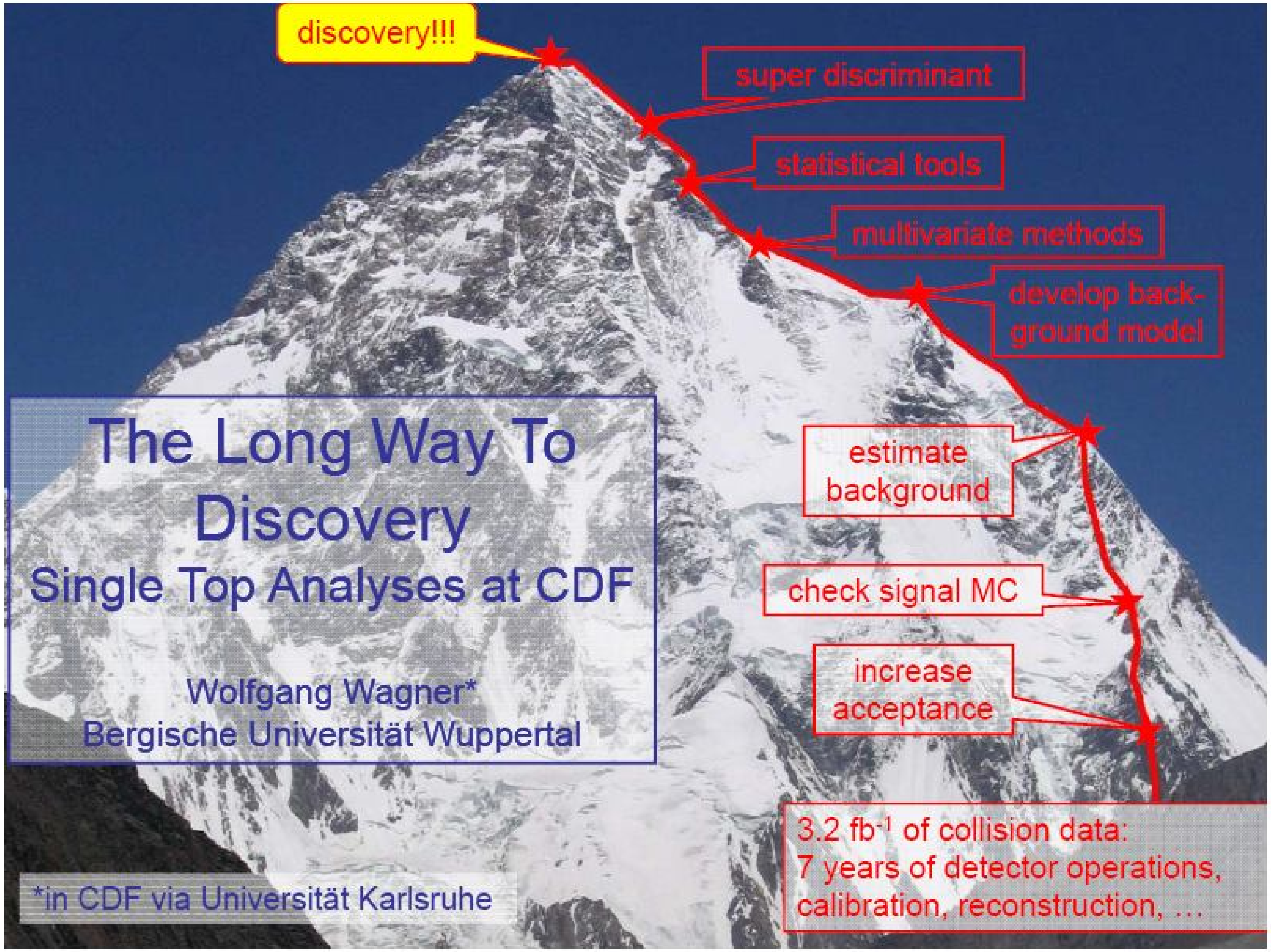
Single-top search: strategies for 1st data at LHC

Workshop on single top physics and fourth generation quarks

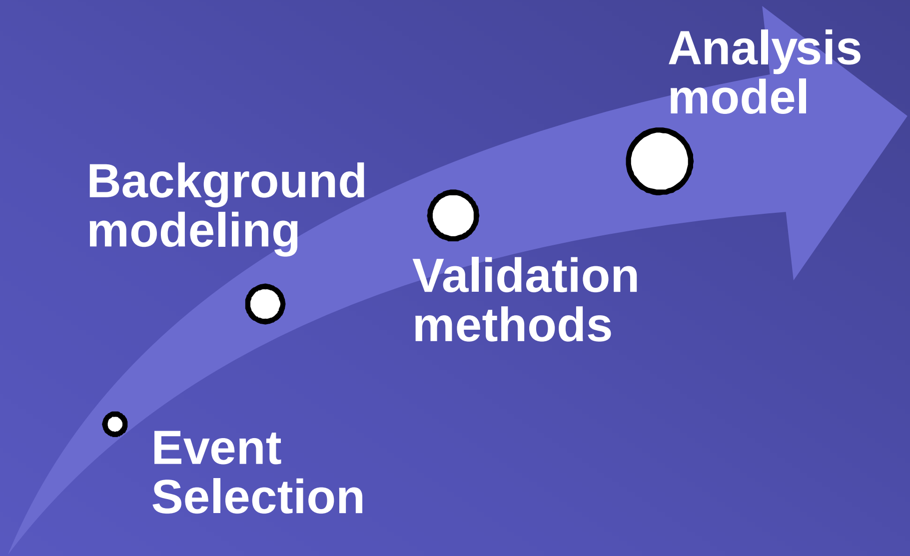
DESY – 15 September 2009

Julien Donini, LPSC Grenoble

Single Top Analyses: From Tevatron …



… to LHC

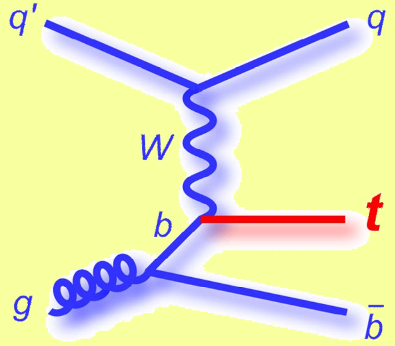


… and search for new physics !

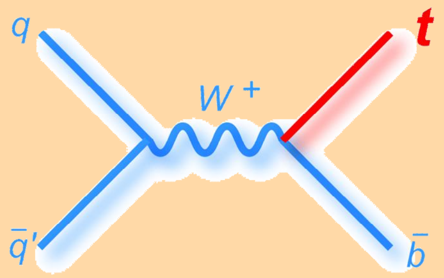
Single Top Quark Production at LHC

- **Single top production**
  - Weak production mechanism: Wtb vertex

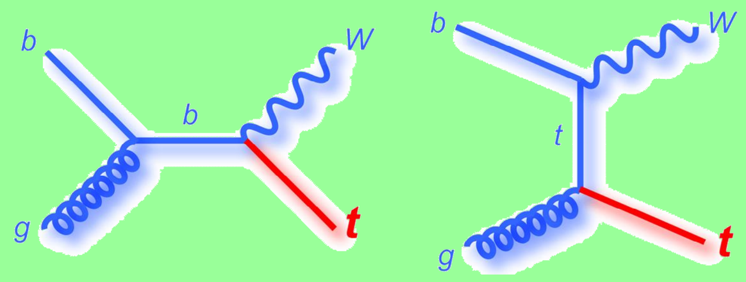
t-channel



s-channel



Wt-channel



ECM	t-ch	s-ch	Wt-ch
14 TeV	246 pb	11 pb	66 pb
10 TeV	124 pb	7 pb	32 pb
7 TeV	59 pb	4 pb	13 pb

LHC (14 TeV) vs Tevatron

Signal: σ_t and σ_{Wt} >×100, σ_s only ×10

BG: σ_{W+jets} ×10, σ_{ttbar} ×100

Single Top Analyses With Early Data

- **Strategy first data**
  - Depend much on ECM and integrated luminosity

10 TeV collisions and ~200 pb

• **t-channel**: few 100 selected events in l+jets channel, S/B~0.5 and a signal significance (including systs.) of 2-3σ can be achieved

• **Wt-channel**: excess in ll and l+jets channels

• **s-channel**: out of reach

7 TeV collisions

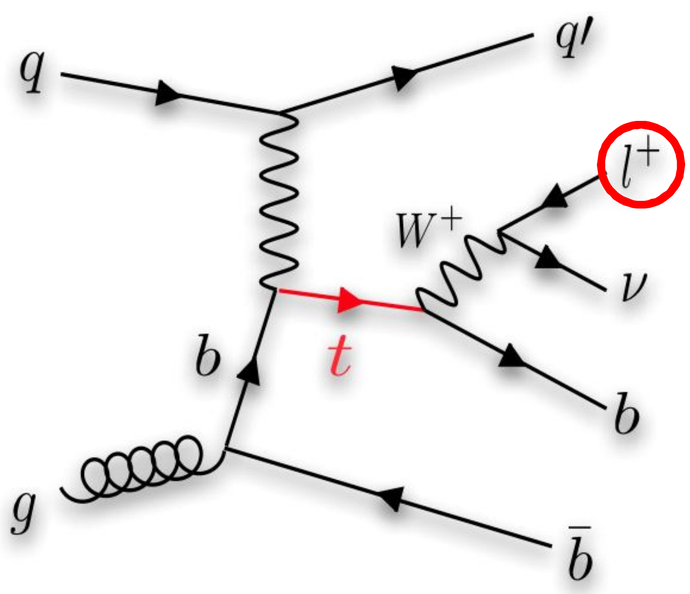
• obtaining similar results should require about 4 times more luminosity

- **Points of view**
  - ➔pessimistic: low stat, no signal excess, calibration data …
  - ➔(very?) optimistic: evidence of t- and Wt- channels, maybe bad agreement with SM (hints for 4th quark generation or new bosons)

Single Top Search With Early Data

- **Search for t-channel signature**
  - $t \rightarrow Wb \rightarrow l\nu b$: lepton+jets selection

Focus first on **cross section measurement** of dominant single top **t-channel** process



Triggers

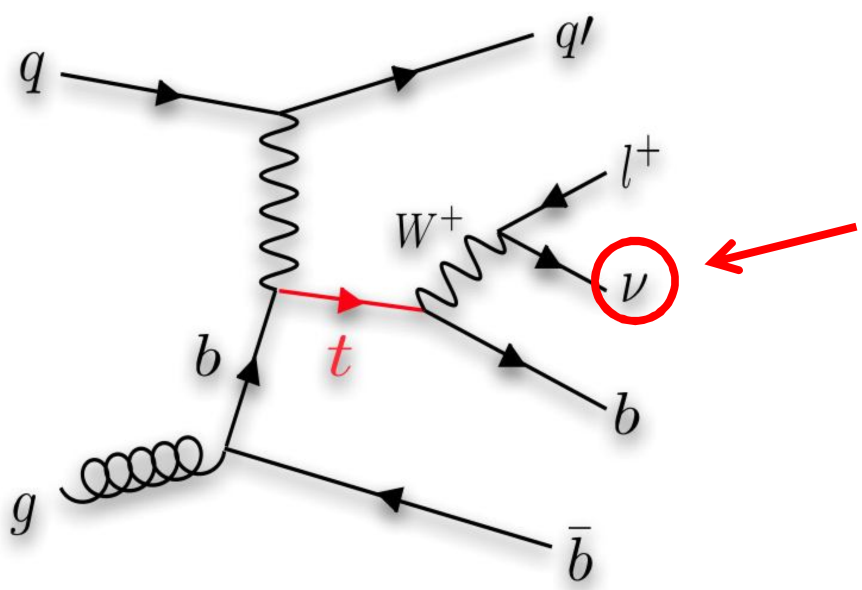
- high p_T inclusive lepton ($e^{\pm}$,$\mu^{\pm}$)
- lepton+jets
- Missing E_T+jets (not for early data)

Lepton ID

- 1 isolated lepton candidate
- loose and tight definitions are useful (estimate of multijet rates)

Single Top Search With Early Data

- **Search for t-channel signature**
  - $t \to Wb \to l\nu b$: lepton+jets selection



Missing E_T

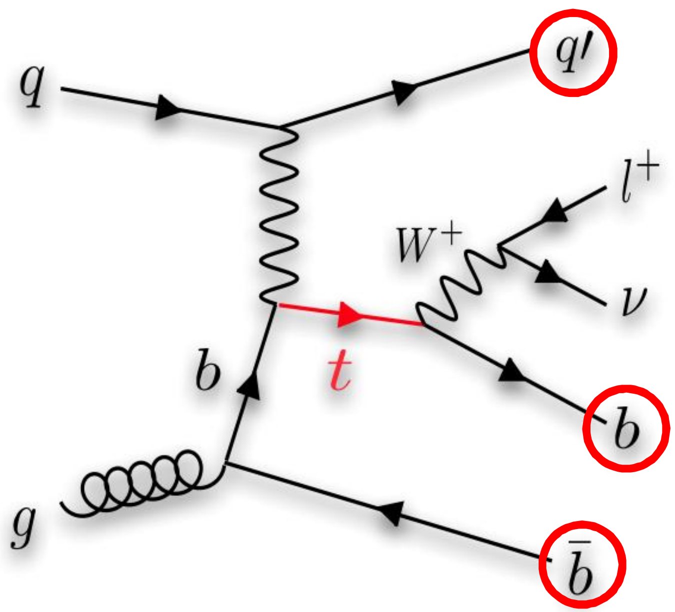
• signal has large missing transverse energy

• cut on MET helps reduce multijet background

• do not trust missing E_T too much during early data commissioning of detectors ….

Single Top Search With Early Data

- **Search for t-channel signature**
  - $t \to Wb \to l\nu b$: lepton+jets selection



Jets

- typically 2-4 jets signature
- best S/B for 2-jets events

b-jets

- 1-2 b-quark jets fall in det. acceptance
- b-tagging is a crucial aspect of single top analyses !
- first data: commisioning of simple b-tagging algorithms
  - → b-tag eff. ~ 40-50% (± 5%)
  - → light jet rej. ~ 100 (10% err.)

Event pre-selection

Trigger
e/mu p_T>10/20 GeV

1 lepton p_T>20 GeV, tight ID criteria

MEt>20 GeV

2 jets p_T>30 GeV

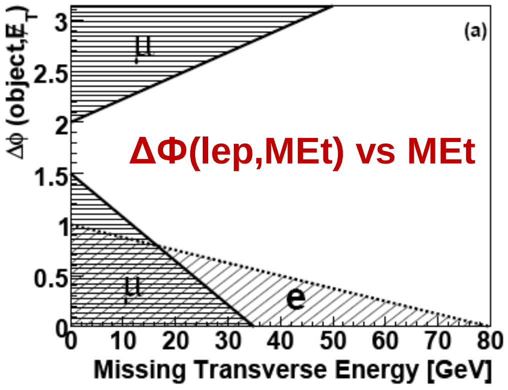
≥1 b-tag p_T>30 GeV

Signal optimization (simple cuts)

Triangular cuts
Reduction QCD

Forward jet (|η|>2.5)
Reduction ttbar

p_T>50 GeV
reduction W+Jets



Expected event yield @10 TeV, 200 pb^{-1}:

S/B ~ 0.3-0.5

~100 signal events

Event Selection

Event pre-selection

Trigger
e/mu p_T>10/20 GeV

1 lepton p_T>20 GeV, tight ID criteria

MEt>20 GeV

2 jets p_T>30 GeV

≥1 b-tag p_T>30 GeV

Backup triggers
b-tag/mistag eff.

Loose ID criteria
QCD & W+jets rates

Efficiencies and BG rates meas.

Event pre-selection

Trigger
e/mu p_T>10/20 GeV

↓

1 lepton p_T>20 GeV, tight ID criteria

↓

MEt>20 GeV → Low MEt region

↓

2 jets p_T>30 GeV → 1-5 jets

↓

≥1 b-tag p_T>30 GeV → Data before tagging, 0-tag

Control and validation samples

Modeling of Signal and Background

Monte Carlo Modeling

Much progresses in recent years

- NLO Monte Carlo generators: MC@NLO, POWHEG
- Prescriptions to match Matrix Element and parton shower jets (MLM matching in ALPGEN)

At LHC: tuning of MC with data will be an important step

- Parton showering and underlying event
- Kinematical distributions of reconstructed objects (lepton, jets …)

➔ Affect single top acceptance

Data-driven models

In particular used for modeling of multijet background

Several useful techniques come from Tevatron

Modeling of Signal and Background

Single top processes

Progresses in simulating and modeling ST processes

• ex: defined procedure for Wt-channel simulation

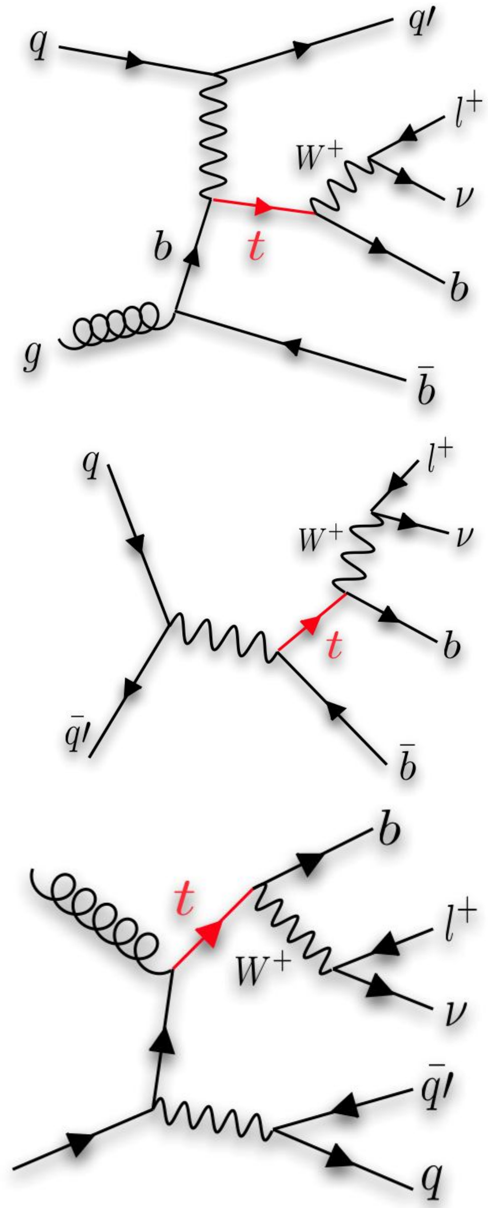
Generators:

• Tevatron: MADEVENT, SINGLETOP

• LHC: now also AcerMC, MC@NLO

In principle single top analyses should benefit from more accurate **NLO** simulation.

Negative event weights can be a problem (for ex in multivariate methods)

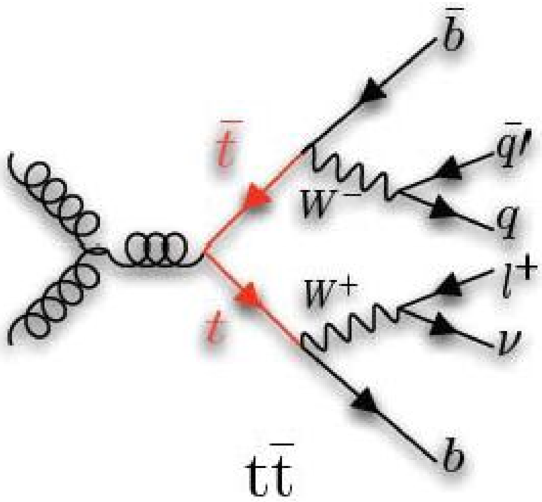


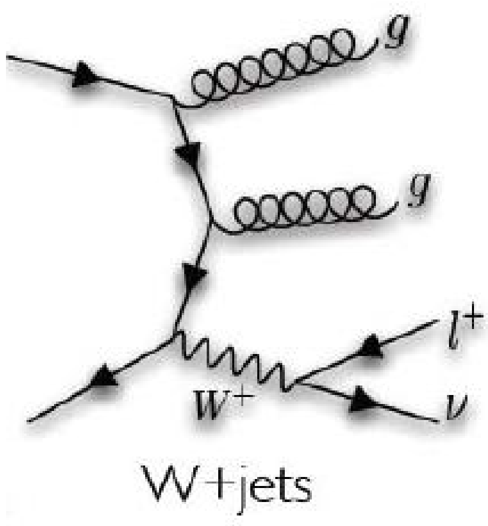
Modeling of Signal and Background

Top pairs

Generators: MC@NLO (ATLAS), MADGRAPH (CMS), …

Normalizing rate using data → better precision than theoretical uncertainties





W+jets

Generator: ALPGEN

- Separate Wjj, Wcc, Wcj and Wbb processes
- Wbb and Wcc fraction are underestimated in ALPGEN. Correction factor derived from data (ex: untagged samples)
- Normalize total rate using data

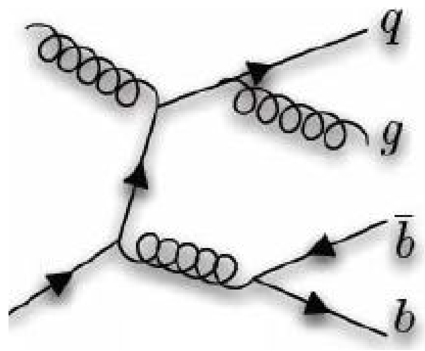
Modeling of Signal and Background

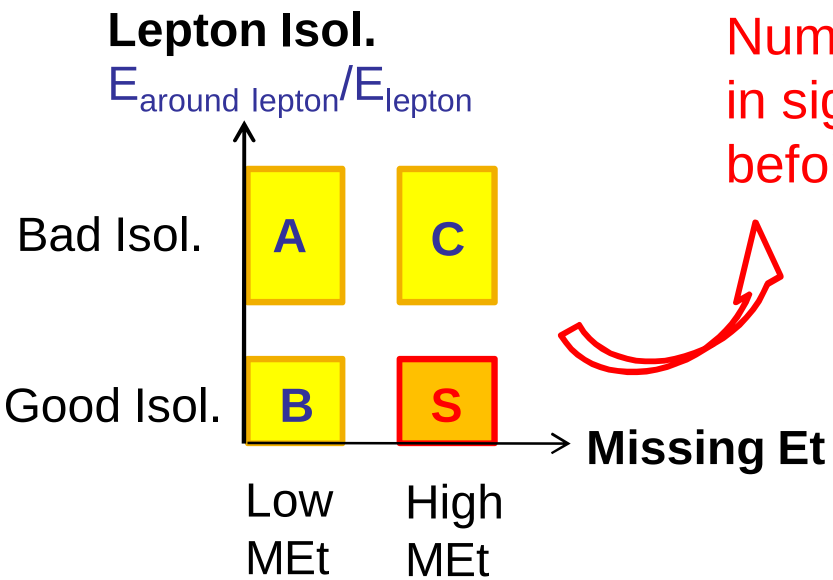
Multijet

Model and normalize QCD background using data (before b-tagging) in sideband regions.

- define region in two (assumed) uncorrelated variables

  → ex: lepton isolation and Missing Et





Number of QCD events in signal region (S) before tagging

$$\boxed{N_{QCD} = \frac{N_B \times N_C}{N_A}}$$

- correction from contribution of W and ttbar events (estimated from MC)
- N_{QCD} for ≥1 tag data is estimated by appling QCD tagging rate (measured in multijet data)

Matrix Method

Normalize W+jets and multijet rates using data before b-tagging

Two data samples:

- **tight** sample: all event selections except b-tagging
- **loose** sample: same but leptons selected with loose identification criteria

Solve following equations:

$$N_{\text{loose}} = \boxed{N_{\text{loose}}^{\text{fake}-\ell}} + \boxed{N_{\text{loose}}^{\text{real}-\ell}}$$

W-like events (W+jets, ttbar)

QCD events in loose sample

$$N_{\text{tight}} = N_{\text{tight}}^{\text{fake}-\ell} + N_{\text{tight}}^{\text{real}-\ell}$$
$$= \boxed{\varepsilon_{\text{fake}-\ell}} N_{\text{loose}}^{\text{fake}-\ell} + \boxed{\varepsilon_{\text{real}-\ell}} N_{\text{loose}}^{\text{real}-\ell}$$

Prob. for real isolated lepton to pass tight selection: measured in Z→ll data

Prob. for fake isolated lepton to pass tight selection: measured in multijet events

Background Normalization

Matrix Method

Normalize W+jets and multijet rates using data before b-tagging

Two data samples:

- **tight** sample: all event selections except b-tagging
- **loose** sample: same but leptons selected with loose identification criteria

Determine from equation solutions:

$N_{\rm tight}^{\rm fake-\ell}$ $\longrightarrow$ **Fraction of multijet background**

$N_{\rm tight}^{\rm real-\ell}$ $\longrightarrow$ **Normalization factor for W+jets background (after subtraction of exp. ttbar contribution)**

Then the rate of these background in the ≥1 b-tag signal sample is determined by applying (data-driven) tag rate functions

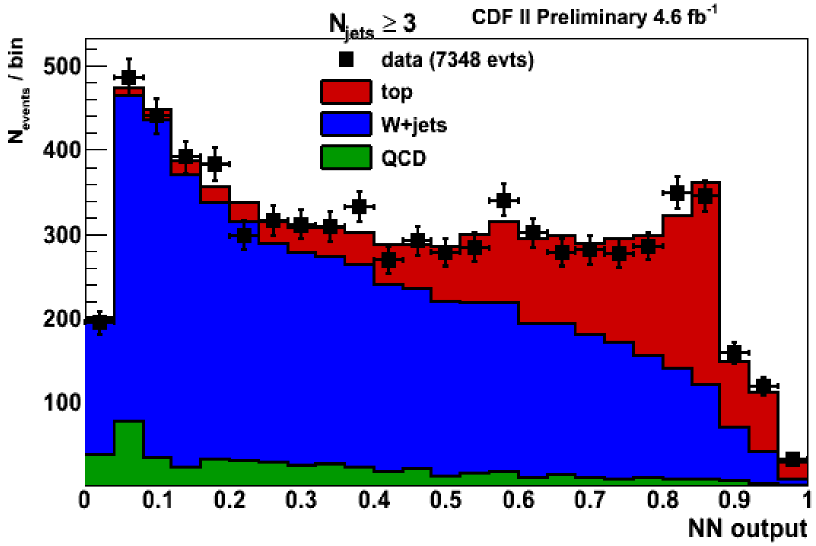
Background Normalization

Fitting Discriminant Shape

Construct (multivariate) discriminant from data before tagging

• Likelihood fit allows measurement of W+jets and ttbar background (QCD contribution is fixed and estimated from other data samples)

• then fraction of ttbar, mistag and W+HF is extrapolated in tagged data



Example: binned likelihood fit to NN output (CDF)

At LHC preliminary studies using NN (see D. Hirschbuehl talk):

→ a precision better than theoretical uncertainties should be achievable for ttbar and W+jets rates with 200 pb^{-1}

Correction Factors

Monte Carlo corrections

Correct simulated samples so that reconstruction and selection efficiencies match those found in data.

Correction factor that will need to be estimated with LHC first data are:

- **Trigger** efficiency: measured from different triggered datasets
- **Lepton ID** efficiency: Z→ll data (tag and probe method)
- **Jet energy scale and resolution**: Z/γ+jets data
- **b-tagging** efficiencies: heavy flavor multijets data, but also ttbar events
- **mistagging** efficiencies: multijet data (inclusive sample of jet triggers)

Contacts with dedicated performance and reconstruction groups (lepton ID, b-tagging …) are a very important part, in particular for early data !

Background Validation

Data/MC validation

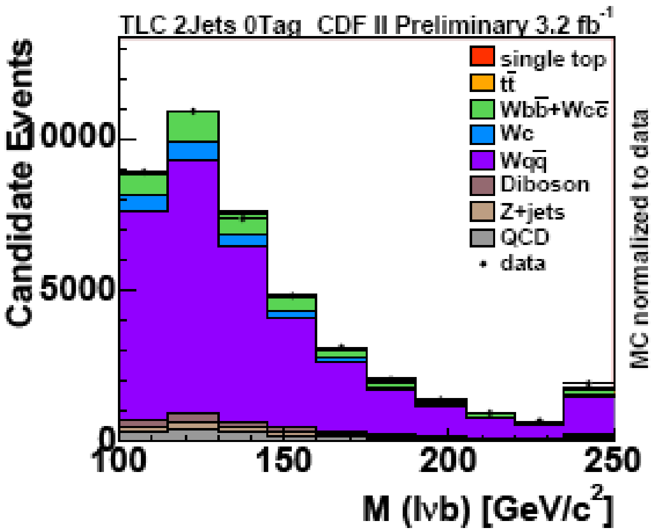
Check that simulated (and data-driven) backgrounds fit data in control samples

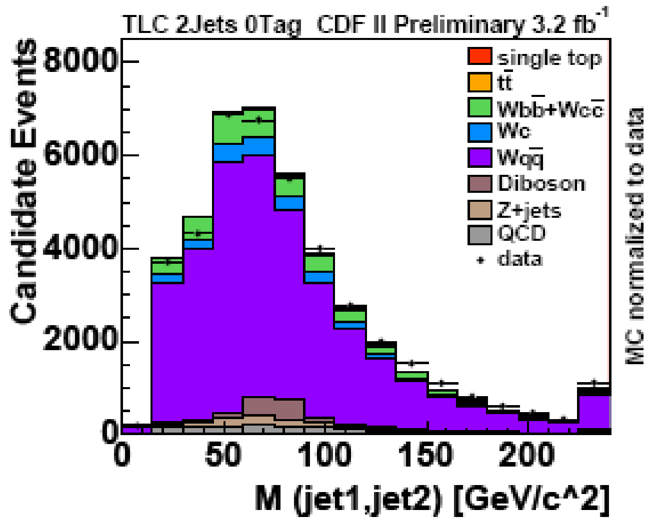
• Control region orthogonal to signal region: sample before tagging (W+3jets), 0-tag sample (W+2jet) …

• Shape comparison of kinetic variables, angular distributions, correlation between objects, etc

Example from CDF:

• **Validation with W+2jets 0-tag data**





Event Selection Optimization

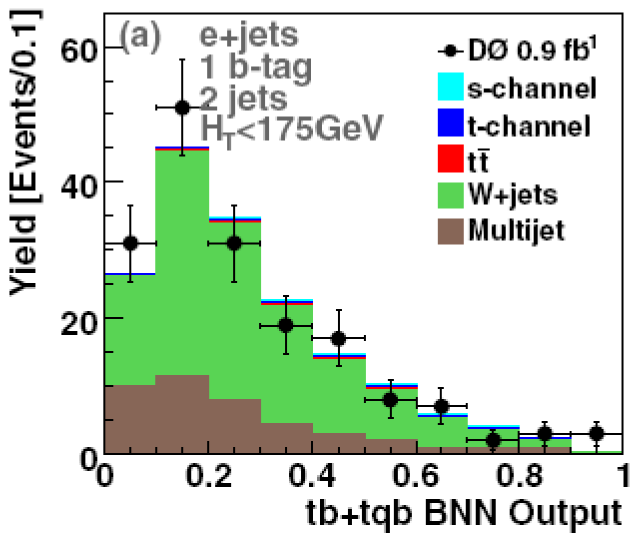
Multivariate Analyses

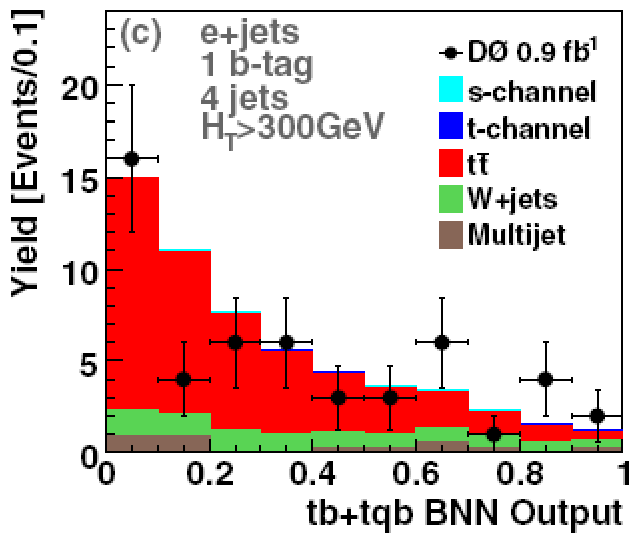
MVA: Boosted Decision Trees, Neural Networks … → increase sizably sample purity by exploiting many input variables and their correlations

- Complex methods: several internal parameters (number of trees, nodes, etc) that need to be carefully tuned → avoid overtraining
- Validation of input variables with data
- Cross check of multivariate discriminants in well defined control regions

Example from D0:

- **2j, low H_T: W+jets dominated**
- **4j, high H_T: ttbar dominated**





Systematic Uncertainties

Evaluation of systematics

In general three types of errors:

• **acceptance**: trigger, lepton ID, JES, b-tagging, Monte Carlo …

• **background**: W+jets modeling (Q^2 scale), multijet normalization (matrix method) …

• **luminosity**

Some systematics can be evaluated from MC, but other need data.

First single top x-section measurements dominated by systematics

→ b-tagging and mistagging: depend on algorithm, tuning with data

→ JES: can be large, in particular for low p_T jets

→ luminosity: effect is reduced if main backgrounds are data-driven

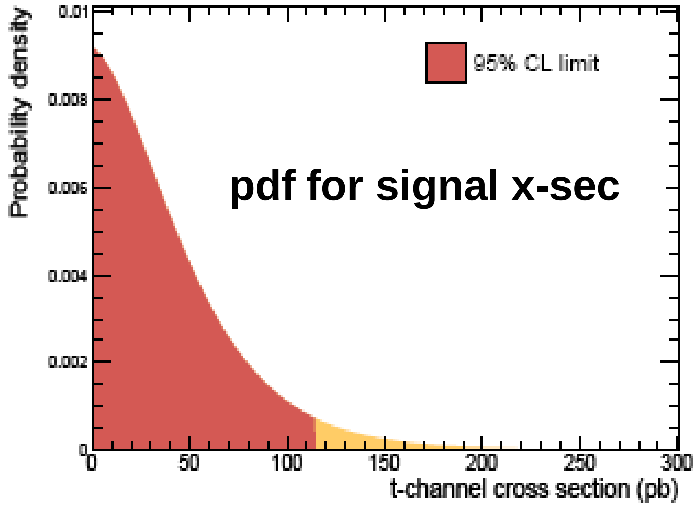
Cross Section Measurement Method

Statistical tools

Posterior probability density function of signal cross section:

• Ensemble tests: propagate uncertainties (stat, syst) on measurement

• construct p.d.f with Markov Chain Monte Carlo, Bayesian methods …



Interpretation of result

Signal excess ? ➔ significance for excluding BG only hypothesis, interpret result within SM and BSM scenarios

No excess ? ➔ derive limits … will be luckier next time (with more data !)

Conclusion

Single top analysis are complex

• Need to reconstruct and identify precisely different type of objects: leptons, jets, missing Et, b-tagged jets

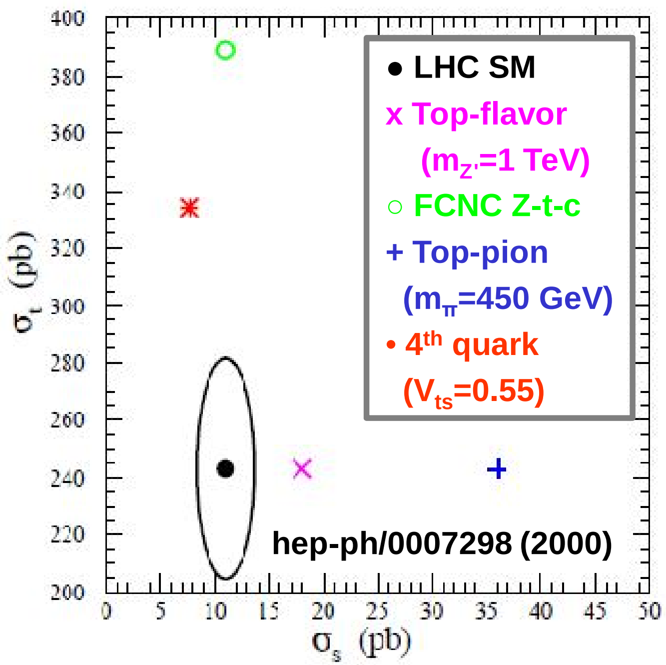
• Lot of effort to improve signal purity, control systematics, etc

➔ At LHC single top analyses will be among the first to use b-tagging and perform multivariate analysis

Results can be very rewarding

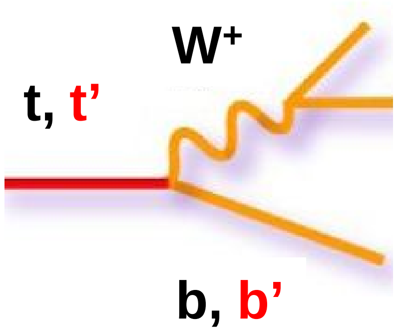
Direct probe for new physics !

Already with 200-1000 pb^{-1} can expect interesting results for t- and Wt-channels

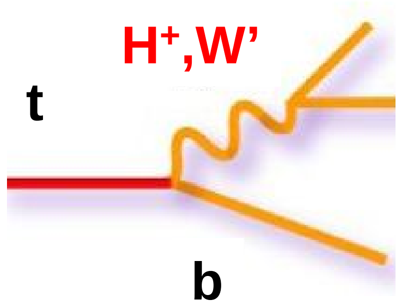


BACKUP

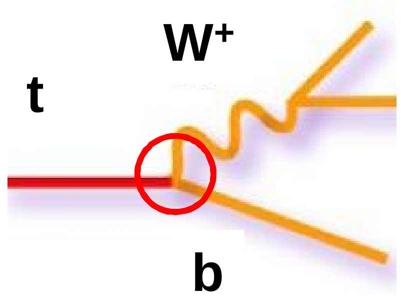
- **Window to new physics**
  - Electroweak interaction
  - → Cross section proportional to $|v_{tb}|^2$
  - → Sensitive to any new particle that can modify the top weak coupling



→ New quarks 4th generation



→ New bosons



→ Modified couplings

→ Anomalous polarisation

Single Top Observation ! (at Tevatron …)

- **Single top 5σ observation: CDF/D0**
  - D0: http://arxiv.org/abs/0903.0850
  - CDF: http://arxiv.org/abs/0903.0885
  - See also results shown at winter conferences (Moriond EWK)
- **Combination of several multivariate techniques**
  - D0: DBT, Bayesian NN, ME → $\sigma_{st} = 3.94 \pm 0.88$
  - CDF: Likelihood, NN, ME, Likelihood opt. s-ch, BDT, MEt+jet

→ $\sigma_{st} = 2.3 \pm 0.6$