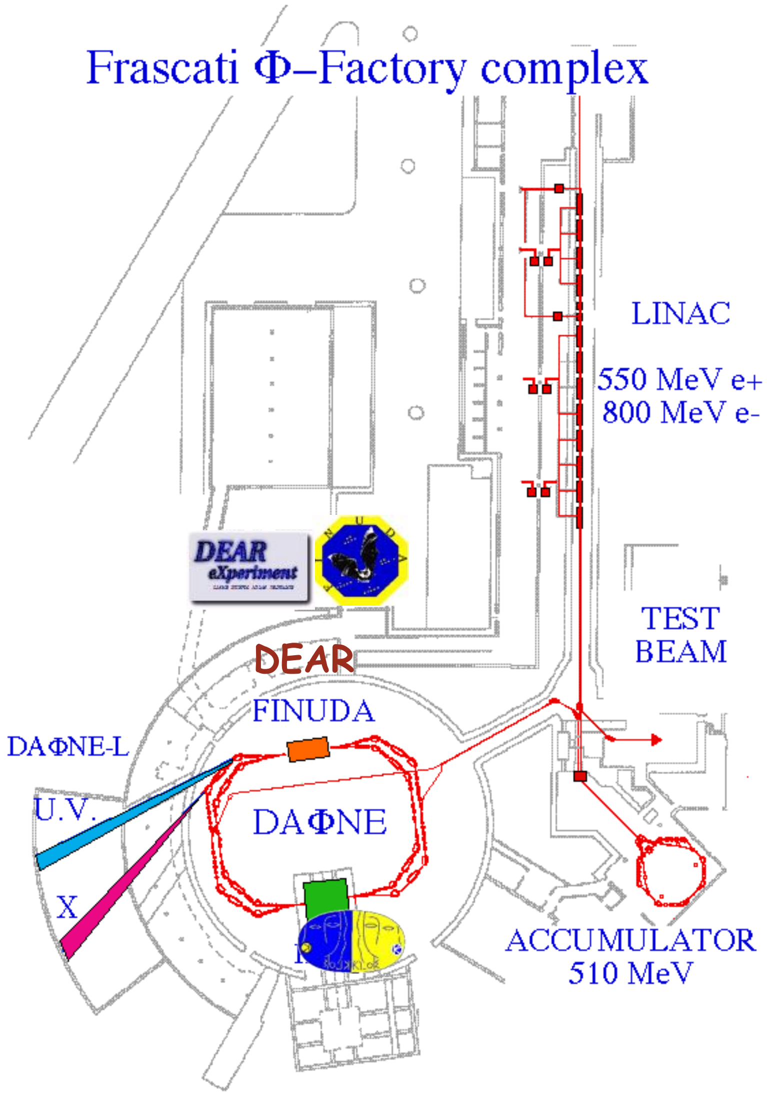
Instrumental effects: examples

The Frascati ϕ-factory: DAΦNE

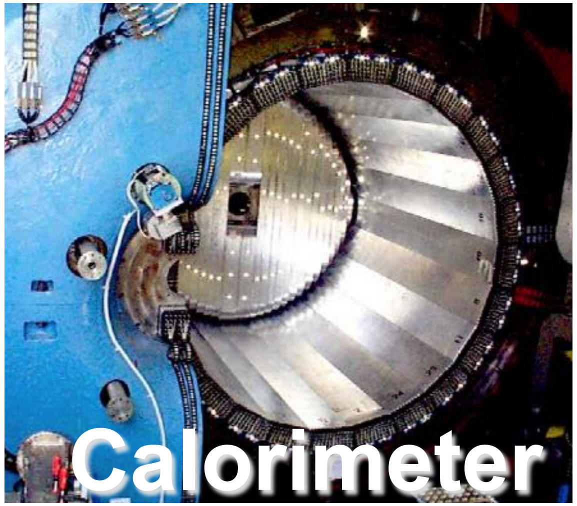
e+e- collider at √s = 1020 MeV
TRF = 2.7 ns, up to 120 bunches
Topping-up injection
Worked for KLOE (2000-2006):

15 mrad crossing angle
Max peak lumi: 1.5 10^{32} $cm^{-1}s^{-1}$
Best daily int. lumi: 8.5 pb^{-1}





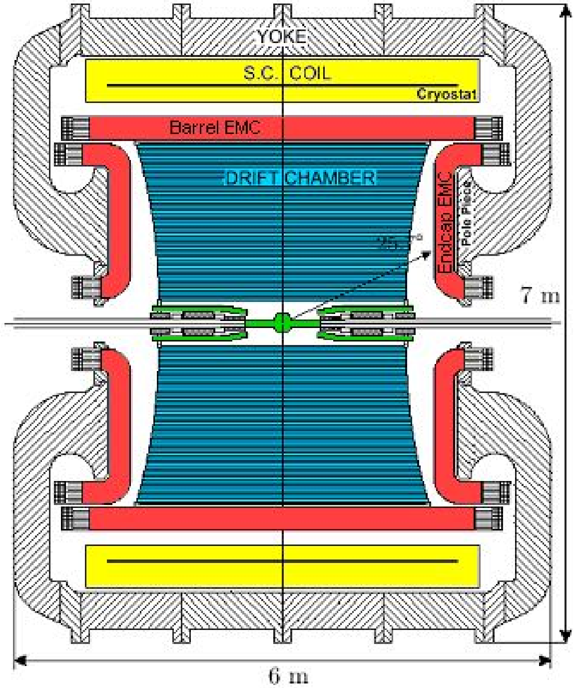
The KLOE detector at DAΦNE



Lead/scintillating fiber
4880 PMTs
98% coverage of solid angle

Superconducting coil

B = 0.52 T



4 m diameter × 3.3 m length
90% helium, 10% isobutane
12582/52140 sense/total wires
All-stereo geometry

$\sigma_E/E \cong 5.7\% / \sqrt{E(\mathrm{GeV})}$

$\sigma_t \cong 54\ \mathrm{ps} / \sqrt{E(\mathrm{GeV})} \oplus 50\ \mathrm{ps}$
(relative time between clusters)

$\sigma_{\gamma\gamma} \sim 2\ \mathrm{cm}$ (π^0 from $K_L \to \pi^+\pi^-\pi^0$)

$\sigma_p/p \cong 0.4\ \%$ (tracks with $\theta > 45°$)

$\sigma_x^{hit} \cong 150$ mm (xy), 2 mm (z)

$\sigma_x^{vertex} \sim 1$ mm

Events in KLOE

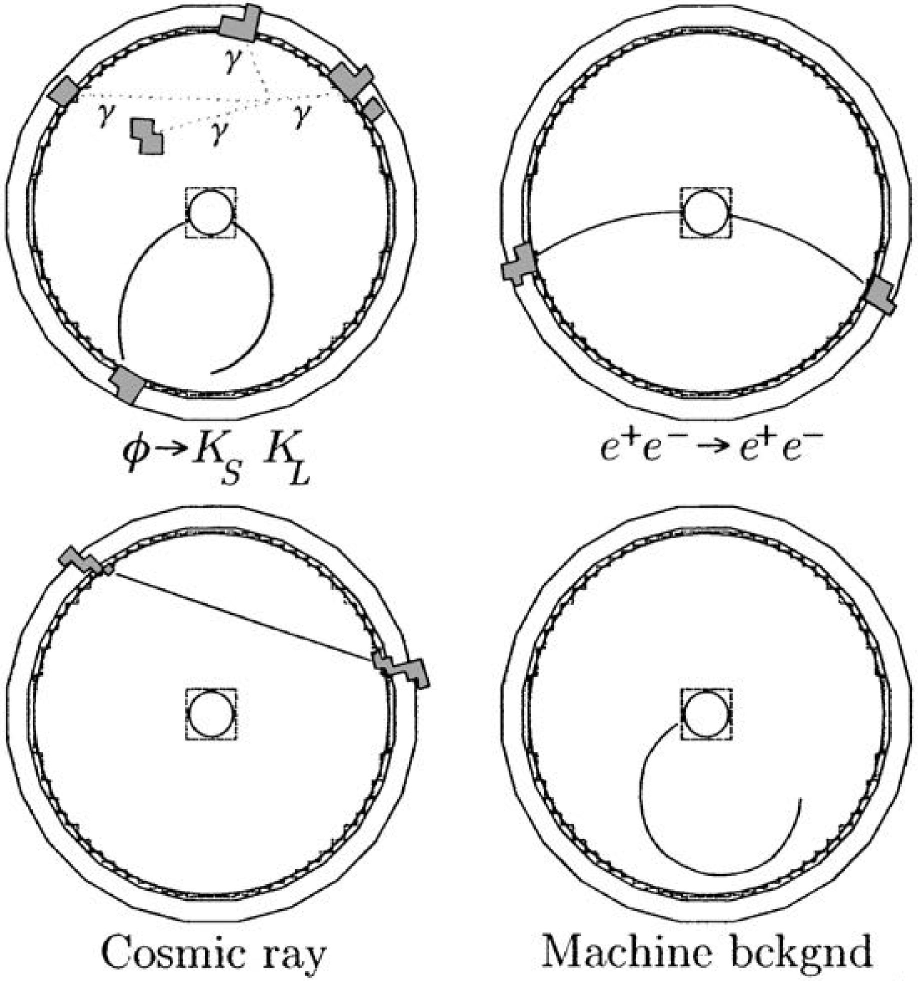


Fig. 12. Example of events. The grey areas indicate energy deposits in the calorimeter.

Trigger selection logic in KLOE

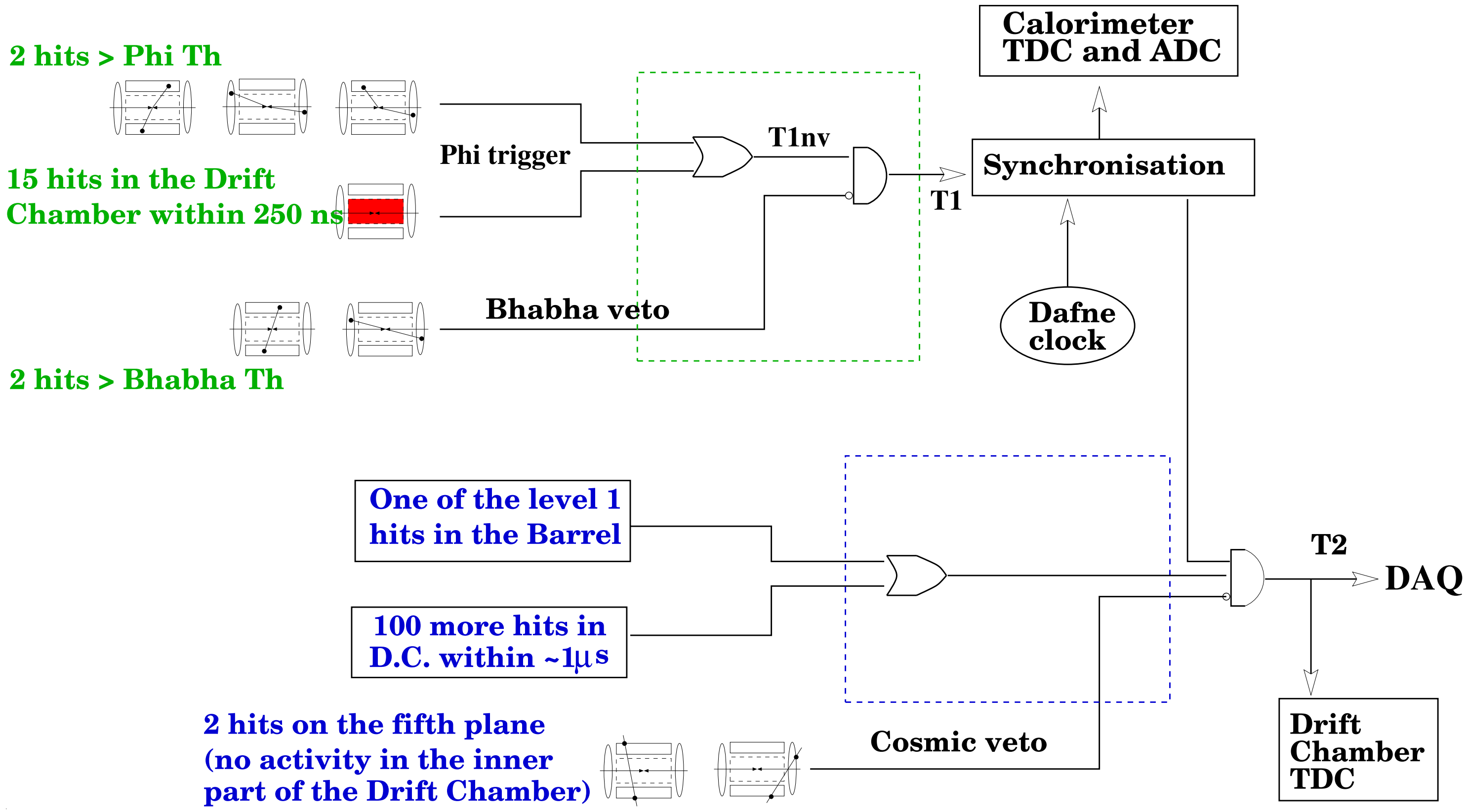


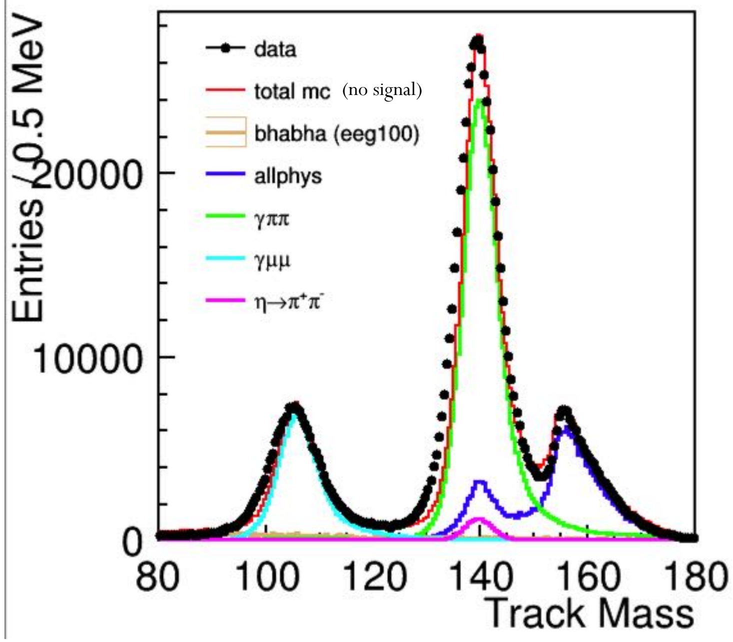
Diagram of the two-level trigger logic. It has been optimized to preserve the majority of $e^+e^- \to \phi$ decays, and provide efficient rejection of the two main sources of background: small angle Bhabha scattering and particles lost from DAΦNE beams. Both T_1 and T_2 triggers are based on the topology of energy deposits in the EMC and on the hit multiplicity in the DC. Figure adapted from Ref. [39].

Search for $\eta \rightarrow \pi^+\pi^-$ decay

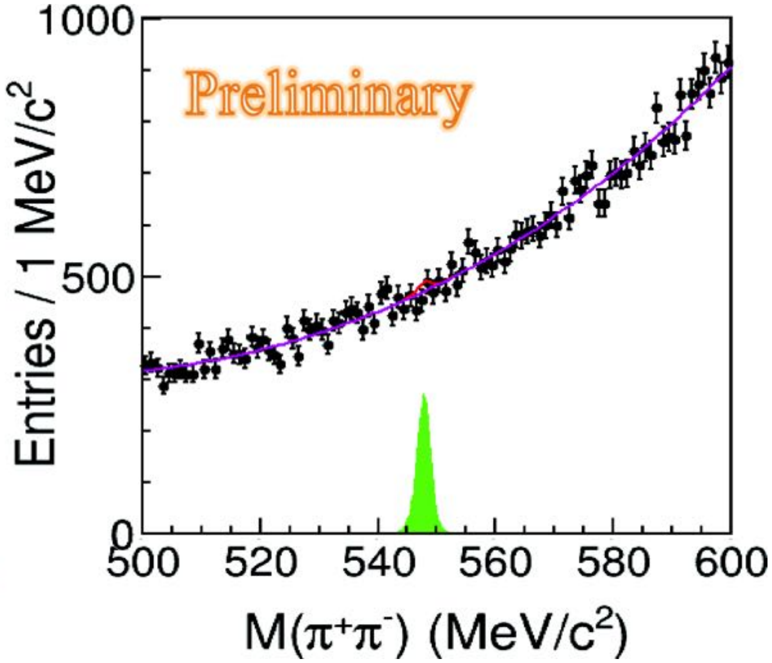
- P and CP violating, Br expected of order 10^{-27} in the SM
- Detection at any accessible level would be signal of CP viol. beyond the SM

Best limit $Br<1.3\times10^{-5}$ @ 90% C.L. ($L = 350\ pb^{-1}$) [KLOE, PLB606(2005)276]

LHCb recent measurement: $Br<1.6\times10^{-5}$ @ 90% C.L. [PLB764(2017)233]



After cut: $129 < M_{tr} < 149$ MeV



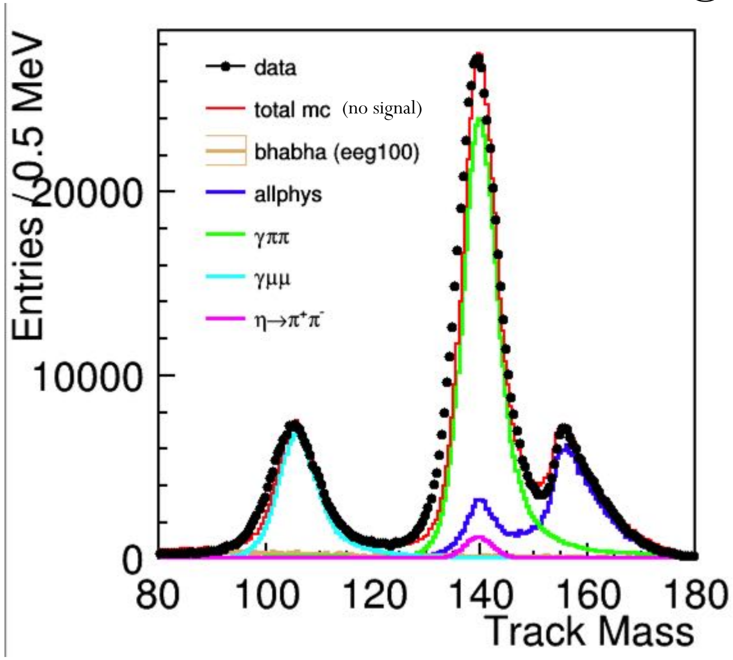
- $L = 1.7\ fb^{-1}$ (KLOE data) ⇒ preliminary U.L.: $Br < 5.8\times10^{-6}$ @ 90% C.L.
- Combining KLOE + KLOE-2 statistics ($8\ fb^{-1}$) ⇒ U.L. expected $\sim 3\times10^{-6}$

Search for $\eta \to \pi^+\pi^-$ decay

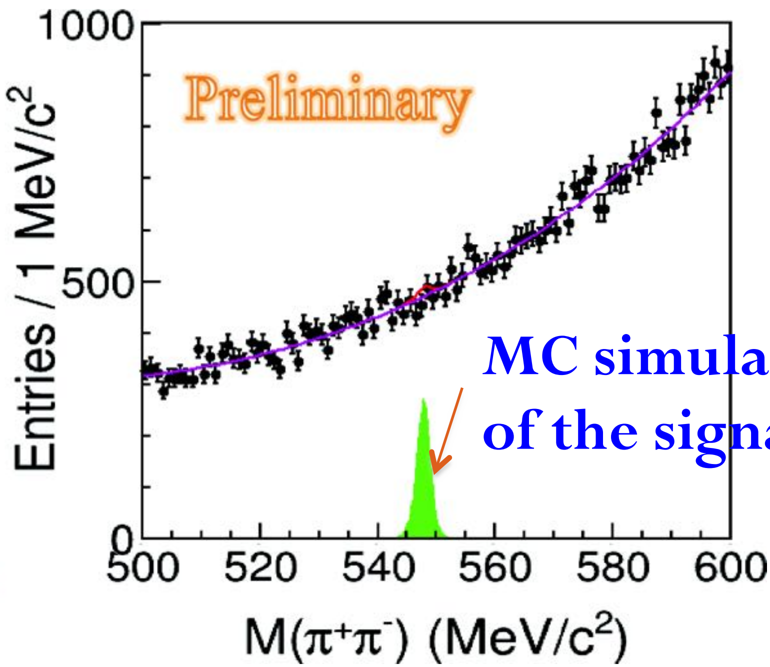
- P and CP violating, Br expected of order 10^{-27} in the SM
- Detection at any accessible level would be signal of CP viol. beyond the SM

Best limit $Br<1.3\times10^{-5}$ @ 90% C.L. ($L = 350\ pb^{-1}$) [KLOE, PLB606(2005)276]

LHCb recent measurement: $Br<1.6\times10^{-5}$ @ 90% C.L. [PLB764(2017)233]



After cut: $129 < M_{tr} < 149$ MeV



MC simulation of the signal

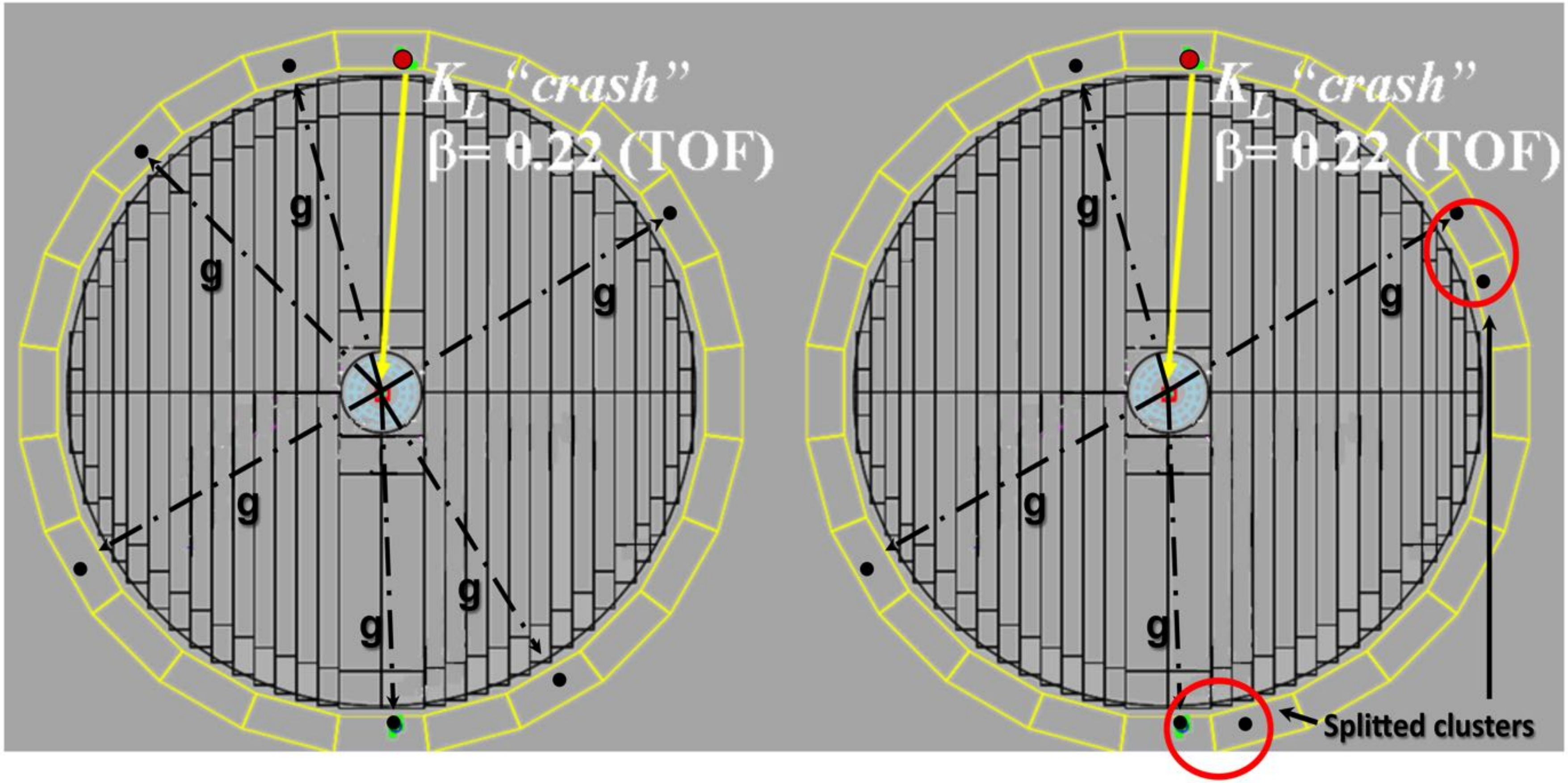
- $L = 1.7\ fb^{-1}$ (KLOE data) ⇒ preliminary U.L.: $Br < 5.8\times10^{-6}$ @ 90% C.L.
- Combining KLOE + KLOE-2 statistics ($8\ fb^{-1}$) ⇒ U.L. expected $\sim 3\times10^{-6}$

Search for the CP violating $K_S \to \pi^0\pi^0\pi^0$ decay

Standard Model prediction: $BR(K_S \to 3\pi^0) = 1.9 \cdot 10^{-9}$ PLB 723 (2013) 54

Best upper limit by KLOE with $1.7\ fb^{-1}$ **$BR(K_S \to 3\pi^0) < 2.6 \times 10^{-8}$ @ 90% CL**

SIGNAL **BACKGROUND**



$K_S \to 3\pi^0 \to 6\gamma$

$K_S \to 2\pi^0$ + accidental/splitted clusters

$K_L \to 3\pi^0$, $K_S \to \pi^+\pi^-$ („fake K_L crash")

Event information structure in a MC simulation

The event information structure in a MC simulation is EXACTLY the same as for the data. In ADDITION there is the event information of the MC "truth".

Block info example

EVENT INFORMATION:

```
***********************************************************************
nrun .........  run #
nev  .........  event #
pileup .......  pileup event_flag (0: no pileup/1:pileup)
gcod  ........  generation code (1:PHI/2:BHA/3:COSM/4:mach/5:mumugam/6:mumu)
phid .........  Phi-decay(0:NOPHI/1:K+K-/2:KsKl/3:RhoPi/4:pi+pi-gam/
                5:etagam/6:pi0gam/7:f0gam/8:pi0pi0gam/9:pi+pi-gam)
a1typ ........  particle type #1 from Kl
a2typ ........  particle type #2 from Kl
a3typ ........  particle type #3 from Kl
b1typ ........  particle type #1 from Ks
b2typ ........  particle type #2 from Ks
b3typ ........  particle type #3 from Ks
***********************************************************************
```

Event information structure in a MC simulation

CALORIMETER CLUSTERS:

Block info example

```
*****************************************************************************
nclu ........... # of reconstructed clusters
EneCl(nclu)  ... Reconstructed Energy (MeV)
Tcl(nclu)  ..... Reconstructed average Time (ns)
Xcl(nclu) ...... Centroid position X (cm)
Ycl(nclu) ...... Centroid position Y (cm)
Zcl(nclu) ...... Centroid position Z (cm)
Xacl(nclu)...... Shower Apex position X (cm)
Yacl(nclu)...... Shower Apex position Y (cm)
Zacl(nclu)...... Shower Apex position Z (cm)
XRmCl(nclu) .... Cluster RMS in X (cm)
YRmsCl(nclu).... Cluster RMS in Y (cm)
ZrmsCl(nclu).... Cluster RMS in Z (cm)
FlagCl(nclu) ... Cluster Flag
A specialized sub-block:
Npar(nclu)...... Particles beloging to the cluster (<=10 )
Pnum1(nclu) .... First  particle in cluster related number in KINE block
Pid1(nclu)  .... First  particle in cluster related number in GEANT
Pnum2(nclu) .... Second particle in cluster related number in KINE block
Pid2(nclu)  .... Second particle in cluster related number in GEANT
Pnum3(nclu) .... Third  particle in cluster related number in KINE block
Pid3(nclu) ..... Third  particle in cluster related number in GEANT
*****************************************************************************
```

Event information structure in a MC simulation

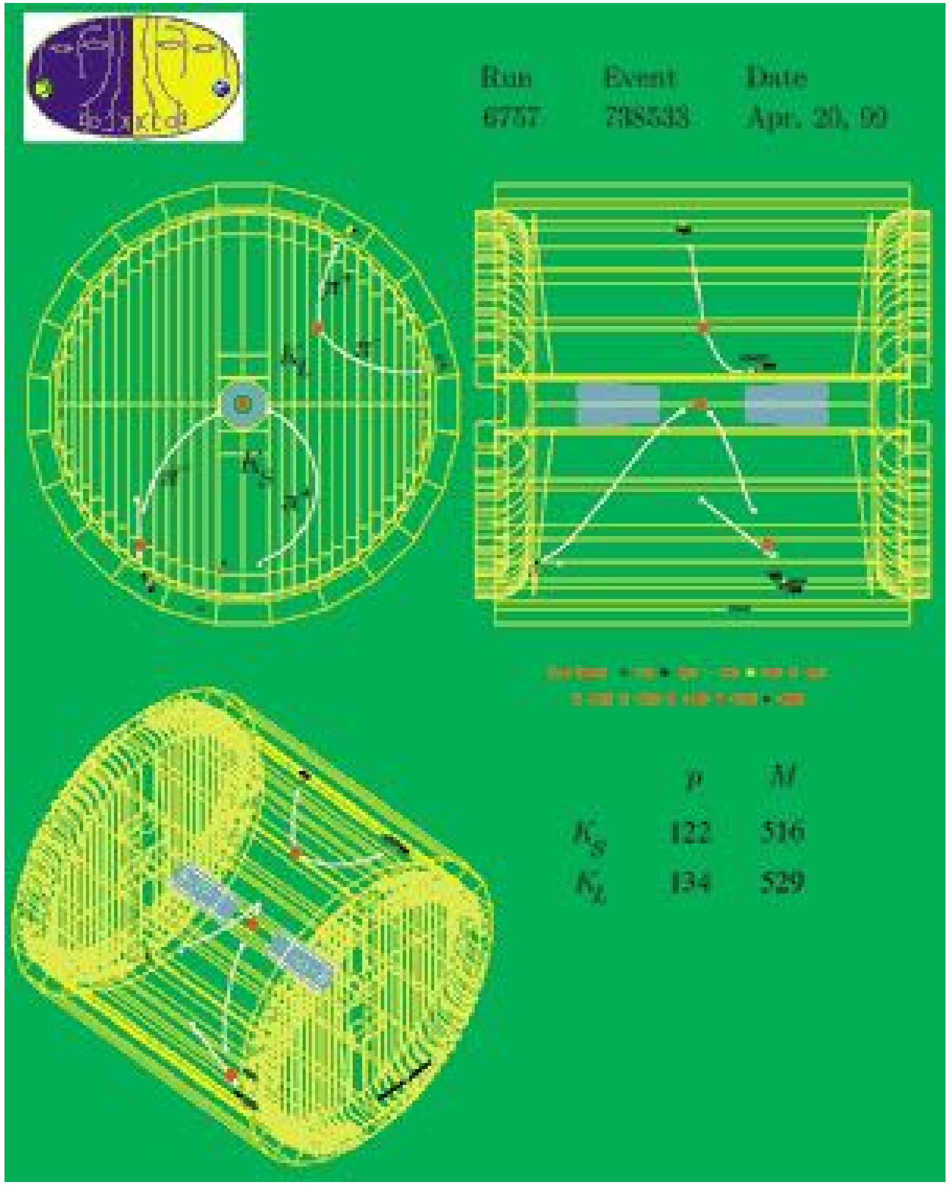
CALORIMETER CELLS:

Block info example

```
*********************************************************************
Ncel ........ # of fired cells.
ICL(Ncel) ... Cluster # of the selected cell
DET(Ncel) ... Detector(1:Ecapa,2:Barrel,3:Ecapb)
WED(Ncel) ... Wedge Number (1:24 Barrel, 1:64 EndCap)
PLA(Ncel) ... Plane Number 1:5
COL(Ncel) ... Column Number
ENE(Ncel) ... Reconstructed Energy for the cell as in bank CWRK
T   (Ncel) ... Reconstructed Time as in CWRK
X   (Ncel) ... Reconstructed X as in CWRK
Y   (Ncel) ... Reconstructed Y as in CWRK
Z   (Ncel) ... Reconstructed Z as in CWRK
EA(Ncel) .... Deposited Energy side A
ta(ncel) .... Timing side A
eb(ncel) .... Deposited Energy side B
tb(ncel) .... Timing side B
Emc(ncel) ... True MC energy deposited in the fiber
Tmc(ncel) ... True MC arrival time
Xmc(ncel) ... True MC X position
Ymc(ncel) ... True MC Y position
Zmc(ncel) ... True MC Z position
Ptyp(ncel) .. Geant Particle Type firing the cell
Knum(ncel) .. Kine number of particle firing the cell
Nhit(ncel) .. #of hit per cells (1 single hit per cell)
              (>1 there are replica of the cell)
              (0 are the cells' replica)
*********************************************************************
```

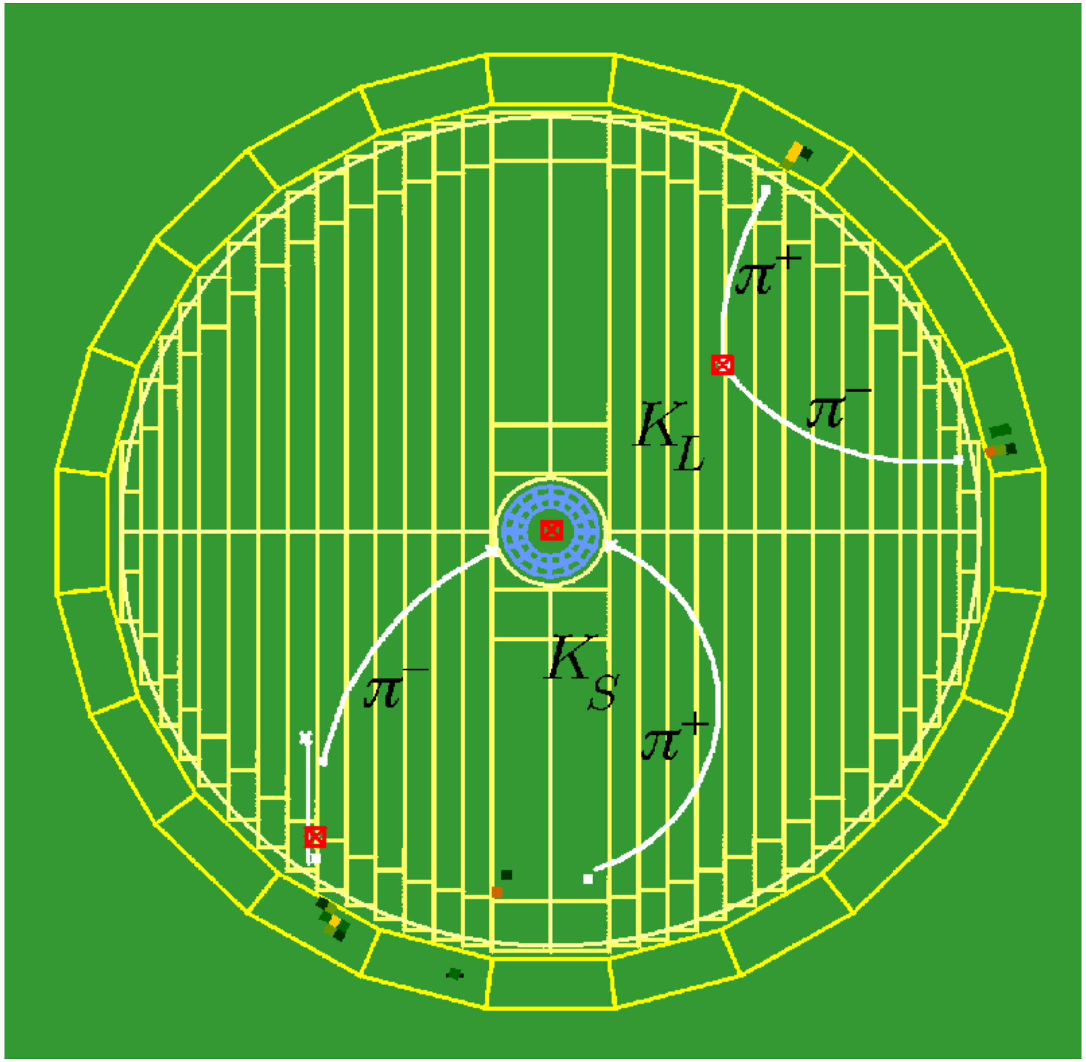
Instrumental effects: importance of resolution effects

$\phi \rightarrow K_S K_L \rightarrow \pi^+\pi^- \pi^+\pi^-$: test of quantum coherence



KLOE event:
$\phi \rightarrow K_S K_L \rightarrow \pi^+\pi^-\pi^+\pi^-$

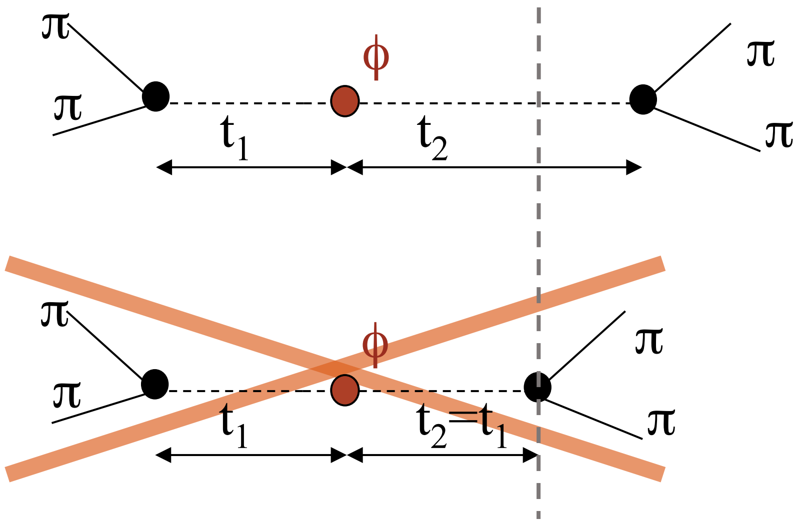
$\phi \rightarrow K_S K_L \rightarrow \pi^+\pi^- \pi^+\pi^-$: test of quantum coherence



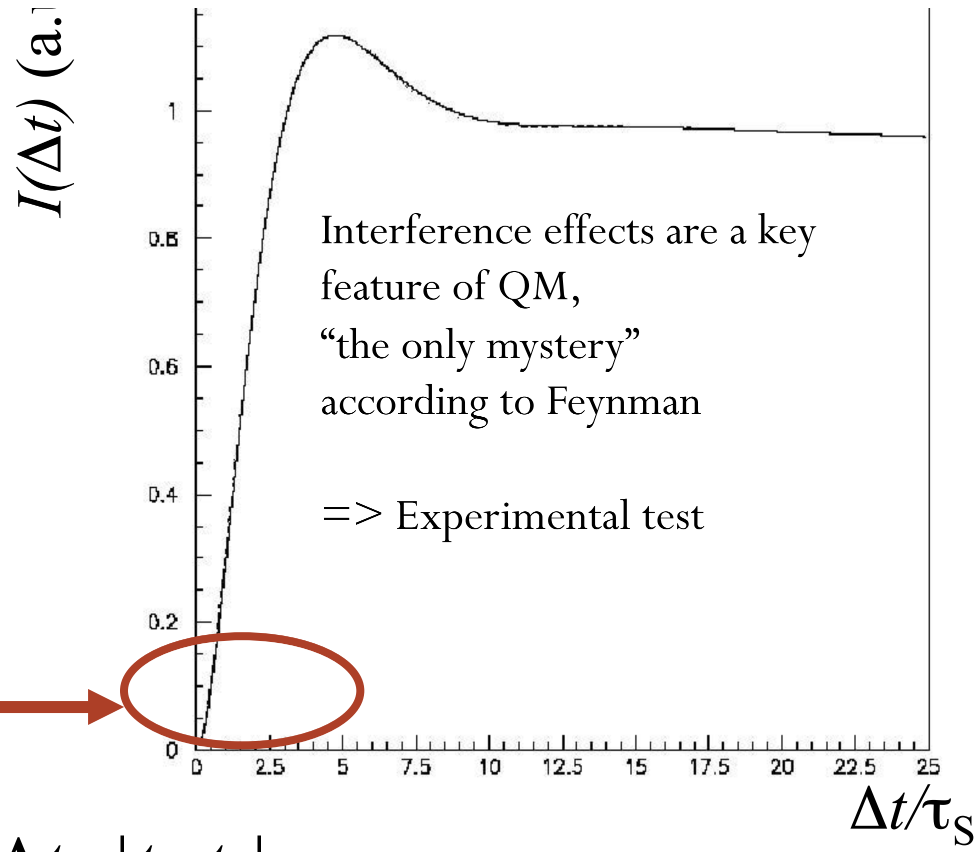
KLOE event:
$\phi \rightarrow K_S K_L \rightarrow \pi^+\pi^-\pi^+\pi^-$

$\phi \rightarrow K_S K_L \rightarrow \pi^+\pi^- \pi^+\pi^-$: test of quantum coherence

$$|i\rangle = \frac{1}{\sqrt{2}}\left[|K^0\rangle|\bar{K}^0\rangle - |\bar{K}^0\rangle|K^0\rangle\right]$$



Same final state for both kaons: $f_1 = f_2 = \pi^+\pi^-$
(this specific channel is suppressed by CP viol.
$|\eta_{+-}|^2 = |A(K_L \text{->} \pi^+\pi^-)/A(K_S \text{->} \pi^+\pi^-)|^2 \sim |\varepsilon|^2 \sim 10^{-6}$)



EPR correlation:

no simultaneous decays (Δt=0) in the same final state due to the fully destructive quantum interference

$\Delta t = |t_1 - t_2|$

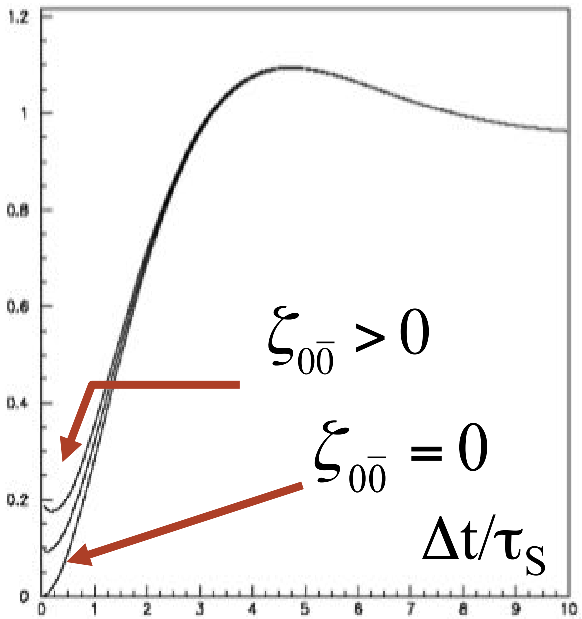
$\phi \to K_S K_L \to \pi^+\pi^- \, \pi^+\pi^-$: test of quantum coherence

$$|i\rangle = \frac{1}{\sqrt{2}}\left[\left|K^0\right\rangle\left|\overline{K}^0\right\rangle - \left|\overline{K}^0\right\rangle\left|K^0\right\rangle\right]$$

$$I\left(\pi^+\pi^-,\pi^+\pi^-;\Delta t\right) = \frac{N}{2}\left[\left|\left\langle\pi^+\pi^-,\pi^+\pi^-\left|K^0\overline{K}^0(\Delta t)\right.\right\rangle\right|^2 + \left|\left\langle\pi^+\pi^-,\pi^+\pi^-\left|\overline{K}^0K^0(\Delta t)\right.\right\rangle\right|^2\right.$$

$$\left. -\left(1-\zeta_{0\bar{0}}\right)\cdot 2\Re\left(\left\langle\pi^+\pi^-,\pi^+\pi^-\left|K^0\overline{K}^0(\Delta t)\right.\right\rangle\left\langle\pi^+\pi^-,\pi^+\pi^-\left|\overline{K}^0K^0(\Delta t)\right.\right\rangle^*\right)\right]$$

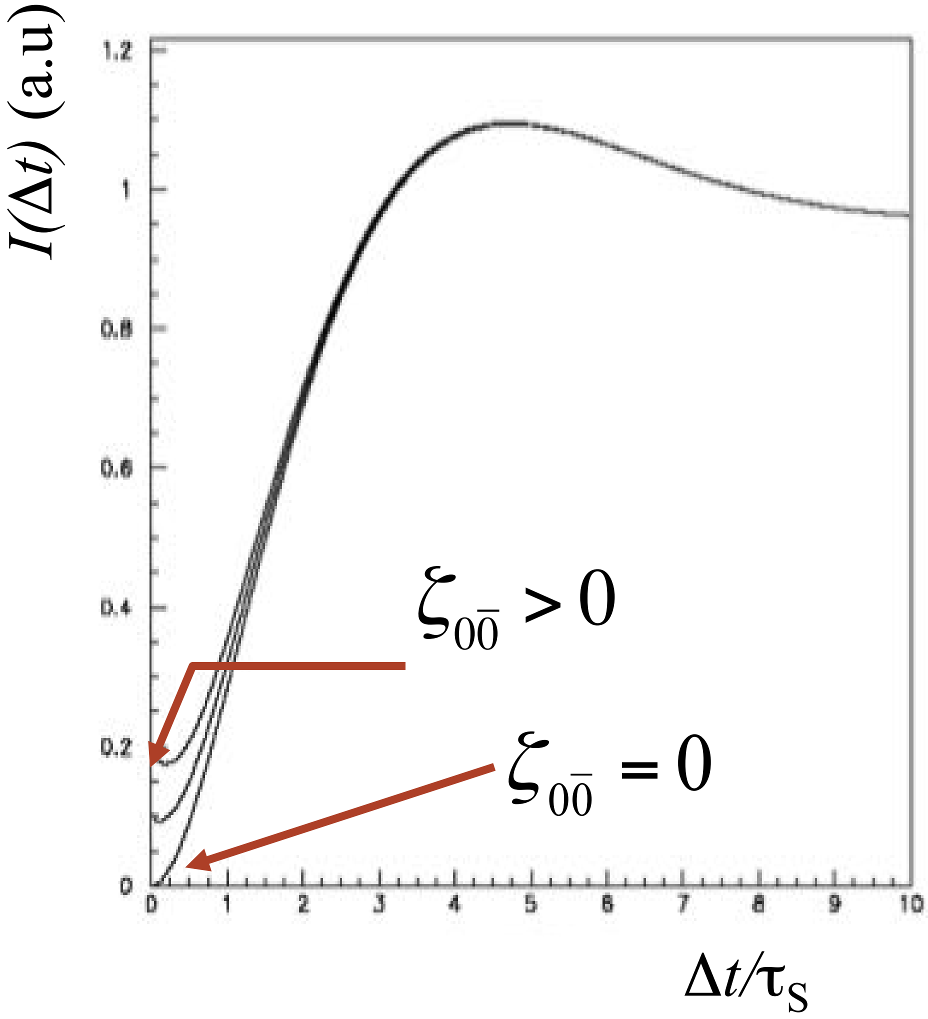
I(Δt) (a.u.)

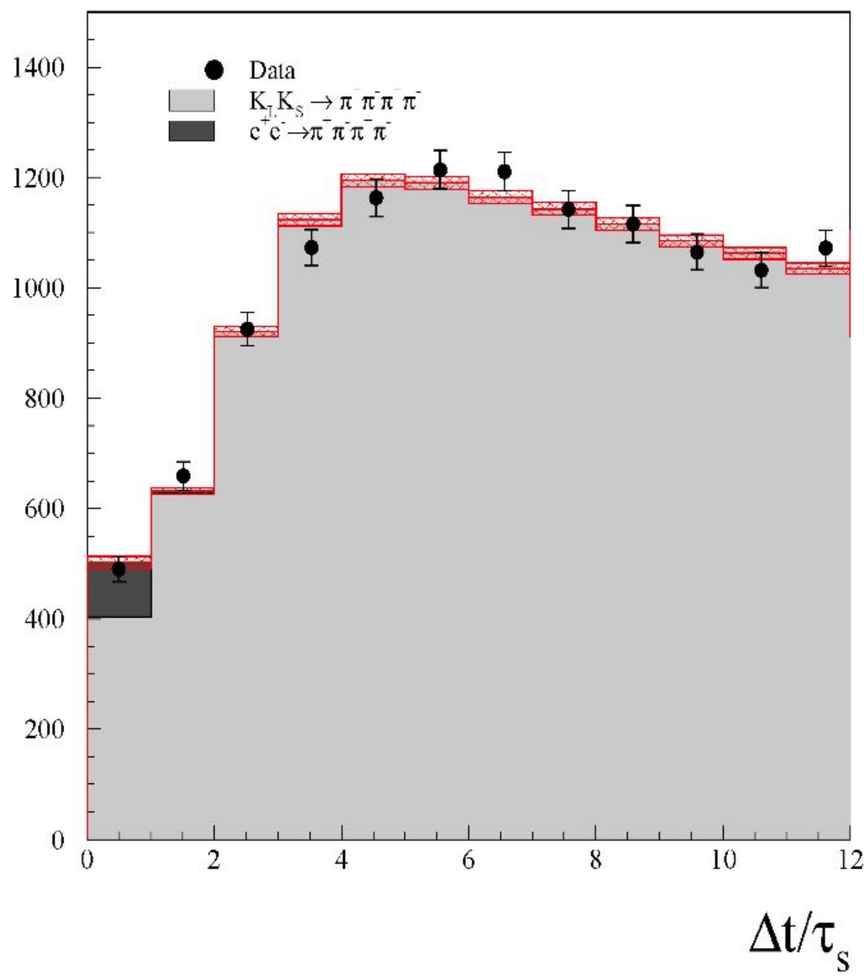


Decoherence parameter:

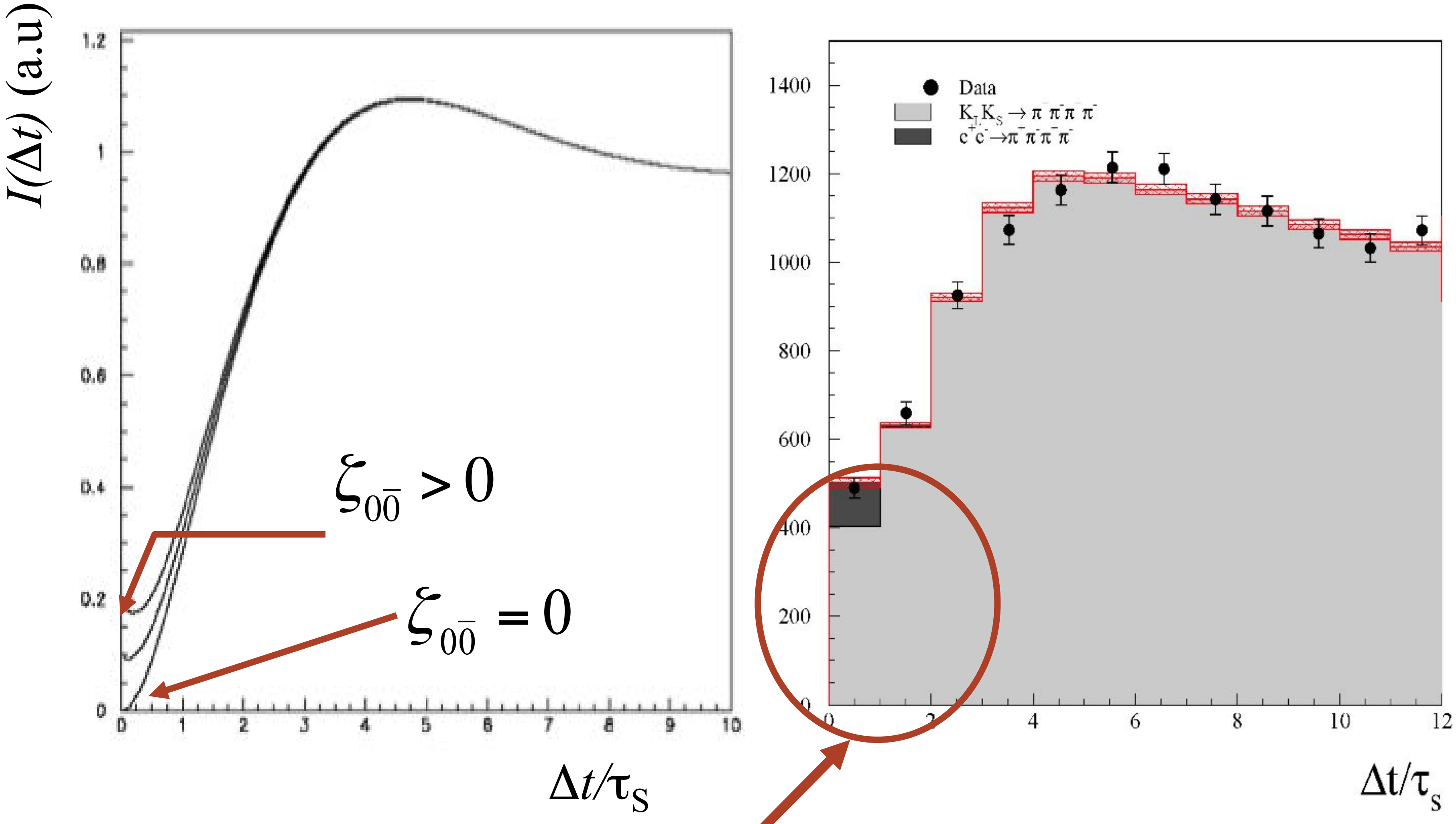
$\zeta_{0\bar{0}} = 0 \quad \to \quad$ QM

$0 < \zeta_{0\bar{0}} \leq 1 \quad \to \quad$ Violation of QM!





$\Delta t = |t_1 - t_2|$



Violation of QM?

$$\Delta t = |t_1 - t_2|$$

$\phi \rightarrow K_S K_L \rightarrow \pi^+\pi^- \pi^+\pi^-$: test of quantum coherence

- Analysed data L=1.5 fb[-1]
- Fit including Δt resolution and efficiency effects + regeneration

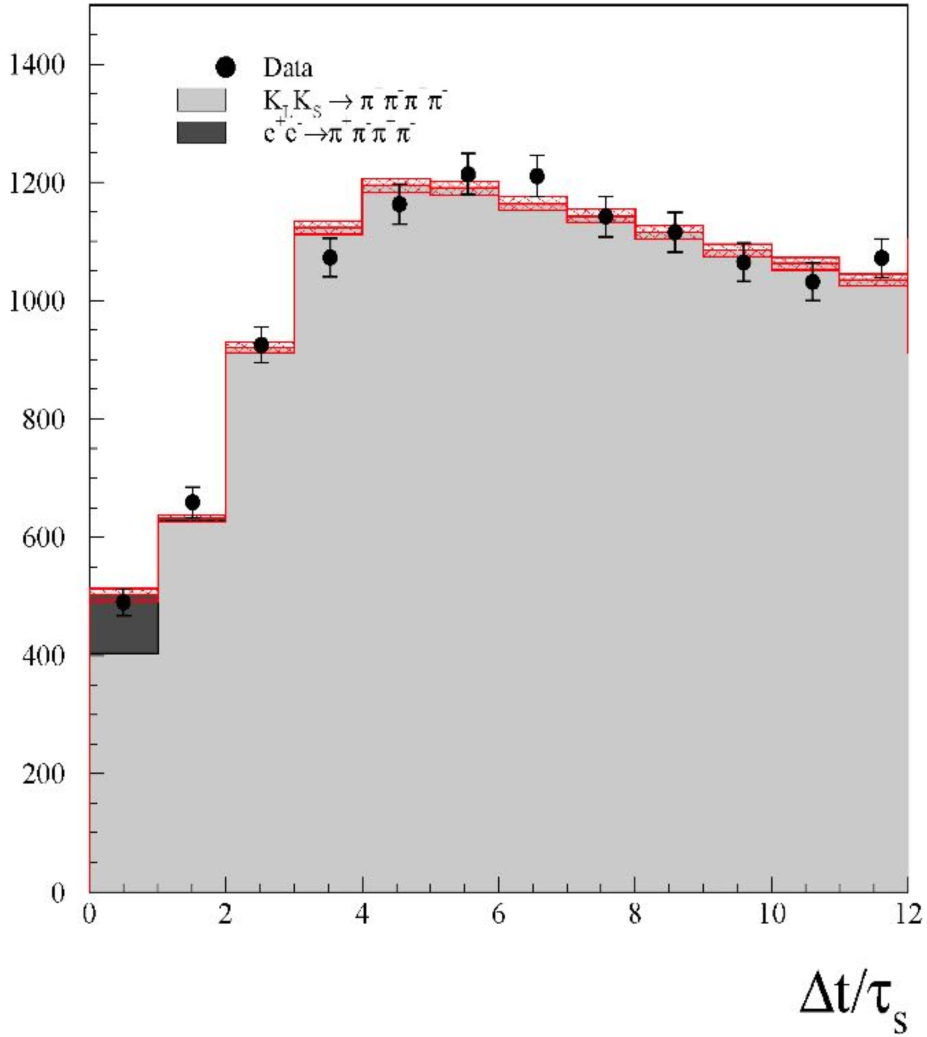
KLOE result: PLB 642(2006) 315 Found. Phys. 40 (2010) 852

$$\zeta_{0\bar{0}} = \left(1.4 \pm 9.5_{STAT} \pm 3.8_{SYST}\right) \times 10^{-7}$$

CP violation: $|\eta_{+-}|^2 \sim |\varepsilon|^2 \sim 10^{-6}$

=> terms $\zeta_{00}/|\eta_{+-}|^2$ => **high sensitivity to ζ_{00}**

=> **Amplification mechanism due to CPV**



Instrumental effects: importance of resolution

→ "physics" distribution: f(x)
→ efficiency ε(x)
→ resolution R(x-x')
→ measured distribution: g(x)

Take into account instrumental effects with the convolution integral:

$$g(x) = \int \varepsilon(x')R(x-x')f(x')dx'$$

If efficiency effects are negligible:

$$g(x) = \int R(x-x')f(x')dx'$$

Do homework n.1 !

γ Spectroscopy -I

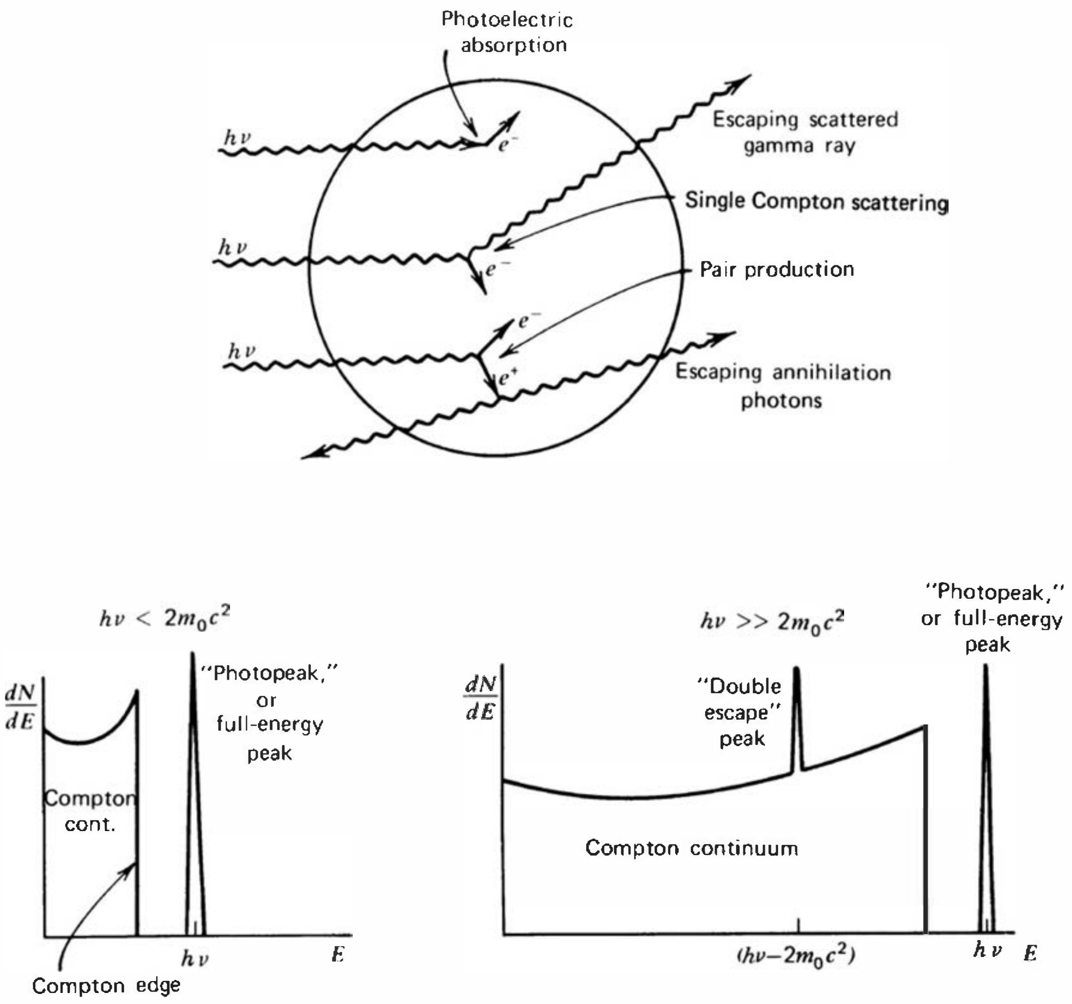


Figure 10.2 The "small detector" extreme in gamma-ray spectroscopy. The processes of photo-electric absorption and single Compton scattering give rise to the low-energy spectrum at the left. At higher energies, the pair production process adds a double escape peak shown in the spectrum at the right.

γ Spectroscopy -II

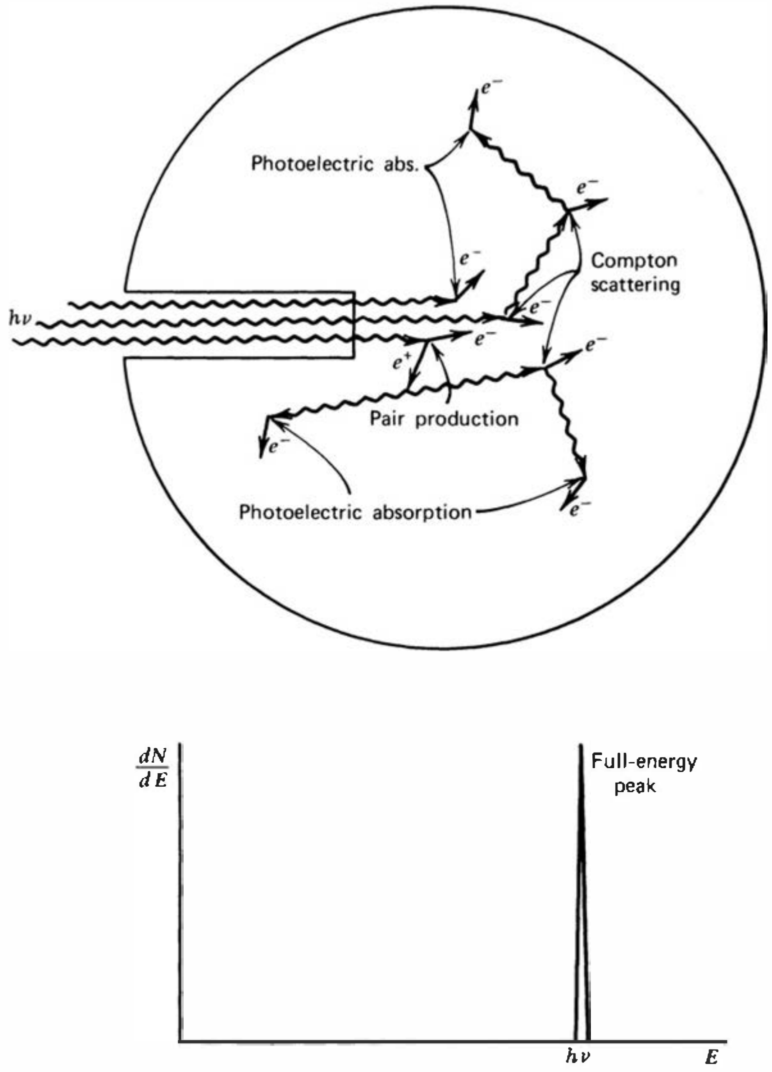


Figure 10.3 The "large detector" extreme in gamma-ray spectroscopy. All gamma-ray photons, no matter how complex their mode of interaction, ultimately deposit all their energy in the detector. Some representative histories are shown at the top.

γ Spectroscopy -III

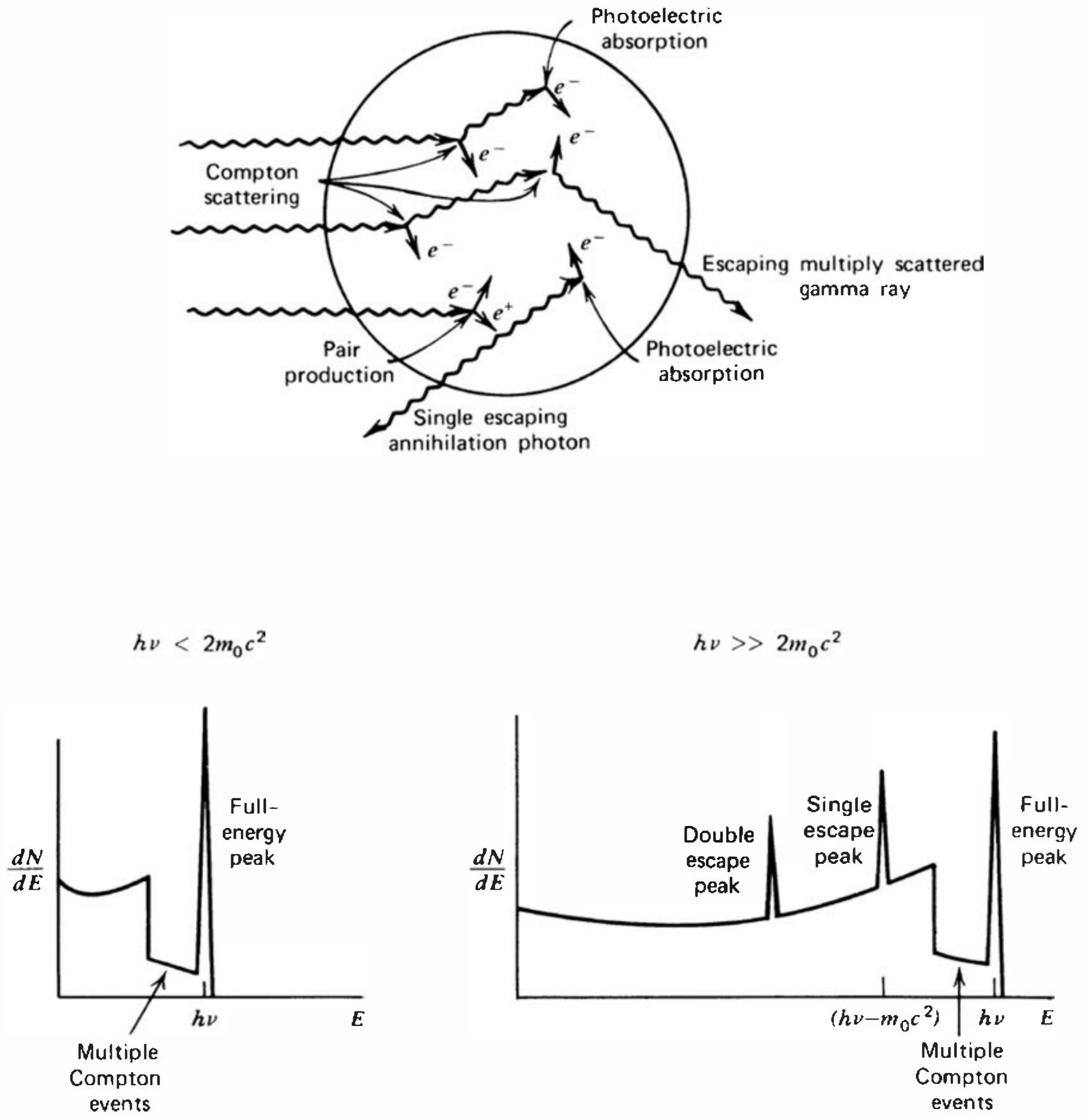
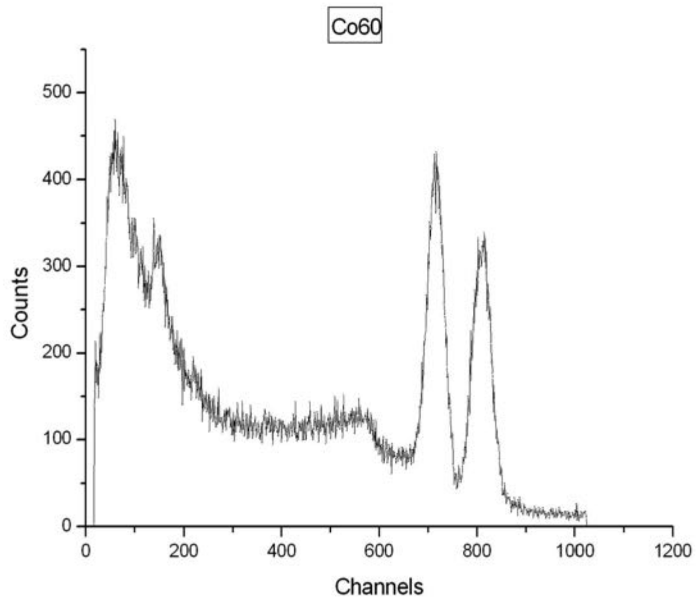
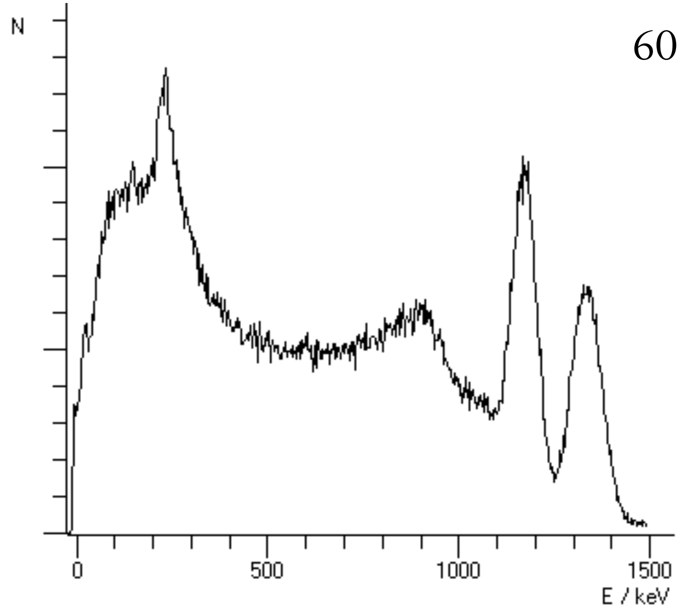
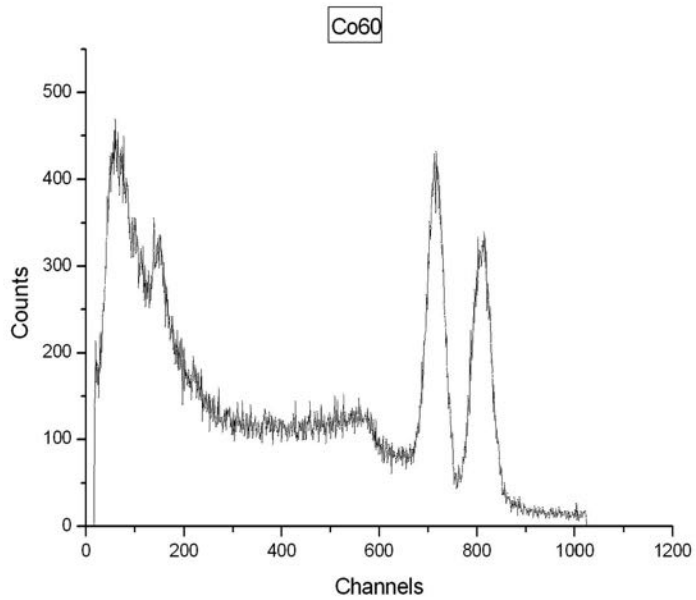


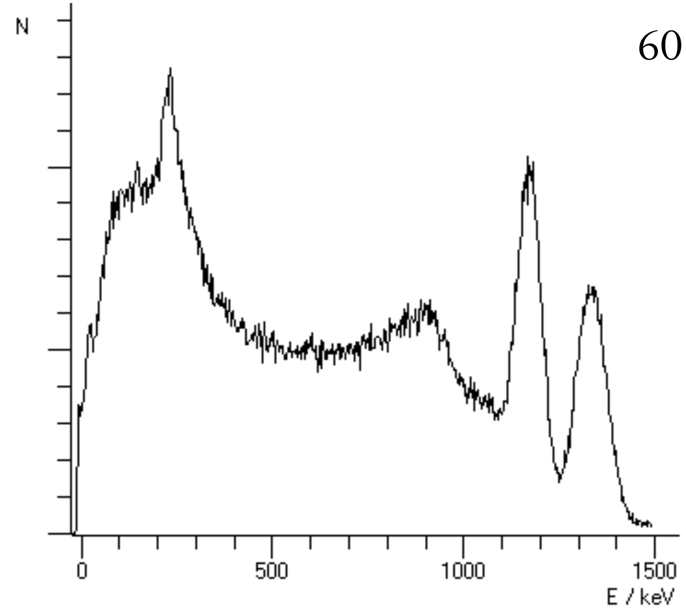
Figure 10.4 The case of intermediate detector size in gamma-ray spectroscopy. In addition to the continuum from single Compton scattering and the full-energy peak, the spectrum at the left shows the influence of multiple Compton events followed by photon escape. The full-energy peak also contains some histories that began with Compton scattering. At the right, the single escape peak corresponds to initial pair production interactions in which only one annihilation photon leaves the detector without further interaction. A double escape peak as illustrated in Fig. 10.2 will also be present due to those pair production events in which both annihilation photons escape.


Co60
Counts
500
400
300
200
100
0
0
200
400
600
800
1000
1200
Channels


N
0
500
1000
1500
E / keV

^{60}Co





^{60}Co

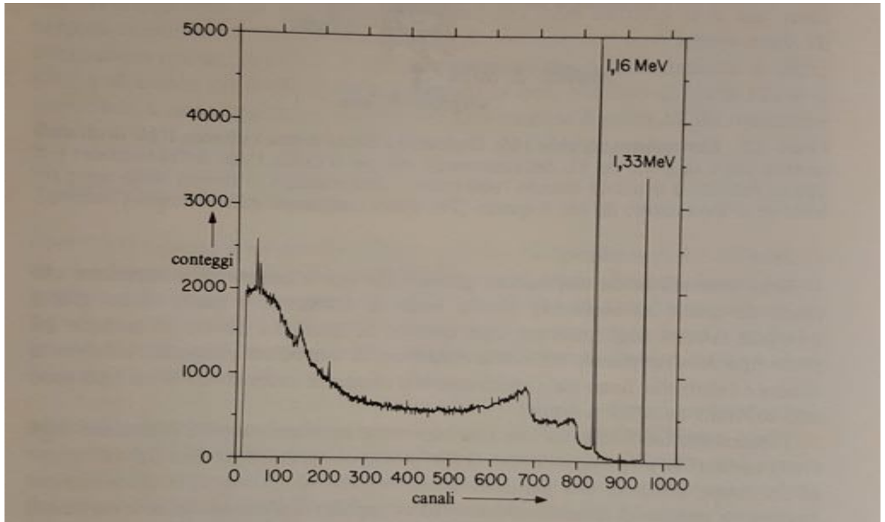
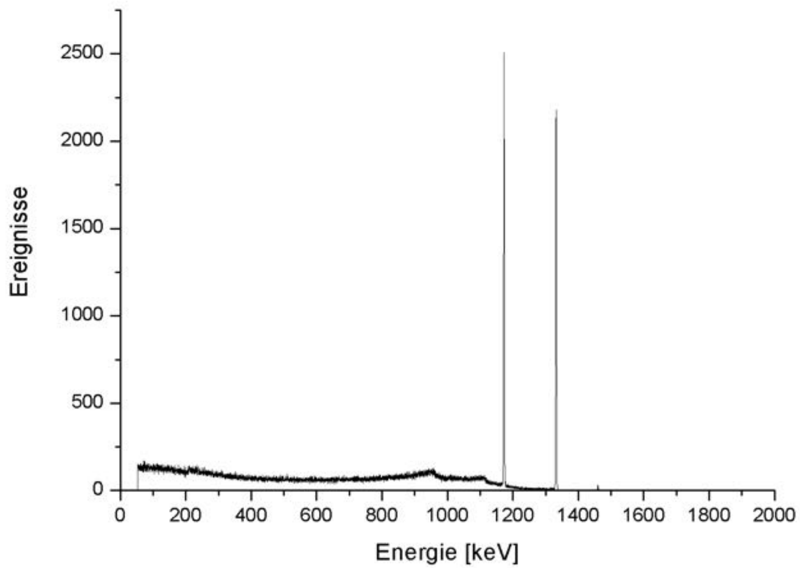
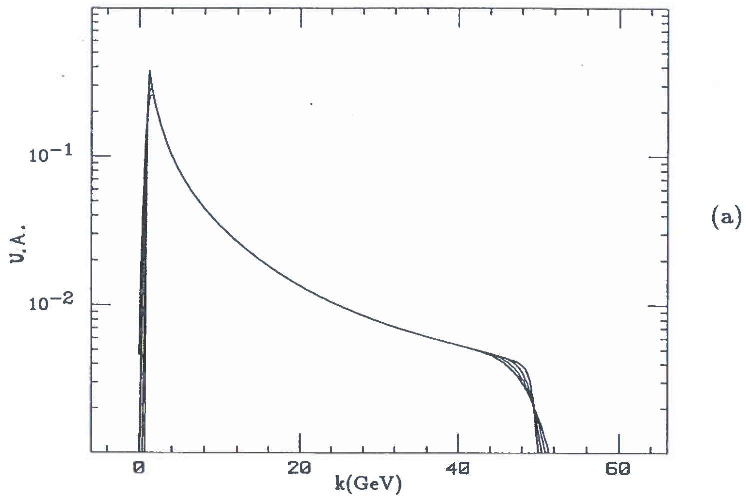


Figura 3.5. Spettro dei raggi gamma del ^{60}Co ottenuto con un rivelatore al germanio. Si notino i due picchi dell'energia totale dovuti ai raggi gamma di 1,16 MeV e di 1,33 MeV, e la distribuzione Compton ad energie inferiori a quelle dei picchi. Le discese nette nella curva ai canali 680 e 790 sono i cosiddetti salti Compton corrispondenti alla massima energia ceduta nell'urto agli elettroni dai raggi gamma di 1,16 MeV e 1,33 MeV. [Goulding e Stone, *Science*, **170**, 280, 1970].

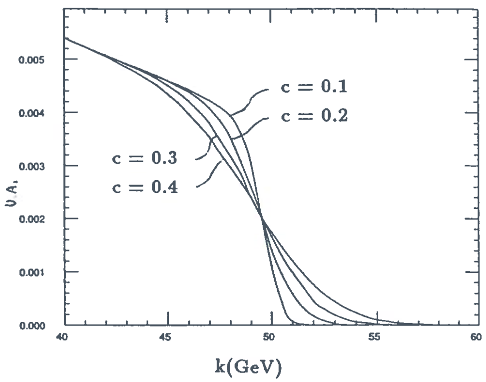


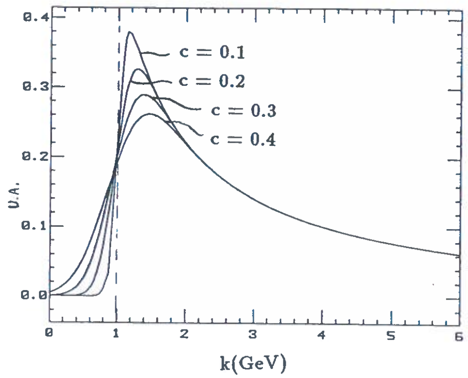
Single Bremsstrahlung photon spectrum at LEP

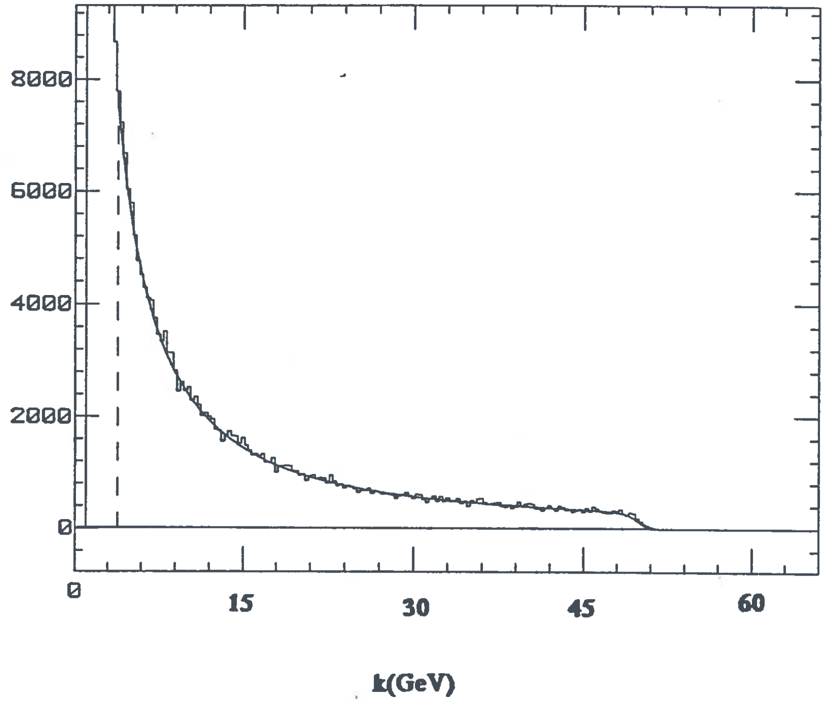


$$\dot{N}_i^{teo} = L' \int_{k_i}^{k_{i+1}} dk \int_0^{\infty} \frac{d\sigma}{dk'} g(k-k',c)dk'$$

$$L' = AL$$







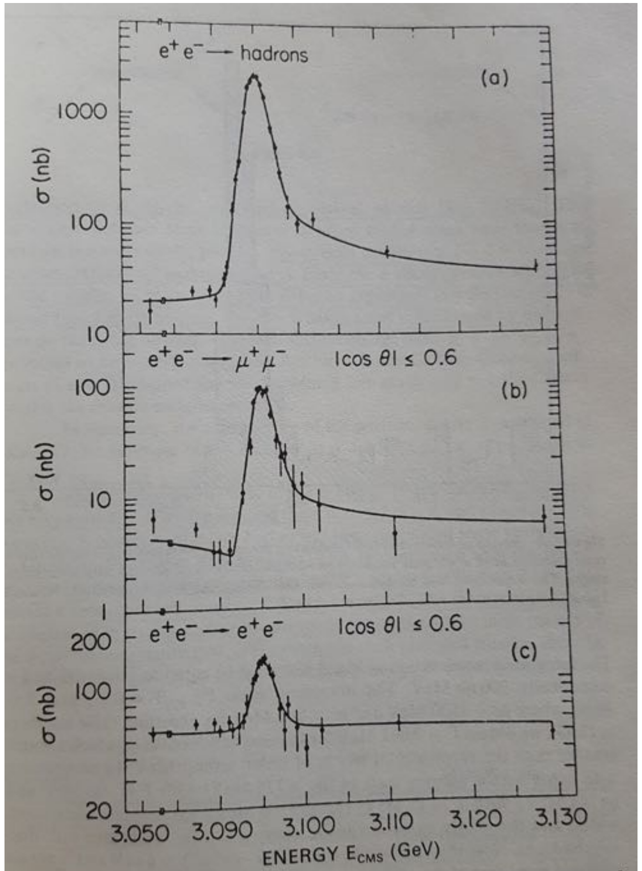


Figure 5.9 Results of Augustin *et al.* (1974) showing the observation of the ψ/J resonance of mass 3.1 GeV, produced in e^+e^- annihilation at the SPEAR storage ring, SLAC.

SLAC: $e^+e^- \to \psi \to$ hadrons
$\to e^+e^-, \mu^+\mu^-$

BNL: $p + Be \to \psi/J +$ anything
$\to e^+e^-$

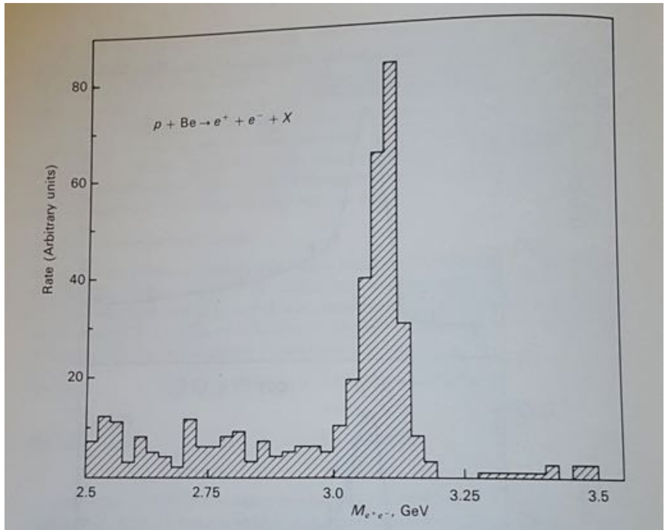


Fig. 5.11 Results of Aubert *et al.* (1974) indicating the narrow resonance ψ/J in the invariant-mass distribution of e^+e^- pairs produced in inclusive reactions of protons with a beryllium target. The experiment was carried out with the 28-GeV AGS at Brookhaven National Laboratory.

$J/\psi(1S)$ $I^G(J^{PC}) = 0^-(1^{--})$

Mass $m = 3096.900 \pm 0.006$ MeV
Full width $\Gamma = 92.9 \pm 2.8$ keV (S = 1.1)
$\Gamma_{ee} = 5.55 \pm 0.14 \pm 0.02$ keV

$J/\psi(1S)$ DECAY MODES		Fraction (Γ_i/Γ)	Scale factor/ Confidence level	p (MeV/c)
hadrons		(87.7 ± 0.5) %		–
virtual $\gamma \to$ hadrons		(13.50 ± 0.30) %		–
ggg		(64.1 ± 1.0) %		–
γgg		(8.8 ± 1.1) %		–
e^+e^-		(5.971 ± 0.032) %		1548
$e^+e^-\gamma$	[*rraa*]	(8.8 ± 1.4) $\times 10^{-3}$		1548
$\mu^+\mu^-$		(5.961 ± 0.033) %		1545

σ_E(BNL)~25 MeV
σ_E(SLAC)~ 2 MeV

Folding – Unfolding - I

- Folding: convolution integral
- Unfolding: e.g. by Fourier Transform techniques

(see later for folding-unfolding techniques using directly MC)

Fourier transformation and Convolution

Convolution Theorem:

assume $\mathcal{F}\{f(t)\} = F(u), \mathcal{F}\{h(t)\} = H(u)$

then $\mathcal{F}\{f(t) * h(t)\} = F(u)H(u)$

$$f(t) * h(t) \Leftrightarrow H(u)F(u)$$

$$f(t)h(t) \Leftrightarrow H(u) * F(u)$$

Proof

$$f(t) * h(t) = \int_{-\infty}^{\infty} f(x)h(t-x)dx$$

$$\mathcal{F}\{f(t) * h(t)\} = \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} f(x)h(t-x)dx]e^{-i2\pi ut}dt$$

$$= \int_{-\infty}^{\infty} f(x)[\int_{-\infty}^{\infty} h(t-x)e^{-i2\pi ut}dt]dx$$

$$= \int_{-\infty}^{\infty} f(x)[H(u)e^{-i2\pi ux}]dx = H(u)\int_{-\infty}^{\infty} f(x)e^{-i2\pi ux}dx = H(u)F(u)$$

Instrumental effects: importance of resolution

Resolution effect:
Smearing of spectrum structures,
i.e.enlarging peaks, smoothing sharp edges, filling holes or gaps

“Logic” of an EPP experiment: end of selection => candidate events

"Logic" of an EPP experiment - VI

- End of the selection: CANDIDATES sample N_{cand}
- Which relation is there between N_{cand} and N_X ?
  - ***Efficiency***: not all searched final states are selected and go to the candidates sample.(Trigger efficiencies are particularly delicate to treat.) Efficiency includes also the **acceptance.**
  - ***Background***: few other final states are faking good ones and go in the candidates sample.

$$\varepsilon N_X = N_{cand} - N_b$$

- where:
  - ε = efficiency ($0<\varepsilon<1$); $\boldsymbol{\varepsilon} = \boldsymbol{A} \times \boldsymbol{\varepsilon_d}$
  - N_b = number of background events
- Estimate ε and N_b is a crucial work for the experimentalist and can be done either using simulation (this is tipically done before the experiment and updated later) or using data themselves.

Counting

- So we do collisions at a given $\sqrt{s}$. What do we actually measure ?
- We "count" the number of times a final state is obtained. This frequency is somehow related to the probability of that final state and so it allows to measure the cross-section/decay width/branching ratios
- Connection btw probability and frequency:
  - Population → probability
  - Sample → frequency
- Sampling fluctuations

Random variables – Outline - I

- Concept of PDF
  - Meaning and connection to actual probabilities
  - Discrete vs. real variables
  - Single vs. multiple variables: factorization
- Definitions/properties
  - Physical dimension, positivity, normalization
  - Momenta → "functional"
  - Mean, variance, standard deviation, skewness, kurtosys
  - Covariance matrix
  - Propagation

Random variables – Outline - II

- The average and the RMS: two particular and interesting random variables, functions of random variables
- Few random variables which provide good statistical models of typical situations in experimental physics:
  - Binomial
  - Poissonian
  - Exponential
  - Gaussian
  - χ^2
- BUT: up to here only "populations"
- =>Statistical inference (see slides on Probability and Statistics: recap 1&2)

Binomial or Poissonian ?

- ***N*** initial states prepared ***n*** final states observed → inference on ***p***. So binomial ?Yes BUT:
- ***N*** is not known exactly
- If ***N*** → ∞ and ***p*** → ***0*** → **n follows a poissonian distribution** (easy to prove)

Quantities to be measured

- In order to estimate N_X we need to measure:
  - N_{cand}
  - ε
  - N_b
- We already know that each of these variables have a fluctuation model:
  - N_{cand} is described by a Poisson process
  - ε is described by a Bernoulli process
  - N_b

N_{cand}: a Poisson variable - I

- If events come in a random way (without any time structure) the event count N is a Poisson variable.
- ➔ if I count N, the best estimate of λ is N itself and the uncertainty is $\sqrt{N}$

$$E[\lambda] = N$$
$$\mathrm{var}[\lambda] = N$$

- If N is large enough ($N>20$) Poisson → Gaussian. ➔ $N \pm \sqrt{N}$ is a 68% probability interval for N.
- If N is small (close to 0) the Gaussian limit is not ok, a specific treatment is required (see later in the course).

N_{cand}: a Poisson variable - II

- If events come in a random way (without any time structure) the event count N is a Poisson variable.
- ➔ if I count N, the best estimate of λ is $N+1$ and the uncertainty is $\sqrt{N+1}$ *(Bayes' theorem, uniform prior)*

$$P(N,\lambda) = \lambda^N e^{-\lambda}/N! \Rightarrow P(\lambda \mid N) = \lambda^N e^{-\lambda}/N!$$

$$E[\lambda] = N+1$$

$$\mathrm{var}[\lambda] = N+1$$

- If N is large enough ($N>20$) Poisson → Gaussian. ➔ $N\pm\sqrt{N}$ is a 68% probability interval for N.
- If N is small (close to 0) the Gaussian limit is not ok, a specific treatment is required (see later in the course).

Efficiency: a binomial variable - I

- Bernoulli process: success/failure ***N*** proofs, 0<***n***<***N***, ***p*** = success probability. ***p*** == ***ε***

$$P(n/N,p) = \binom{N}{n} p^n (1-p)^{N-n}$$
$$E[n] = Np$$
$$\mathrm{var}[n] = Np(1-p)$$

- Inference: given ***n*** and ***N*** which is the best estimate of ***p*** ?

And its uncertainty ? *(see previous lectures)*

$$\varepsilon = \hat{p} = \frac{n}{N}$$
$$\sigma(\varepsilon) = \frac{\sigma(n)}{N} = \frac{1}{\sqrt{N}}\sqrt{\hat{p}(1-\hat{p})}$$

Efficiency: a binomial variable - II

- Bernoulli process: success/failure ***N*** proofs, 0<***n***<***N***, ***p*** = success probability. ***p*** == ***ε***

$$P(n/N,p) = \binom{N}{n} p^n (1-p)^{N-n}$$

$$E[n] = Np$$

$$\text{var}[n] = Np(1-p)$$

- Inference: given ***n*** and ***N*** which is the best estimate of ***p*** ?

And its uncertainty ? *(see previous lectures)*

$$\varepsilon = \hat{p} = \frac{n+1}{N+2}$$

(Bayes' Theorem, uniform prior)

$$\sigma(\varepsilon) = \frac{\sigma(n)}{N} = \frac{1}{\sqrt{N+2}}\sqrt{\hat{p}(1-\hat{p})}$$

Efficiency: a binomial variable - III

- How measure it ?
  - From data: Sample of N true particles and I measure how many, out of these give rise to a signal in my detector
  - From MC: I generate N_{gen} "signal" events. If I select N_{sel} of these events out of N_{gen}, the efficiency is (assume N_{gen} and N_{sel} large numbers):

$$\varepsilon = \frac{N_{sel}}{N_{gen}}$$

$$\sigma(\varepsilon) = \frac{\sigma(N_{sel})}{N_{gen}} = \frac{1}{\sqrt{N_{gen}}}\sqrt{\frac{N_{sel}}{N_{gen}}\left(1-\frac{N_{sel}}{N_{gen}}\right)}$$

Background N_b

- Simulation of N_{gen} "bad final states"; N_{sel} are selected. What about N_b ?
- We define the "rejection factor" $R = N_{gen}/N_{sel} > 1$
- We also need a correct normalization in this case: we need to know N_{exp} = total number of expected "bad final states" in our sample (N_{exp} related to luminosity and cross-section).

$$N_b = N_{sel}\frac{N_{\exp}}{N_{gen}} = \frac{N_{\exp}}{R}$$

$$\sigma(N_b) = \sigma(N_{sel})\frac{N_{\exp}}{N_{gen}} = \sqrt{N_{sel}}\frac{N_{\exp}}{N_{gen}} = \frac{N_{\exp}}{\sqrt{RN_{gen}}}$$

Statistical Errors

- In all cases there is an irreducible error on $\boldsymbol{N_X}$ given by limited statistics. It is a random error, coming from the procedure of "sampling" that is intrinsic in our experiments.
- In all cases increasing the statistics, the error decreases

$$\frac{\sigma(N_{cand})}{N_{cand}} = \frac{1}{\sqrt{N_{cand}}}$$

$$\sigma(\varepsilon) \approx \frac{1}{\sqrt{N_{gen}}}$$

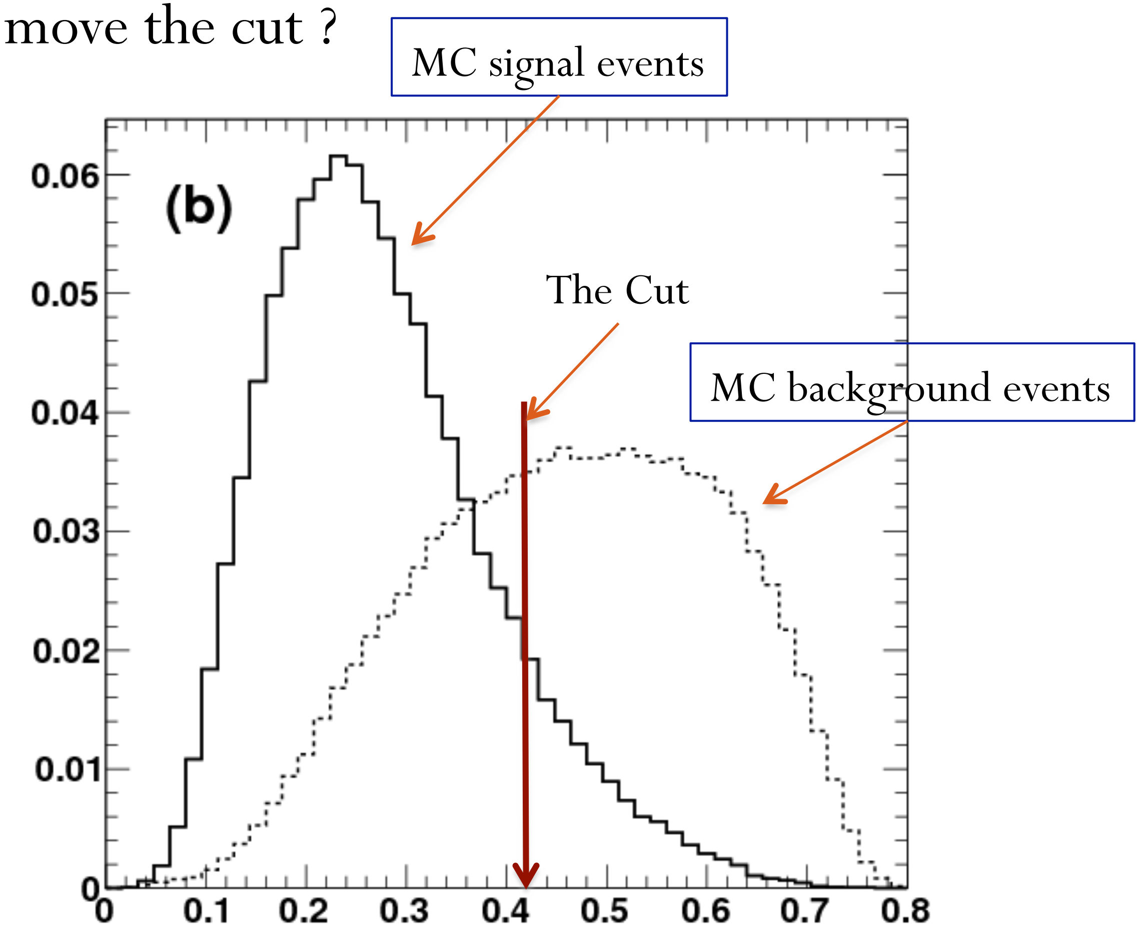
$$\sigma(N_b) \approx \frac{1}{\sqrt{N_{gen}}}$$

Summarizing

- N_{cand}: poissonian process ➔ the higher the better
- ε: binomial process ➔ high N_{gen} and high $\boldsymbol{\varepsilon}$
- N_b: normalized ≈poissonian process ➔ high R and high N_{gen}, low N_{exp}
- Moreover: unfortunately efficiency and background cannot be both improved simultaneously…

Efficiency vs. background

What happens if I move the cut ?



Efficiency-background relation

Example: selection of b-jets in ATLAS.

"b-jet" is the signal;

"light jet" is the background.

MC samples of ***b-jets*** and ***light-jets***

Application of 5 different ***selection recipes***

each with a "***free-parameter***".

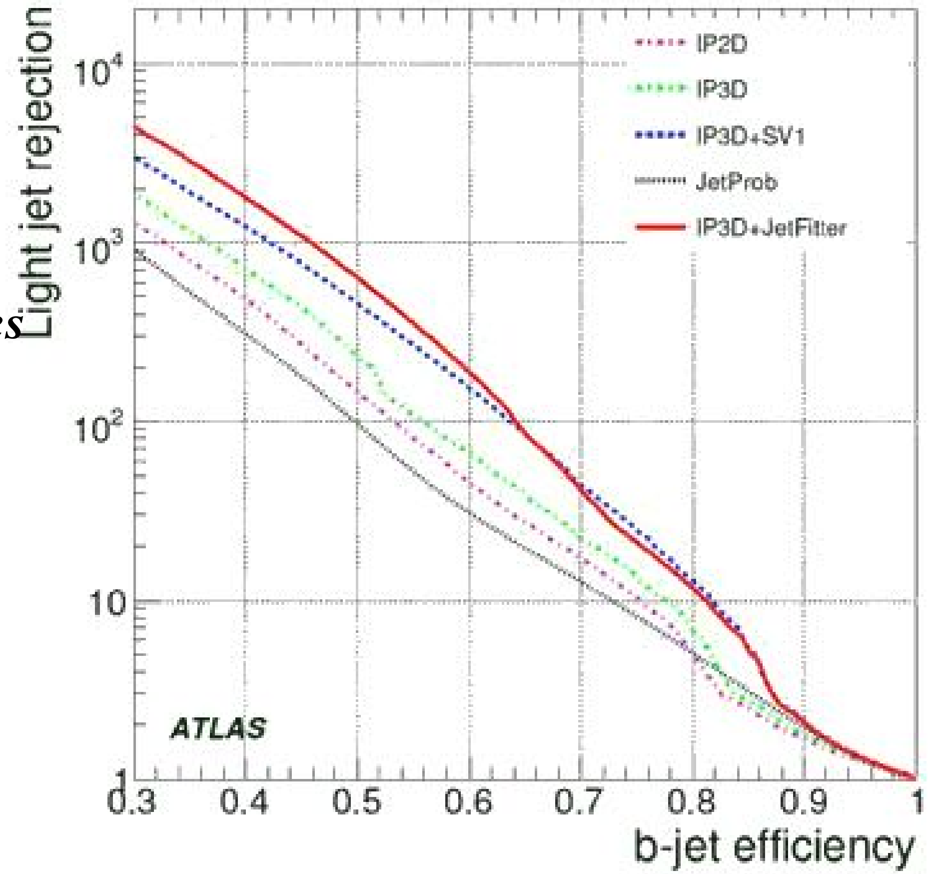
For each point I evaluate

- b-jet efficiency

  $= N_{sel}/N_{gen}$ (b-jet sample)

- light-jet rejection

  $= N_{gen}/N_{sel}$ (light-jet sample)



Choice of a working point, "compromise".

Unlucky situation: if you gain in efficiency you increase your bckg and viceversa…

Combining uncertainties

- Given the uncertainties on $\boldsymbol{N_{cand}}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{N_b}$, how can we estimate the uncertainty on $\boldsymbol{N_X}$?
- ➔ Uncertainty Propagation. General formulation

$$\left(\frac{\sigma(N_X)}{N_X}\right)^2 = \left(\frac{\sigma(\varepsilon)}{\varepsilon}\right)^2 + \frac{\sigma^2(N_{cand}) + \sigma^2(N_b)}{(N_{cand} - N_b)^2}$$

Assumption: three indipendent contributions

NB: if $\boldsymbol{N_{cand}} \approx \boldsymbol{N_b}$ the relative uncertainty becomes very large (the Formula cannot be applied anymore…)

Can we say we have really observed a signal ???

Or we are simply observing some fluctuation of the background ?