A Measurement of the Gluon Splitting Rate into $c\bar{c}$ Pairs in Hadronic Z Decays with the ALEPH detector

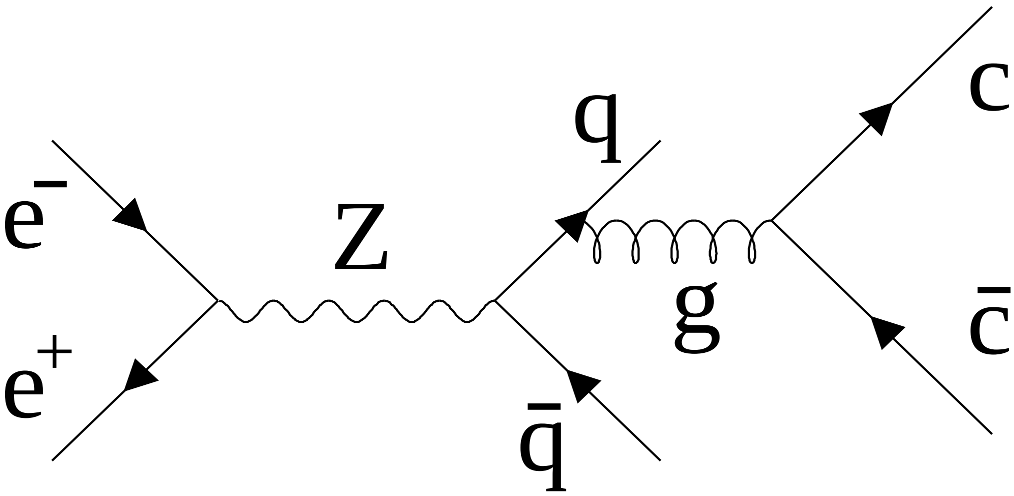
Andrea Giammanco

Scuola Normale Superiore Pisa

- The process
- The theory behind
- $g \rightarrow c\bar{c} \rightarrow l\nu X$
- Event selection
- Background composition
- Neural networks
- Extraction of the $g_{c\bar{c}}$ rate

The physical process:

$$Z \to q\bar{q}\, g \quad , \quad g \to c\bar{c}$$



$$\boxed{g_{c\bar{c}} = \frac{\text{N. of events with } g \to c\bar{c}}{\text{Total number of } Z \to q\bar{q}}}$$

The Standard Model particle content:

Leptons			Quarks		
Name	Charge/e	Mass	Name	Charge/e	Mass
e	-1	511 KeV	u	$+2/3$	$2 \div 8$ MeV
ν_e	0	< 3 eV	d	$-1/3$	$5 \div 15$ MeV
μ	-1	105 MeV	c	$+2/3$	$1.0 \div 1.4$ GeV
ν_μ	0	< 0.19 MeV	s	$-1/3$	$100 \div 300$ MeV
τ	-1	1.77 GeV	t	$+2/3$	$167 \div 181$ GeV
ν_τ	0	< 18.2 MeV	b	$-1/3$	$4.0 \div 4.5$ GeV

Name	Charge/e	Mass	Spin
photon (γ)	0	0	1
W^+	$+1$	80 GeV	1
W^-	-1	80 GeV	1
Z	0	92 GeV	1
gluon (g)	0	0	1
graviton (G)	0	0	2

The **Standard Model** is a gauge field theory based on the group

$$G = SU(3)_c \times SU(2)_L \times U(1)_Y$$

with a spontaneous breakdown of symmetry (via the *Higgs mechanism*) which leaves $SU(3)_c \times U(1)_{em}$ as "visible" symmetry.

The only missing particle is the Higgs boson.

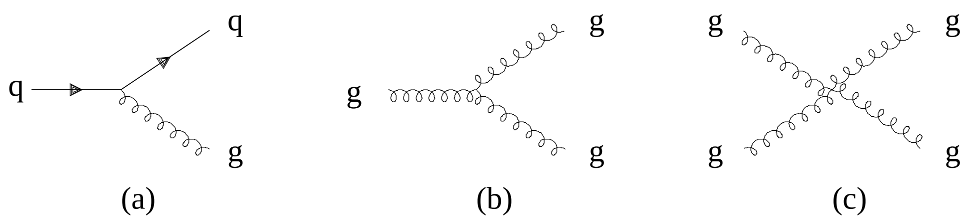
Strong interaction:

$$\mathcal{L}_{QCD} = \bar{q}_a(i\gamma^\mu D_\mu - m_q)_{ab} q_b - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu}$$

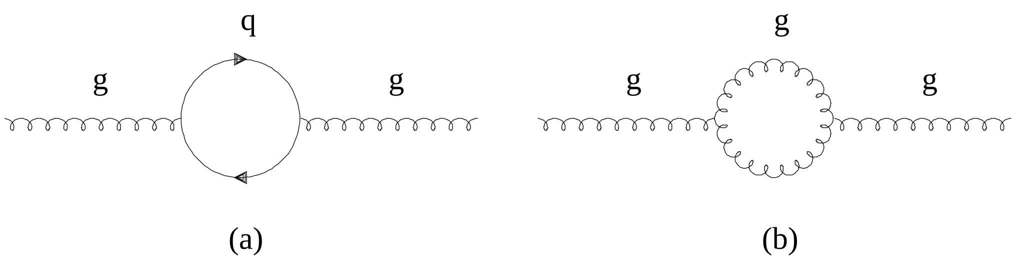
QCD is a non-abelian field theory:

$$F^A_{\alpha\beta} = [\partial_\alpha \mathcal{A}^A_\beta - \partial_\beta \mathcal{A}^A_\alpha - g_s f^{ABC} \mathcal{A}^B_\alpha \mathcal{A}^C_\beta]$$

This implies gluon self-interaction:

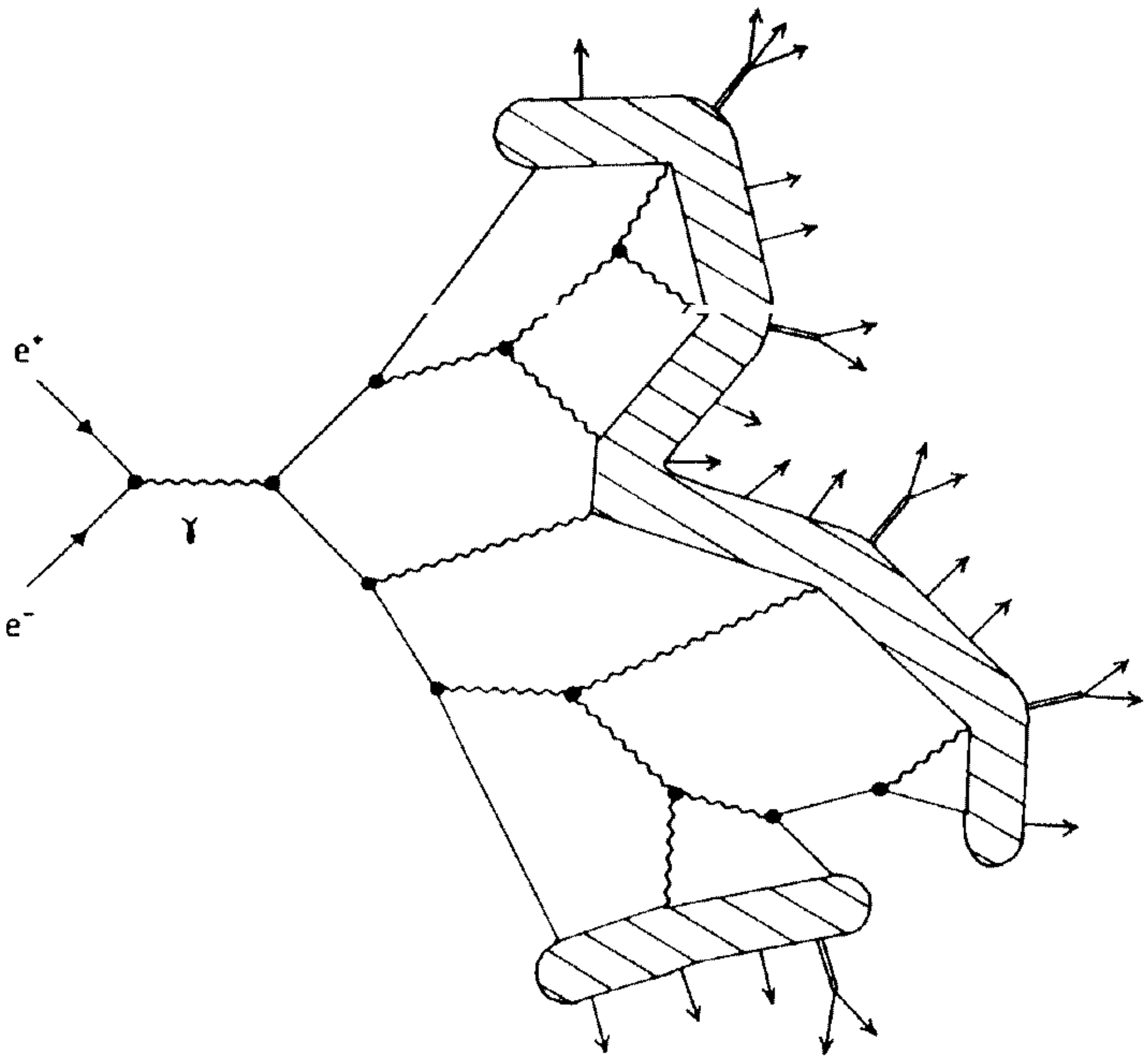


This makes $\alpha_s \equiv \frac{g_s^2}{4\pi}$ a **decreasing** function of Q^2, due to the gluon loop:



Hadronization:

The least understood aspect of strong interactions is the passage from partons (quarks and gluons) to hadrons.



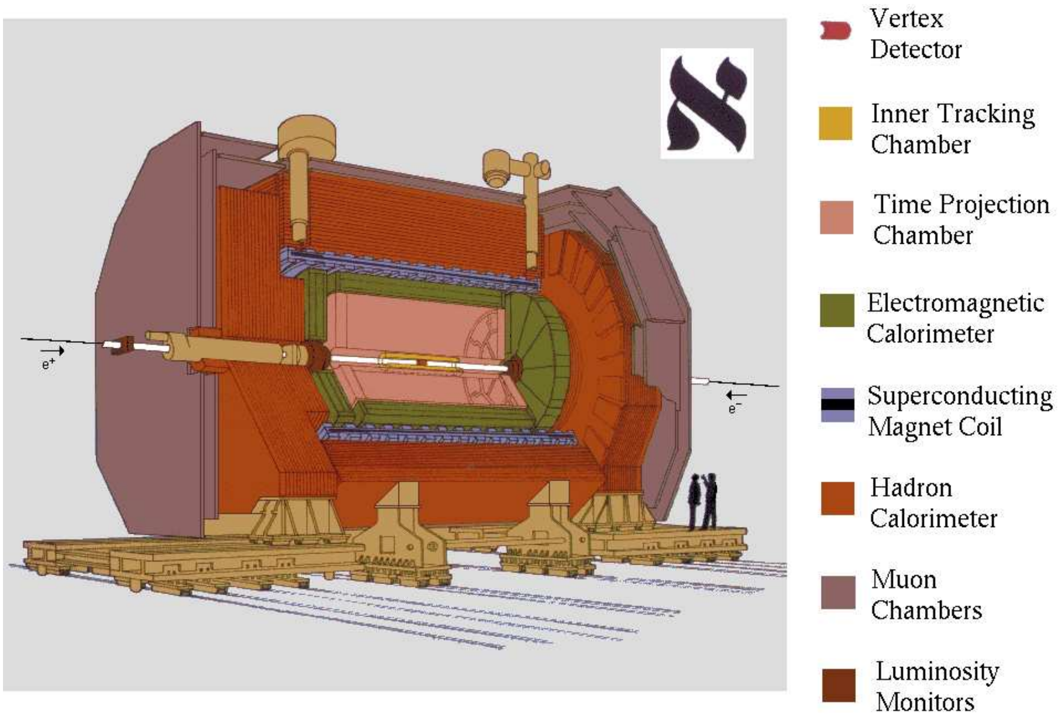
Perturbative QCD breaks down at large distances (small momenta).

Hadronization is a **non-perturbative** effect, to be treated via **phenomenological models** (e.g. *string models*...).

LEP & ALEPH:

LEP was an e^+e^- symmetric collider.
It created almost 5 million Z's
between 1989 and 1995.

ALEPH was a multipurpose "4π" detector
built around one of the beam crossing points.



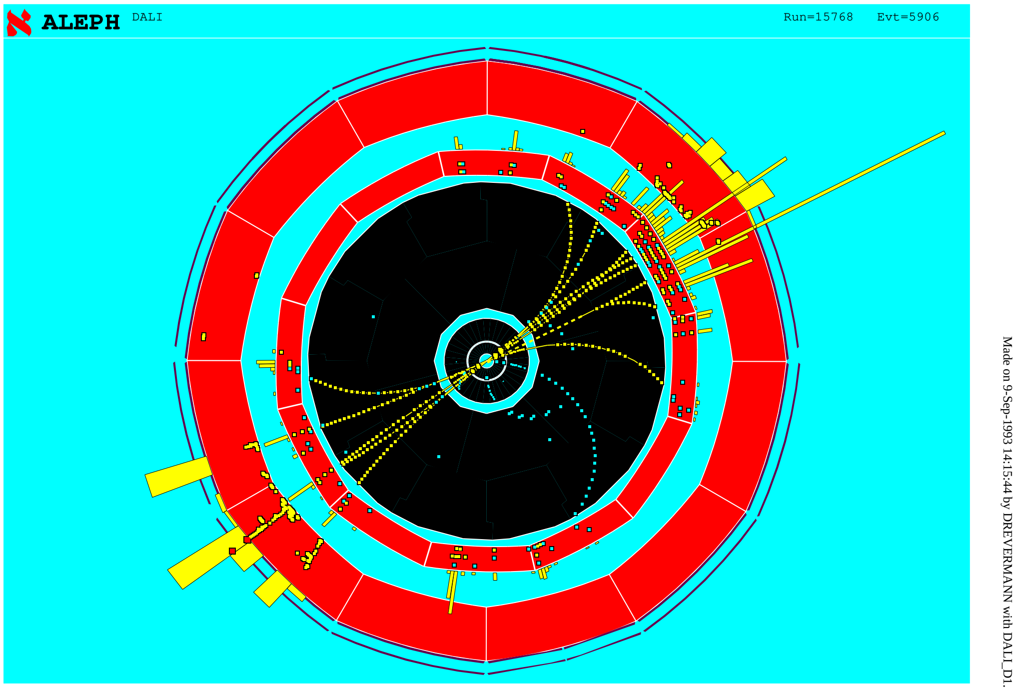
The ALEPH Detector

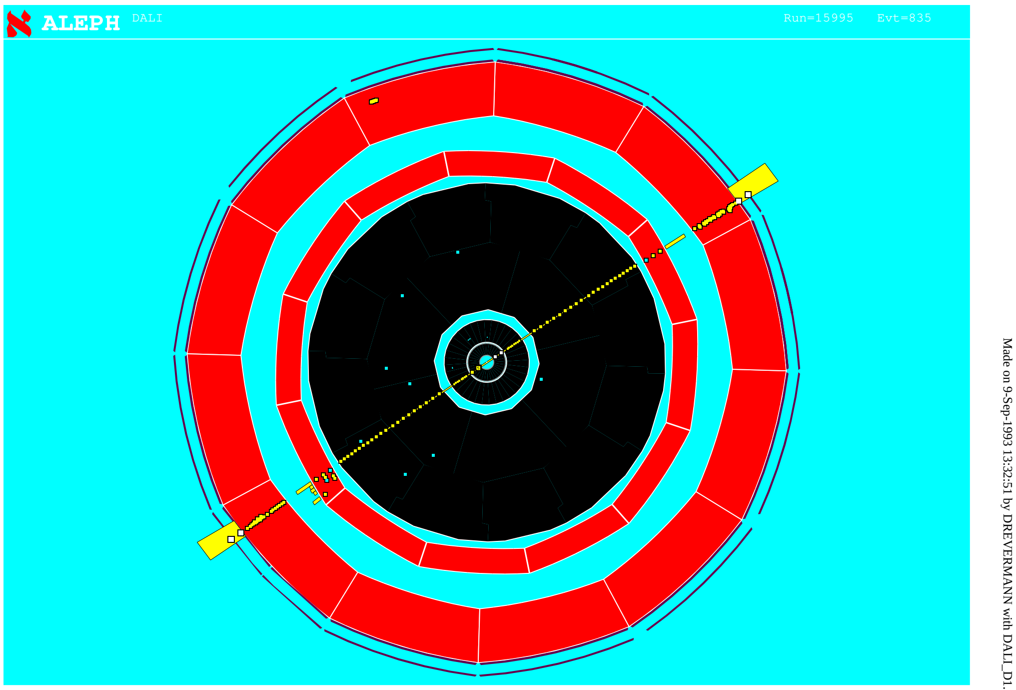
Subdetectors: VDET, ITC, TPC,
ECAL, HCAL, μ-chambers
(They will be described in the following)

Z peak:

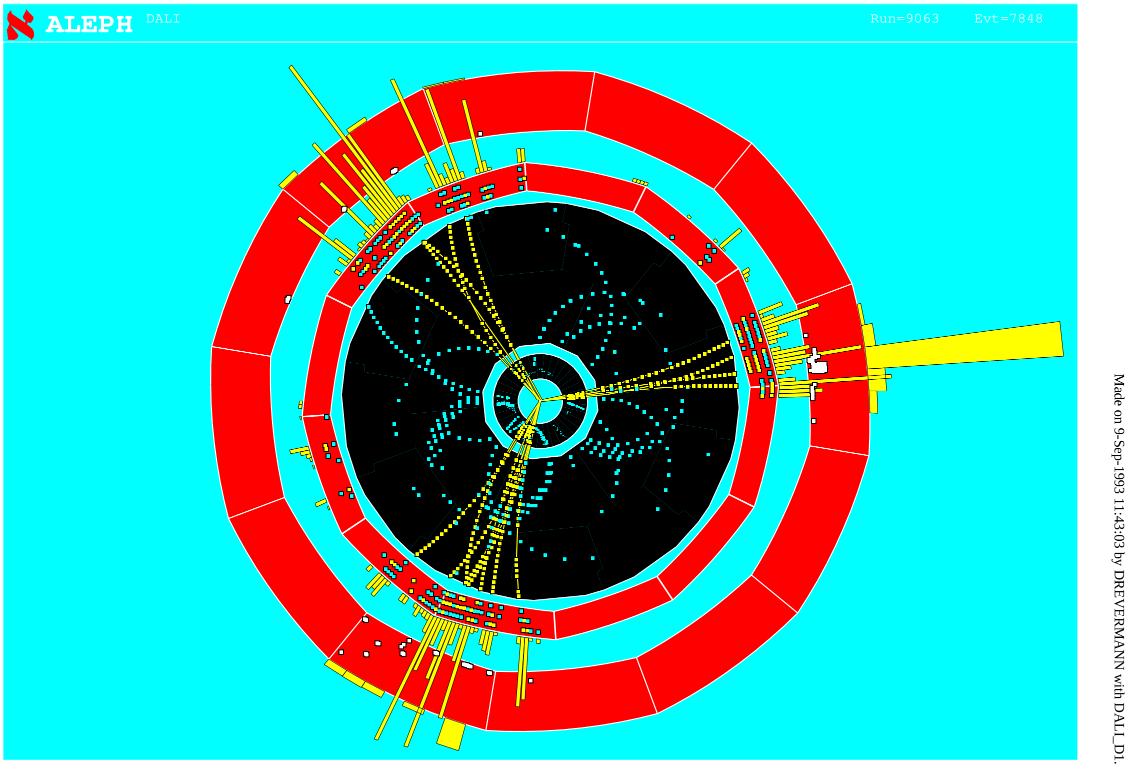
At the peak (**91.2 GeV**),
$\sigma(e^+e^- \rightarrow Z \rightarrow anything) \approx 40$ nb

70% $Z \rightarrow hadrons$, 10% $Z \rightarrow l^+l^-$,
20% $Z \rightarrow \nu\bar{\nu}$





Multi-jet events:



This event looks like $Z \longrightarrow q\bar{q}g$ ("hard" g).

Ideally, one primary parton $\Leftrightarrow$ one jet.

In practice, **there is no univoque definition of jet**.

In my analysis, the **JADE algorithm** has been chosen.

1. take the particles i and j having the minimum invariant mass $M_{ij}^2 = 2E_iE_j(1 - \cos\theta_{ij})$ (null mass approx.);
2. if $(M_{ij}/E_{vis})^2 < y_{cut}$, merge i and j;
3. repeat from #1 until no couple fulfils #2.

Theoretical predictions for $g_{c\bar{c}}$

The framework is **perturbative QCD**
($\alpha_s \approx 0.12$ in this energy domain)

- The heavy quark mass provides a natural infrared cut-off
- Tree level + leading logarithmic calculations are available
- The leading logarithmic terms are large
- No next-to-leading order calculation is available

Leading Order calculation

At LO, $g_{c\bar{c}}$ can be exactly calculated:

$$\begin{aligned} g_{c\bar{c}} \;=\; & \left(\tfrac{\alpha_s}{4\pi}\right)^2 \tfrac{4}{3} C_F \int_{x^2}^{1} dz \left\{ \tfrac{1}{z}\sqrt{1-\tfrac{x^2}{z}}\left(1+\tfrac{x^2}{2z}\right) \times \right. \\ & \times \left[(1+z)^2 \log^2 z + 5 - 5z^2 + (3+4z+3z^2)\log z\right] + \\ & + \tfrac{\log z}{1+z}\left[\sqrt{1+\tfrac{x^2}{z}}\left(-\tfrac{20}{3} - 8x^2 - \tfrac{4x^2}{3z} + \right.\right. \\ & \left. + 8z + x^2 z + 2z^2\right) - (8-3x^4)\log x + \\ & \left.\left. + (8-3x^4)\log(\sqrt{z}+\sqrt{z-x^2})\right]\right\} \end{aligned}$$

For $x \equiv \dfrac{2m_c}{\sqrt{s}} \approx 0$ this can be written as:

$$\begin{aligned} g_{c\bar{c}} \;=\; & \left(\tfrac{\alpha_s}{4\pi}\right)^2 C_F \tfrac{4}{3} \left\{ -\tfrac{1}{3}\log^3 x^{*2} - \tfrac{3}{2}\log^2 x^{*2} - \right. \\ & \left. - \left(\tfrac{76}{9} - \tfrac{2}{3}\pi^2\right)\log x^{*2} + 2.359\ldots \right\} \end{aligned}$$

where $x^* \equiv \frac{1}{2} e^{5/6} x \approx 1.15\ x$.

These results hold assuming that interference between $Z \to q\bar{q}g \to q\bar{q}c\bar{c}$ and $Z \to c\bar{c}g \to c\bar{c}q\bar{q}$ is negligible.

Luckily this is the case, by a consequence of **Furry's theorem**.

Resummation

Every time two widely different mass scales m_1, m_2 are present, large logarithmic terms like $[\alpha_s \log^2(m_2/m_1)]^n$ appear to **all orders** in perturbative expansion, spoiling the reliability of the LO results.

Intuitively: for every gluon radiated one has to consider the collinear and the soft gluon limits, each contributing with a logarithm.

In the present case, the relevant mass scales of the problem are m_c and M_Z.

Several techniques exist to "resum" these large logarithms.

This is usually done in terms of **evolution equations** that describe how the quantity of interest changes with the mass scale.

Another approach is to formulate the evolution as a **branching chain**, which can be simulated numerically using Monte Carlo methods, choosing an appropriate **evolution variable** (e.g. Q^2, or the opening angle between partons produced in a branching).

Numerical results:

After the matching of the resummed calculation and of the LO (exact) result, theoretical calculations predict:

$$g_{c\bar{c}} = 2.3 \times 10^{-2} \quad (m_c = 1.2 \text{ GeV}, \alpha_s = 0.120)$$
$$g_{c\bar{c}} = 1.7 \times 10^{-2} \quad (m_c = 1.5 \text{ GeV}, \alpha_s = 0.120)$$
$$g_{c\bar{c}} = 2.0 \times 10^{-2} \quad (m_c = 1.2 \text{ GeV}, \alpha_s = 0.118)$$

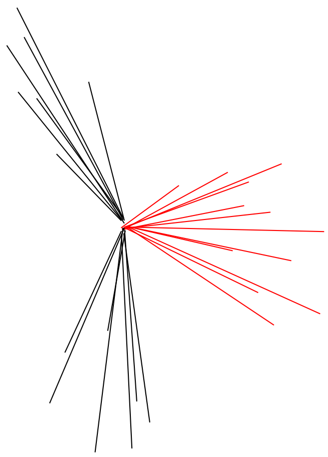
($15 - 30\%$ uncertainty due to the truncation of the perturbation series)

References:

M.H. Seymour, *Heavy quark pair multiplicity in e^+e^- events*, 1995, Nuclear Physics B 436, 163–183

D.J. Miller, M.H. Seymour, *Secondary heavy quark pair production in e^+e^- annihilation*, Phys.Lett. **B435** (1998) 213

General strategy:



$q\bar{q}g/q\bar{q}$: three jets topology; the gluon jet is usually the lowest momentum jet.

bc/uds: presence of a "prompt" lepton in the "gluon jet" ($BR(c \to l\nu X) \approx 10\%$).

$c/udsb$: variables related to the presence of vertices; c has intermediate properties between uds and b.

Backgrounds:

For $Z^0 \to q\bar{q}g, g \to Q\bar{Q}$, the main background is $Z^0 \to Q\bar{Q}g, g \to q\bar{q}$.

$g \to b\bar{b}$ is a background for $g \to c\bar{c}$.

π,K,$\pi \to \mu$,$K \to \mu$,$\gamma \to e^+e^-$ can fake a "prompt" lepton ($\Rightarrow uds$ contamination).

Event selection:

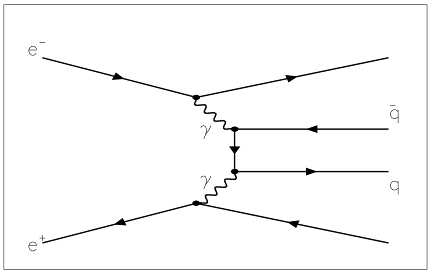
- Standard hadronic selection (3.9 M events); gives the denominator of the $g_{c\bar{c}}$ ratio.
- Preselection:
  - thrust cut
  - clusterization forced in 3 jets (ordered by momentum)
  - rejection of non-planar events ($\Leftrightarrow$ sum of jet angles $\approx 360°$)
  - request of one prompt lepton track (e, μ) in the third jet, with 1 or 2 VDET hits and a d_0 cut.
- Cut on the output of a Neural Network based on:
  - lepton properties
  - event topology
  - b-tagging variables for hemispheres and lepton track
  - missing momentum along the third jet

Hadronic selection:

At least 5 "good" tracks in the TPC.

"Good" track means:

- At least 4 TPC coordinates.
- The track goes through a cylinder of radius 2 cm and lenght 10 cm surrounding the beam axis in the crossing point.
  (This rejects spurious tracks and cosmics.)
- Angle between the track and the beam axis larger than 18.2° ($\Leftrightarrow |cos\theta| < 0.95 \Leftrightarrow > 6$ TPC rows of sense wires).



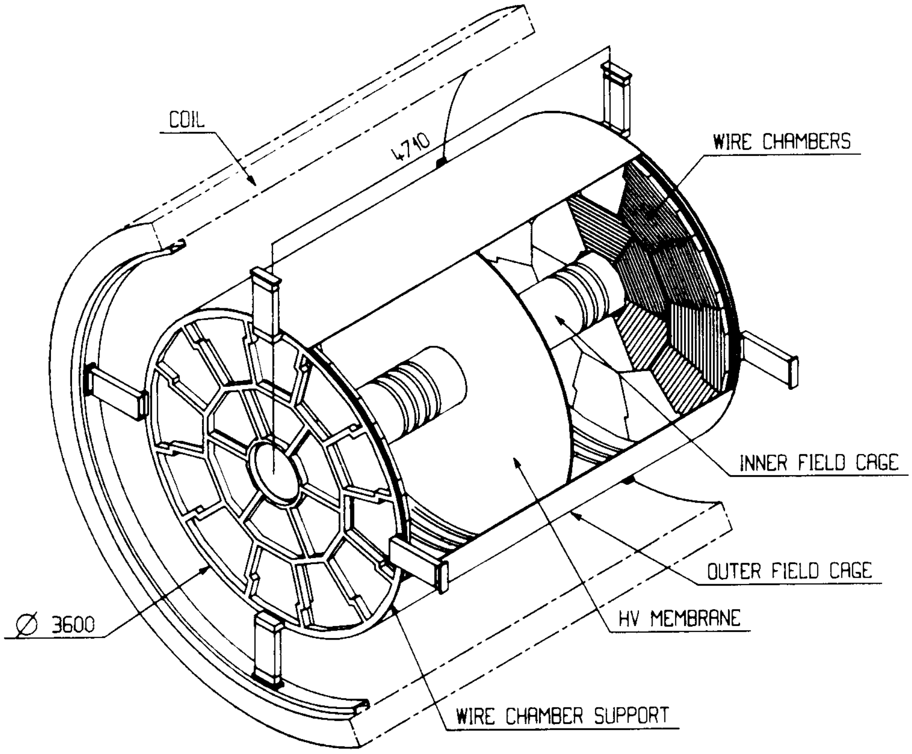
To further reject $\gamma\gamma$ background and interactions of the e^+e^- beams with gas molecules in the LEP pipe,

$$E_{charged} > 0.1 E_{CM}$$

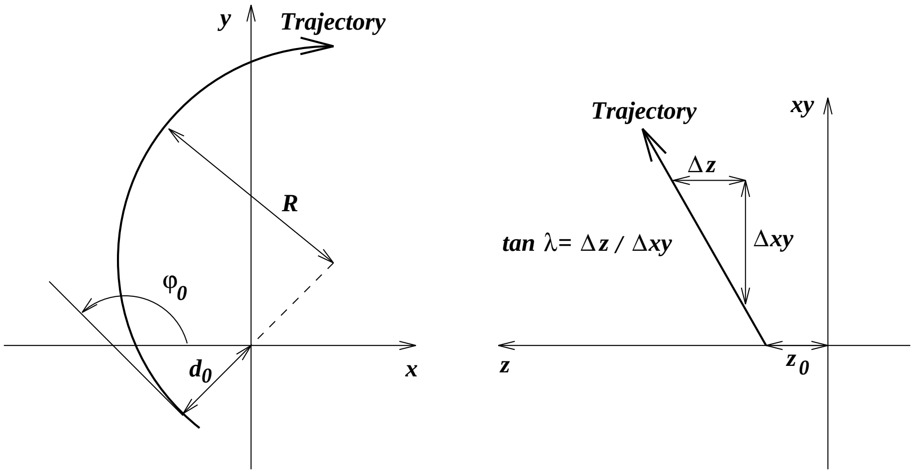
Selection efficiency
for hadronic events: 97.5%

$\gamma\gamma$ **and** $Z \rightarrow \tau^+\tau^-$ **contamination:**
$< 0.3\%$ **of the selected sample**

Track reconstruction:



The first step is the reconstruction in the TPC, with a fit to a helix (5 parameters).



Then extrapolation (and refitting) is made to the ITC, and then to VDET.

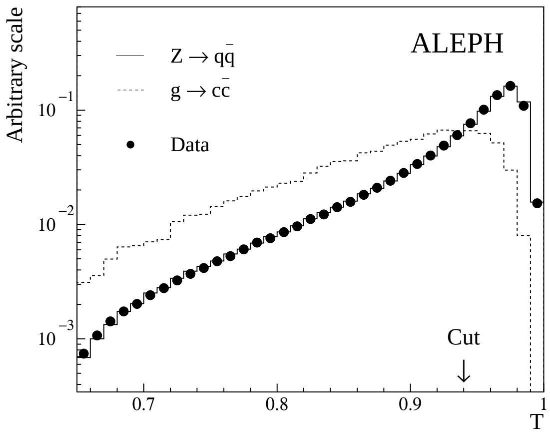
$$p_T = qBR \quad , \quad p = p_T/\sin\theta$$

Thrust:

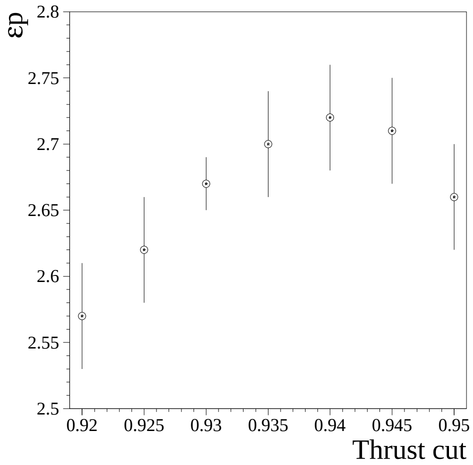
$$T = \max_{\vec{n}} \left(\Sigma_i \frac{\vec{p_i} \cdot \vec{n}}{|\vec{p_i}|} \right)$$

T is related to the event **“sphericity”**:

- 2 back-to-back tracks: $T = 1$
- 3 tracks: $2/3 \leq T \leq 1$
  (collinear: $T = 1$; 60°: $T = 2/3$)
- 4 tracks: $1/\sqrt{3} \leq T \leq 1$
- In general: $1/2 < T \leq 1$
  (∞ isotropic tracks: $T = 1/2$)



(a)



(b)

Preselection cut: $T < 0.94$

Clusterization in jets:

The usual way to apply the JADE algorithm is by choosing a y_{cut} value giving the maximum invariant mass of the jets.

This analysis makes use of a different algorithm, inspired by JADE, which doesn't make use of y_{cut}:

1. take the particles i and j having the minimum invariant mass $M_{ij}^2 = 2E_iE_j(1 - \cos\theta_{ij})$ (null mass approx.);
2. if $N_{jets} > 3$, merge i and j;
3. repeat from #1 until only 3 jets are left.

Jet momenta (defined as $\vec{p}_{jet} = \sum_a \vec{p}_a$) have relatively poor resolution.

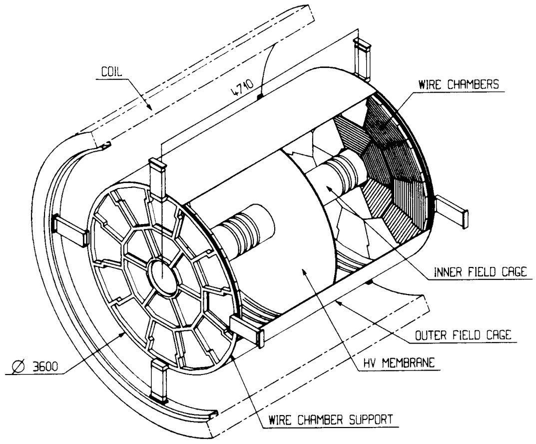
Jet momenta are re-defined:

$$p_i = E_{cm}\frac{\sin\psi_{jk}}{\sin\psi_{jk}+\sin\psi_{ij}+\sin\psi_{ik}}$$

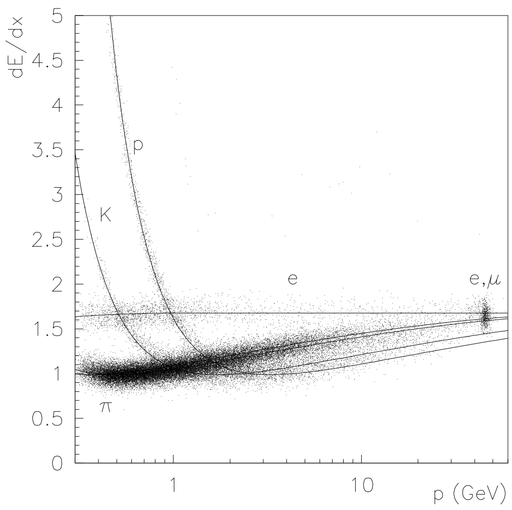
This formula holds for massless particles and planar events. Events with $\psi_{12} + \psi_{23} + \psi_{31} < 358°$ are rejected.

Electron identification:

Ionization in the TPC and electron shower in ECAL.

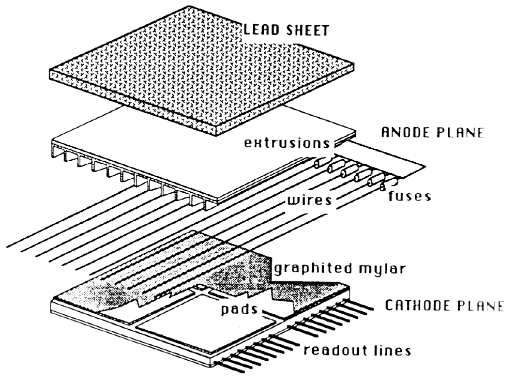


(c)

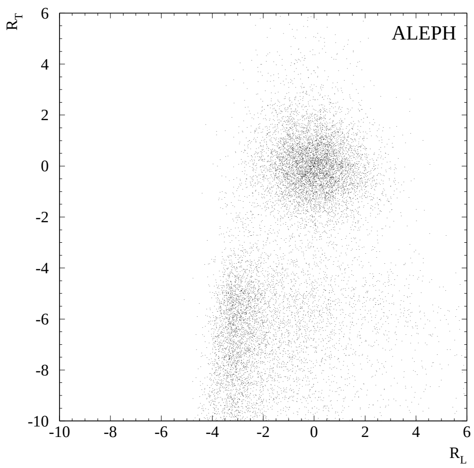


(d)

$$R_I \equiv \frac{(dE/dx) - \overline{dE/dx}_{el}}{\sigma_{dE/dx}} > -2.0$$



(e)



(f)

$$-1.8 < R_L \equiv \frac{X_L^{-1} - \overline{X_L^{-1}}_{el}}{\sigma_{X_L^{-1}}} < 3.0 \quad , \quad R_T \equiv \frac{E_4/p - \overline{E_4/p}_{el}}{\sigma_{E_4/p}} > -1.6$$

X_L: depth of the maximum energy release along the shower
E_4: energy released in the four ECAL segments nearest to the extrapolated track

Muon identification:

Small energy loss in HCAL, deep penetration.

A *road* is defined around the extrapolated tracks, wide $3\sigma_{ms}$.

$\sigma^{\mu}_{ms}(45 \text{ GeV}) = 3$ cm , $\sigma^{\mu}_{ms}(3 \text{ GeV}) = 15$ cm

The extrapolated track is tagged as μ if:

- it intersects at least 10 planes in HCAL;
- > 4 planes give signal inside the *road* in the last 10 expected planes;
- the ratio between the number of planes which give signal in the *road* and the total number of expected ones is ≥ 0.4.

The cut $X_{mult} < 1.5$ is applied, where X_{mult} is the number of hits within 30 cm from the extrapolated track in the last 11 HCAL planes divided by the number of planes giving signal.

In the Muon Chambers, $< 4\sigma_{ms}$ between the extrapolated track and a reconstructed hit.

Other lepton preselection cuts:

"Prompt" leptons have harder spectra than background. **Momentum cut:**

$$p_e > 2.0 \text{ GeV} \quad , \quad p_\mu > 2.5 \text{ GeV}$$

For b-tagging we need a reliable impact parameter information on the lepton track, so at least one VDET hit is required.

Detector	$r\phi$ (μm)	rz (μm)
TPC	310	808
TPC+ITC	107	808
TPC+ITC+VDET	23	28

Distance in the $x - y$ plane from the interaction point is larger for non-prompt leptons ($\pi, K \rightarrow \mu\nu$; $\gamma \rightarrow e^+e^-$):

$$d_0 < 0.1 \text{ cm}$$

Sample composition:

Before any selection (3.9 M hadronic events):

$g \to c\bar{c}$	2.96 %
$g \to b\bar{b}$	0.254 %
uds	58.8 %
c	16.5 %
b	21.4 %

After preselection (5639 e, 7724 μ):

	e	μ
$g \to c\bar{c}$	17.0 ± 0.3 %	12.8 ± 0.3 %
$g \to b\bar{b}$	3.4 ± 0.2 %	2.5 ± 0.1 %
uds	10.9 ± 0.3 %	27.7 ± 0.4 %
c	21.5 ± 0.4 %	20.0 ± 0.3 %
b	47.2 ± 0.4 %	37.0 ± 0.4 %

uds are a serious trouble only for μ:

Muon candidates	
prompt μ	56%
$\pi \to \mu\nu$	25%
misidentified π	10%
$K \to \mu\nu$	5%
misidentified K	3%
others	1%

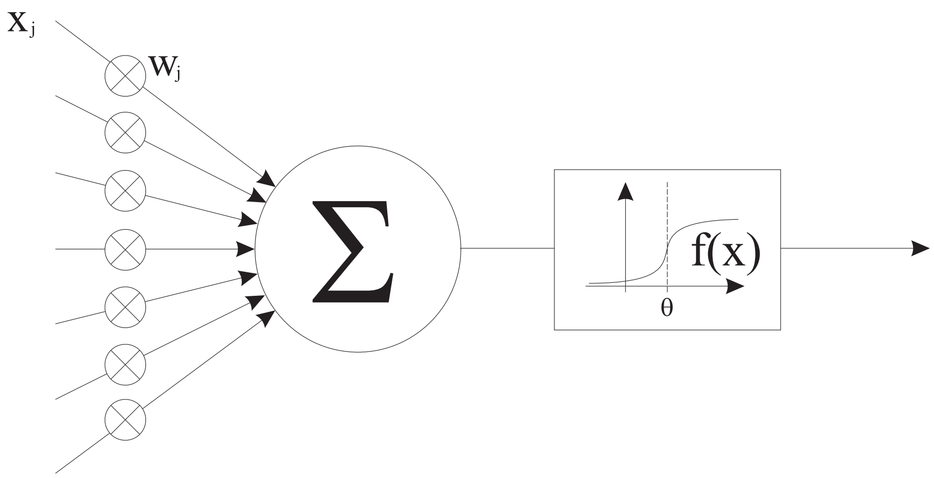
Electron candidates	
prompt e	69%
$\gamma \to ee$	26%
$\pi \to e\nu$	4%
others	1%

Neural Networks:

In principle, neural networks can compute any *computable function* (i.e. they can do everything a digital computer does), under some assumptions.

In practice, NNs are especially useful for classification and function approximation/mapping problems which are tolerant of some imprecision, which have lots of training data available, but to which hard and fast rules cannot easily be applied.

The basic unit is the **neuron**, or "Threshold Logic Unit":

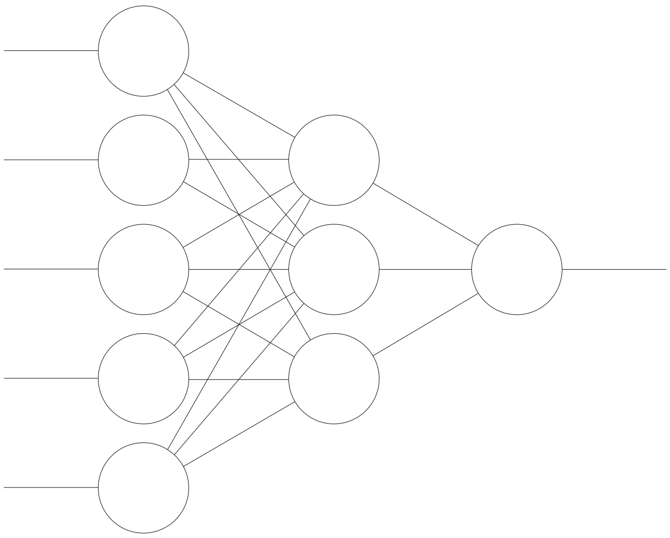


The simplest NN is the "perceptron", constituted by a single neuron.

NNs can have any topology suitable to the problem to which are tailored (bi- or three-dimensional, layered or unlayered...)

Multi-Layer Perceptrons:

The MLP is an example of "feedforward" network, i.e. the information flows only in one direction, without cycles.



MLP's structure consists of:

- one **"input layer"**: neurons have one single input each from the environment
- one **"output layer"**: neurons communicate their outputs to the user
- one or more **"hidden layers"**: neurons have as input the outputs of the preceding layer and give their outputs as inputs of the following layer

Supervised learning:

For every event i the user provides both the inputs $\vec{I}_i$ and the desired output T_i.

With the current weights (random, at the beginning) the NN associates $\vec{I}_i \rightarrow O_i$, and a **distance** or **error** $d(O_i - T_i)$ is computed.

Usually, $d(O_i - T_i) = |O_i - T_i|$ or $d(O_i - T_i) = (O_i - T_i)^2$

Errors are then **propagated back** through the NN, and the weights are modified attempting to have a smaller error at the next iteration.

Examples of supervised training:

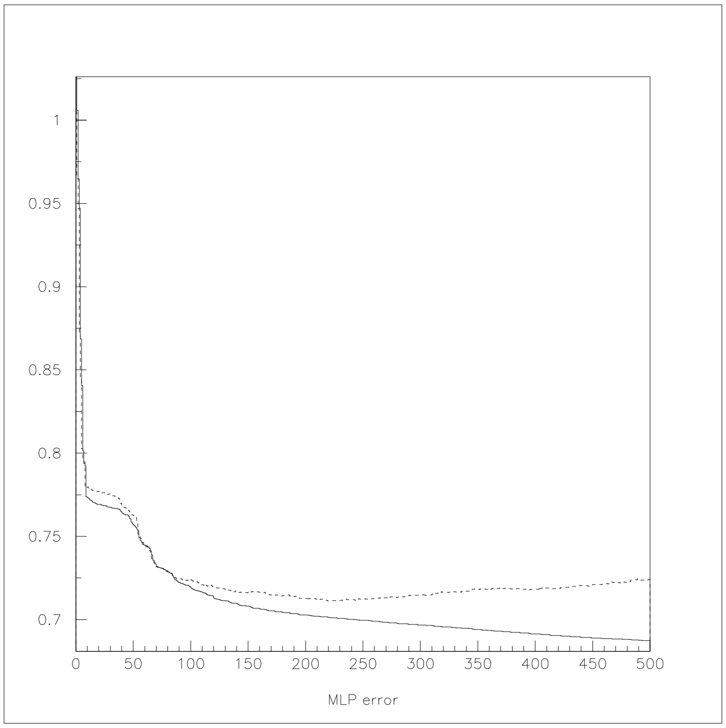
- Forecasts: the NN tries to guess the next values of the input variables, $T_i = I_{i+1}$.
- Fitting: $T_i = f(I_i)$, f unknown. NN can approximate with arbitrary accuracy any continuous (and quasi-continuous) function.
- Classification: f is discrete (e.g. signal$\rightarrow 1$, background$\rightarrow 0$).

How to train a NN (or anything else) to classify things:

- **Show examples of both "signal" and "background"**
  (to teach what a dog is, show dogs and non-dogs).
- **Examples should be as various as possible for every class**
  (people in isolated amazonic tribes aren't able to recognize white men as humans, since they are too different from any other human being that they know).
- **In principle, examples should be as many as possible...**
- **...but many examples $\Rightarrow$ many learning iterations.**
  If available learning time is not ∞, a balance has to be found.
- **Don't iterate too much**
  to avoid "overlearning" or "overfitting" (see next transparency).

Overfitting:

If the training has gone too far the weights can be optimally adapted to the data set used to learn but the network can have poor performance on any other analogous data set.



Solution: use two data sets.

- *Training sample*: examples seen by NN.
- *Test sample*: at every learning iteration, the NN is applied to the test sample and $d(O_i - T_i)$ is computed.

Learning is stopped at the minimum of $d(O_i - T_i)$ for the test sample. This is the point of maximum generalization for the NN.

Training algorithm:

$d(O_i - T_i)$ is treated as a function of the synaptic weights of all the neurons.

Training is a minimization problem in the space of the weights.

The most popular training algorithms are variants of *Standard backpropagation*, which is an NN implementation of the **heavy ball method**:

the minimum is looked for by following the gradient of the "error surface", $\frac{\partial E}{\partial w_{ij}}$

Minimization algorithms are designed to avoid, as far as possible,

- to be trapped in a local ("false") minimum;
- to skip the absolute ("true") minimum.

The success of an algorithm depends on the features of the particular error surface given by the problem.

RPROP (Resilient Propagation):

RPROP was chosen both for its performances in minimum finding and for its speed.

RPROP belongs to the family of Backpropagation algorithms:

$$\Delta w_{ij} = \eta \delta_i o_j$$

o_j is an output of unit j and an input to i

δ_i is a function of the outputs of the preceding layer's neurons:

- if i is an output neuron: $\delta_i = o_i(1 - o_i)E \quad (E = d(O - T))$
- if i is a hidden neuron: $\delta_i = o_i(1 - o_i) \sum_k \delta_k w_{ki}$

η is the **learning rate**.

The special feature of RPROP is the way to update η. Only the direction of $\frac{\partial \vec{E}}{\partial w_{ij}}$ is used.

It begins with an initial small update value, and then:

- increases, if $\frac{\partial \vec{E}}{\partial w_{ij}}(t)$ and $\frac{\partial \vec{E}}{\partial w_{ij}}(t-1)$ have angle $< 180°$;
- decreases otherwise.

How many hidden neurons?

Question: what is the optimal number of hidden neurons for a given problem?
Answer: nobody knows.

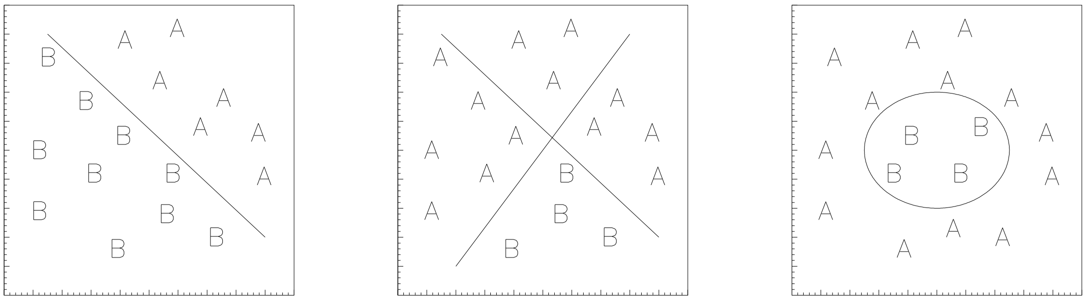
NN literature is full of "rules of thumb", but their limits of application are often tight or/and mostly unknown.

The recommended way to proceed is by trial and error:
beginning with only one unit per layer, the number of units is increased until the performance of the NN ceases to improve.

When the number of units is too large, the number of connections, and consequently the number of weights to variate, makes the training very slow and a larger training sample becomes necessary to avoid over-fitting.

How many hidden layers?

The first question is:
do we need hidden layers?



(g) (h) (i)

- (a) **Linearly separable problem:** finding a single hyper-plane separating A/B.

  A **perceptron** is able to do that. Output ≈ 0 on the "A" side, ≈ 1 on the other side, and intermediate values in the proximity of the separating line.

- (b) **Non-linearly separable problem:** no straight line can properly separate the two categories of events.

  The introduction of a hidden layer overcomes the problem.

- (c) **Non-linearly separable problem with closed surface:** a second hidden layer can greatly improve the performances of the NN.

Structure chosen:

The optimal topology was found to be:

1. Input layer (7 neurons)
2. First hidden layer (7 neurons)
3. Second hidden layer (7 neurons)
4. Ouput layer (1 neuron)

Input variables:

- $P^3_{\text{miss}} = \vec{P}_{\text{miss}} \cdot \vec{n}_3$
- P_l
- $|\cos\theta_l|$
- B_1 (b-tag prob. for the "1^{st}" hemisphere)
- B_2 (b-tag prob. for the "2^{nd}" hemisphere)
- b_l (b-tag prob. for the lepton track)
- T

P^3_{miss}:

$$\vec{P}_{\mathrm{miss}} = -\sum_i \vec{P}_i$$

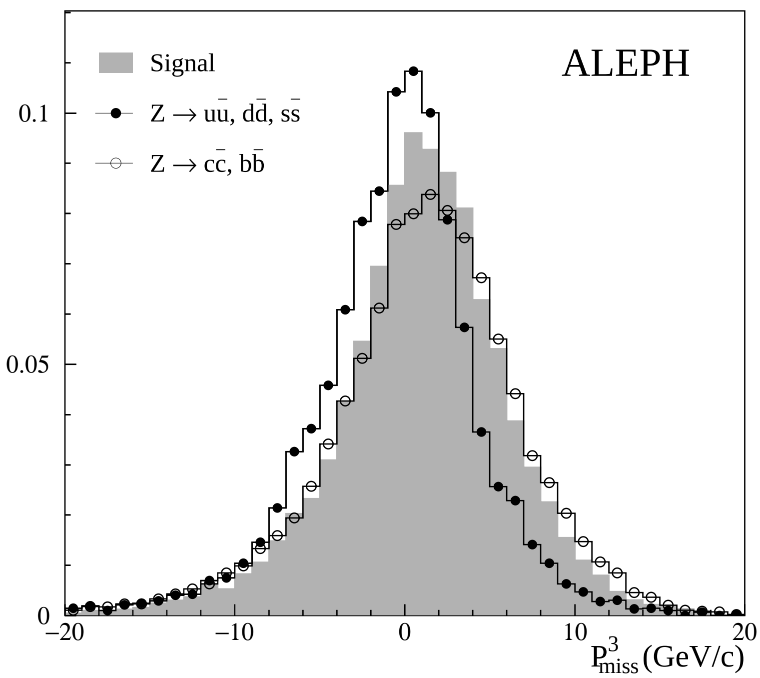
i runs on all the reconstructed particles

(including photons and neutral hadrons)

$$P^3_{\mathrm{miss}} = \vec{P}_{\mathrm{miss}} \cdot \vec{n}_3$$

Neutrinos from $g \to c\bar{c} \to l\nu X$ are mostly near to the gluon jet.

In uds events, $\langle P^3_{\mathrm{miss}} \rangle = 0$
(broadness is due to the energy resolution)

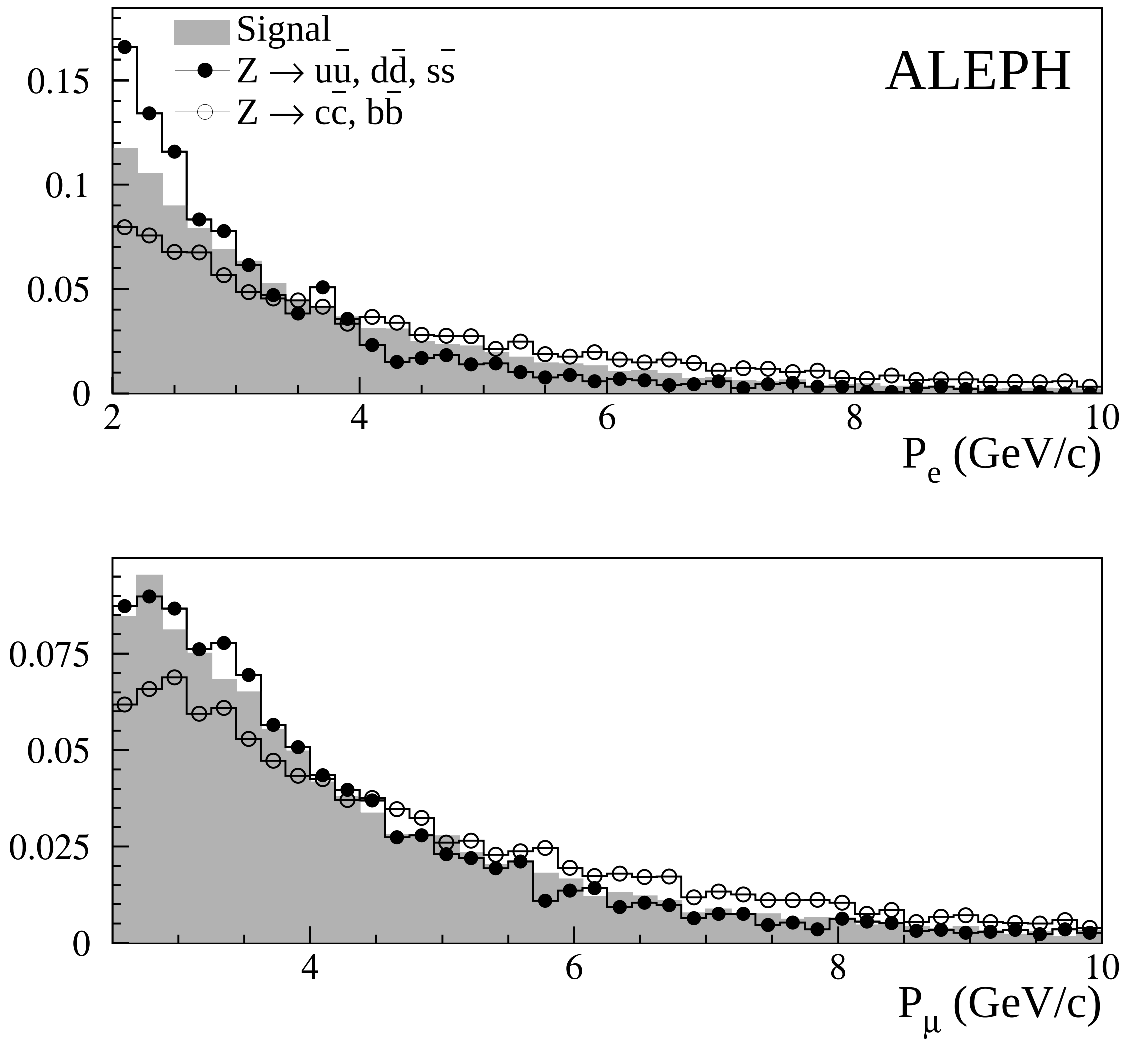


This variable has never been used before in gluon splitting measurements.

P_l:

Momentum spectrum is harder for **prompt leptons** than for **non-prompt** and **fake** leptons, which give rise to the *uds* bkg.

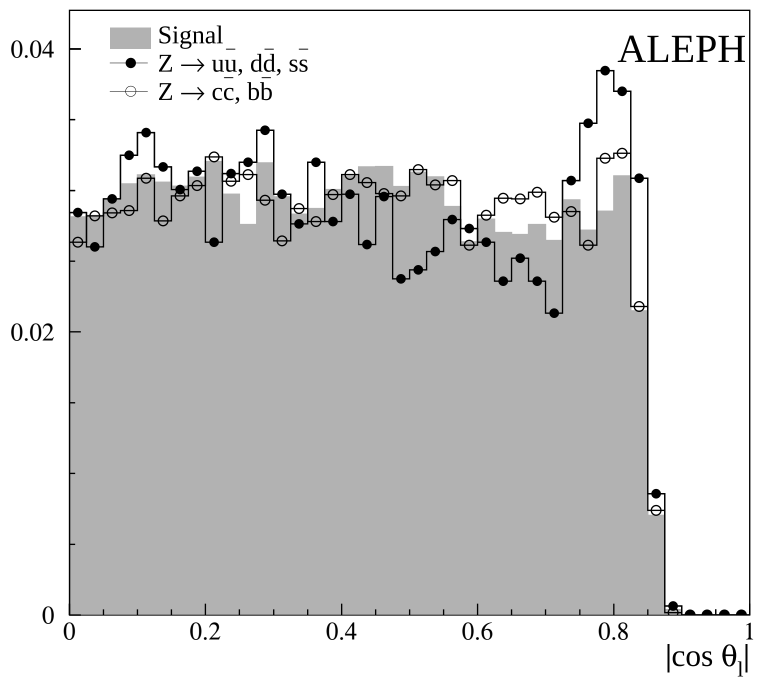
But $g \to c \to l$ spectrum is softer than $Z \to c \to l$ (and $Z \to b \to l$):



$|\cos\theta_l|$:

The cosine of the angle between the lepton track and the beam line is correlated with the quality of the lepton reconstruction.

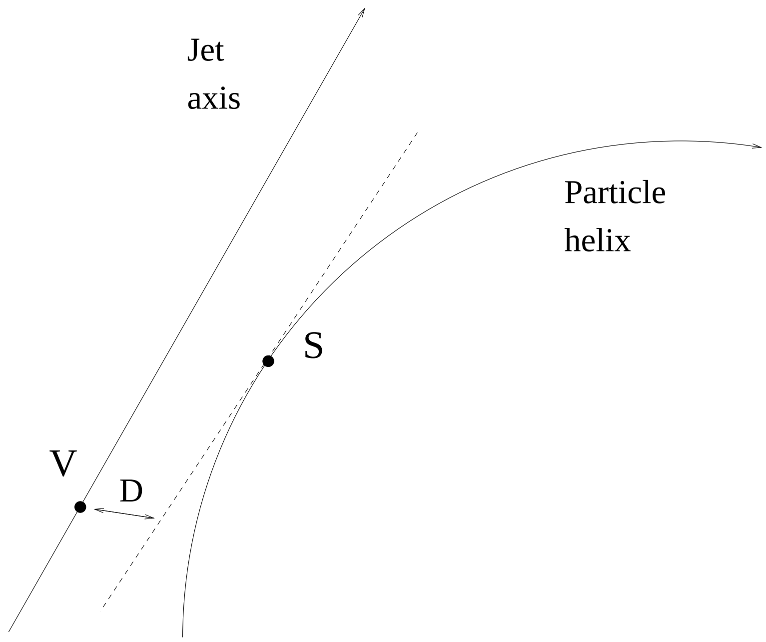
It gives to the NN an information on the reliability of some of the other inputs (e.g. lepton momentum and impact parameter).



The *uds* distribution is higher in the angular region corresponding to the **transitions between barrel and end-caps** of the calorimeters.

Impact parameter (D):

After the clusterization in jets, the trajectories of the charged tracks are linearized in their point of maximum approach to the jet axis (point S):



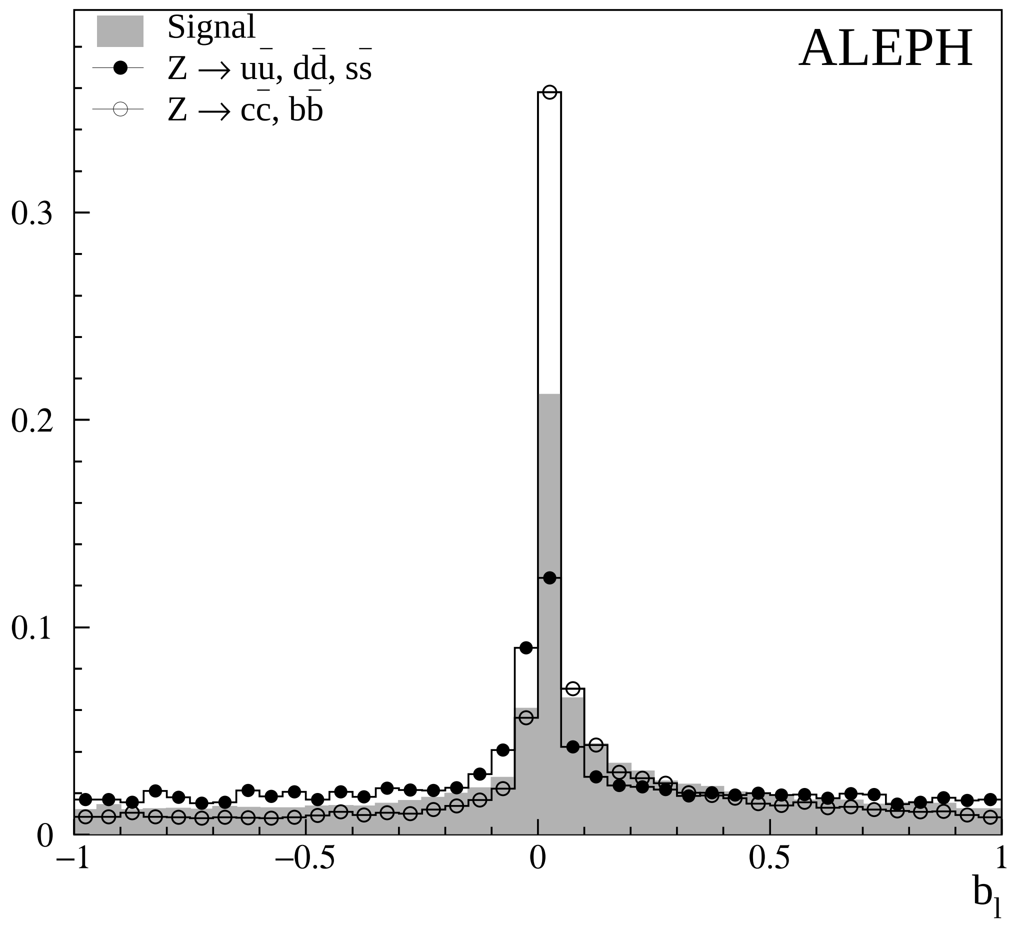
$D > 0$ if the vector joining V with S points in the direction of the jet, $D < 0$ otherwise.

The quantity of interest is $\Delta \equiv D/\sigma_D$

b_l:

b_l is the probability, based on $\Delta = \frac{D}{\sigma_D}$, that the lepton comes from the interaction point.

For leptons from $b, c \to l$, in general, the b_l distribution is peaked near 0:



Leptons from the $g \to c \to l$ chain are less peaked in $b_l \approx 0$ due to their smaller momentum $\Rightarrow$ worst D resolution.

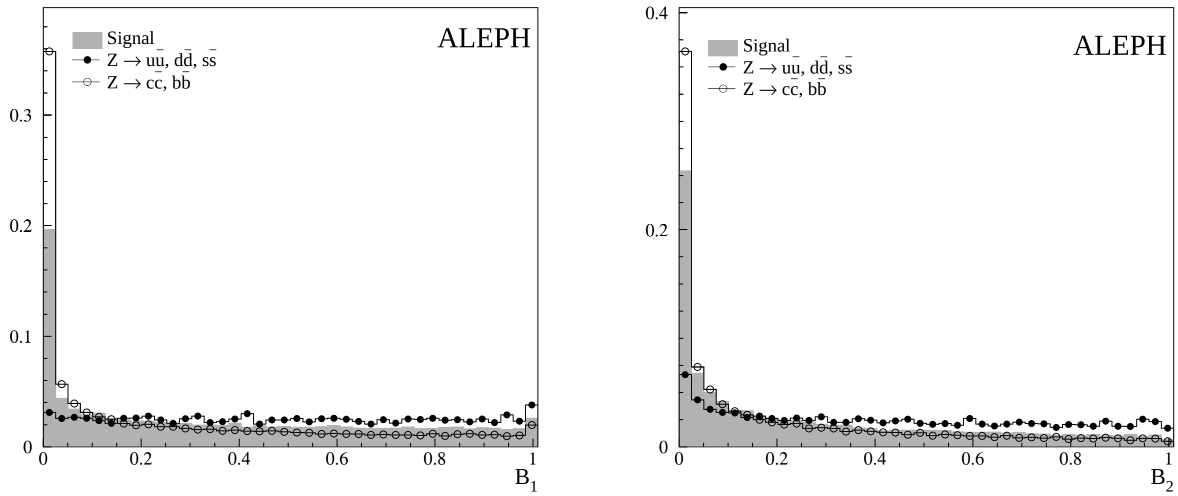
B_1, B_2:

For a group of N tracks, the combined probability to come from the primary vertex is:

$$B \equiv \Pi \cdot \sum_{j=0}^{N-1} \frac{(-\ln \Pi)^j}{j!}$$

with $\Pi \equiv \prod_{i=1}^{N} b_i$

where b_i is the probability for a single track)



(j) (k)

Hemispheres are defined by a plane perpendicular to the **thrust vector**

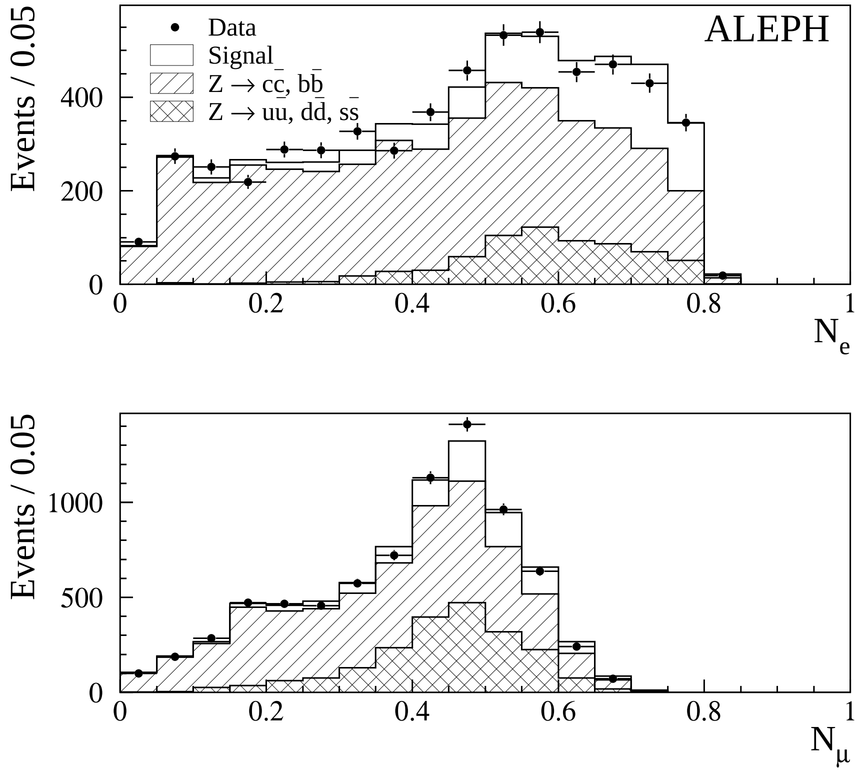
(the vector $\vec{n}$ which minimizes $\Sigma_i \frac{\vec{p_i}\cdot\vec{n}}{|\vec{p_i}|}$ in the definition of T)

1: light hemisphere; **2**: heavy hemisphere.

NN output:

MC sample: 8.7 M $Z \to q\bar{q}$, 1.8 M $g \to c\bar{c}$

One half was used as "training sample", the other half as "test sample".



Cut	σ^e_{stat}	σ^e_{syst}	σ^e_{tot}	σ^μ_{stat}	σ^μ_{syst}	σ^μ_{tot}
0.30	0.26	0.54	0.80	0.33	0.80	1.13
0.35	0.26	0.52	0.78	0.34	0.77	1.11
0.40	0.26	0.51	0.77	0.36	0.75	1.10
0.45	0.27	0.48	0.75	0.38	0.72	1.10
0.50	0.27	0.45	0.72	0.45	0.65	1.10
0.55	0.28	0.42	0.70	0.58	0.62	1.20
0.60	0.29	0.41	0.70	0.88	0.73	1.61
0.65	0.32	0.39	0.71			

Extraction of $g_{c\bar{c}}$:

The NN cut selects 2258 e events and 3332 μ events:

	e	μ
$g \to c\bar{c}$	25 %	18 %
$g \to b\bar{b}$	4 %	2 %
uds	18 %	35 %
c	28 %	22 %
b	25 %	23 %

$$N_S = N_H \left[(1 - g_{c\bar{c}} - g_{b\bar{b}})\, \epsilon_Q + g_{c\bar{c}}\, \epsilon_C + g_{b\bar{b}}\, \epsilon_B \right]$$

$$g_{c\bar{c}} = \frac{f - (1 - g_{b\bar{b}})\, \epsilon_Q - g_{b\bar{b}}\, \epsilon_B}{\epsilon_C - \epsilon_Q} \qquad (f \equiv \frac{N_S}{N_H})$$

- ϵ_Q: efficiency for $Z \to q\bar{q}$ (without g.s.)
- ϵ_C: efficiency for $g \to c\bar{c}$
- ϵ_B: efficiency for $g \to b\bar{b}$

The efficiencies are extracted from MC, excluding the "training sample" (which would introduce a bias) and including 5.1 M $Z \to b\bar{b}$ and 2.3 M $Z \to c\bar{c}$.

Extraction of $g_{c\bar{c}}$:

	electrons (%)	muons (%)
f	0.0600 ± 0.0013	0.0885 ± 0.0015
ϵ_Q	0.0431 ± 0.0006	0.0747 ± 0.0009
ϵ_C	0.491 ± 0.005	0.473 ± 0.005
ϵ_B	0.81 ± 0.06	0.80 ± 0.06

$$g_{b\bar{b}} = (2.54 \pm 0.51) \times 10^{-3}$$

Results:

$$g^{e}_{c\bar{c}} = (3.32 \pm 0.28) \times 10^{-2}$$

$$g^{\mu}_{c\bar{c}} = (2.99 \pm 0.38) \times 10^{-2}$$

(statistical error only)

Source of error	$\Delta g^{e}_{c\bar{c}}$ (%)	$\Delta g^{\mu}_{c\bar{c}}$ (%)
Statistical error	0.28	0.38
ϵ_Q, ϵ_C, ϵ_B	0.13	0.21
$g_{b\bar{b}}$	0.09	0.10
Thrust cut efficiency	0.07	0.11
BR($b \to l$)	0.05	0.06
BR($b \to c \to l$)	0.02	0.03
BR($c \to l$)	0.23	0.20
$b \to l$ spectrum	0.01	< 0.01
$b \to c \to l$ spectrum	0.02	0.02
$c \to l$ spectrum	0.03	0.12
Lifetimes of b hadrons	0.03	0.03
$\langle x_E^{(c)} \rangle$	0.04	0.05
$\langle x_E^{(b)} \rangle$	0.02	0.03
Lepton identification	0.18	0.01
Lepton background	0.09	0.50
Generators, θ^*	0.06	0.12
Generators, $E_{\mathrm{g}}, m^*_{\mathrm{g}}$	0.05	0.26
m_{c}	0.19	0.24
Systematic error	0.42	0.72

Systematic errors:

MC stat

$\Delta g^{e}_{c\bar{c}} (\%)$	$\Delta g^{\mu}_{c\bar{c}} (\%)$
0.13	0.21

Thrust cut efficiency

year	data (%)	MC (%)
1991	39.73 ± 0.10	39.96 ± 0.08
1992	39.60 ± 0.07	40.11 ± 0.04
1993	39.98 ± 0.06	40.09 ± 0.04
1994	40.00 ± 0.04	40.28 ± 0.03
1995	40.47 ± 0.06	40.32 ± 0.04

The MC is reweighted to the data, on a year by year basis.

If no reweighting is performed, a shift is observed:

$\Delta g^{e}_{c\bar{c}} (\%)$	$\Delta g^{\mu}_{c\bar{c}} (\%)$
$+0.07$	$+0.11$

Systematic errors

Semileptonic c and b decays

	$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
$\mathrm{BR}(b \to l) = (10.65 \pm 0.23)\%$	± 0.05	± 0.06
$\mathrm{BR}(b \to c \to l) = (8.04 \pm 0.19)\%$	± 0.02	± 0.03
$\mathrm{BR}(c \to l) = (9.73 \pm 0.32)\%$	± 0.23	± 0.20
$b \to l$ modelling	-0.01	< 0.01
$b \to c \to l$ modelling	$+0.02$	$+0.02$
$c \to l$ modelling	$+0.03$	$+0.12$

Lifetimes of B hadrons

B hadron	Mean lifetime (ps)
B_d	1.55 ± 0.03
B_u	1.66 ± 0.03
B_s	1.47 ± 0.06
Λ_b	1.19 ± 0.05

$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
0.05	0.05

Systematic errors

Gluon splitting in $b\bar{b}$

	$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
$g_{b\bar{b}} = (0.254 \pm 0.051)\ \%$	0.09	0.10

Lepton identification efficiency

- The ECAL selection efficiency for electrons in MC is corrected to the data in **6 bins in** $|\cos\,\theta|$ and 2 bins in p_T $(p_T < 0.25\ \mathbf{GeV}/c,\ p_T > 0.25\ \mathbf{GeV}/c)$
- The efficiency of the dE/dx selection in the TPC is corrected in **5 angular bins**.
- The μ id. efficiency is corrected in **40** $|\cos\,\theta|$ **regions** and 3 momentum bins $(2.5 < p < 3\ \mathbf{GeV}/c,\ 3 < p < 4\ \mathbf{GeV}/c,\ p > 4\ \mathbf{GeV}/c)$

$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
0.18	0.01

Systematic errors:

Fragmentation function

Heavy quark fragmentation is simulated using the **Peterson model**:

$$D_Q^H(z) \propto \frac{1}{z[1-1/z-\varepsilon_Q/(1-z)^2]^2}$$

where $z = \frac{(E+p_\parallel)_{hadron}}{(E+p)_{original\ quark}}$, $\varepsilon_Q \approx \frac{m_q^2}{m_Q^2}$

ε_b and ε_c are adjusted to reproduce the measured value of the mean scaled energy of weakly-decaying heavy hadrons:

$$\langle x_E^c \rangle = 0.484 \pm 0.008$$
$$\langle x_E^b \rangle = 0.702 \pm 0.008$$

where $x_E = \frac{E_{\mathrm{hadron}}}{E_{\mathrm{beam}}}$

	$\Delta g^e_{c\bar{c}}(\%)$	$\Delta g^\mu_{c\bar{c}}(\%)$
$\langle x_E^c \rangle$	0.04	0.05
$\langle x_E^b \rangle$	0.02	0.03

Systematic errors

Lepton background

In the Monte Carlo:

- an uncertainty of $\pm 20\%$ is assigned to the rate of hadrons misidentified as electrons
- an uncertainty of $\pm 5\%$ to the rate of photon conversions $\gamma \to e^+e^-$
- a correction factor of $(17 \pm 5)\%$ is assigned to fake and non-prompt muons

$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
0.09	0.50

Fake and non-prompt muons:

A sample **enriched in hadrons** is selected in MC and data.
Real hadrons and non-prompt μ flagged as prompt μ are counted in MC.

In real data, the percentage of **real** prompt μ's (from the MC) is subtracted from the hadron sample.

To select *uds*-fragmentation hadrons we use a *uds*-tag variable.

	Simulation	Data
Number of tracks (N_{tot})	564396	332103
Number of prompt μ (N_{μ})	432	—
Tracks identified as μ (n_{tot})	3072	2077
Prompt μ identified as μ (n_{μ})	351	—
Mistag (10^{-3})	4.83 ± 0.08	5.64 ± 0.13

$$\epsilon_{\mathrm{MC}}^{\mathrm{mistag}} = \frac{n_{tot}^{MC} - n_{\mu}^{MC}}{N_{tot}^{MC} - N_{\mu}^{MC}}$$

$$\epsilon_{\mathrm{data}}^{\mathrm{mistag}} = \frac{n_{tot}^{DT} - n_{\mu}^{DT}}{N_{tot}^{DT} - N_{\mu}^{DT}}$$

where

$$N_{\mu}^{DT} = N_{tot}^{DT} \times C_{\mu} \quad , \quad C_{\mu} \equiv N_{\mu}^{MC} / N_{tot}^{MC}$$

Correction Factor: 1.17 ± 0.03

Fake and non-prompt muons:

A second pion enriched sample, with different kinematic properties, is selected by tagging the $K^0_S \to \pi^+\pi^-$ decay.

I used $\approx 300k$ K^0_s candidates selected with a mass windows of width $(0.014 + 0.15 P_K/E_b)$ GeV, decay lenght > 1 cm, and a uds-tag cut, and applying to their "daughters" the same kinematical cuts used in the $g_{c\bar{c}}$ analysis ($P > 2.5$, $d_0 < 0.1$, $N_{VDET} > 0$) I found:

	Simulation	Data
Number of tracks (N_{tot})	113068	50162
Number of prompt μ (N_μ)	199	—
Tracks identified as μ (n_{tot})	662	326
Prompt μ identified as μ (n_μ)	156	—
Mistag (10^{-3})	4.48 ± 0.20	5.13 ± 0.31

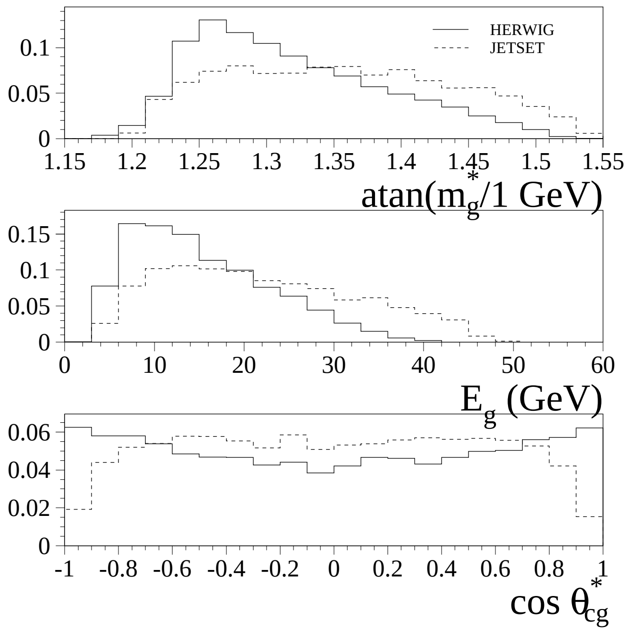
Correction Factor: 1.15 ± 0.09

Systematic errors:

Generators

Three variables are chosen to reweight **JETSET** to **HERWIG**:

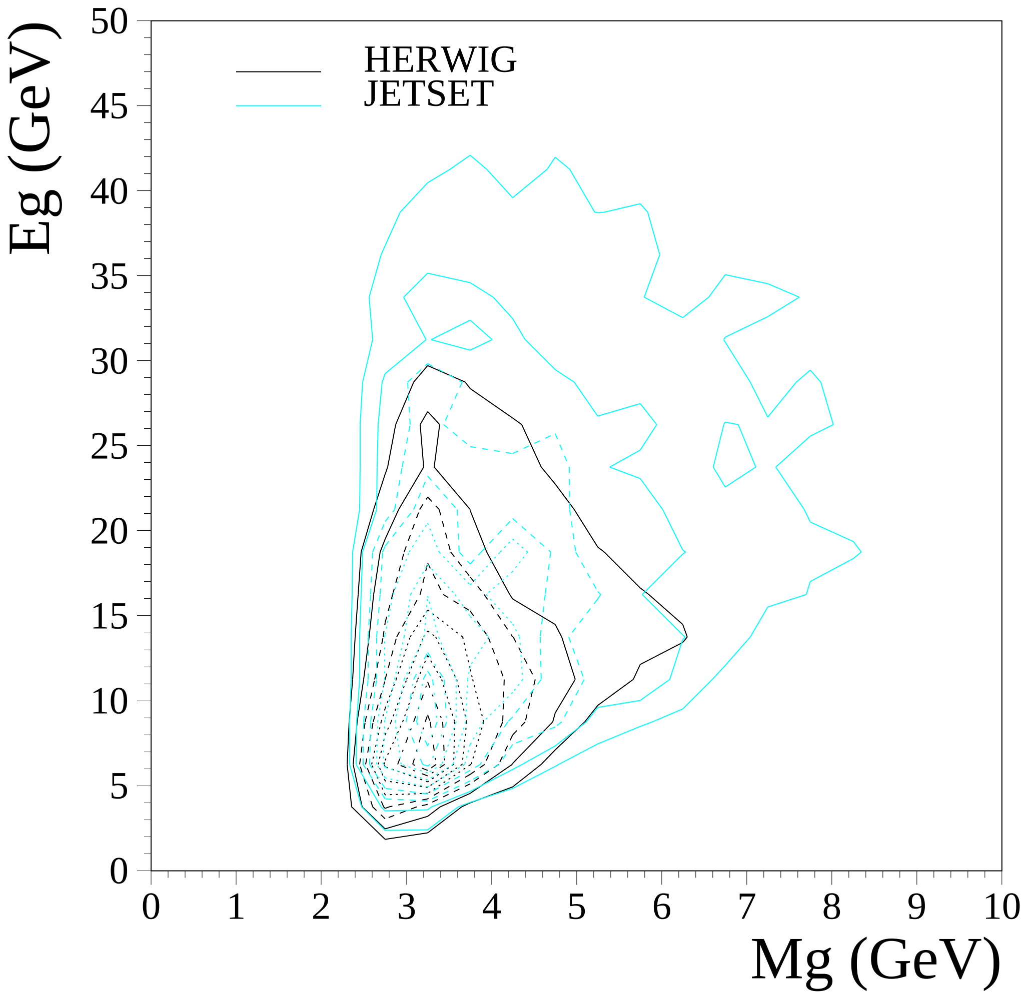
- the energy E_g of the gluon
- its effective mass m_g^*
- the angle θ^* between the gluon direction and the c quark direction, measured in the gluon rest frame.



The difference observed in the measured value of $g_{c\bar{c}}$ is taken as an estimate of the systematic uncertainty:

	$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
E_g and m_g^*	$+0.05$	$+0.26$
θ^*	-0.06	-0.12

E_g, m_g^*:



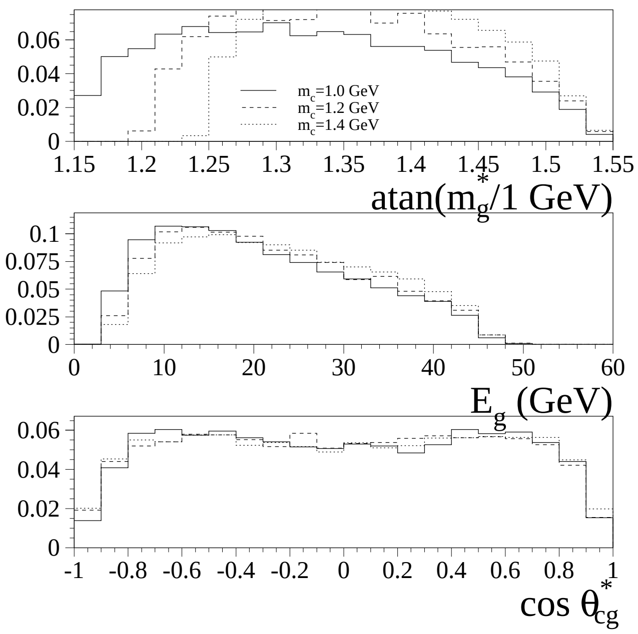
E_g vs M_g^*, for HERWIG and JETSET

Systematic errors:

Charm mass

The mass of the c quark is taken to be
1.2 ± 0.2 GeV$/c^2$

A shift of m_c in the MC results in a variation of the E_g and m_g^* distributions:



	$\Delta g^{e}_{c\bar{c}}(\%)$	$\Delta g^{\mu}_{c\bar{c}}(\%)$
$\Delta m_c = 0.2$ GeV$/c^2$	0.19	0.24

Systematic errors vs P_{min}:

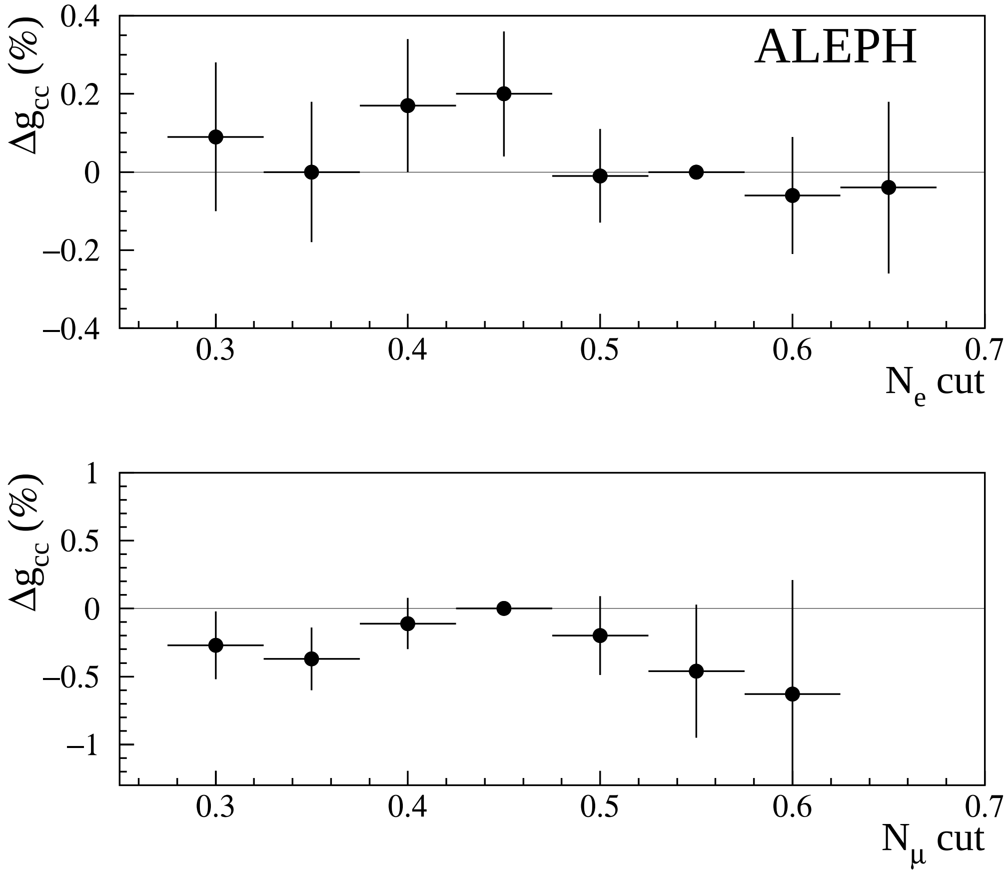
Some of the systematic errors have very different values for the electronic and muonic channels, even if important correlations exist between the two channels. Why?

Answer: e and μ selection imply different P cuts

$p > 2.0$ GeV$/c$	$\Delta g^{e}_{c\bar{c}}$ $(\times 10^{-2})$	$\Delta g^{\mu}_{c\bar{c}}$ $(\times 10^{-2})$
$c \rightarrow l$ spectrum	0.03	-
Generators, θ^*	0.06	-
Generators, $E_{\mathrm{g}}, m^*_{\mathrm{g}}$	0.05	-
$p > 2.5$ GeV$/c$	$\Delta g^{e}_{c\bar{c}}$ $(\times 10^{-2})$	$\Delta g^{\mu}_{c\bar{c}}$ $(\times 10^{-2})$
$c \rightarrow l$ spectrum	0.06	0.12
Generators, θ^*	0.10	0.12
Generators, $E_{\mathrm{g}}, m^*_{\mathrm{g}}$	0.26	0.26
$p > 3.0$ GeV$/c$	$\Delta g^{e}_{c\bar{c}}$ $(\times 10^{-2})$	$\Delta g^{\mu}_{c\bar{c}}$ $(\times 10^{-2})$
$c \rightarrow l$ spectrum	0.09	0.15
Generators, θ^*	0.14	0.19
Generators, $E_{\mathrm{g}}, m^*_{\mathrm{g}}$	0.42	0.44

The gluon energy is strongly correlated with the lepton energy.

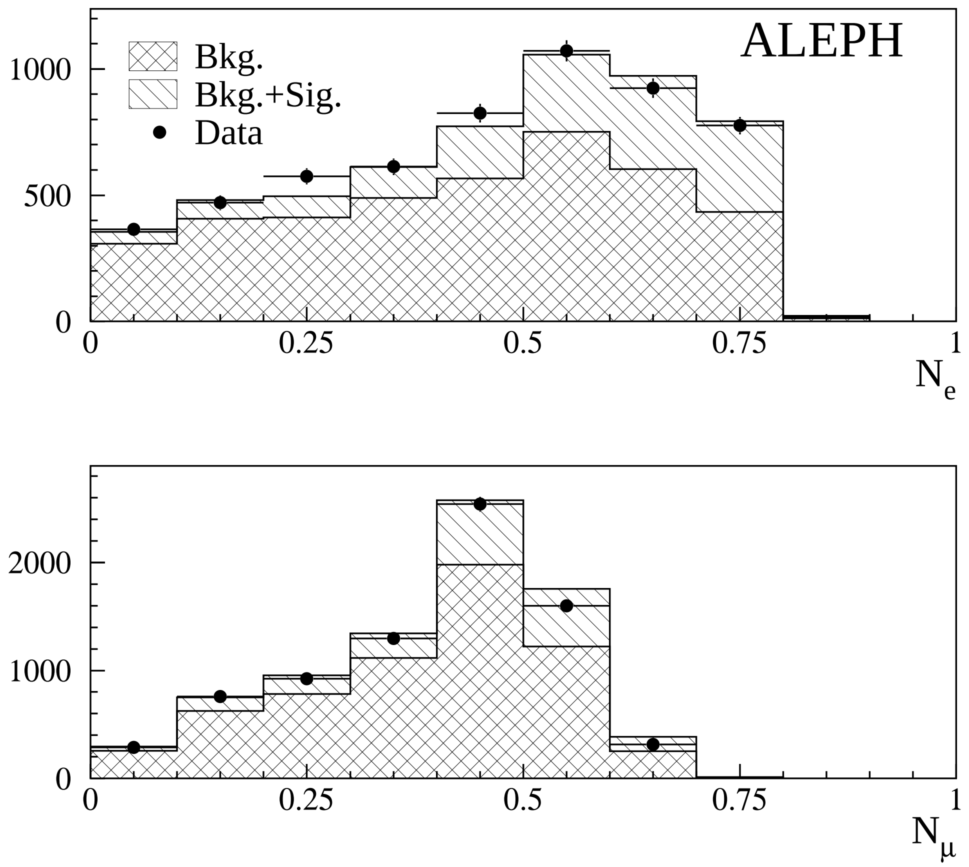
Check - Stability of the NN cuts:



Variation $\Delta g_{c\bar{c}}$ vs the cut on N_e, N_μ

The errors bars indicate the uncorrelated statistical uncertainties only.

Check - Shape of the NN outputs:



The normalization of the NN is such to reproduce the data in the region $N_{e,\mu} < 0.2$

In fact, the signal is less well known than the background (which by far dominates over signal for low values of N_e and N_μ).

Equalizing simulated and real data in the regime of low neural network output implies that **any deviation for high N_e and N_μ values is probably to be imputed to the simulation of the signal**.

Extraction of $g_{c\bar{c}}$:

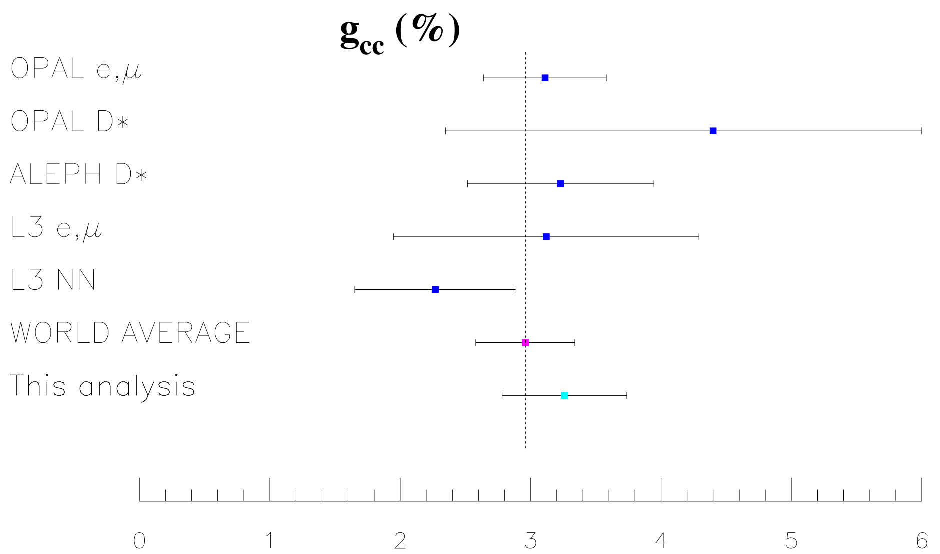
The final results:

$$g^{e}_{c\bar{c}} = (3.32 \pm 0.28(\mathrm{stat}) \pm 0.42(\mathrm{syst})) \times 10^{-2}$$

$$g^{\mu}_{c\bar{c}} = (2.99 \pm 0.38(\mathrm{stat}) \pm 0.72(\mathrm{syst})) \times 10^{-2}$$

The two results are combined taking in account the correlations between systematic uncertainties using the **B**est **L**inear **U**nbiased **E**stimator technique:

$$g_{c\bar{c}} = (3.26 \pm 0.23(\mathrm{stat}) \pm 0.42(\mathrm{syst})) \times 10^{-2}$$



BLUE:

n measurements y_i of the observable y

The best estimate $\hat{y}$ is looked for as

$\hat{y} = \sum_i \alpha_i y_i$

Unitarity requirement: $\sum_i \alpha_i = 1$

We minimize the variance of $\hat{y}$, given by

$\sigma^2 = \sum_i \sum_j E_{ij} \alpha_i \alpha_j$

E_{ij}: elements of the *error matrix*

E_{ii}: variance of the single measurement y_i

E_{ij}, $i \neq j$: covariance of i-th and j-th measurements

In the special case of two measurements:

$$E = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where r is the *correlation coefficient.*

Every source of uncertainty has its own error matrix, and the global error matrix is the sum of the single error matrices.

New World Average:

All the $g_{c\bar{c}}$ measurements are combined taking into account all the correlations.

An iterative method is used:

- $g_{c\bar{c}}$ and $g_{b\bar{b}}$ measurements are combined separately to obtain $g_{c\bar{c}}^{best}$ and $g_{b\bar{b}}^{best}$
- $g_{c\bar{c}}^{best}$ and $g_{b\bar{b}}^{best}$ are used to obtain a new estimate from each $g_{c\bar{c}}$ and $g_{b\bar{b}}$ measurement

Given the strong correlation between $g_{c\bar{c}}$ and $g_{b\bar{b}}$ it is convenient to write every single result as $g_{c\bar{c}}^{i} = A_i - B_i \times g_{b\bar{b}}^{best}$

In our case,

$$g_{c\bar{c}}^{e} = 3.76 \times 10^{-2} - 1.71 \times g_{b\bar{b}}^{best} \,,$$
$$g_{c\bar{c}}^{\mu} = 3.46 \times 10^{-2} - 1.83 \times g_{b\bar{b}}^{best} \,.$$

After a few iterations (3 are sufficient):

$$g_{c\bar{c}} = (3.01 \pm 0.33) \times 10^{-2}$$

$$g_{b\bar{b}} = (2.54 \pm 0.51) \times 10^{-3}$$

New World Average:

	old	**new**
$\Delta g_{c\bar{c}}$	0.38×10^{-2}	0.33×10^{-2}

Weights of the single $g_{c\bar{c}}$ measurements:

OPAL e	22.6%
OPAL μ	3.0%
OPAL D^*	1.0%
L3 e	2.3%
L3 μ	0.4%
L3 ev.sh.	22.2%
ALEPH D^*	14.4%
ALEPH e	29.5%
ALEPH μ	4.6%

Correlations with the other $g_{c\bar{c}}$ measurements:

	ALEPH e	ALEPH μ
OPAL e	28%	23%
OPAL μ	18%	15%
OPAL D^*	5%	7%
L3 e	11%	6%
L3 μ	12%	7%
L3 ev.sh.	2%	1%
ALEPH D^*	4%	7%
ALEPH e	100%	34%
ALEPH μ	34%	100%

Comparison with theory:

The new average has a 2σ distance from:

$$g^{th}_{c\bar{c}} = 2.3 \times 10^{-2} \quad (m_c = 1.2 \text{ GeV}, \alpha_s = 0.120)$$

The theoretical uncertainty due to the truncation of the perturbation series ($30\% \Leftrightarrow \Delta g^{th}_{c\bar{c}} = \pm 0.7 \times 10^{-2}$) is enough to account for the discrepancy, but it is certainly overestimated (a more realistic guess is 15%)

A shift of $+0.002$ in α_s yields an increase of $0.3 \times 10^{-2} \approx 1\sigma$ in $g_{c\bar{c}}$

A shift of -0.2 GeV$/c^2$ in m_c would yield a positive shift of 0.4×10^{-2} on the $g_{c\bar{c}}$ prediction, partially compensated by a shift of 0.2×10^{-2} on the experimental result, for an overall reduction of the discrepancy between theory and data by $2/3\ \sigma$.

Summary

- A new $g_{c\bar{c}}$ measurement is available.

- Distinctive features of the new analysis are:
  - lepton tag;
  - preselection based on thrust;
  - missing momentum along the "gluon jet";
  - use of a neural network to combine the discriminant variables.

- The result is consistent with the previous measurements.

- The electron result has the largest weight on the world average among all the available measurements.

- $g_{c\bar{c}}$ average is 2σ far from the theoretical prediction, and tends to favour low m_c or high α_s. But large uncertainties affect the predicted value.