

## **SPRG comments on the first draft of CDF 10960, the top mass in events with MET+jets,**

**By Jeff Appel, Diego Tonelli, Peter Renton, Jonathan Rosner and giorgiob, January 23, 2013**

### **General comments**

The paper needs several refinements. The motivation of the analysis is not incisive enough. First of all, the text is not clear enough about the method, referring to semileptonic decays without noting clearly enough that this is despite using events with no detected electron or muon. When the lepton veto is mentioned, it is only in the context of using orthogonal data samples without directly addressing the surprising tack of using semileptonic decays with no detected lepton. There is almost no discussion on where the "missing" lepton goes i.e. why is it missed. This deserves a paragraph in the paper. More details should be devoted to the discussion on how a fit procedure that includes an a posteriori MC-based 10% correction (P6L16) can yield a result that is better than 1% precise. Do you really know your MC-based correction so well?

A number of wording changes are suggested in the line-by-line comments to address this complaint.

The description of the analysis is not very clear. There is no real introduction/discussion on the combinatorial problem of assigning jets. The case for the indirect constraints on the Higgs mass is made, but looks a little marginal after the direct discovery. One sentence or two on the experimental status of the top-quark mass is in order between the motivation and the description of the analysis. In general, the paper is too reliant on references to previous papers (in particular the previous version of this analysis). There are several pointers to later discussions (page 4, L3, L10, etc...) that make the paper confuse to understand. Suggest making an effort at reshuffling the structure in such a way that these pointers to the future are minimized, if not avoided.

Please make this paper readable as a stand-alone paper. Some specific suggestion on this issue is added line-by-line below.

The use of the top-quark width (1.5 GeV) in the  $\chi^2$  function seems strange as the resolution on these quantities is surely much larger (see also a comment line-by-line). It is not clearly stated what topology is being searched for and which selection criteria achieve this goal. In this respect there is too much emphasis on this analysis being independent of others.

One statement (page 3, line 25) is inaccurate and should be corrected.

Remove vertical lines by the Tables (AIP style). Please delete all instances of "our", which are colloquial and unnecessary. There is a general lack of hyphens for compound adjectives, e.g., in W-boson mass and Higgs-boson mass (see a number of specific suggestions line-by-line). Please check the APS (and SPRG) guidelines eg add all authors in references up to 10 authors (a number of specific suggestions are made line-by-line below).

In the conclusions, please add how this measurement improves the CDF average on  $M_t$  so as to justify the method.

### **Line-by-line**

## TITLE.

"Data set" (two words, and not *date*)

## Page 1, Abstract

This needs largely to be rewritten. Firstly the physics channel being studied should be given, then the method used to select it should be outlined.

**Line 2.** "...with large missing transverse energy..."

**Line 3.** Suggest "...corresponding to a Tevatron integrated luminosity of  $8.7 \text{ fb}^{-1}$  at a center of mass energy  $\sqrt{s} = 1.96 \text{ TeV}$ ."

**Lines 4, 5, 6.** Suggested rewording: "Events used in this analysis come from the channel with one semileptonic  $t$  or  $t\bar{b}$  decay, but without detection of the electron or muon. In this way, the data are not explored by other CDF top-quark mass measurements.". The statement "The method uses two fully-reconstructed top-quark mass values ..." is very confusing. The reader might well wonder why? If this is to be included please explain better.

**Line 6.** The statement about the  $W$  boson  $\rightarrow jj$  decays is unclear. Please say first why you do this then how you do this. "...two jets from one  $W$ -boson decay to compare..." Note the hyphen in " $W$ -boson".

## Page 2.

**Line 2.** "weak-isospin" (add hyphen). Also "...approximately 40 times larger..." [the mass is not heavy, it's the particle that's heavy]

**Line 5.** " $W$ -boson mass".

**Line 6.** Suggest "hence, the combination..."

**Line 7.** Delete "the" before  $M_W$ . Also " $Higgs$ -boson mass". Is this argument of the "important constraint" to the Higgs boson mass still so relevant after the LHC discovery? Consider rewording or repurposing (some room for SUSY?) this motivation.

**Lines 9,10.** One sentence is unnecessary. Suggest "Since the top quark decays...and a  $b$  quark[3], we shall be looking for  $t\bar{b} \rightarrow \dots$  events."

**Line 11.** "...or a charged lepton and its neutrino.". Suggest new paragraph starting with "In this paper". The material after that is rather different from that in the preceding text.

**Line 12** "ref [4]", **and line 25** "...on line selection requirements..."

There is no detector description although there appears to be room for a brief overview of the relevant parts of the detector, in particular the trigger and calorimeter. Thus the mention of calorimeter towers in reference [4] and mention of the trigger late in the paper comes without preparation. Suggest adding text including "...the pointing-tower-geometry electromagnetic and hadronic calorimeters..." and "...three-level online event-selection system (trigger)...", with appropriate references cited.

**Line 13.** "Drop "intend to" to have".  $\rightarrow$ "We select top-quark candidates...". We assume you mean that this analysis aims to select events in the channel where one  $W$  decays to a lepton + neutrino and the other to 2 jets. However this is not clear. Please expand to make this clear and discuss where the lepton goes and how it is not detected. Note that "top quark that comes from" a decay mode is poor wording. Consider "We select top-quarks in their decays into "lepton+jets" final states..." And afterwards you should define and explain what the "lepton + jets" decay mode is.

**Line 14.** Hyphenate compound adjective "high-transverse-momentum". Suggest replacing "We intend to ... events" → "We select top-quark candidates"

**Lines 14-17.** Please define what a multijet event is. Also, consider "...events that come from the "lepton + jets" decay mode even though we veto events with identified high transverse momentum ( $p_t$ ) electrons or muons (*Why was that sample studied first? give a (Ref). Also relevant to page 3, lines 7, 8).* We also veto multijet events where both W bosons decay hadronically. This ensures a result which is statistically independent of the other CDF top-quark mass measurements [5-8] and allows for easy combination of results."

**Line 15.** "...this ensures THAT our result..."

**Lines 18-19.** "event-selection criteria" (– hyphen for compound adjective).

**Line 20, 21.** Suggest remove "of reconstructed ....mass" as this is implicit

**Line 22.** "scaling for the amount of data used" However, it is unclear what this sentence means. Is the 18% in addition to the effect of increasing the sample size or not? If yes, just say "...of about 18% in statistical precision over the improvement expected from increasing the sample size alone", or similar.

**Lines 24, 25.** The phrase "The sample ... online selection requirements" is unclear. Please be more specific.

**Line 30.** "parametrized" [AIP suggested spelling]

### Page 3.

**Line 1.** "b-quark hadronization" (– hyphen for compound adjective).

**Line 2.** Suggest "...through the presence of...", or "through measuring". Note that you do not "measure" the presence. You may consider "detecting the presence" or "inferring the presence".

**Line 3.** Please organize the event selection better. Start with defining the channel you are searching for and how the W bosons decay in this decay chain. Please then say how the criteria you use achieve selecting the channel you want. In particular how the criteria select the W → lν decay mode.

Suggest (b-tagging was already defined): "We require at least one jet to be b-tagged. In order to improve the statistical power of the analysis, we divide the candidate events into two samples, one of which contains a single b-tagged jet (1-tag) and the other contains two or more b-tagged jets (2-tag)."

**Line 6.** "...to have four to six jets..."

**Line 7.** "...measurements [give reference,...]"

**Lines 8-12.** Suggest "...and  $|\eta| < 1.0$ . In order to reject multijet backgrounds from QCD processes, we require the events to have MET significance (defined as ...) to be  $< 3 \text{ GeV}^{1/2}$ , where the sum in the denominator runs over all identified jets in an event, and that  $\Delta\phi(\text{vector MET; vector } p_t) < 2:0$ , where  $\Delta\phi(\text{vector MET; vector } p_t)$  is the azimuthal angle between vector MET and missing transverse momentum vector  $p_t$  [17]." Next, suggest mentioning that this cut is applied in order to make this sample orthogonal to the top mass measurement in the all-hadronic sample.

A number of changes are incorporated in this suggestion.

Note that both items in the sentence seem to relate to the QCD background and should be treated in parallel, and independent of what came before. Also, note the use of vector symbols for MET and  $p_t$  here, and that  $\Delta\phi$  is not a distance, but an angle (see the comment on page P10, lines 7-11 about defining vector and scalar  $p_t$ ,  $E_t$ , and MET).

**Line 9.** Add comma before "where". There does not appear to be a selection made on MET itself. Is this correct? If so how do you "select" the neutrino?

Perhaps you can modify the statement. Even without assuming unitarity, it isn't now experimentally established that the top width is almost completely saturated by  $W \rightarrow b\nu$ .

**Line 12.** Suggest "THE selection on.....replaces..."

Referring to [17] is not an easy way to tell the reader what  $P_{t,miss}$  is. Suggest adding a short footnote.

**Line 14.** Delete "good".

**Lines 15-16.** Actually, b-tagged events come also from mistags. In this sentence you seem to allude to genuine b-tags only. Suggest rephrasing.

**Lines 18-19.** Suggest "...using input variables proposed in Refs [12, 18]. Please briefly summarize the input variables globally, to reduce the need to check references [12] and [18].

**Line 19.** Drop "of". Also "...we added new input variables such as..."

**Line 20.** "...as detailed in Ref [18]."

**Line 21.** "aiming to" sounds better than "aiming for"

**Line 24.** "Following Ref. [9], we estimate the background..."

**Line 25.** "presence" -> "component", and "negligible t-tbar contamination". However, the t-tbar signal in events with exactly 3 jets may not be negligible. You must justify this statement and possibly replace "negligible" with an upper limit.

**Line 26.** "parametrization" is favored US spelling

**Line 29.** Delete "as done in" and just keep the reference.

#### Page 4.

**Table 1, caption.** First line: "Number of expected signal and background events in the signal region COMPARED TO the number of events observed in data..." In line 2, remove "corresponding to ... signal region". Third line: what does "simulation of online selections" mean?. This is not discussed in the text. Please explain why online selections are relevant. In lines 5, 6, remove "The t-tbar ....7.45 pb" (both these are given in the text). Last line: "cross-section" --> "cross section".

**Line 3.** Suggest starting with "Signal events are...", and "...number of jets for reasons explained..."

**Line 4.** Suggest dropping "in the process of forming estimators". What do "x" and "y" mean? Are these Cartesian coordinates? Please explain (referring to [23] is not good enough).

**Line 6.** "...them, we reconstruct..." (comma after two introductory prepositional phrases).

**Lines 9, 10.** Suggest ending the sentence at "are used." By varying  $\Delta(\text{JES})$  you do not *reduce the systematic effects*.

**Line 11.** "...we reconstruct three quantities using..."

**Line 13.** "hadronically-decaying"

**Line 15.** "...whose invariant mass value is closest...". Suggest "...from the two non-b-tagged jets". What about events with only one b-tag? Please add a general discussion first about the combinatorial problem of associating jets to the underlying parton hypotheses.

**Line 17.** "...using a fitter that...", and "...used in Ref. [23]."

**Line 18.** Suggest "...the kinematic properties of each event" ("kinematics" is jargon)

#### Page 5.

Are natural widths the right thing to use for the last three terms of Eq. 1? Doesn't detector resolution play a role here as in the first two terms? Has this been studied? It needs to be discussed briefly. Consider using the expected uncertainties in the third line rather than the top-quark width. The resolution on these quantities is surely larger than 1.5 GeV.

**Line 3.** Delete "chosen from the event", unnecessary and does not facilitate reading.

**Line 4.** Use of vector  $E_t$  allows the text to speak of x and y components. "unclustered energy" is not a vector. Please rephrase.

**Line 6.** "W-boson decay-width"

**Line 10.** Add reference for top-quark width prediction. Also "...imposes the same constraint..."

**Line 11.** "...regarded as W-boson decay products..."

**Lines 12-13.** Please do not start a sentence with a symbol; Suggest, e.g. "The free parameter to be fitted in the  $\chi^2$  function is  $m_t^{\text{reco}}$ ."

**Line 14.** Delete "of jets", which is redundant

**Line 16.** "2nd lowest" [no superscript]

**Line 17.** Delete "from the event", which is redundant Please add the number of combinations possible in the various cases.

**Line 20.** Events with 5 or 6 jets are mentioned here, while only events with 4 or 5 jets were used to form templates in the simulation (page 4, line 12). This is confusing.

**Lines 21 and 22.** "top-quark masses".

**Line 24.** "(p.d.f.)" (period after "f")

## Page 6.

**Line 1.** "three-dimensional" (hyphen)

**Line 2.** Suggest "...we maximize a likelihood on unbinned data using the p.d.f's of the signal and background samples to extract  $M_{\text{top}}$ . The likelihood fits..."

**Line 6.** Suggest "...in order to maximize the statistical information..."

**Line 16.** Be consistent with the "meas" superscript. It's roman and italic in the same line.

**Lines 17-18.** The scope of the last sentence of this paragraph is unclear. Please be more specific. Also add "the" for "on the ratio". How good is the linear regression fit?

**Line 23.** "b-jet energy scale" (hyphen for compound adjective)

**Line 24.** "...b-hadron..."

**Line 26.** "light-quark-jet energy scale" (hyphen for compound adjective)

## Page 7.

**Table II.** It is unclear what "calibration" means. Capitalize only the first word of each of the phrases in the first column. Also, it would probably be clearer to left justify the entries in that column to avoid any appearance of one element looking like a subset of the above element. Also, rather than "Systematic sources", suggest "Source", or "Systematic uncertainty source", and "gg fraction" (italicize gg)

**Line 6.** "up and down"? Suggest rewording.

**Line 7.** Need a definition of "trigger" in the suggested detector text additions. Suggest here "...to account for the trigger simulation mismodeling..."

**Line 12.** "...different configurations of pythia. (sc)

**Lines 12, 13.** "...summarizes all systematic uncertainties that total 0.9 GeV."

## Page 8.

**Fig. 1.** Suggest using the same mass scale in the two  $m_{jj}$  plots. Caption: 4→four.

**Line 4.** Please transform numerals into words, and add some conclusions on what you learn from these figs 1 and 2.

**Line 7.** "...to be statistically independent from samples used in others..."

**Line 10.** Drop "and"

**Line 11.** Suggest "contribute significantly". Also "values" (plural, since there are two), and add serial comma before "and".

## Page 9.

**Caption** of Fig 2. Spell jet-multiplicity numerals

## Page 10, references.

The references have several issues (see <http://gate.hep.anl.gov/abw/reader/check.pdf> when in doubt). From the PRL and PRD "Information for Contributors": "The names of all authors of cited papers should normally be given in the references except when the number of authors is very large (say, more than 10)."

Proper J. High Energy Phys. format in references has not been used. E.g. Ref. [13] should be "J. High Energy Phys. 07 (2002) 012." The first two digits are the month; then (year) in parentheses; then article number. Also applies to Ref. [16].

Several specific suggestions are as follows.

**Line 3.** Include all authors: H. Flacher, M. Goebel, J. Haller, A. Hocker, K. Monig, and J. Stelzer.

**Lines 7-11.** Start with vector  $p_t$ ,  $E_t$ , and MET, and then define the moduli. That will allow using the vectors as needed on page 3, line 13. Note, need to describe towers somehow here if not included in an added detector description as suggested on page 2, line 11.

**Line 26.** Add "(CDF Collaboration)" after "et al."

## Page 11.

**Line 1.** Fewer than ten authors. Include all: T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, and E. Norrbin

**Line 2.** Drop "Phys. Res., Sec." to make "Nucl. Instrum. Methods A 566, 375 (2006)." As in ref. [15].

**Line 11.** Fewer than ten authors. Include all: G. Corcella, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, and B.R. Webber. Note that J. High Energy Phys. format does not use bold face for first two digits. Please correct Refs. [28,30,33]. Example: "[28] ... 01 (2001) 010."

**Line 13.** Fewer than ten authors, include all: J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P.M. Nadolsky, and W.K. Tung.

**Line 15.** Fewer than ten authors, include all: A.D. Martin, R.G. Roberts, W.J. Stirling, and R.S. Thorne

**Line 16.** Fewer than ten authors – include all: M. Cacciari, S. Frixione, M.L. Mangano, P. Nason, and G. Ridolfi

