

MonteCarlo Produktion und die Suche nach dem Higgsboson



U. Schwickerath, CERN

DELPHI Collaboration

Karlsruhe, Jan. 2002



Übersicht

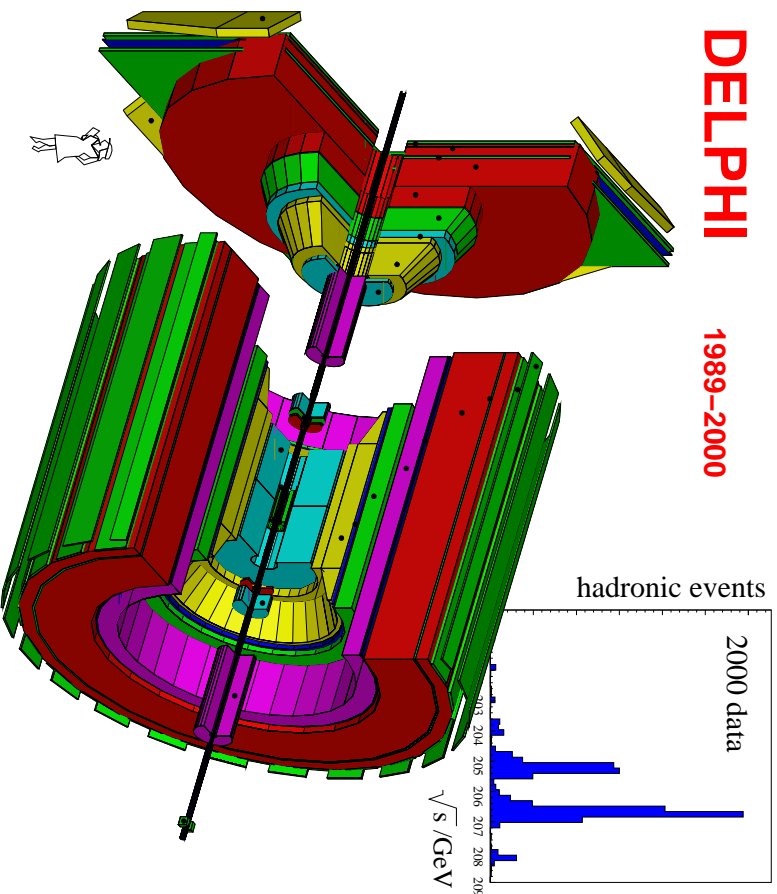
- Higgssuche mit DELPHI
 - Datennahme bei LEP2
 - Produktion und Zerfall des SM Higgs
 - Vorläufige Ergebnisse
 - Suchen jenseits des SM
- Massive Monte Carlo Produktionen
 - Motivation nach Ende der Datennahme
 - Simulationsbedarf, Beispiel Higgsanalyse
 - Anforderungen und Lösungsansätze
 - Erfahrungen
- Zusammenfassung



Datennahme bei LEP2

DELPHI

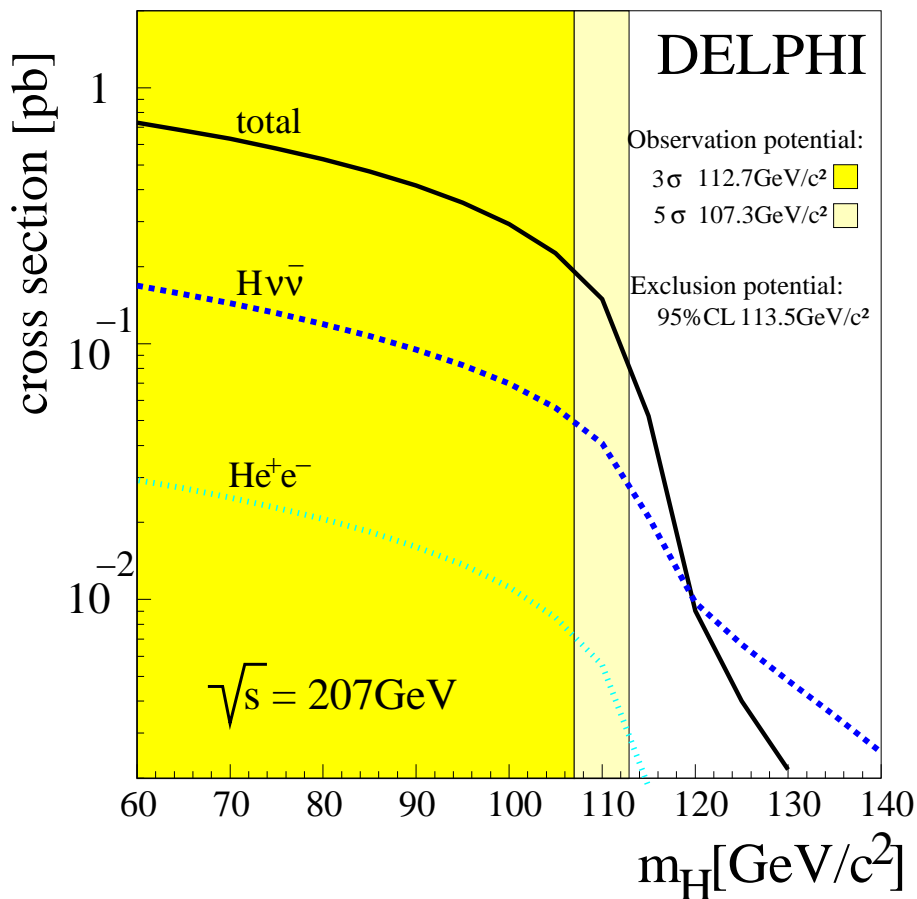
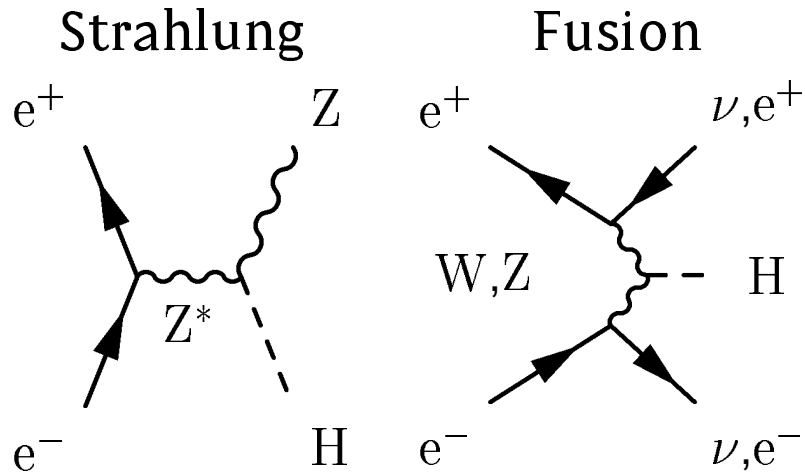
1989–2000



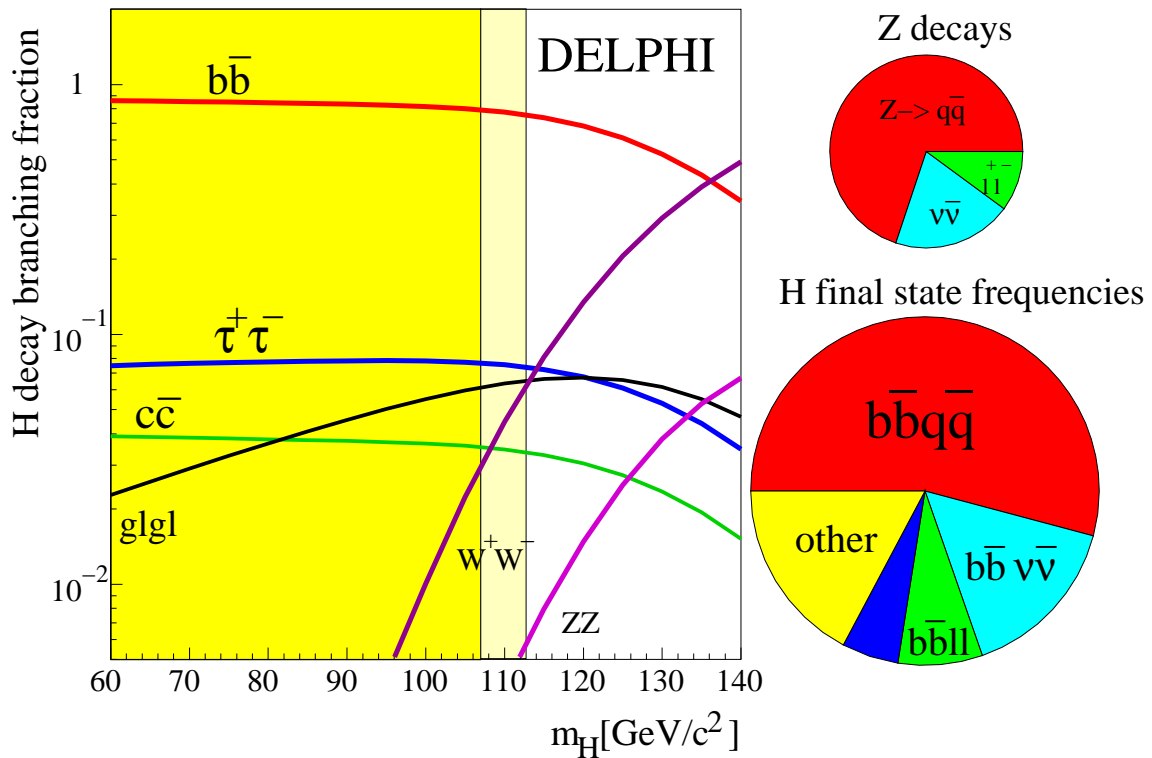
- etwa 225 pb^{-1} Daten mit \sqrt{s} zwischen 200 and 209 GeV
- etwa 688 pb^{-1} oberhalb der WW Schwelle



Produktionsmechanismen des SM Higgs bei LEP



Zerfallseigenschaften des SM Higgs

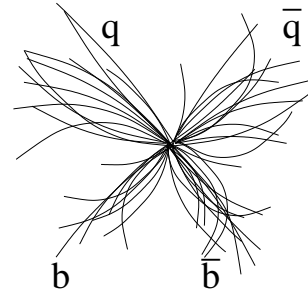


- Suche nach $H \rightarrow b\bar{b}$
- gegebene Effizienzen enthalten kleine Anteile anderer Zerfallskanäle



Der 4-jet Kanal

$H \rightarrow \text{any but } \tau, Z \rightarrow q\bar{q}$



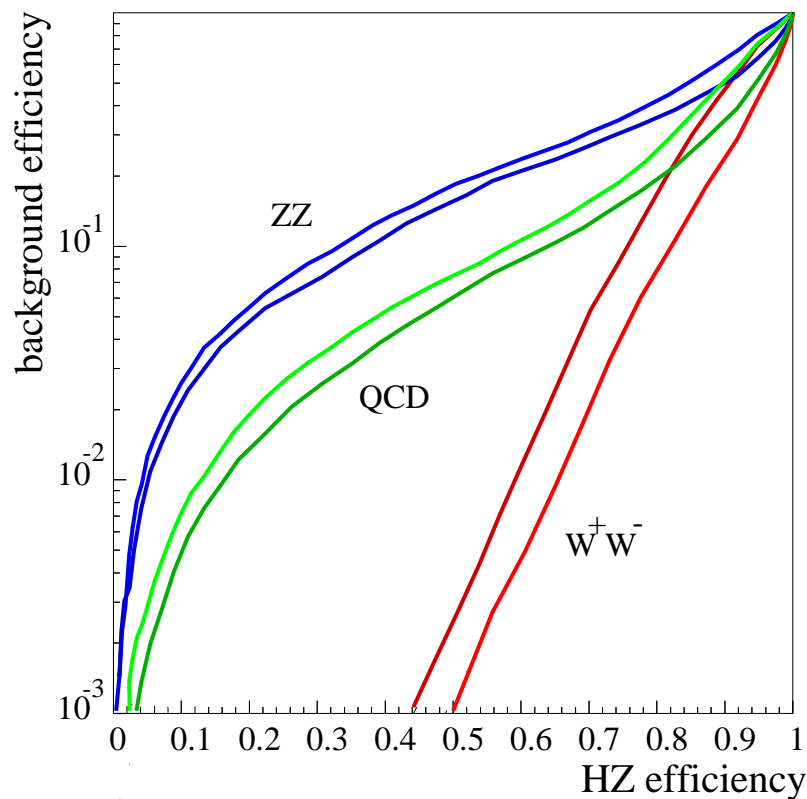
- sequentielle 4-Jet Ereignisvorselektion
- NNW mit 13 Variablen:
 - Anti-QCD (8)
 - WW and ZZ Ähnlichkeit (4)
 - Summe der beiden höchsten Jet b-tags
- Identifikation des Jetpaares aus dem Higgszerfall durch LLH

Schnitt	data	background	eff. [%]
Presel.	2266	2342	85.0
intermed.	398	423.7	79.0
tight	8	7.4	36.0



b-Erkennung

- Kombination von Lebensdauer und Topologie
- Deutliche Verbesserung der Untergrundunterdrückung für 4-Jets im Vergleich zur Analyse von 1999



- Korrektur des $g \rightarrow b\bar{b}$ (falsch im MC)
- Gute Beschreibung der Daten



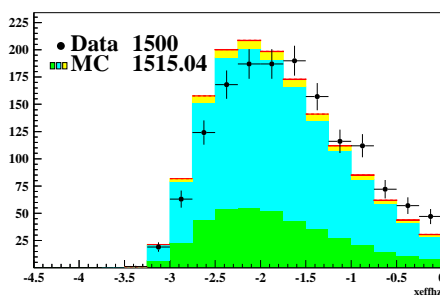
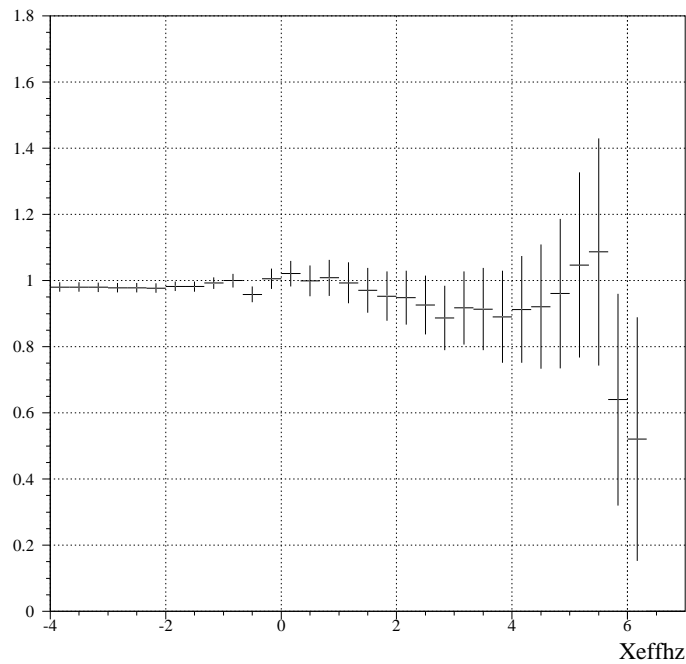
Größen zur b-Erkennung

- Polarwinkel der Jetrichtungen
- Jetenergie
- Geladene Multiplizität der Jets
- Winkel zum nächsten Jet
- mittlerer Transversalimpuls geladener Teilchen zur Jetrichtung
- Anzahl Teilchen mit negativem Impaktparameter
- Invariante Masse des Jets

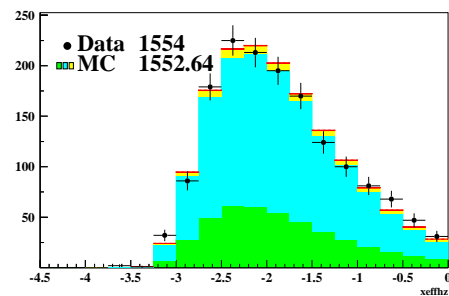


Vergleich 1999-2000 Daten mit MC

DELPHI - $\sqrt{s} = 192\text{-}210$ (450 pb^{-1})



Sommer 2000



derzeitige Analyse



DELPHI Collaboration

4-jet Kanal: Verwendete Größen

- Anti-QCD Variable:
 - Summe des zweiten und vierten Fox-Wolfram Momentes
 - Produkt der minimalen Jetenergie und des minimalen Öffnungswinkels zwischen je zwei Jets
 - Maximaler und minimaler Jetimpuls
 - Summe der Multiplizitäten der beiden Jets mit niedrigstem Impuls
 - Summe der Massen der beiden Jets mit niedrigster Masse
 - Minimale Masse aller Jetpaare
 - Minimale Summe aller $\cos(\alpha)$ der beiden Jetpaare unter Berücksichtigung aller Paare



4-Jet Kanal: Verwendete Größen

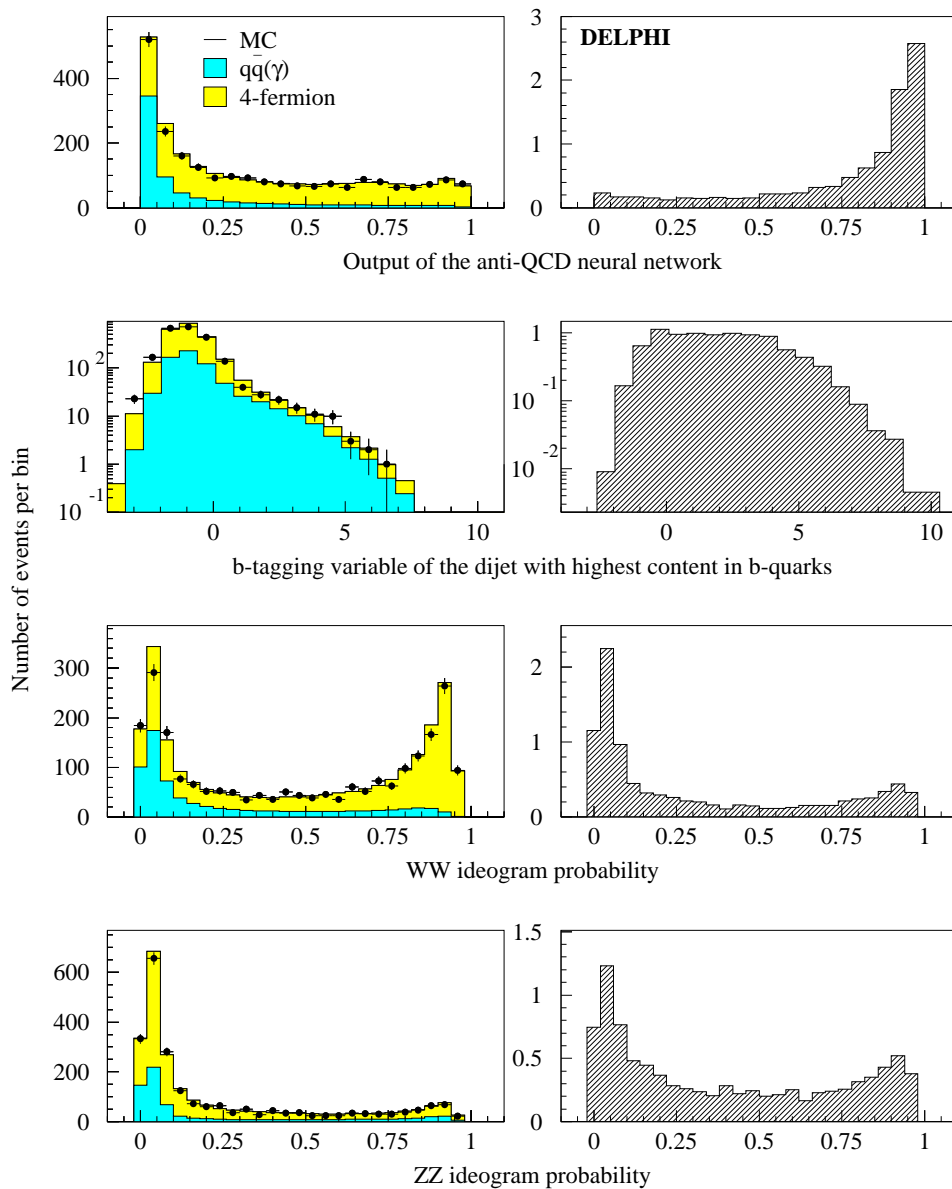
- Ereignis b-Merkmal, definiert als die Summe der b-tags der beiden b-ähnlichsten Jets
- W^+W^- und ZZ
Ideogrammwahrscheinlichkeiten,
interpretiert als 4 oder 5-Jet Ereignis

Unterschiede zur ursprünglichen Analyse

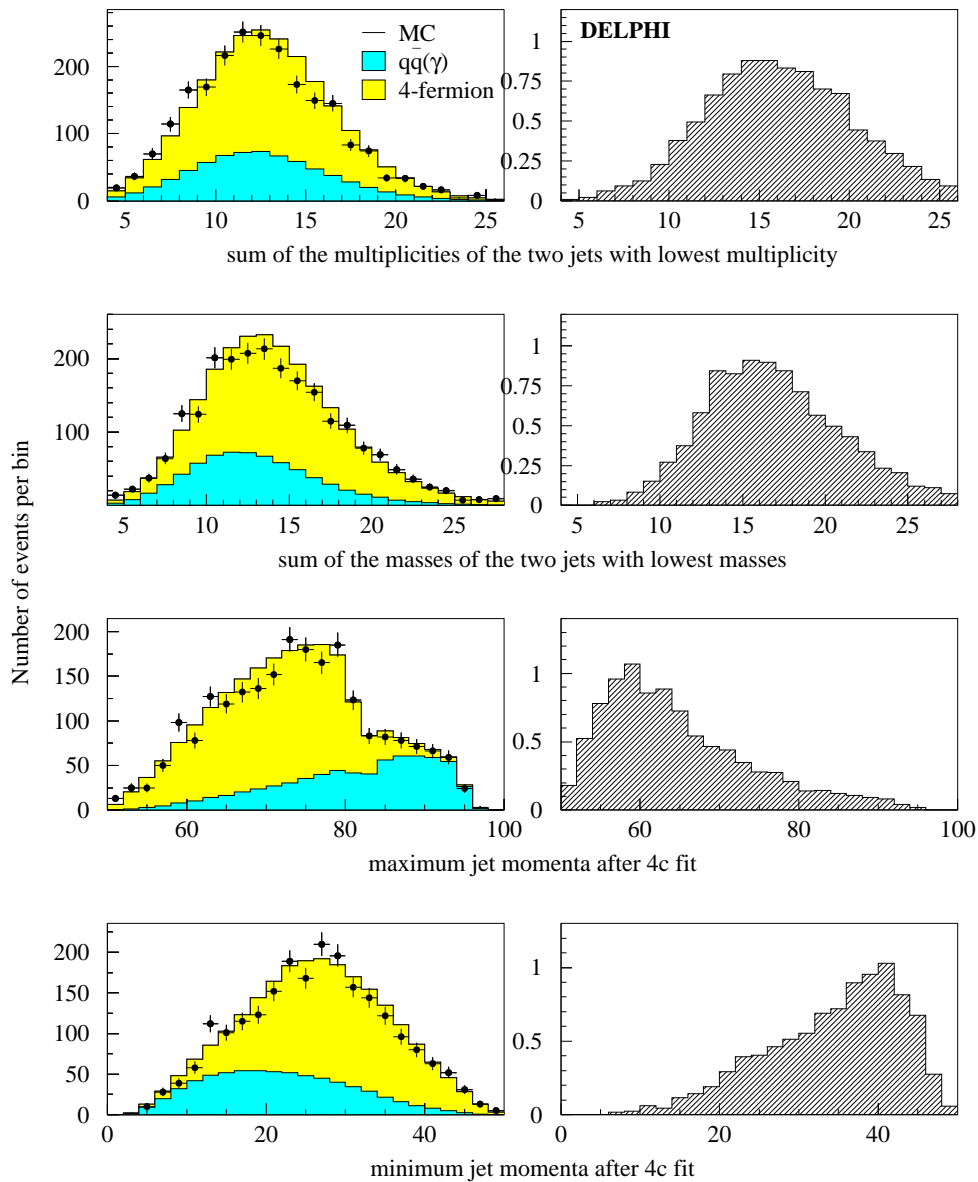
- Anti - QCD Variables vorher in einem NNW kombiniert, optimiert fuer 200 GeV
- 5-Jet Ideogramme geben zusätzliche Information



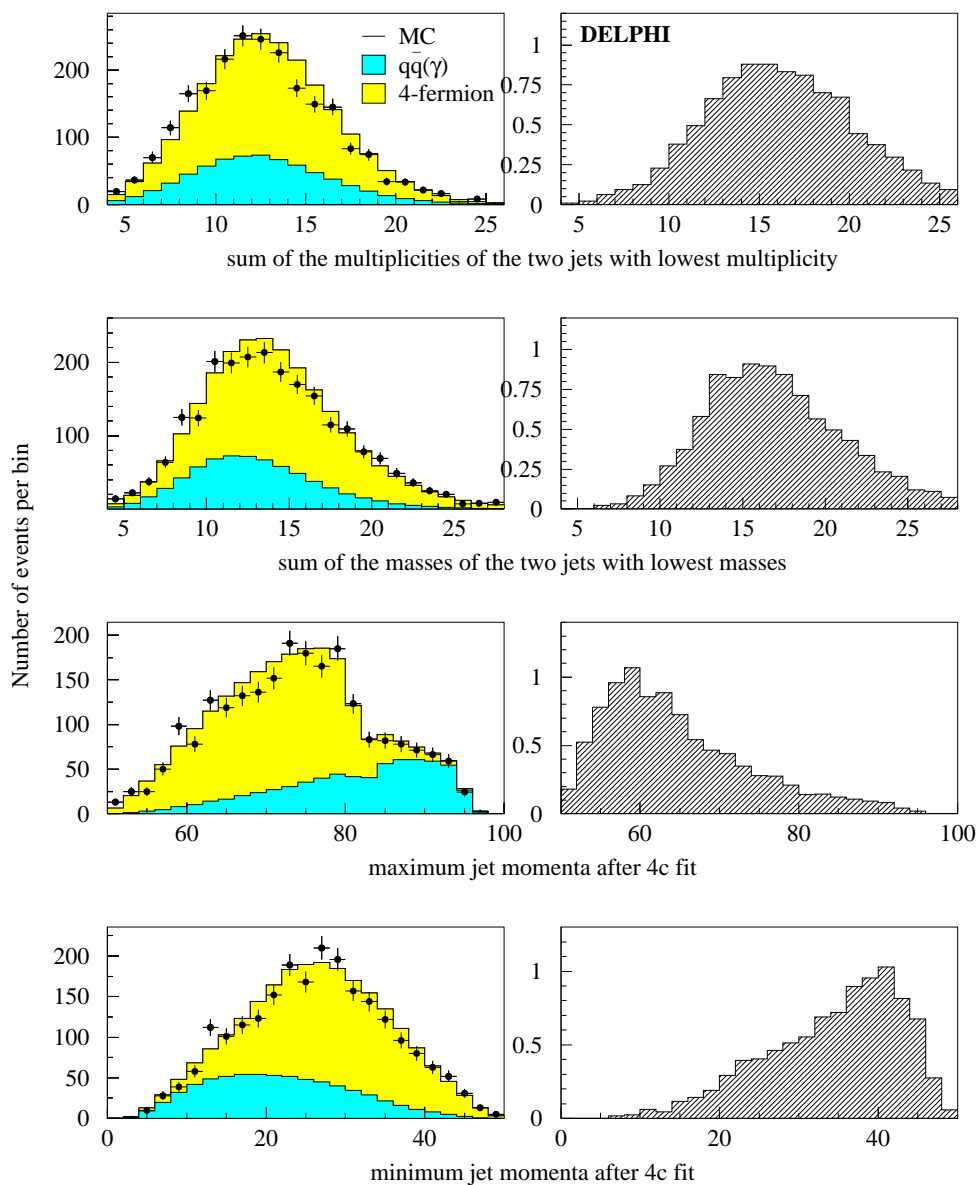
4-Jet Kanal: Verwendete Größen



4-Jet Kanal: Verwendete Größen



4-Jet Kanal: Ideogrammwahrscheinlichkeiten



4-jet Kanal: Massenrekonstruktion

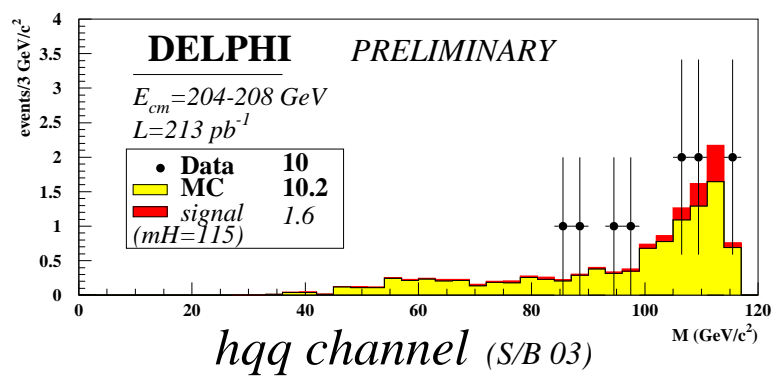
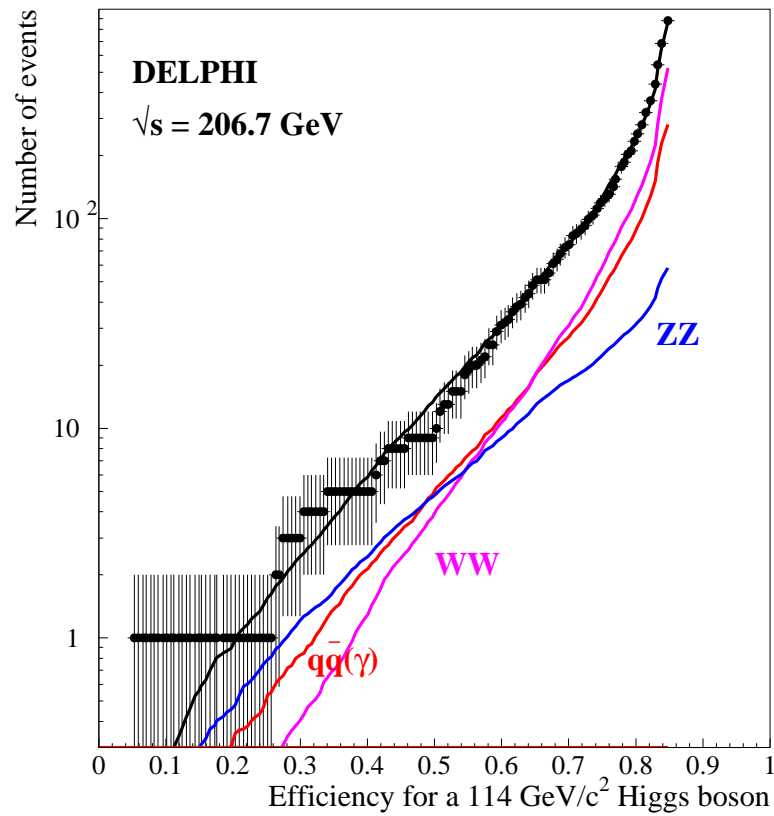
Wahl des Paares durch LLH Funktion:

$$\begin{aligned}
 & ((1 - R_b^Z - R_c^Z) \cdot \mathcal{P}_b^{j_1} \cdot \mathcal{P}_b^{j_2} \cdot \mathcal{P}_q^{j_3} \cdot \mathcal{P}_q^{j_4} \\
 & + R_b^Z \cdot \mathcal{P}_b^{j_1} \cdot \mathcal{P}_b^{j_2} \cdot \mathcal{P}_b^{j_3} \cdot \mathcal{P}_b^{j_4} \\
 & + R_c^Z \cdot \mathcal{P}_b^{j_1} \cdot \mathcal{P}_b^{j_2} \cdot \mathcal{P}_c^{j_3} \cdot \mathcal{P}_c^{j_4} \\
 & \cdot P_{j_3, j_4}^{5C}
 \end{aligned}$$

- R^Z BR des Z Boson in b oder c Quarks
- \mathcal{P} PDF des b -tag, aus MC
- P_{j_3, j_4}^{5C} Constrained Fit Wahrscheinlichkeit, mit Zwangsbedingung $M_{j_3, j_4} = M_Z$



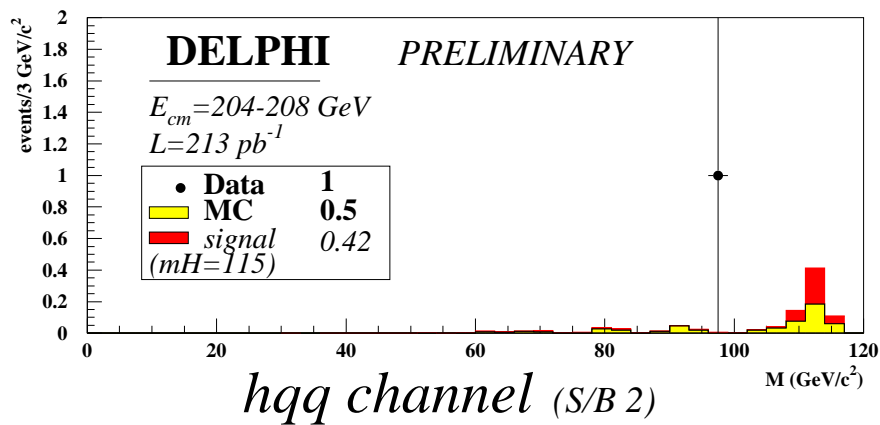
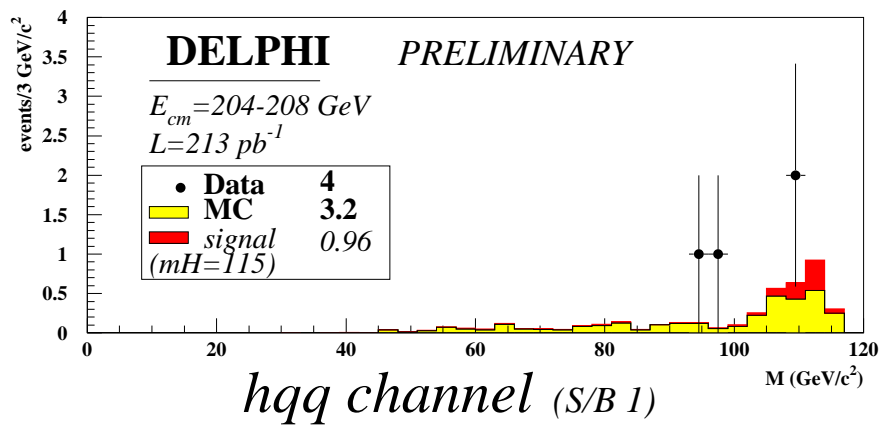
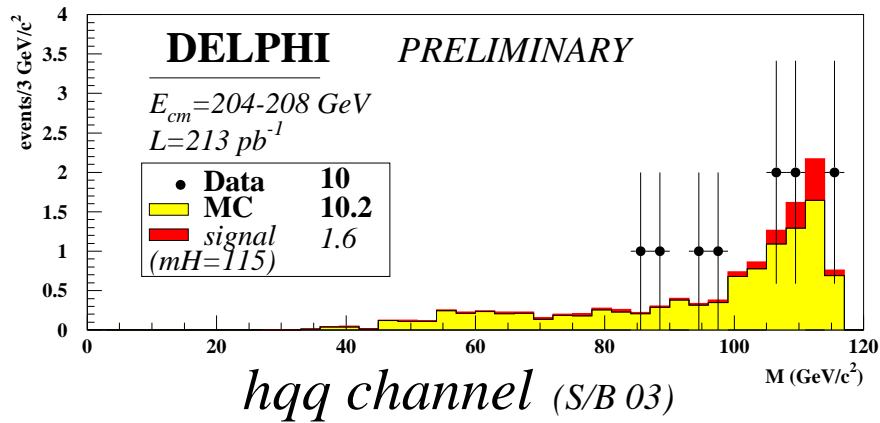
Ergebnisse im 4-Jet Kanal

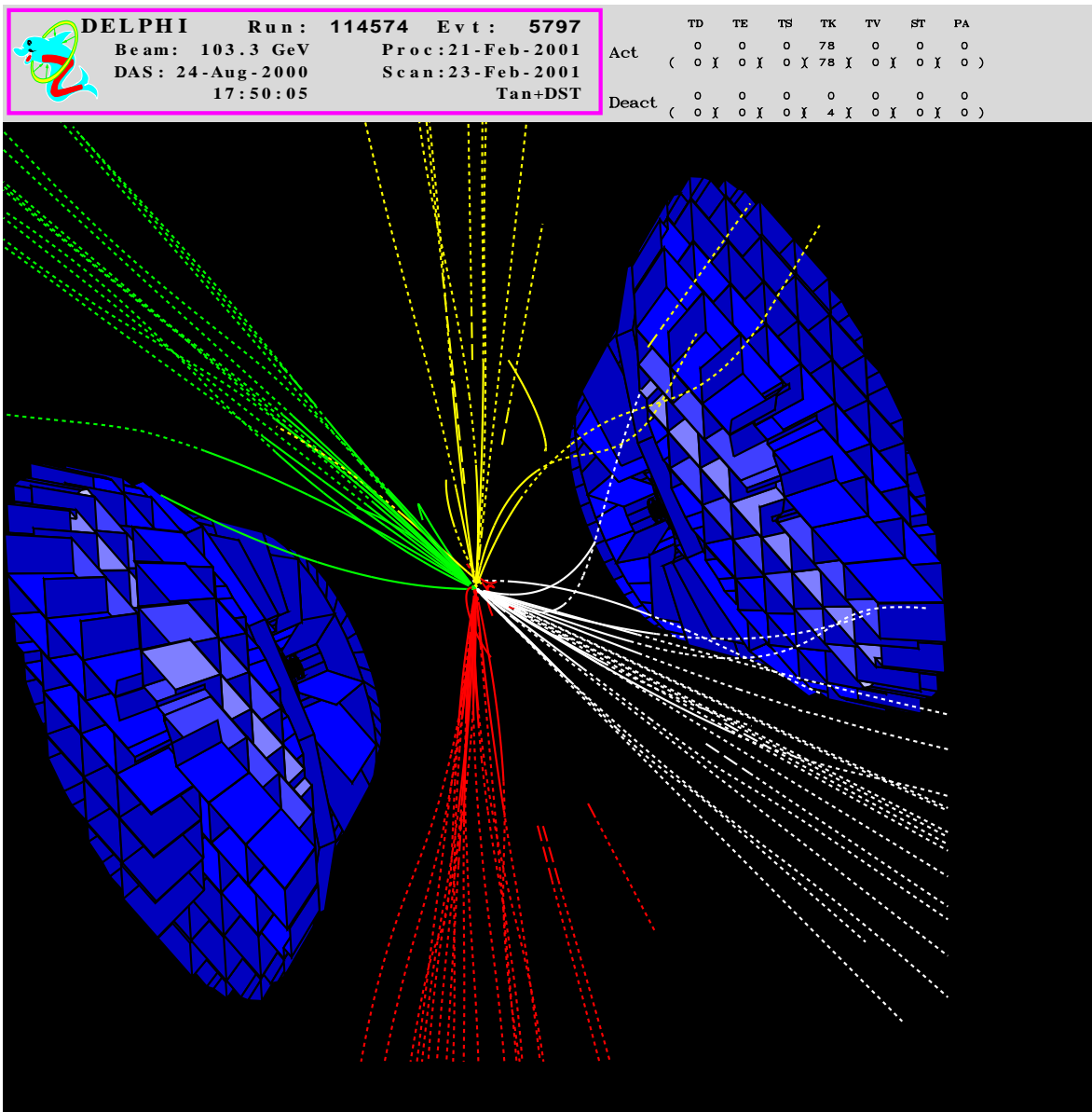


Kein Hinweis auf ein Signal



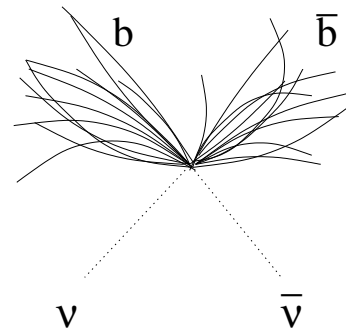
4-jet Kanal: Massenverteilung





Kanal mit fehlender Energie

$$H \rightarrow \text{any}, Z \rightarrow \nu\bar{\nu}$$

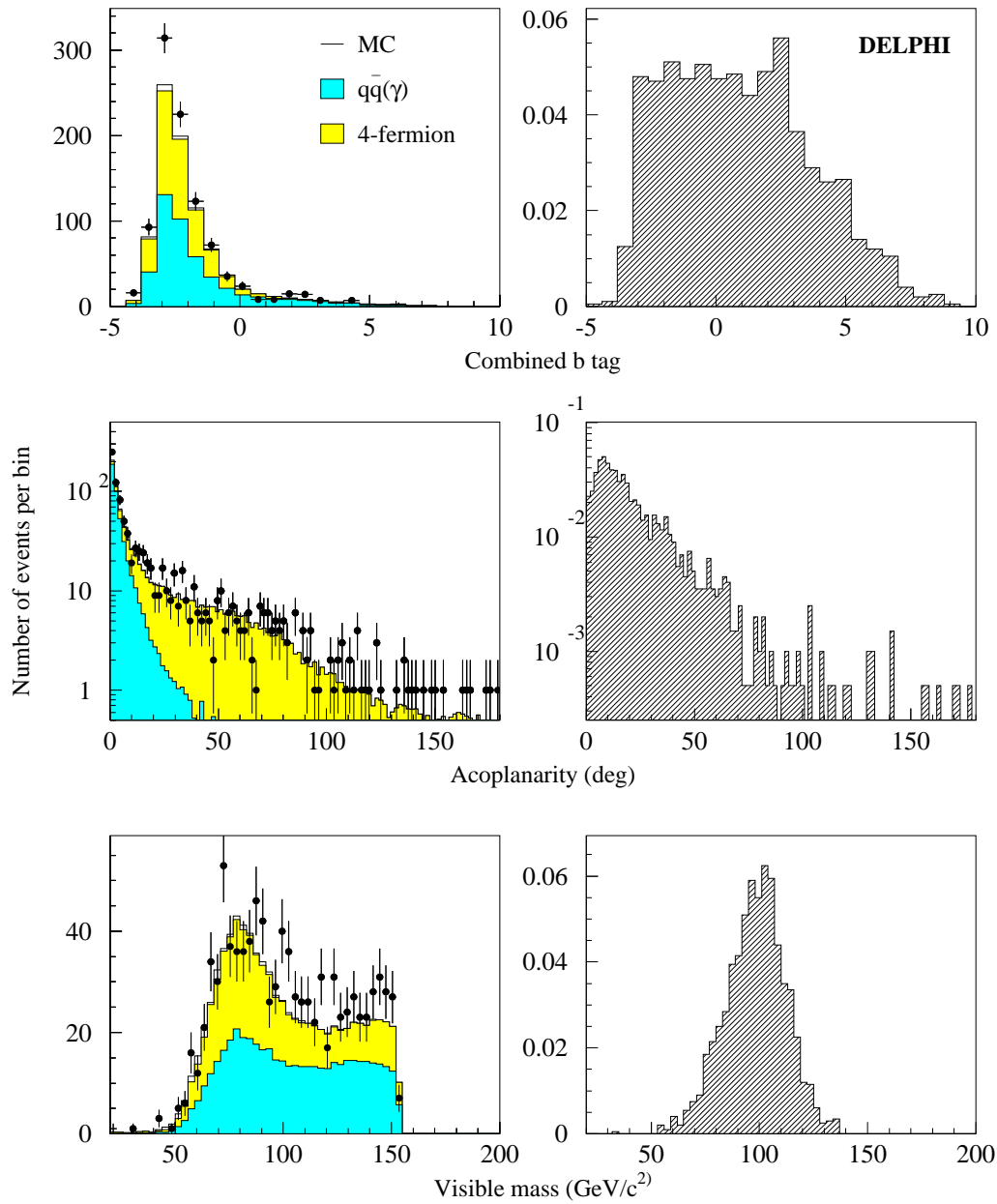


- Vierstufige sequentielle Vorselektion
- likelihood Analyse

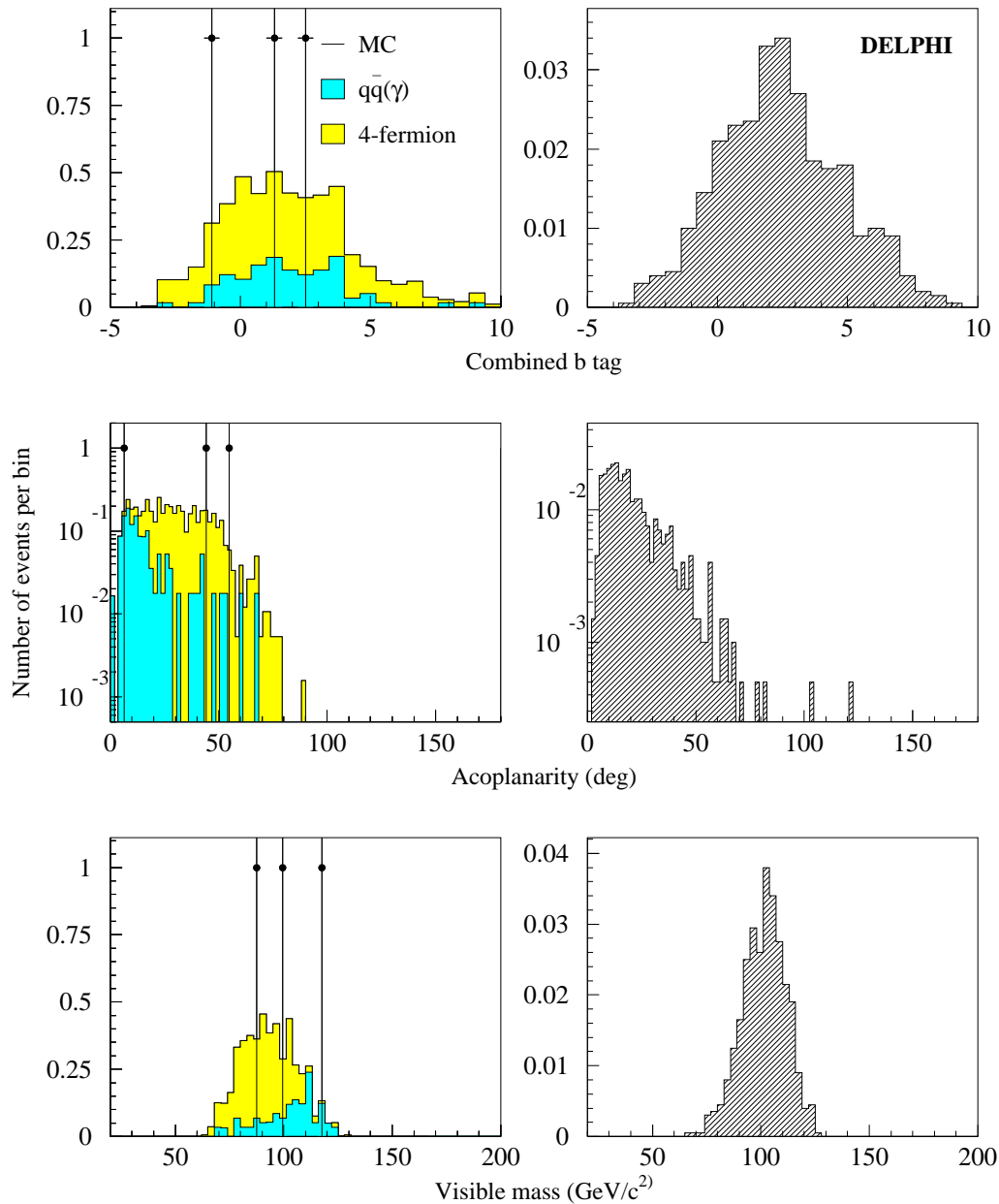
Cuts	data	background	eff. [%]
bhabha $\gamma\gamma$	17503	17753	86.1
$q\bar{q}(\gamma)$	1808	1681	78.3
W^+W^-	1357	1189	75.5
final presel.	970	851	66.9



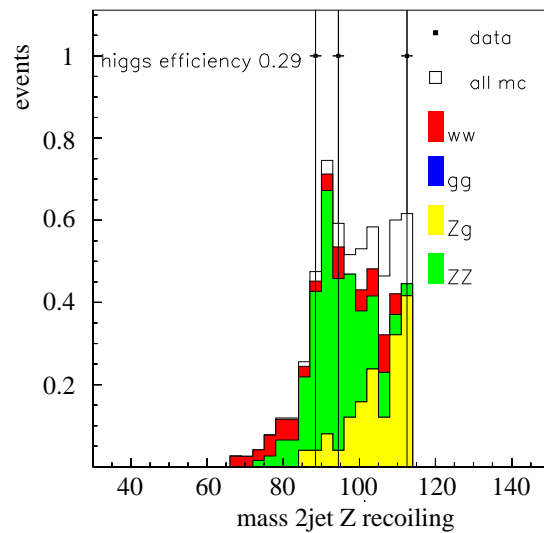
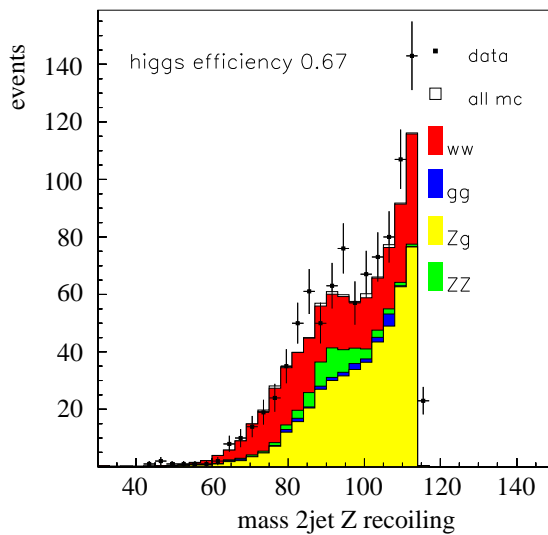
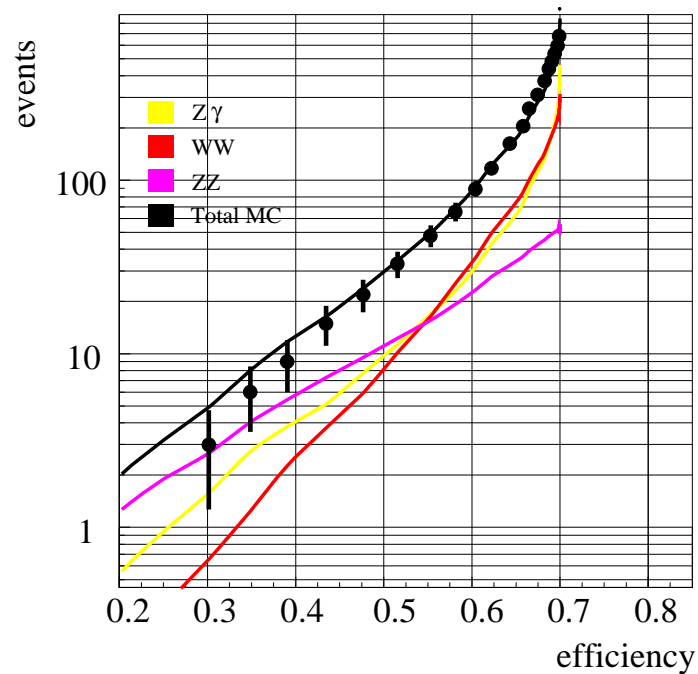
fehlende Energie: Vorselektion



fehlende Energie: harte Schnitte



Ergebnisse:



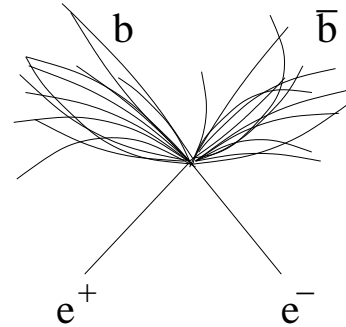
Kein Hinweis auf ein Signal



Elektron - Kanal

$$H \rightarrow \text{any}, Z \rightarrow e^+e^-$$

- Sequentielle Analyse
- Letzter Schnitt im b-tag



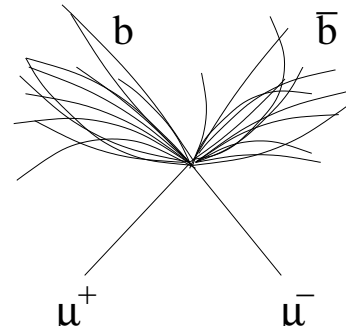
Cut	data	background	eff. [%]
Presel.	1242	1171.6	77.9
Ident.	168	195.9	65.6
El.ene.	68	78.7	63.3
Fit pr.	31	30.2	60.1
Jet Iso.	13	14.6	57.0
Mass sel.	7	11.6	56.7
btag	1	3.5	49.3

Kein Hinweis auf ein Signal



Myon - Kanal

$$H \rightarrow \text{any}, Z \rightarrow \mu^+ \mu^-$$



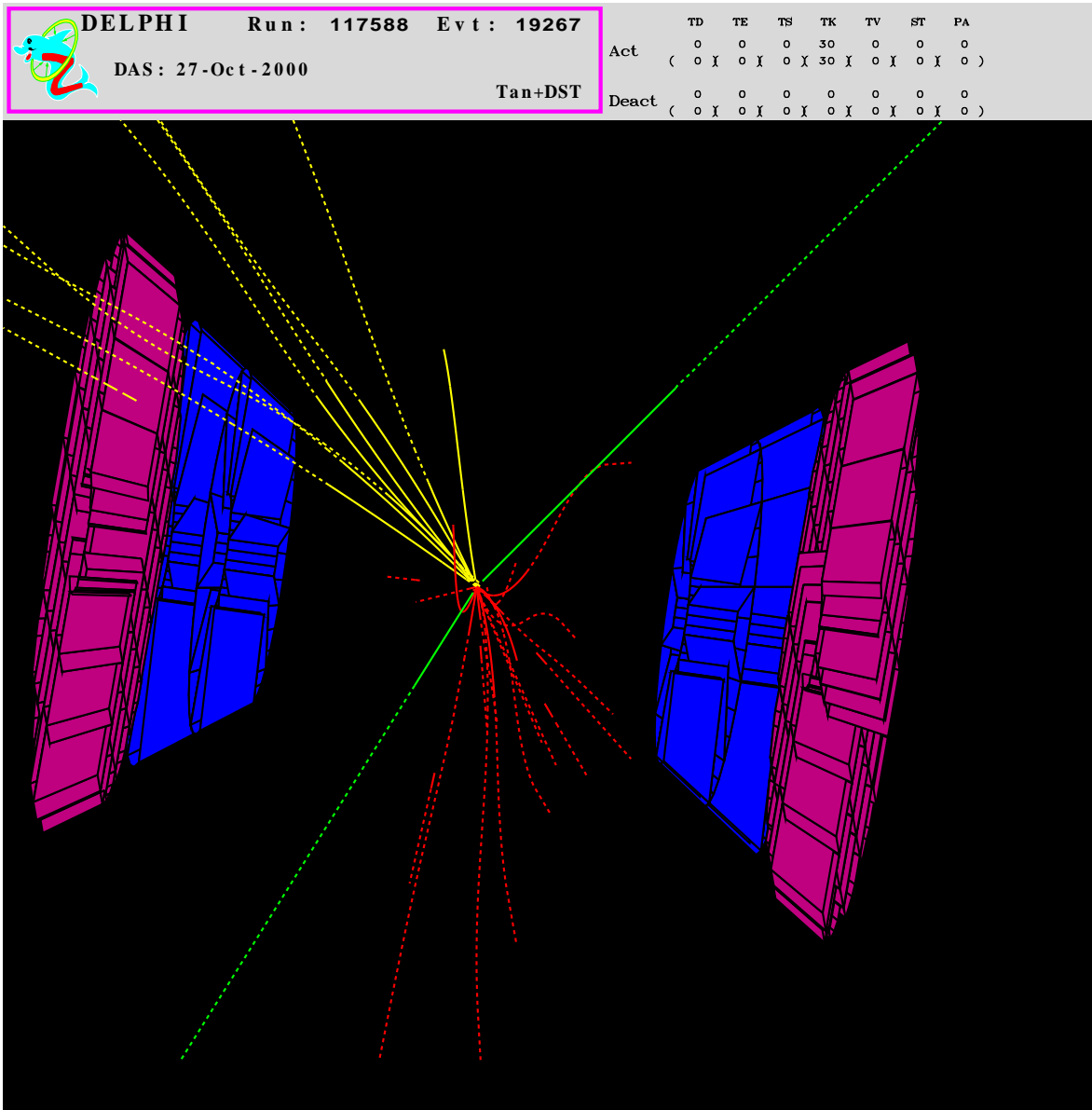
- Sequentielle Analyse
- letzter Schnitt im Btag

Cut	data	background	eff. [%]
Presel	3780	3763.9	81.
Dil. ang.	2858	2812.7	81.
momentum	991	893.6	78.
Muon Id.	26	27.1	73.
Isolation	12	15.4	71.
5C - Fit	7	10.6	68.
B tag	3	5.8	61.

Kein Hinweis auf ein Signal



$H_{\mu\mu}$ Kanal

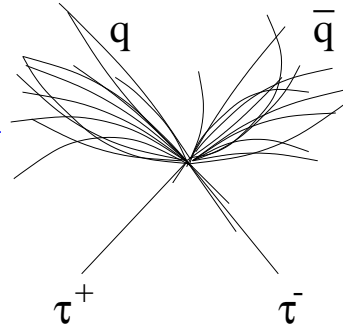


$$M_{qq} = 94 \text{ GeV}/c^2, M_{\mu\mu} = 85 \text{ GeV}/c^2, \text{large btag}$$



$H_{\tau\tau}$ und $\tau\tau Z$ Kanäle

$H \rightarrow \text{any but } \tau, Z \rightarrow \tau^+\tau^-$
 $H \rightarrow \tau^+\tau^-, Z \rightarrow q\bar{q}$



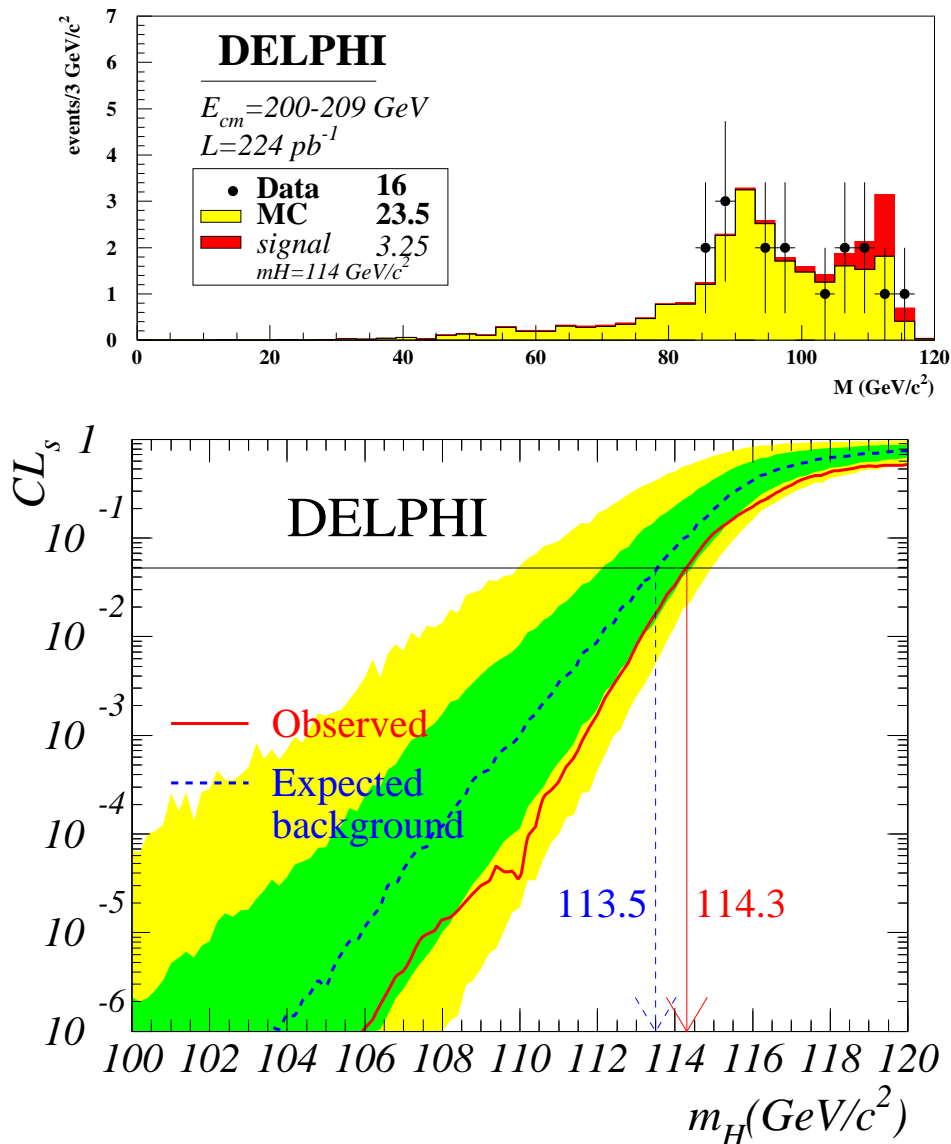
- τ Selektion mit likelihood Funktion
- Sequentielle Analyse
- Auch MSSM $h, A \rightarrow \tau^+\tau^-$ Kanäle

Cut	data	background	eff. [%]
Presel.	9180	8876.0	98.1
final	5	7.15	15.9

Kein Hinweis auf ein Signal



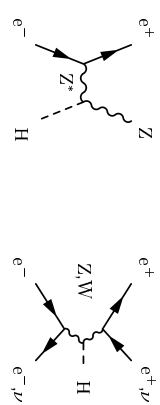
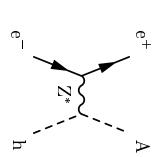
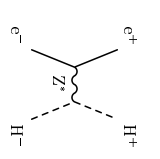
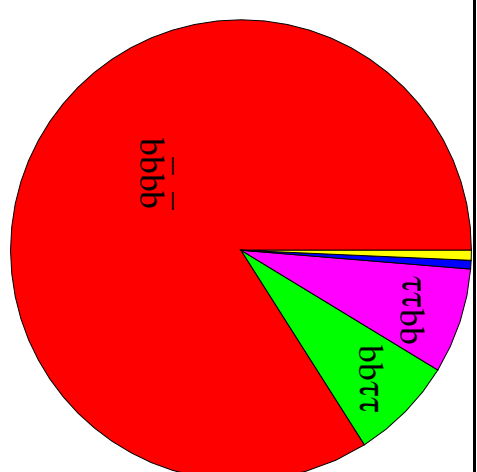
SM Ergebnisse: DELPHI Limits



$M_H > 114.3 \text{ GeV}/c^2$ (113.5 GeV/c^2 exp.)



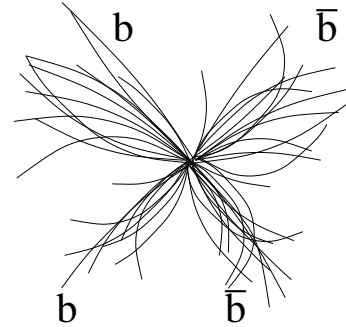
Jenseits des SM: 2-Douplet Modell, MSSM, ...

Produktion	Zerfall	
<p>h, H</p> 	<p>h, H</p> 	<p>H^\pm</p> 
<p>$b\bar{b}$</p> <p>$\tau^+\tau^-$</p> <p>\dots</p> <p>$\tilde{\chi}^0\tilde{\chi}^0$</p>		<p>$c\bar{s}c s$</p> <p>$c s \tau \nu$</p> <p>$\tau^+\nu\tau^-\bar{\nu}$</p> <p>\dots</p>



$$e^+e^- \rightarrow hA \text{ Suchen}$$

$$h \rightarrow b\bar{b}, A \rightarrow b\bar{b}$$

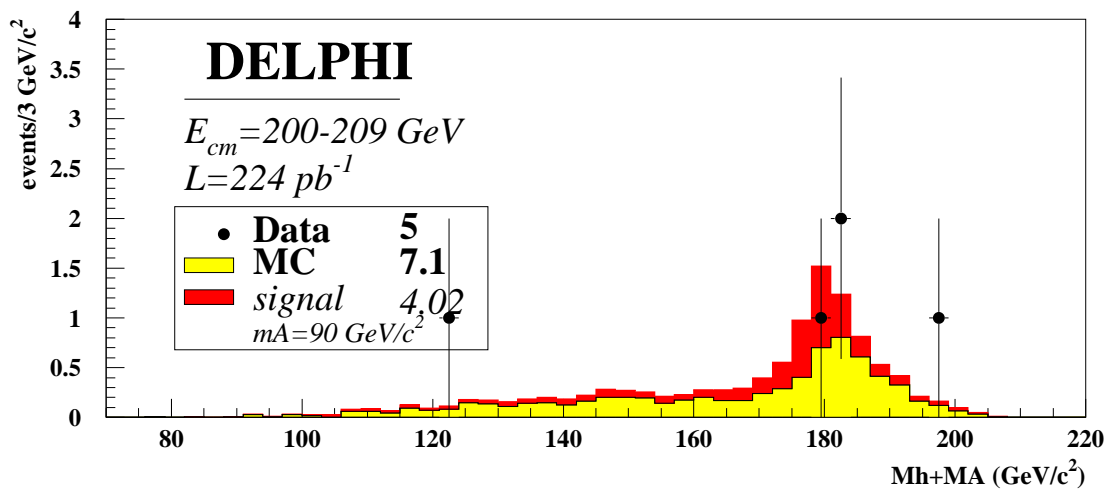
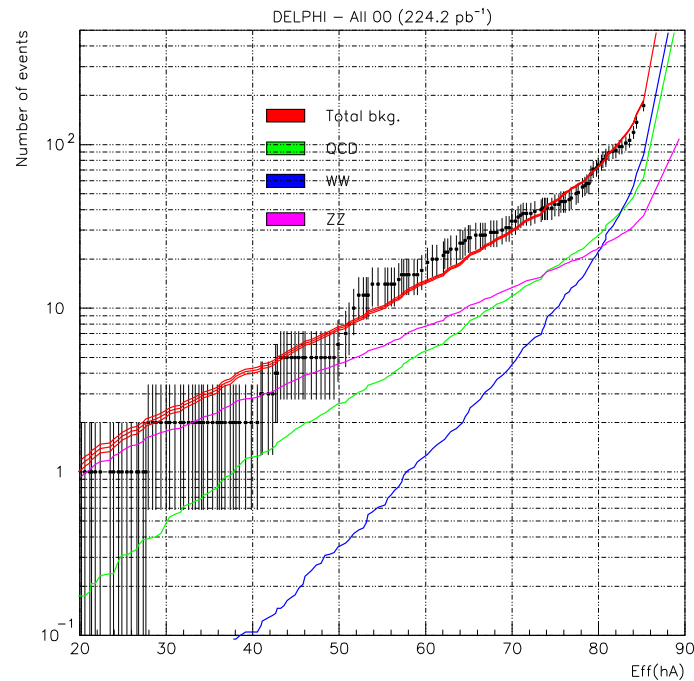


- Sequentielle Vorselektion analog SM $Hq\bar{q}$
- zusätzliche Schnitte gegen $q\bar{q}(g)$
- likelihood Analyse

Cut	data	background	eff. [%]
presel.	1803	1933.5	89.2
intermed.	129	145.2	84.1
s/b=0.3	24	18.7	63.9
s/b=1	2	3.1	35.2



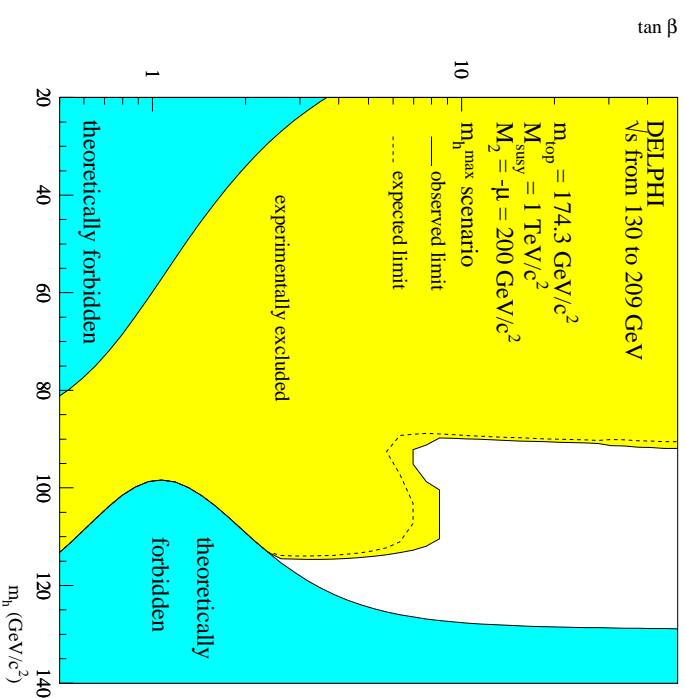
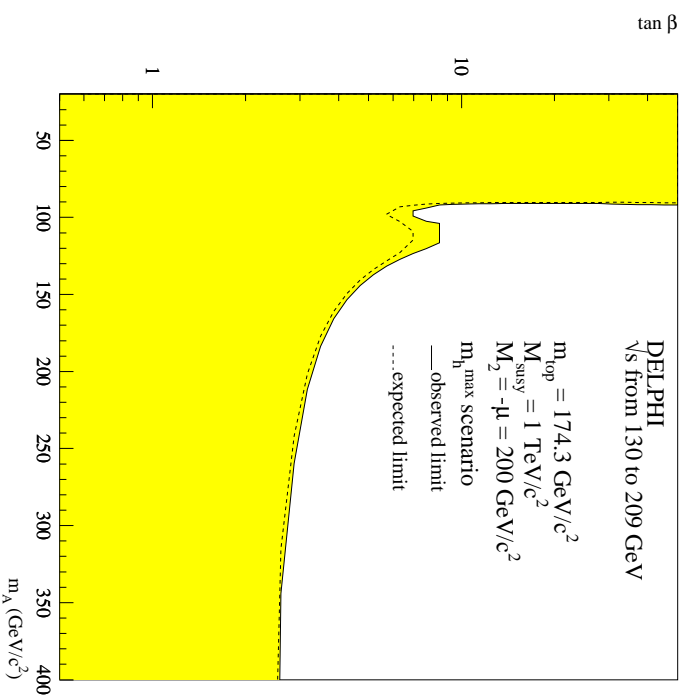
$e^+e^- \rightarrow hA$ Suchen



Verträglich mit SM Untergrund



Limits im MSSM: m_h^{max} Szenario



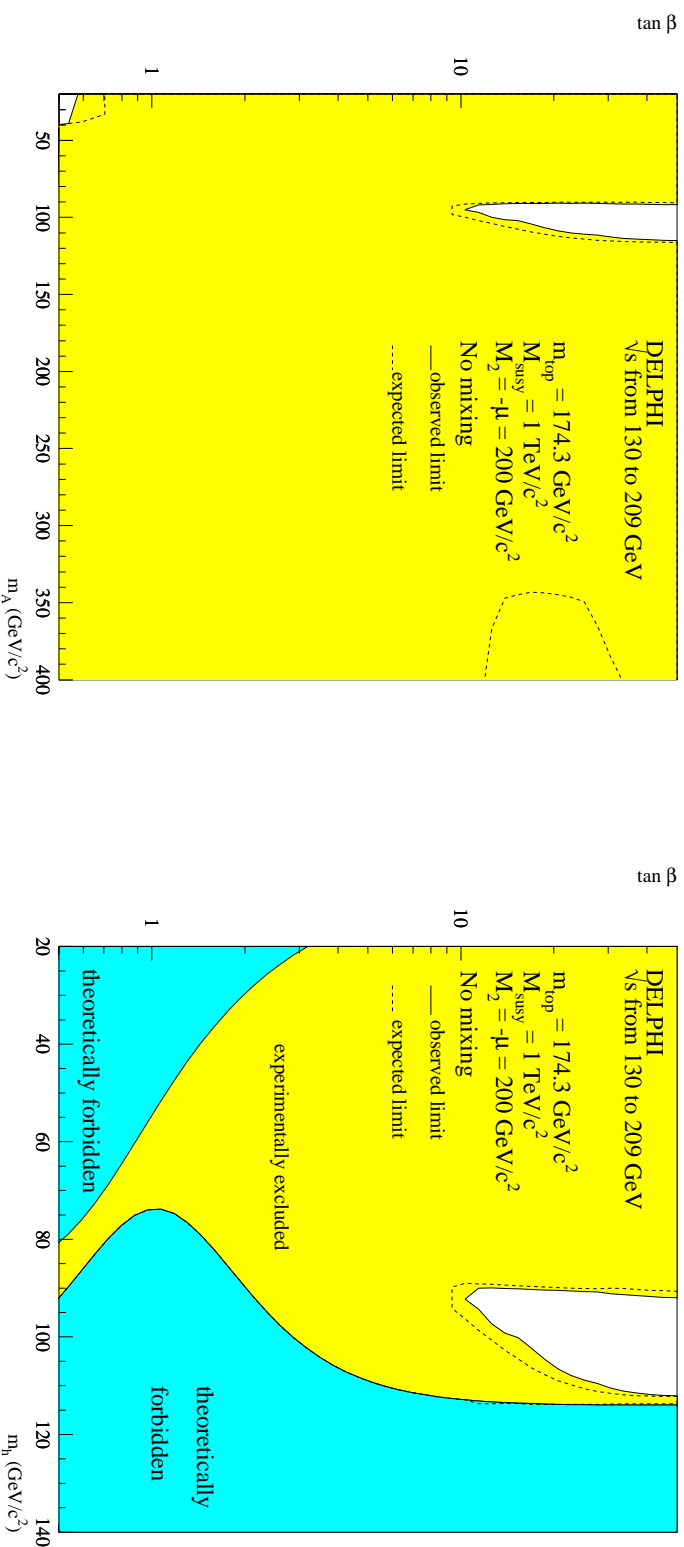
95% CL Ausschluß, m_h^{max} Szenario:

$$m_A > 90.9 \text{ GeV}/c^2 \text{ (90.1 GeV}/c^2), m_h > 89.8 \text{ GeV}/c^2 \text{ (89.0 GeV}/c^2)$$

$$0.49 < \tan\beta < 2.36 \text{ (0.54 < \tan\beta < 2.36)}$$



Ausschlußgrenzen im MSSM 'No mixing Szenario'



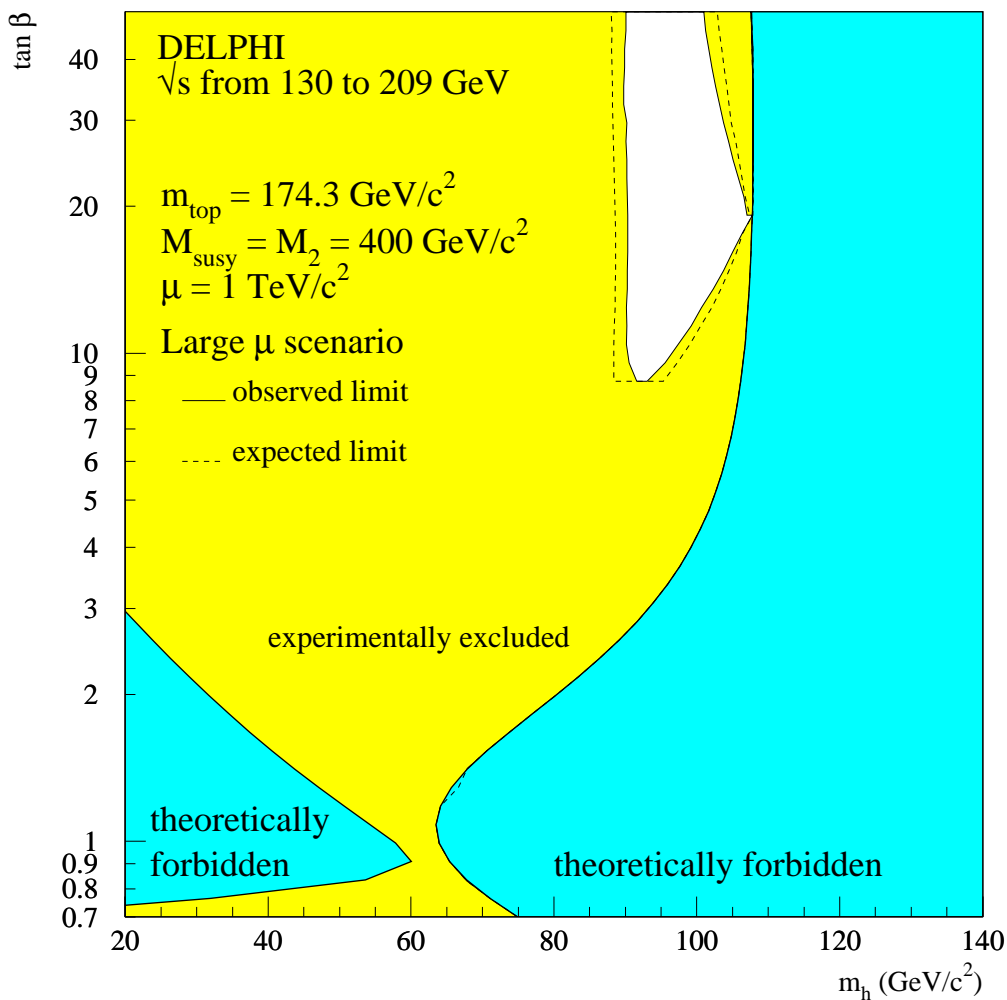
95% CL Ausschluß, no mixing

$$m_A > 90.8 \text{ GeV}/c^2 \text{ (} 90.0 \text{ GeV}/c^2 \text{), } m_h > 90.0 \text{ GeV}/c^2 \\ (89.1 \text{ GeV}/c^2) \text{ } 0.59 < \tan \beta < 9.36 \text{ (} 0.72 < \tan \beta < 9.36 \text{)}$$



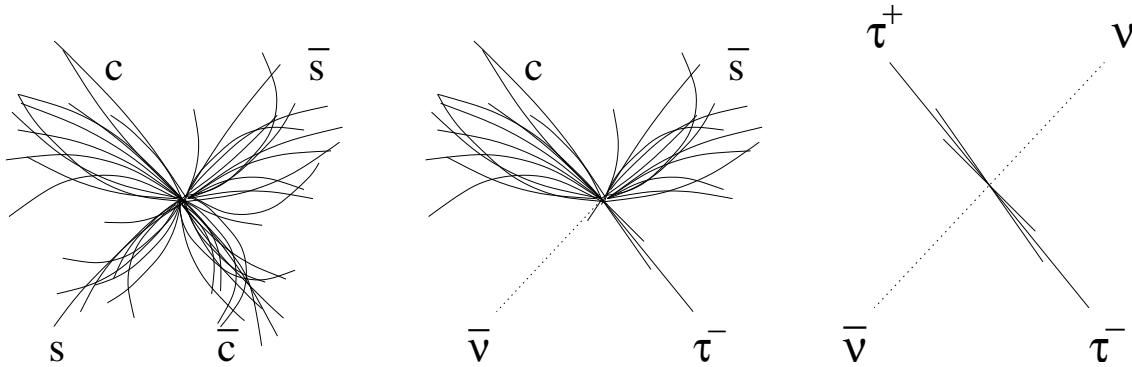
Interpretation in the MSSM: large μ

- Vorschlag von Carena et al.
(CERN-TH/99-374)
- Higgsmasse immer in LEP2 Gebiet
- $h, H \rightarrow b\bar{b}$ können unterdrückt sein



Suche nach geladenen Higgsbosonen

$$H^+ H^- \rightarrow c\bar{s}c\bar{s}, c\bar{s}\tau\nu, \tau\nu\tau\nu$$



- sequentielle Vorselektion
- likelihood Analyse in allen drei Kanälen
- Hauptproblem ist WW Untergrund

Channel	data	background	eff. [%]
$c\bar{s}c\bar{s}$	1040	1058.5	34
$c\bar{s}\tau\nu$	498	498.2	35
$\tau\nu\tau\nu$	64	63.7	34

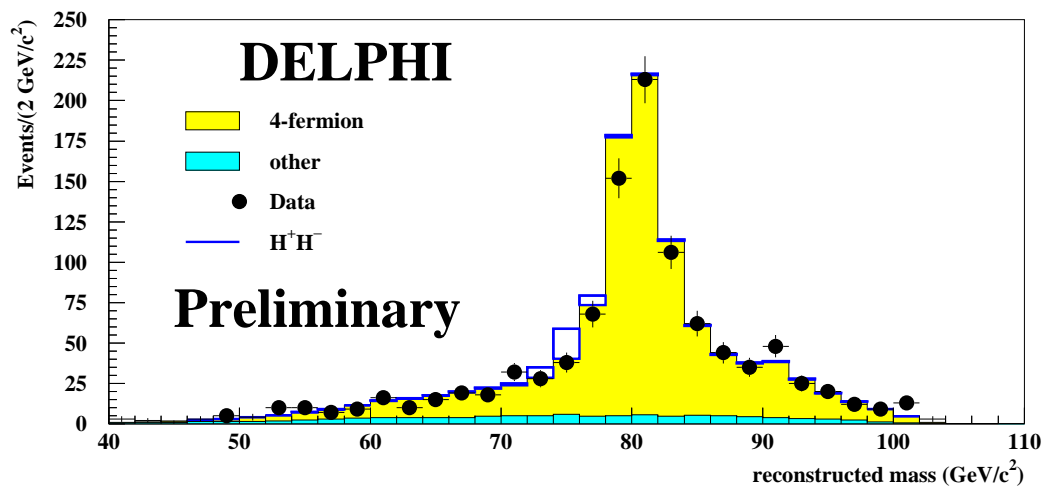
(Effizienzen beziehen sich auf ein $70 \text{ GeV}/c^2$ Signal)

$$(189 \text{ GeV} < \sqrt{s} < 208 \text{ GeV})$$

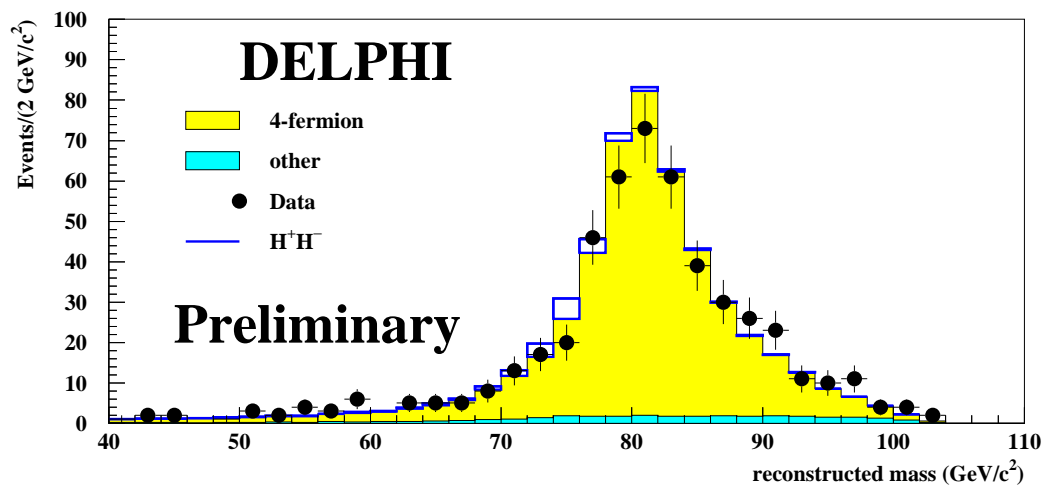


Massenverteilung

hadronic channel

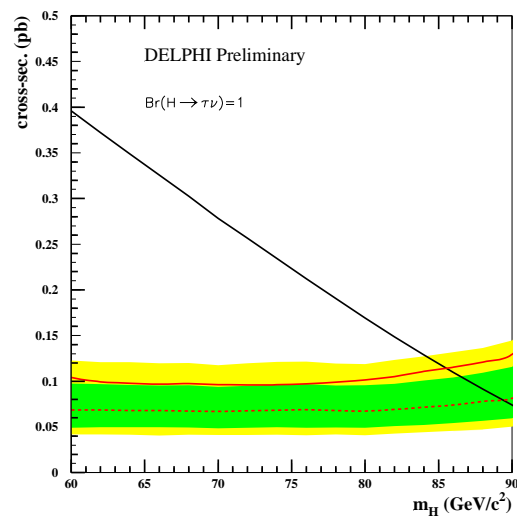
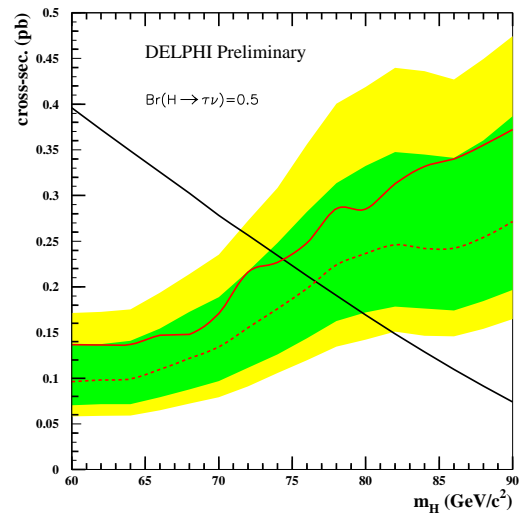
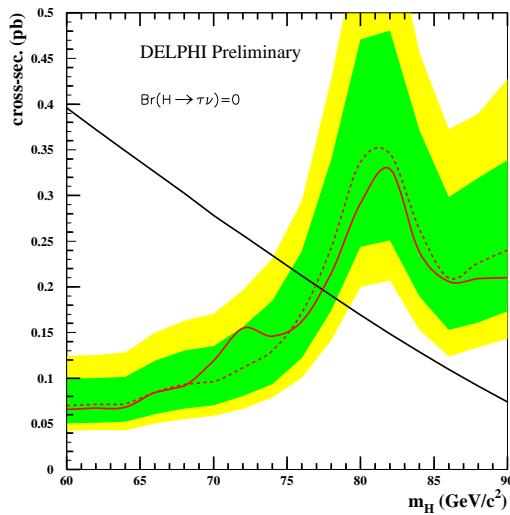


semileptonic channel



DELPHI Collaboration

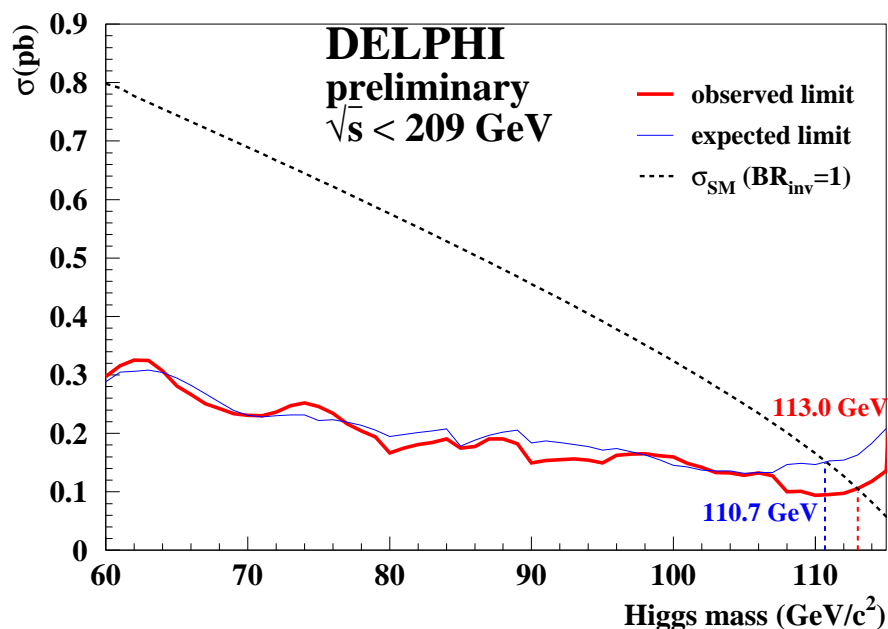
Schranken für den Wirkungsquerschnitt



Unsichtbare Higgszerfälle

$$H \rightarrow inv, Z \rightarrow q\bar{q}, l^+l^-$$

- IDA im Hadronischen Kanal
- Sequentielle Analysen in den leptonischen Kanälen



BR 100%	$m_H > 113.0 \text{ GeV}/c^2$ (110.7 GeV/c^2 exp.)
Any BR	$m_H > 112.6 \text{ GeV}/c^2$ (109.8 GeV/c^2 exp.)



Zusammenfassung der Ergebnisse

framework	observed	expected	remarks
SM	114.3 GeV/c ²	113.5 GeV/c ²	-
$H^+ H^-$	73.8 GeV/c ²	75.4 GeV/c ²	any BR($H^\pm \rightarrow \tau \nu$)
H_{inv}	113.0 GeV/c ²	110.7 GeV/c ²	BR($H \rightarrow inv$)=1
$H_{inv} + H_{vis}$	112.6 GeV/c ²	109.8 GeV/c ²	any BR($H \rightarrow inv$)
MSSM m_A	90.9 GeV/c ²	90.1 GeV/c ²	m_h^{max} scenario
m_h	89.8 GeV/c ²	89.0 GeV/c ²	
$\tan \beta$	0.49 ... 2.36	0.54 ... 2.36	
MSSM m_A	90.8 GeV/c ²	90.0 GeV/c ²	no mixing
m_h	90.0 GeV/c ²	89.1 GeV/c ²	
$\tan \beta$	0.59 ... 9.36	0.72 ... 9.36	

Alle Zahlen sind vorläufig!



weitere Aktivitäten

- $h \rightarrow AA, A \rightarrow Zh$
- $H^{\pm\pm}$
- flavour blind/ nicht $b\bar{b}$ Higgs decays:
 - $h \rightarrow \gamma\gamma$
 - $h \rightarrow W^+W^-$
 - $h \rightarrow$ Gluon Gluon
 - $h \rightarrow s\bar{s}$
- Reprozessierung aller LEP2 Daten und aller MC



Situation am Ende der Datennahme

- **Motivation für Reprozessing**
 - Besseres Verständnis des Detektors am Ende des Experimentes
 - Archivierung: eine Version für alle Jahre
- **Motivation für neue Simulationen**
 - Fortschritt in der Theorie, besseres Verständnis der Generatoren
 - Besseres Verständnis von Teilchenzerfällen (eg. B-Sektor)
- **Motivation für größere Samples**
 - SM Hintergrund Abschätzung bei Suchen
 - Messungen bei LEP2: Modellierung kleiner Effekte (Bose-Einstein, Color-Reconnection etc)
 - Hinterlassenschaft von LEP



Bedarf an Simulationen

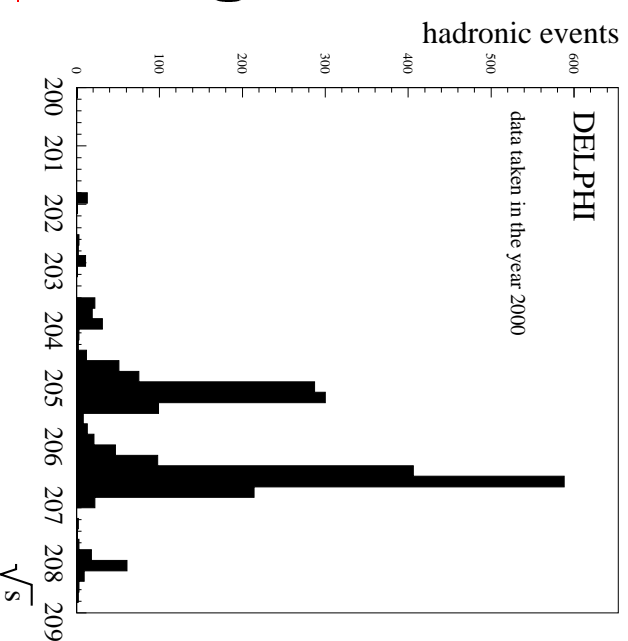
- Messungen, eg. W Masse, Systematische Fehler
 - Fragmentationsmodelle
 - Bose Einstein Korrelationen
 - Color - Reconnection
 - verschiedene M_W
- Suchen, eg. Higgs
 - Untergrundabschätzung
 - feines Gitter in M_H in Signalen
 - exotische oder seltene Higgszerfälle
 - Higgsbosonen jenseits des SM
 - $m_A - m_h$ Scan
- LEP1 Analysen
 - bessere Generatoren
 - Teilchenzerfälle



Bedarf an Simulationen: Beispiel SM Higgs Signal

- 20-120 GeV, 5GeV Schritte
- $\sqrt{s}=192,196,200,202,204,205,207,208$ GeV
- Kanäle: $hq\bar{q}$, $h\bar{e}e$, $h\mu\mu$, $h\tau\tau$, $h\nu\nu$, $\tau\tau Z$
- wenigstens 5000 Ereignisse

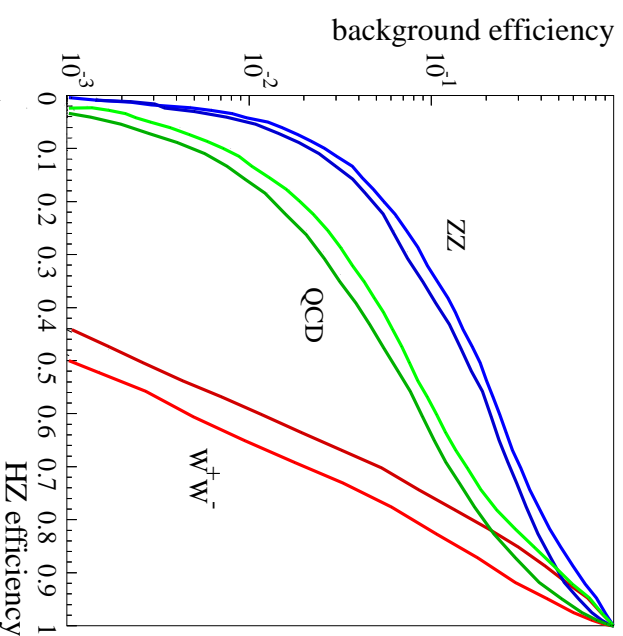
→ 5M Ereignisse, 1000 Samples



Bedarf an Simulationen: Beispiel SM Higgs Untergrund

- WW Unterdrückung von 10^{-3} bei 50% Signaleffizienz
- mindestens 1000 Ereignisse nach diesem Schnitt

→ 1M Ereignisse pro Energiepunkt



Ressourcen

Anfangsphase von LEP	derzeitige Situation
CPU Zeit begrenzt und teuer	(relativ) viel CPU in Form von billigen PC Clustern
begrenzter und teurer Plattenplatz	billige IDE Platten
Datentransport auf Tapes (Post)	schnelle Netze zwischen den Labors
Speicherung auf Exabyte, DLT etc	transparente Speicherung auf Tapes in HPSS bzw. CASTOR



Das CASTOR Projekt

Cern Advanced STORage manager

- Massendatenspeicher
- in Entwicklung für LHC
- Unixartiges Filesystem
- Speicherung auf Tapes im Hintergrund
- Zugriff über RFIO bzw. spezielle Kommandos (Cns Commands)
- Tapeverwaltung im Hintergrund über Oracle Datenbank
- ähnlich wie HPSS



Anforderungen für massive MC Produktionen

- Portierbarkeit
- Parallelisierbarkeit
 - Benutzung von Linuxclustern, Portierung auf Linux
 - Unabhängigkeit vom Environment, verschiedene Labors
 - keine lokale Softwareimplementierung
- Schnelle Verfügbarkeit neuer Ereignisse
 - Automatisierung
 - schnelle Netze
 - Zuverlässige Datenspeicherung
 - transparenter Zugriff durch Anwender
- Systemsicherheit!



Lösungsansätze...

- **Portierbarkeit**
 - Installation benötigter Software “on the fly”
 - Verwendung von perl reduziert Abhängigkeit von Unterschieden in shells
 - Auslagerung laborabhängiger Teile in Modulen
- **Parallelisierung**
 - ein RUN pro Knoten, RUN Nummer und Labor bestimmen Seeds für Zufallszahlen
- **Schnelle Verfügbarkeit**
 - Transport der Ergebnisse übers Netz
 - Verwendung von CASTOR zur Datenspeicherung



Lösungsansätze...

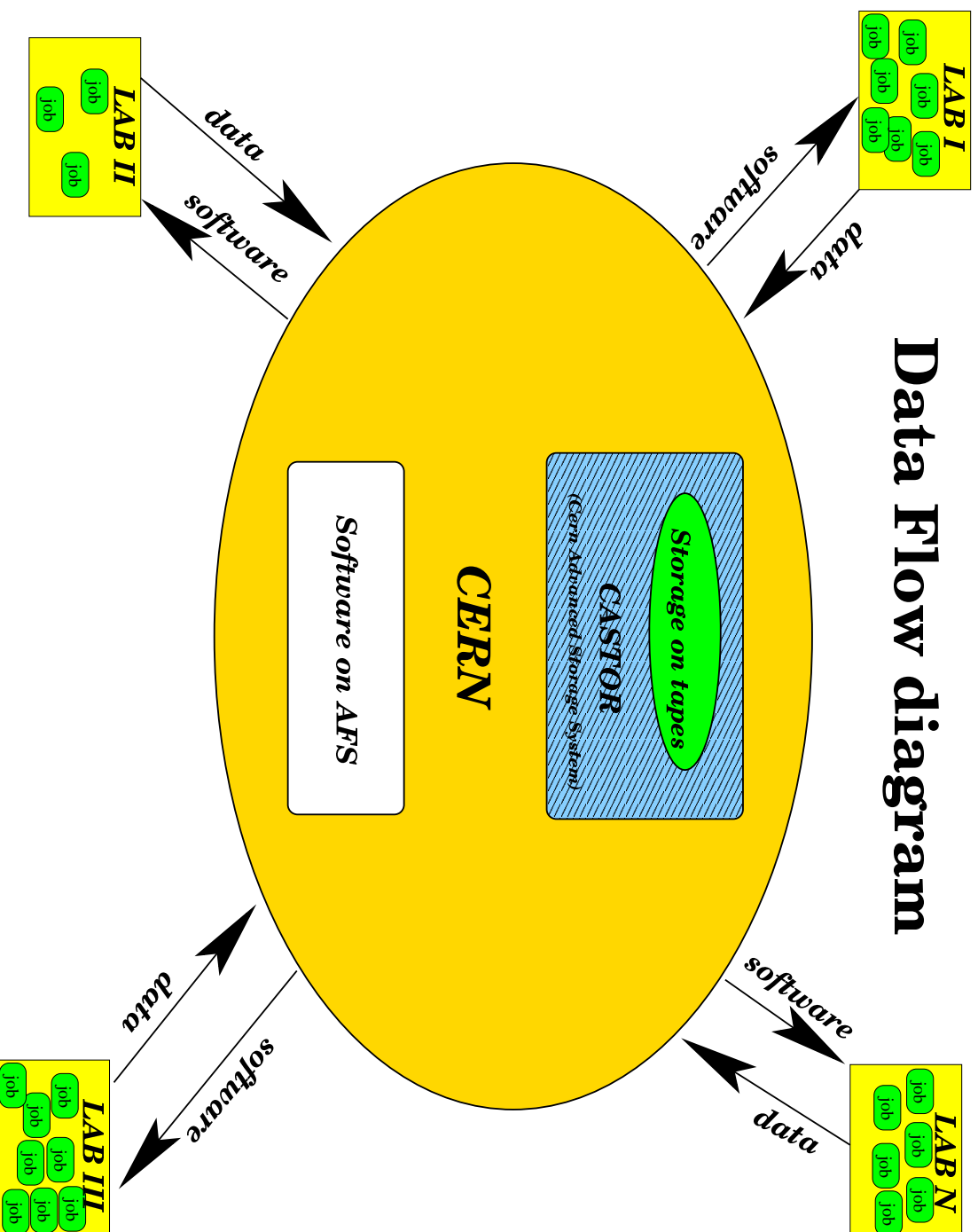
- **Schnelle Verfügbarkeit**
 - weitgehende Automatisierung der Buchführung
 - Zugriff auf Datensets durch *Nicknames*
 - Automatisierung der Dokumentation
- **Systemsicherheit**
 - keine Versendung von Passwörtern in Klartext
 - Verwendung kryptographischer Software, *ssh/scp*
 - Verwendung von *pgp* zum Signieren von Botschaften



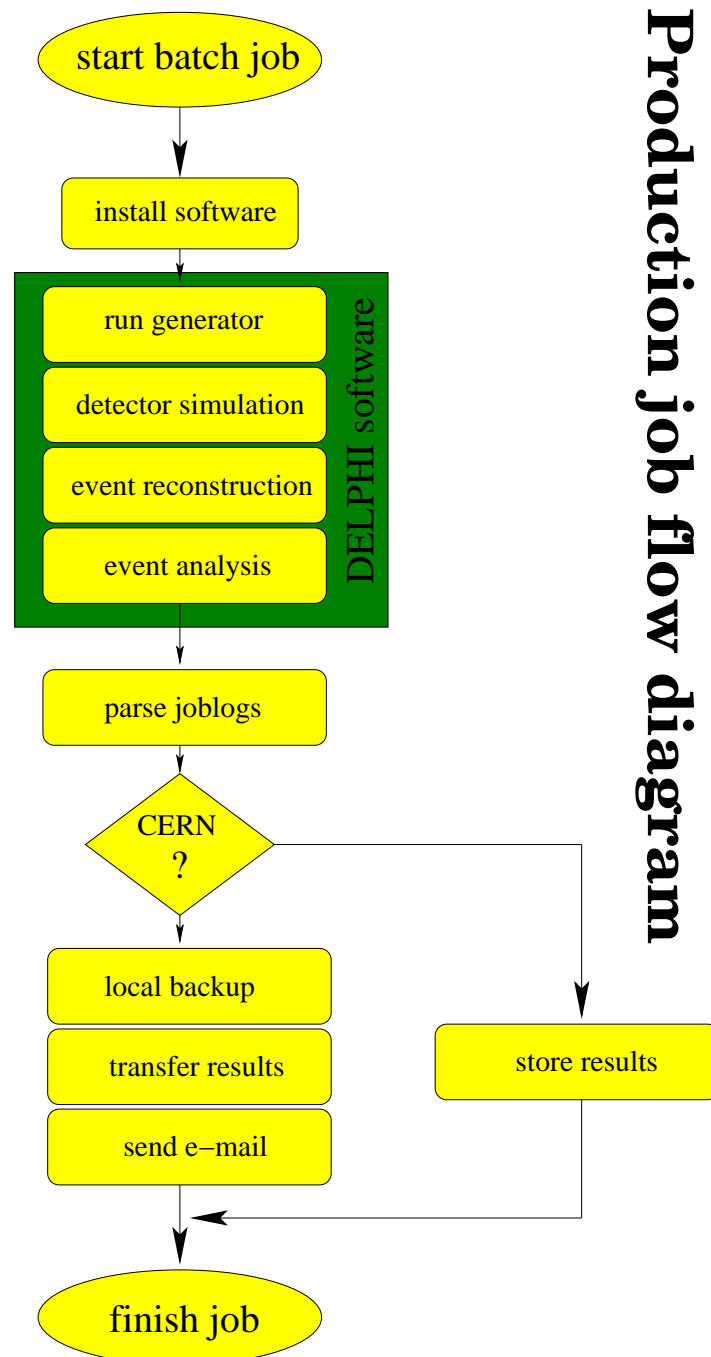
Anforderungen an das Zentrum

- Beschränkung auf Linuxsysteme
- Batch System
- möglichst viele Batchknoten
 - mit genug Arbeitsspeicher
 - großem lokaler Arbeitsplattenplatz
 - lokaler ssh und perl und mail Installation
 - AFS für stets aktuelle Software
- großes global sichtbares Medium zum Speichern von Sicherheitskopien
- schnelle Anbindung ans CERN





Flußdiagramm eines Batchjobs



Benutzerzugriff über Nicknames: Konzept

Charakterisierung eines Samples:

- **Generator:**
 - Art/Modell (eg. Jetset, Pythia, Excalibur, Wphact, HZHA, SUSYGEN)
 - Version
 - Kanal
 - Modellparameter
 - Schwerpunktenergie
- **delsim/delana/dstana**
 - Jahr (Detektorunterschiede!!!)
 - Versionsnummer der Rekonstruktionssoftware (ie.: Prozeßing, Kalibration etc)
 - Versionsnummer der Ereignisanalysesoftware



Beispiel eines Nickname

Beispiel:

xs_hzha03pyth6156ha02_e206.7_m90_ca0_1l_d1/C001

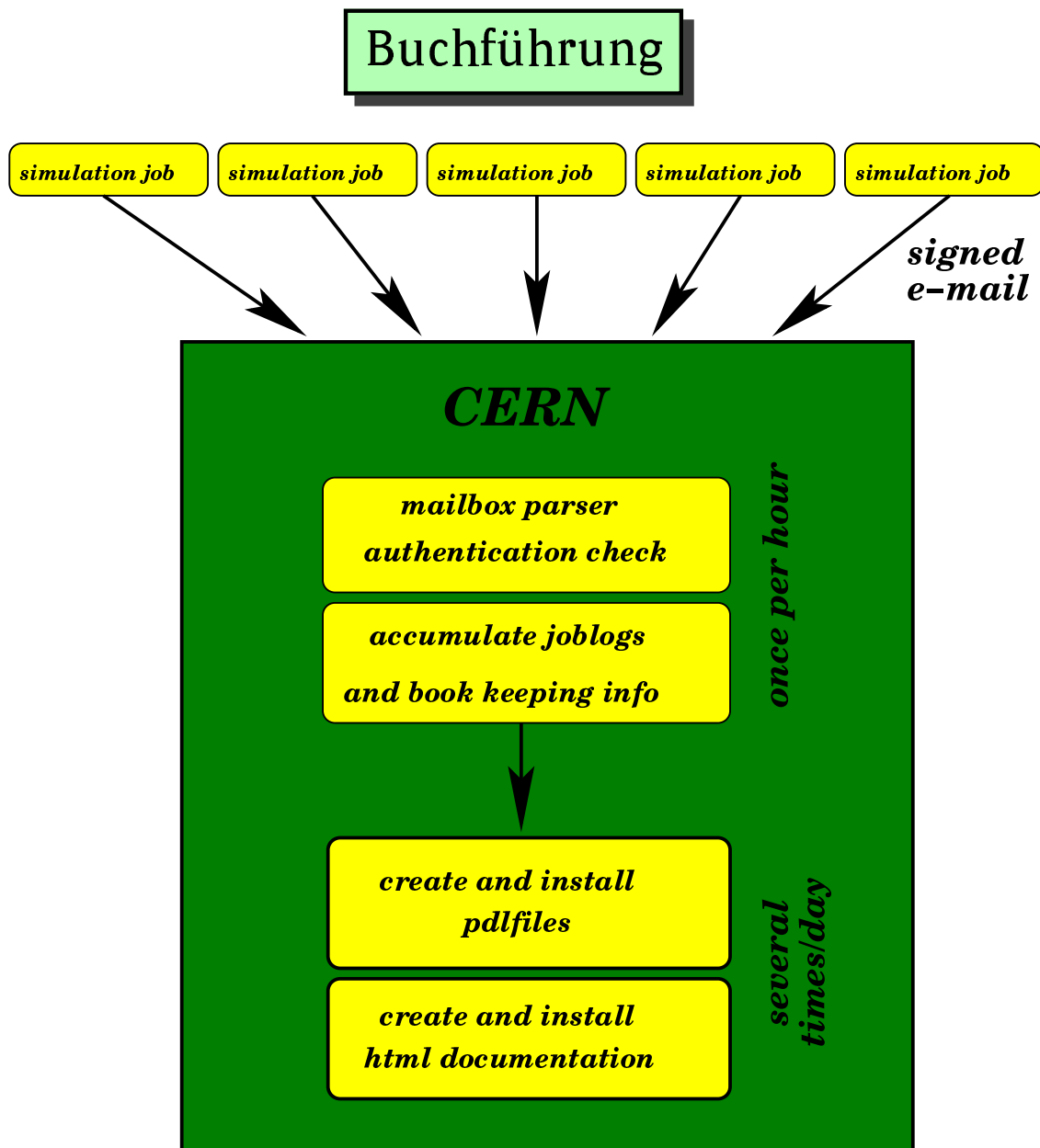
- Datentyp: extended short DST
- Generator HZHA, Version 3, interface Pythia 6.156 Kanal hA mit $\tan \beta = 2$
- $\sqrt{s} = 206.7 \text{ GeV}$
- Masse $m_A = 90 \text{ GeV}/c^2$
- Labor CERN
- Jahr A0 = 2000
- Liquid RICH setup (wichtig für die LEP1 Jahre 93 und 92)
- Prozessing D, erstes DSTANA Prozessing
- ein File angefordert



Automatisierte Buchführung und Dokumentation

- automatische Analyse der Logfiles
- Jobkurzinformation via e-mail
- cron Job sammelt Kurzinfos (1x pro Stunde)
- cron Job erzeugt Nicknamen und Webseite (7x am Tag)
- Einteilung der Nicknamen nach
 - Jahr
 - Prozessing
 - Kategorie
 - Datenformat



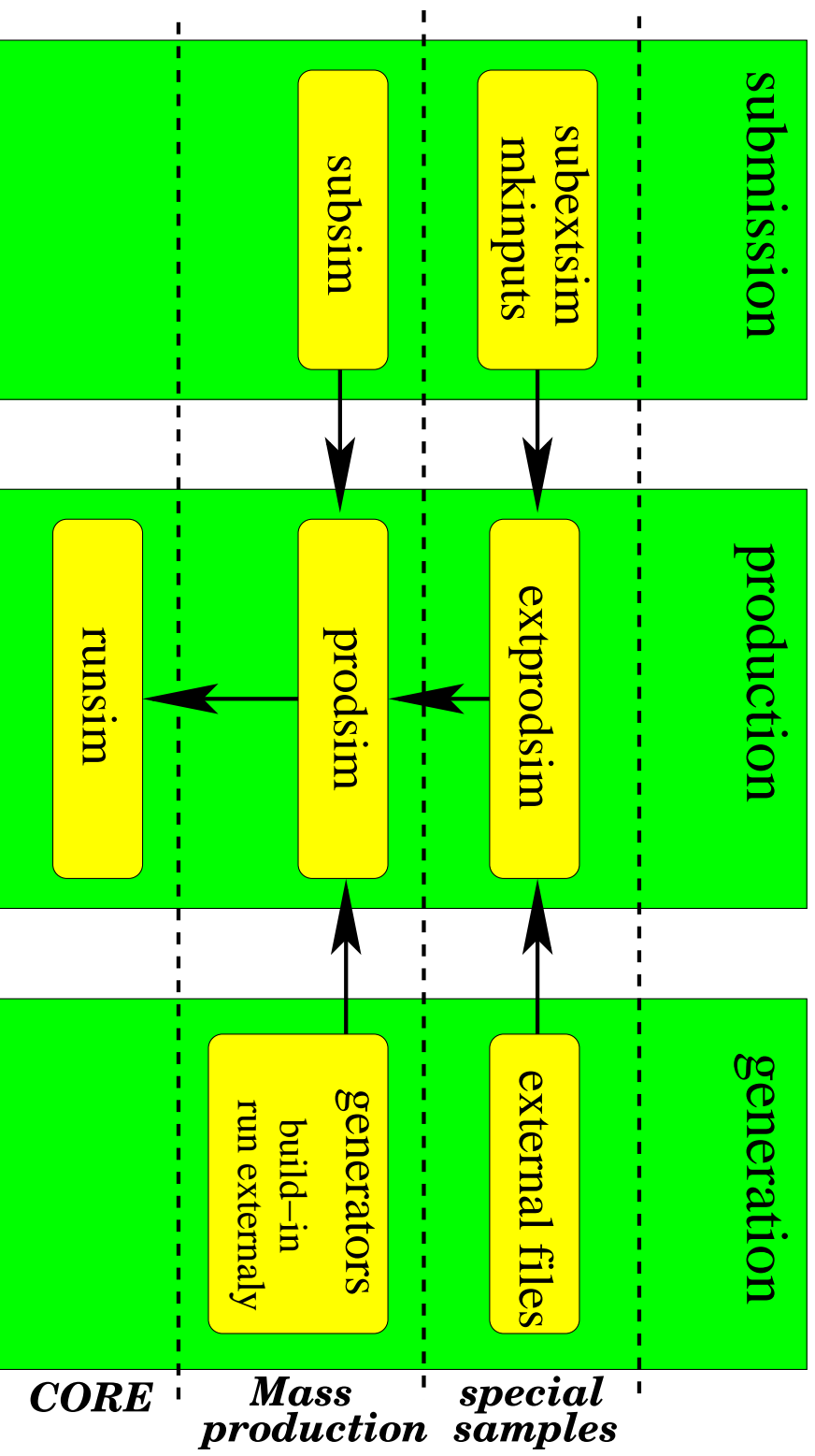


Handarbeit

- Starten der Jobs
- Identifizierung fehlender Jobs
- Kontrolle lokaler Sicherheitskopien
- Pflege der Konfigurationsfiles:
 - Zusatzinformationen wie Wirkungsquerschnitte
 - Kommentare zu speziellen Produktionen
 - Konfiguration bei externen Files
- Archivierung von Logfiles
- Bugfixings etc ...



Zusammenspiel beteiligter Softwarepakete

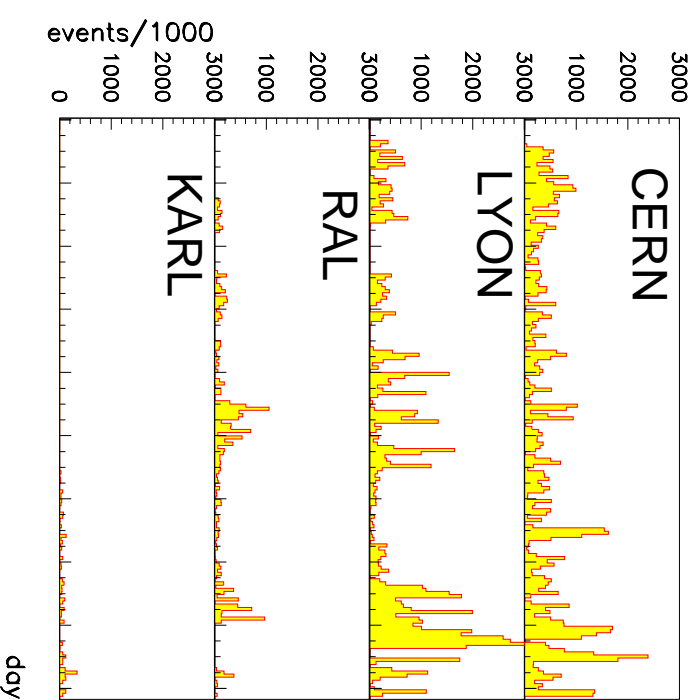
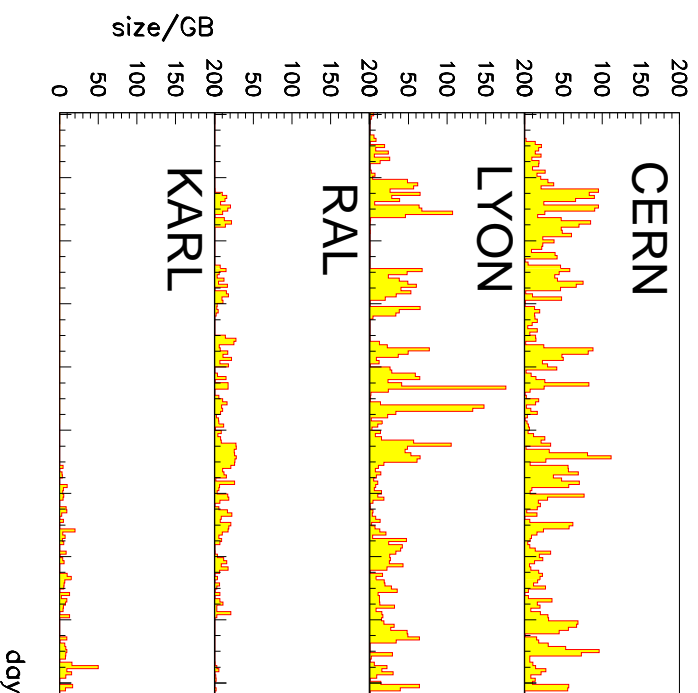


Status (20.1.2002)

- Simulationen für Jahre 1997 bis 2000,
 $\sqrt{s} = 183$ bis 208 GeV,
- vier Zentren
- etwa 170 Mio simulierte Ereignisse
- etwa 16000 verschiedene Nicknames
- etwa 900,000 Ereignisse/Tag
- etwa 5-fache Statistik



Als Funktion der Zeit ...



Zusammenfassung

- Während der LEP2 Phase hat DELPHI etwa 688 pb^{-1} oberhalb der WW Schwelle gesammelt und analysiert.
- Das Higgsboson wurde leider nicht gefunden.
- Neues System zur Produktion großer MC Mengen in kurzer Zeit (900k/Tag)
- Entgültige Ergebnisse im Laufe des Jahres

