Optimized top quark analysis with the decision tree

P. Agrawal,* D. Bowser-Chao,† and J. Pumplin‡
Department of Physics & Astronomy, Michigan State University, East Lansing, Michigan 48824
(Received 10 April 1995; revised manuscript received 7 August 1995)

We present an optimized and physically motivated method for separating top quark signal events from background events at the Fermilab Tevatron. For the top quark signal $t\bar{t} \to e/\mu + 4$ jets, we show how to reject all but 25% of the background in a data sample while retaining 83% of the signal, without introducing bias into the subsequent mass measurement. The technique used is the binary decision tree. Combining this highly efficient procedure for signal identification with a novel algorithm for top quark reconstruction, we propose a powerful new way to measure the top quark mass.

PACS number(s): 14.65.Ha, 12.38.Bx, 13.85.Hd

The Collider Detector at Fermilab (CDF) and D0 Collaborations recently announced the much-awaited discovery of the top quark [1,2]. Both collaborations will next endeavor to study its production and decay properties further, and to improve the measurement of its mass. An important aspect of the analysis is the need to reject a good fraction of the numerous background events, while keeping most of the signal.

In this paper, we employ an artificial-intelligence algorithm, the binary decision tree [3], to arrive at optimized and physically motivated cuts that discriminate signal from background with an efficiency well beyond what is possible using conventional methods [4]. The decision tree has the added advantage of simpler interpretation compared to another standard signal-enhancing method, the neural network.[1] By exploiting differences between the signal and background without relying on explicit reconstruction of the top quark signal, these cuts moreover introduce no bias into measurement of the mass. After presenting the optimized cuts, we propose a new top quark mass reconstruction algorithm in which a peak in a selected three-jet mass distribution reveals $t \to jjj$ and provides a direct measurement of m_t along with a model-independent measurement of the background. With the anticipated integrated luminosity of the current experimental run at the Fermilab Tevatron, there will be enough events not only to see the mass peak clearly, but also to observe the subsequent hadronic decay $W \to jj$, furnishing a new, direct calibration of hadronic calorimetry and the jet-finding algorithm.

In the standard model, the top quark decays electroweakly via $t \to W^+b$. The W boson in turn decays hadronically to two jets ($W^+ \to jj$) approximately 2/3 of the time, and semileptonically ($W^+ \to e^+\nu_e, \mu^+\nu_\mu, \tau^+\nu_\tau$) in the remaining 1/3. At the Fermilab Tevatron, top quarks are mainly produced in pairs via $p\bar{p} \to t\bar{t} + X$. Because of severe QCD backgrounds, reliable detection of a top quark pair requires at least one of the two resulting W bosons to decay semileptonically into e or μ. We will focus on the "single leptonic" signature $\ell + 4$ jets where $\ell = e$ or μ. These events occur with six times the rate for double-lepton events, and have the added virtue of containing only one neutrino, which facilitates the mass measurement.

The main background to this mode is from the direct production of $p\bar{p} \to W + 4$ jets, occurring at about two times the signal rate in the standard model for $m_t \sim 175\,\mathrm{GeV}/c^2$. To suppress this background, one can exploit the fact that two of the four jets in the signal are due to b quarks which can be tagged with some probability, while b jets are rare in the background. Because we seek high signal acceptance, we will eschew a b-tagging requirement, but point out below how it can be used, when available, to complement our analysis.

In the absence of b tagging, the weapon of choice for reducing the background is to impose cuts in appropriate observables. Consider, for example, m_{jj}^6, whose distribution is shown in Fig. 1. (m_{jj}^6 is the lowest of the six invariant masses formed from pairs of the four jets.) The signal peaks near $75\,\mathrm{GeV}/c^2$, while the background (dotted curve) is concentrated at low m_{jj}^6. Requiring each event to have a minimum observed m_{jj}^6 can thus increase the signal/background ratio S/B, without appreciable loss of signal.

Our first improvement over previous analyses comes from introducing new variables, including m_{jj}^6, and showing how the physics of the background and signal makes these variables powerful tools for signal enhancement. The major thrust of our work, however, is toward obtaining cuts in *a set of observables* simultaneously. Before describing how the binary decision tree determines these highly efficient cuts, we review the conventional route to signal vs background discrimination.

*Electronic address: agrawal@msupa.pa.msu.edu
†Electronic address: davechao@msupa.pa.msu.edu
‡Electronic address: pumplin@msupa.pa.msu.edu

[1]The application of the neural network to high-energy physics was suggested in Ref. [5], and carried out for the same top quark signature considered here in Ref. [6] and recently in Ref. [7]. The latter also includes a useful comparison with other algorithms including probability density estimation (PDE) and the H-matrix method. A comparison of binary decision trees with neural networks is given in Ref. [8].

 ©1995 The American Physical Society

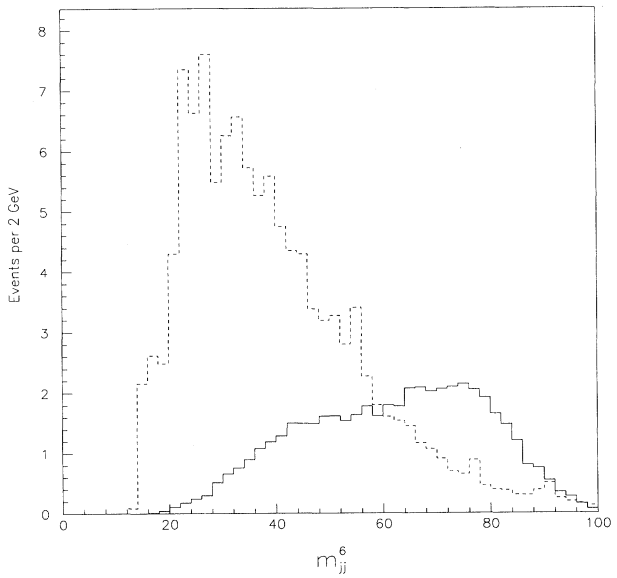

FIG. 1. m_{jj}^6 in 2 GeV/c^2 bins; the dashed curve is the background, the solid curve the signal ($m_t = 175\,\mathrm{GeV}/c^2$); both include only acceptance cuts, at an integrated luminosity of $\int \mathcal{L}dt = 100\ \mathrm{pb}^{-1}$.

Based on comparisons of signal and background distributions like Fig. 1, a list of candidate observables is selected. A simple cut specified by $x_i > x_{i,\min}$ and/or $x_i < x_{i,\max}$ in each variable x_i is arrived at by trial-and-error adjustment, compromising between background rejection and signal acceptance. Each cut is relaxed or tightened in turn to roughly optimize S/B at the desired level of signal acceptance. The virtue of this procedure is that the physical nature of each cut is understandable. For example, the simple cut $m_{jj}^6 > 50\,\mathrm{GeV}/c^2$ enhances signal/background because the jets in background events tend to arise from bremsstrahlung, where the collinear and soft singularities of QCD give rise to low pair masses. If there are two or more variables, however, simple cuts are usually far from optimal. Consider the case of just two observables. One could examine the two-dimensional scatter plot of the signal and background to select an S/B-enhancing cut. Simple cuts would partition the scatter plot along lines running parallel to the coordinate axes, with events in one or three of the resulting quadrants to be accepted and all others rejected. Let us further assume that the signal and background distributions are Gaussian. In this case, an optimal cut is generally along an ellipse or hyperbola which is the contour of constant S/B, and it cannot be written as one or even several simple cuts. Even in the special case where the optimal cut lies along a straight line (which happens in the Gaussian case when the signal and background are identical except for their centroids), that line is generally not a simple cut, because it need not be parallel to a coordinate axis. Furthermore, as Fig. 1 shows, the variables used here are obviously not Gaussian, so the form of the optimal cut is not apparent. It is unlikely, however, that the optimal cut is close to any set of simple cuts. Thus, finding the proper cuts by hand is difficult for two variables, and seemingly impossible for more than two variables.

The neural network approach [5,6] offers an alternative for signal/background classification that eschews the restrictive form of simple cuts in favor of very general decision contours, which can be parametrized recursively in terms of a hyperbolic tangent, or other appropriate transfer function. It has the unfortunate drawback, however, of yielding a "black-box" solution whose cuts are not easy to interpret in physical terms, except in trivial cases or where very few (three or less) variables are employed. Reference [7] applied neural networks to the same signature considered here, giving explicit interpretation of the trained network in the case of two input variables. For the case of six variables, however, the study fell back on examining the distributions in each of the variables separately, which is adequate and informative, assuming that the ideal cuts do not involve complicated correlations. As will be discussed below and summarized in Table I, however, we have found optimized cuts [cuts (b) and (d) in the table] that have clear and powerful physical significance, but that would defy easy interpretation in terms of the individual variables. Some other algorithms that have been considered, including H matrix and probability density estimation [7], also efficiently separate signal from background, but fail to match the transparency of simple cuts.

In this paper we advocate instead the binary decision tree [3], which, compared to the conventional method, yields much higher signal efficiency. The decision tree has been shown to perform at the same level as the neural network in an earlier simple study of the top quark signal [8], but with the distinctions that it yields cuts that are simpler to interpret physically, and makes more modest demands on computer horsepower. The basic decision tree was described in Refs. [3,8]. We outline the algorithm in the form implemented in the program HASTAC [9], which has been tailored for use in high-energy physics signal identification.

Let the set of variables $x = (x_1, \ldots, x_n)$ define the feature space of events, with each x_i an observable such as m_{jj}^6. A generalized cut in $(x_1, \ldots, x_n)$ is the requirement that each event satisfy the inequality $\hat{a} \cdot (x - x^0) > 0$, where $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_n)$ is a vector normalized to $\sum_i \hat{a}_i^2 = 1$. The geometrical interpretation of this expression is clear: the feature space is cut in two by a hyperplane passing through the point x^0, with the hyperplane orientation specified by its normal $\hat{a}$. Simple single-variable cuts are just hyperplanes restricted to normals along one coordinate axis of the feature space. The power of the decision tree derives from its ability to optimally determine $\hat{a}$ and x_0 for one or more generalized cuts. In this paper, we will restrict ourselves to two or three generalized cuts, which are sufficient to strongly suppress the background.

The optimized hyperplane cuts are found by the decision tree as follows [3]. Approximating the step function as $\theta(\lambda) \approx \Theta(\lambda) = (1 + e^{-\lambda/T})^{-1}$, where T is a relatively small number, the number of signal events S_A falling on

TABLE I. Effect of HASTAC-derived generalized cuts on signal and background, in events at $\int \mathcal{L}dt = 100\ \mathrm{pb}^{-1}$. The percentage of events relative to acceptance cuts only is given in parentheses.

Cuts	Signal ($m_t = 175\,\mathrm{GeV}/c^2$)	Signal ($m_t = 190\,\mathrm{GeV}/c^2$)	Background
acceptance cuts only	49.0 (100%)	32.3 (100%)	116.0 (100%)
(a)	40.8 (83%)	28.5 (88%)	29.2 (25%)
(a) and (b)	38.6 (79%)	25.5 (79%)	19.8 (17%)
(a), (b), and (c)	30.0 (61%)	20.4 (63%)	10.0 (9%)
(a), (b), and (d)	22.4 (46%)	16.5 (51%)	6.0 (5%)
(a)	$[0.76\,p_T^4(j) + 0.14\,m_{jjj}^4 + 0.65\,m_{jj}^6 + 0.13\,p_T(W)]/m_W > 1.0$		
(b)	$0.43\,m_{jjj}^4/m_W - 0.32\,m_{jj}^6/m_W + 0.089\,\lvert\eta^4(j)\rvert + 0.097\,\lvert\eta^3(j)\rvert < 1.0$		
(c)	$(-0.13\,m_{jjj}^1 + 0.79\,m_{jjj}^4)/m_W > 1.0$		
(d)	$-0.31\,m_{jjj}^1/m_W + 1.4\,m_{jjj}^4/m_W - 0.63\,\lvert\eta^4(j)\rvert > 1.0$		

the "accepted" side of the hyperplane is approximated by

$$S_A(\hat{a}, x^0) = \sum_\alpha \Theta(\hat{a} \cdot [x(\alpha) - x^0])\,, \tag{1}$$

with a sum over all signal events α. $B_A(\hat{a}, x^0)$ is defined analogously for the background. With S_A and B_A thus transformed into differentiable functions of $\hat{a}$ and x^0, we employ conjugate gradient optimization [10] to maximize

$$Q_N(\hat{a}, x^0) = S_A(\hat{a}, x^0)/[B_A(\hat{a}, x^0)]^N\,. \tag{2}$$

The parameter N can be chosen to assign primary importance to S/B enhancement ($N = 1 \Rightarrow Q = S/B$) or to high signal acceptance ($N \to 0 \Rightarrow Q \to S$). The value $N = 0.5 \Rightarrow Q = S/\sqrt{B}$ makes the optimized function Q equal to the approximate statistical significance S/σ_B of the signal, assuming S and B to be Poisson distributed. After optimization, each cut is specified by $\{\hat{a}, x^0\}$, or more concisely by a form $a \cdot x < c$, where c is a number. Qualitative interpretation of each cut is through the relative signs of a_i, which indicate positive or negative correlation in each variable with the likelihood of an event being signal.

Next, we describe the physical features of the signal and background on which our efficient cuts are based. The primary background to the top quark signature $\ell + 4$ jets is the set of processes leading to direct production of $W + 4$ jets. After minimal acceptance cuts given below, about 40% of the background is due to $q\bar{q} \to Wgggg$ processes. The other major sources of background are $qg \to Wgggq$, $qq \to Wggqq$, and $qg \to Wgqqq$ (q = quark or antiquark), with contributions ranging from 15% to 30%. The background is thus characterized by processes with multiple gluon jets in the final states. The structure of the matrix elements dictates that much of the cross section will lie in regions in phase space close to collinear and/or infrared divergences. Near-collinear radiation of jets with respect to the incoming $p, \bar{p}$ leads to jets with low transverse momentum p_T (due to quark and gluon bremmstrahlung) or high pseudorapidity $\eta = -\ln\tan\theta/2$. Collinear and infrared divergences influence gluon bremmstrahlung and splitting, leading to production of $q + g$ or $g + g$ with small relative angle and low dijet mass m_{jj}. The trijet masses m_{jjj} similarly tend to be low.

In strong contrast, the large mass of the top quark pair implies that it is produced with low velocity (50% of the time with $v/c < 0.32$) at the Fermilab Tevatron energy $\sqrt{s} = 1.8$ TeV. (The bulk of top quark pair production is through $q\bar{q} \to t\bar{t}$, with the next largest contribution $gg \to t\bar{t}$ representing only about 10%.) The velocities of t and $\bar{t}$ are also small. The two-body top quark decay $t \to bW^+$ is roughly isotropic in the top rest frame, giving the b jet a characteristic maximum transverse momentum scale $\sim m_t/2$. In fact, we find, for a top quark of mass $175\,\mathrm{GeV}/c^2$, that the b-jet p_T distribution peaks at $52\,\mathrm{GeV}/c$ with mean $71\,\mathrm{GeV}/c$. The jets from hadronic W decay share the W momentum, so their average p_T is somewhat smaller but still peaks at $32\,\mathrm{GeV}/c$ with mean $56\,\mathrm{GeV}/c$. One expects large trijet masses ($m_{jjj} \sim m_t$), and also large dijet masses, $m_{jj} \sim m_W$, or $m_{jj} \sim m_t/\sqrt{3}$ in view of the kinematic relation $m_{123}^2 = m_{12}^2 + m_{13}^2 + m_{23}^2$.

Before making a detailed comparison of signal and background, we list the minimal acceptance cuts we impose to simulate detector acceptance, and describe our calculation of the signal and background. The acceptance cuts are

$$\begin{aligned} &p_T(j) > 17.0\,\mathrm{GeV}/c, \quad p_T(\ell) > 20.0\,\mathrm{GeV}/c, \\ &\not{p}_T > 25.0\,\mathrm{GeV}/c, \\ &\lvert\eta(j)\rvert < 2.0, \quad \lvert\eta(\ell)\rvert < 2.0, \\ &R(j,j') > 0.7, \quad R(j,\ell) > 0.4\,. \end{aligned} \tag{3}$$

Here $R(j,j') = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ where $\Delta\phi$ and $\Delta\eta$ are the differences in azimuthal angle ϕ and pseudorapidity η between jets j and j'. To simulate detector resolution, the η and ϕ of each parton was smeared from its true value by Gaussian random amounts with standard deviation 0.05 in each. The missing transverse momentum $\not{p}_T$, which is taken as a measurement of p_T^ν, was calculated by smearing each parton energy by a Gaussian random amount with $\sigma(E)/E = C/\sqrt{E_T}$ where $C = 0.6$ for jets and 0.15 for ℓ, before calculating the transverse momentum imbalance. To simulate the effects of hadronization, we further smeared the jet energies so that $\sigma(E)/E = 1.0/\sqrt{E_T}$.

We employed helicity amplitude techniques to compute top quark production, keeping all top quark and W boson decay correlations. To calculate the background, we

used the Monte Carlo package VECBOS [11]. We used the CTEQ2 set five parton distributions [12], which are leading-order fits and hence appropriate for our leading-order calculation. Similarly, we used a leading-order form for α_s, with $\Lambda_{\rm QCD}$ given by the parton distributions. Factorization and renormalization scales were chosen as $\mu_R = \mu_F = m_t$ for the signal and $\mu_R = \mu_F = m_W$ for the background. The background rate in particular has theoretical uncertainties, so its direct measurement described below is most welcome. We will discuss the specific case of $m_t = 175\,\mathrm{GeV}/c^2$ in considerable detail, but also include results for $m_t = 190\,\mathrm{GeV}/c^2$ in the table for comparison. These results show that the efficacy of both our cuts and our top mass reconstruction depend very weakly on the true value of m_t.

Assuming the projected integrated luminosity $\int \mathcal{L}dt = 100\mathrm{pb}^{-1}$ for "Run I" at the Tevatron, we expect a total of 49 top quark signal events (for $m_t = 175\,\mathrm{GeV}/c^2$), and 116 background events, to pass the minimal acceptance cuts (3). Thus we begin with $S/B \sim 0.42$ before our discrimination cuts.

Some variables that we have tried as input to the decision tree program HASTAC are ordered versions of the observables discussed above. The jet transverse momenta are $p_T^1(j) > \cdots > p_T^4(j)$. The jet pseudorapidities are $|\eta^1(j)| < \cdots < |\eta^4(j)|$. The dijet masses are $m_{jj}^1 > \cdots > m_{jj}^6$ and the trijet masses are $m_{jjj}^1 > \cdots > m_{jjj}^4$.

Even before application of the decision tree, several of these variables point out significant differences between signal and background. In the signal, one pair of jets comes from the decay of a W, so the minimum dijet mass m_{jj}^6 is less than m_W except for smearing effects and the W width. The pair masses otherwise tend to be large, so as shown in Fig. 1 the signal climbs steadily with m_{jj}^6 to a peak near m_W, after which it drops sharply. In contrast, the background falls quickly from its largest value at $m_{jj}^6 \approx 2\,p_T^{\min}(j) = 34\,\mathrm{GeV}/c$. A simple cut $m_{jj}^6 > 40\,\mathrm{GeV}/c^2$ passes $(S,B) = (43, 50.2)$ events. A tighter cut could even raise S/B above 1: $m_{jj}^6 > 56\,\mathrm{GeV}/c^2$ passes $(S,B) = (30.8, 18.2)$ events. The power of this variable reflects the qualitative differences between signal and background described above.

Similarly, the distribution in lowest trijet mass m_{jjj}^4 for the signal rises to a peak near m_t and then falls sharply because the signal always contains at least one trijet combination with $m_{jjj} = m_t$ (modulo jet resolution and width of the top quark), so that the minimum trijet mass cannot rise above m_t. Meanwhile the background distribution falls steadily with m_{jjj}^4. The simple cut $m_{jjj}^4 > 120\,\mathrm{GeV}/c^2$ would accept $(S,B) = (42.0, 44.0)$ events. However, unlike the m_{jj}^6 distribution, a tighter cut would not yield any further significant enhancement in S/B. It is interesting to note that m_{jjj}^4 by itself could serve as a crude but effective method to directly detect the top quark without recourse to any fitting procedure or assumptions about the value of m_t (other than that it lies somewhere above $100\,\mathrm{GeV}/c^2$). Despite uncertainties in the background due to scale uncertainties and possible large radiative corrections, the $W+4$ jets background *must* be smoothly falling in m_{jjj}^4, even with background-suppressing simple cuts in such variables as $p_T^4(j)$. The signal, on the other hand, would necessarily peak around m_t, even in the case of a *non-standard-model* decay of the top quark such as $t \to bH^+$, with a charged Higgs boson decaying $H^+ \to c\bar{s}$.

As expected from the infrared enhancement in the QCD background, the minimum jet transverse momentum $p_T^4(j)$ also distinguishes well between signal and background. The cut $p_T^4(j) > 25\,\mathrm{GeV}/c$ keeps $(S,B) = (35.2, 36.2)$ events. An extremely tight cut of $p_T^4(j) > 35\,\mathrm{GeV}/c$ will raise S/B to more than 2, at the cost of signal acceptance, with $(S,B) = (17.2, 8.2)$ events. We note in passing that D0 Collaboration employed a related variable, the scalar sum of the jet transverse momenta $H_T = \sum_i |p_T^i(j)|$. A high signal acceptance cut in H_T is as good as $p_T^4(j)$ — taking $H_T > 210\,\mathrm{GeV}/c$, the events passing this cut are $(S,B) = (37.4, 39.4)$ — but no amount of tightening the cut on H_T will obtain S/B significantly over 1.

The three observables just discussed, m_{jj}^6, m_{jjj}^4, and $p_T^4(j)$, are the most powerful discriminators we have found, as judged by their solo performances. We used them as input to the HASTAC optimization, in concert with four additional variables whose individual distributions do not so clearly separate signal from background, but which prove useful in correlation with the first three. We should note that several other observables [e.g., the other m_{jj}^i, m_{jjj}^i, and $p_T^i(j)$] are similarly helpful, so that our choice of variables was dictated largely by taste and the ease in interpreting the final cuts.

The first two of these four additional variables are the two largest jet pseudorapidities $|\eta^4(j)|$ and $|\eta^3(j)|$, which complement $p_T^4(j)$ in recognizing jet radiation that is collinear with the incoming beams. The third variable is the leptonic W transverse momentum $p_T(W)$, which peaks toward $\sim m_t/2$ in the signal and tends to be smaller in the background. The fourth added observable is the maximum trijet mass m_{jjj}^1, which can serve to close a "high-mass" loophole for the background. The QCD collinear enhancement can give rise to beam jets, which have high energy but low p_T and lie close to the beam. If, for example, a pair of narrowly separated beam jets is produced in each direction, two beam jets in the same direction can have a relatively high mass (despite their small separation) due to their high energy. Much higher dijet masses result from picking two jets in opposite directions. This high-m_{jj} loophole can be closed by further requiring low trijet masses (low m_{jjj}^1 or m_{jjj}^4), since any trijet mass will necessarily be dominated by the larger dijet masses. We note that the range of trijet mass for the signal will be relatively small, centering around be relatively small, centering around m_t.

Having selected the observables, one would conventionally make cuts in several of these variables individually, and by trial and error adjust the cuts for the best discrimination. That route would not only be laborious; it would also totally miss any useful *correlations* between the variables, because it permits only "rectangular" cuts. We therefore presented the variables to HASTAC for automatic generation of efficient generalized cuts. We de-

tail our generalized cuts in Table I. Because two of the cuts involve variables of different dimension, we scale all momenta and masses by m_W for convenience. We compare the background with a signal for $m_t = 175\,\mathrm{GeV}/c^2$ in the following, but note that very similar results for $m_t = 190\,\mathrm{GeV}/c^2$ are indicated in Table I.

The first generalized cut (a), which drastically shrinks the background, can be understood by examining the coefficients in Table I. For an event to pass this cut, the left hand side of the inequality (a) must be sufficiently large and positive. Since all four observables have positive coefficients, the cut simultaneously requires high $p_T^4(j)$, m_{jj}^6, m_{jjj}^4, and $p_T(W)$ — precisely as anticipated in the discussion above. The advantage of generalized cuts shows up in the extra 25% decrease of background relative to cuts in any one variable for the same signal efficiency. The power of the variables m_{jj}^6 (Fig. 1) and $p_T^4(j)$ is reflected in their large coefficients, while m_{jjj}^4 and $p_T(W)$ have relatively smaller weight, and thus less discriminating power (for this particular cut).

The second generalized cut (b) functions differently. The coefficient of m_{jjj}^4 is now the largest, and has a negative sign relative to that of m_{jj}^6. The hyperplane describing this cut is thus somewhat orthogonal to that of the first cut. Indeed, taking into acccount the reversed inequality sign, the negative coefficient demands a high m_{jj}^6, and the positive coefficient of m_{jjj}^4 imposes an upper limit on this trijet mass. This cut serves to close the "high-m_{jj} loophole" described above. Cut (b) also clamps down further on accidental high-dijet masses by suppressing beam jets, since the positive coefficients of $|\eta^4(j)|$ and $|\eta^3(j)|$ force the jets to be more central.

Together, cuts (a) and (b) pass 79% of the signal, but only 17% of the background, giving S/B almost as high as the tight cut in $p_T^4(j)$ described above, but with *twice* the signal acceptance. This set of high-acceptance cuts (a) and (b) will serve as the starting point for our reconstruction of the top quark mass. But first we comment on the more stringent third and fourth cuts.

Both of these cuts function similarly to cut (b), and are intended only to illustrate how an even higher S/B can be obtained without explicit top quark reconstruction (though clearly the latter may also be used to increase S/B). Indeed, in the more extreme case [cuts (a), (b) and (d)], the signal/background ratio is almost 4, which is unattainable through any of the variables taken individually. Their interpretation is straightforward. Since $m_{jjj}^4 \leq m_{jjj}^1$, cut (c) explicitly requires $m_{jjj}^4 > 1/(0.79 - 0.13)\, m_W \approx 3/2\, m_W$, and puts a m_{jjj}^4-dependent upper bound on m_{jjj}^1, which, like cut (b), should help shrink the high-dijet mass loophole. The more severe cut (d) strengthens the requirement for centrally located jets.

Finally, we remark that, although we have discussed above only $m_t = 175\,\mathrm{GeV}/c^2$, Table I shows that all of the cuts have almost identical effect on a signal with $m_t = 190\,\mathrm{GeV}/c^2$, which reflects the relatively small dependence on the exact value of m_t for which the cuts were optimized.

We next present a new top quark reconstruction algorithm that, applied to events passing the high acceptance cuts (a) and (b), can measure m_t directly. Our first key observation is that the measurement should be based on the hadronic decay $t \to bq\bar{q}$, since the rather poor measurement of the neutrino momentum significantly degrades the mass resolution for $t \to b\ell\nu$. Our goal is to form a histogram of m_{jjj} for three-jet systems that are tagged as coming from $t \to bq\bar{q}$, using the $t \to b\ell\nu$ mass only for the tagging, i.e., to recognize which of the four jets came from the leptonically decaying top, leaving the leftover trio as the hadronic decay. The location of the peak in m_{jjj} will measure m_t (with Monte Carlo calculations needed only to assess instrumental effects).

Fitting the distribution to the standard model signal plus a smooth but uncalibrated background will also allow for direct measurement of the backgrounds from QCD and from incorrect jet assignment. For such a measurement, it is crucial that the reconstruction procedure does not artificially induce a peak in the background. Because of the possibility that leading-order models of the background such as VECBOS may be quite unreliable, a significant background peak would greatly reduce the usefulness of such a fit. We will show below that our algorithm [including cuts (a) and (b)] produces no perceptible peak in the background, and thus makes a model-independent measurement of the background possible.

Unlike other analyses, we do not attempt to fully reconstruct the event by trying to identify which pair of the three jets in the hadronic decay came from the W. This keeps the "combinatoric problem" under control, since it cuts down the possible jet assignments from 12 per event to just four. Also, since we treat the three jets in $t \to jjj$ symmetrically, at the end of the analysis we can plot a histogram of dijet pair masses from $t \to jjj$ candidates (three combinations per event) and, without reconstruction-induced bias, observe the $W \to jj$ peak in it. This will give an important independent calibration of jet energy measurement and jet-finding algorithms.

Our partial reconstruction is carried out as follows. For each event that passes the S/B enhancement cuts (a) and (b), we assign each of the four jets in turn to go with the lepton. Let m_{jjj} be the invariant mass of the remaining three jets. We select the assignment if (1) $120\,\mathrm{GeV}/c^2 < m_{jjj} < 240\,\mathrm{GeV}/c^2$; (2) $|m_{j\ell\nu} - m_{\mathrm{trial}}| < 20\,\mathrm{GeV}/c^2$; (3) $|m_{j\ell\nu} - m_{\mathrm{trial}}|$ is the smallest of the four possibilities that pass (1) and (2). We took the trial top quark mass $m_{\mathrm{trial}} = 175\,\mathrm{GeV}/c^2$, but show below that this choice affects only the height, and not the location, of the mass peak in m_{jjj}. In practice, of course, a range of m_{trial} may be swept to optimize the signal peak. The mass range for m_{jjj} is kept very broad, so there is ample room to separate peak from background. The mass range for $m_{j\ell\nu}$ was chosen to keep $\sim 70\%$ of the true signal. We have checked that this algorithm does not produce fake peaks due to either the QCD or combinatoric backgrounds.

Measurement of the neutrino momentum is crucial for measurement of $m_{j\ell\nu}$. The transverse momentum of ν is taken to be the negative of the total $\vec{p}_T$ observed in the calorimeter, giving it an uncertainty due to the uncertainties of all four jet $\vec{p}_T$'s added in quadrature; plus

contributions from inaccurate measurement of the many low-p_T particles in the event, the possibility of other neutrinos (e.g., from semileptonic decays in one or both b jets), and instrumental effects due to gaps in the detector coverage. The longitudinal momentum of the neutrino can be computed from $m_{\ell\nu} = m_W$, with a twofold ambiguity in addition to uncertainties due to the width of the W and the error in $\vec{p}_T^{\,\nu}$. That computation is usually expressed by a quadratic equation for p_L^ν, but it is much clearer to think of it as follows. The invariant mass $m_{\ell\nu}$ is given by

$$m_{\ell\nu}^2 = 2\,p_T^\nu\,p_T^\ell\,[\cosh(\eta_\nu - \eta_\ell) - \cos(\phi_\nu - \phi_\ell)]. \qquad (4)$$

By assuming $m_{\ell\nu} = m_W$ one determines $\cosh(\eta_\nu - \eta_\ell)$ and hence $|\eta_\nu - \eta_\ell|$. The twofold solution ambiguity is due to the undetermined sign of $\eta_\nu - \eta_\ell$: *the two solutions for η_ν lie on either side of η_ℓ and equidistant from it.* There will be considerable uncertainty in $|\eta_\ell - \eta_\nu|$ due to errors in p_T^ν and ϕ_ν, the finite W width, and because $\cosh(\eta_\ell - \eta_\nu)$ is usually close to 1, where $m_{\ell\nu}$ is rather insensitive to $\eta_\ell - \eta_\nu$. It can even happen ($\sim 20\%$ of the time) that there is no solution, in which case $\eta_\nu = \eta_\ell$ is the best guess. When there are two solutions, we choose the sign of $\eta_\ell - \eta_\nu$ to be that of η_ℓ (the solution with the smaller W energy), which most of the time is correct at the Tevatron, since the W's are produced rather centrally in rapidity due to the limited total energy. Even for the $\sim 22\%$ of events where the wrong solution is chosen, this rule is often adequate since (1) the two solutions are often close to each other, (2) we only need the neutrino momentum to compute $m_{b\ell\nu}$, which is not always very sensitive to η_ν, and (3) we only need $m_{b\ell\nu}$ measured accurately enough to tag the correct one of the four jets.[2]

Figure 2 shows the resultant plot for m_{jjj}, with $m_t = 175$ and $190\,\mathrm{GeV}/c^2$. We have plotted only the events passing cuts (a) and (b) with $|m_{j\ell\nu} - m_{\mathrm{trial}}| < 20\,\mathrm{GeV}/c^2$, $m_{\mathrm{trial}} = 175\,\mathrm{GeV}/c^2$. This includes 32.6/38.6 signal events for $m_t = 175\,\mathrm{GeV}/c^2$, 18.2/25.5 signal events for $m_t = 190\,\mathrm{GeV}/c^2$, and only 13.7/19.8 background events. The resulting clear peak has suffered almost no shift away from m_t, despite simulated detector smearing effects and, importantly, nonoptimal choice of m_{trial} in the case of $m_t = 190\,\mathrm{GeV}/c^2$. (The peak for $m_t = 190\,\mathrm{GeV}/c^2$ increases by 10% if $m_{\mathrm{trial}} = 190\,\mathrm{GeV}/c^2$ is used, but its location is unchanged.) This result provides verification

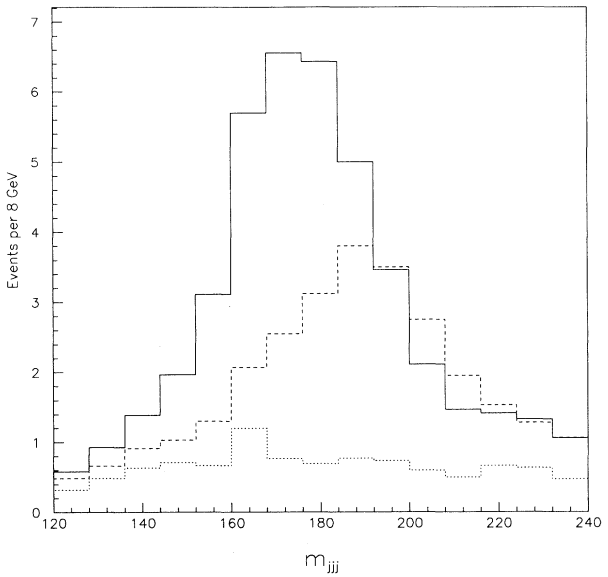

FIG. 2. The reconstructed top quark mass m_{jjj} in $8\,\mathrm{GeV}/c^2$ bins, with cuts (a) and (b) and the requirement that $|m_{j\ell\nu} - m_{\mathrm{trial}}| < 20\,\mathrm{GeV}/c^2$, for $m_{\mathrm{trial}} = 175\,\mathrm{GeV}/c^2$. The solid curve indicates the sum of signal, with $m_t = 175\,\mathrm{GeV}/c^2$, and background; the dashed curve gives the sum of signal, with $m_t = 190\,\mathrm{GeV}/c^2$, and background; the dotted curve gives the background alone. All are presented for $\int \mathcal{L}dt = 100\ \mathrm{pb}^{-1}$.

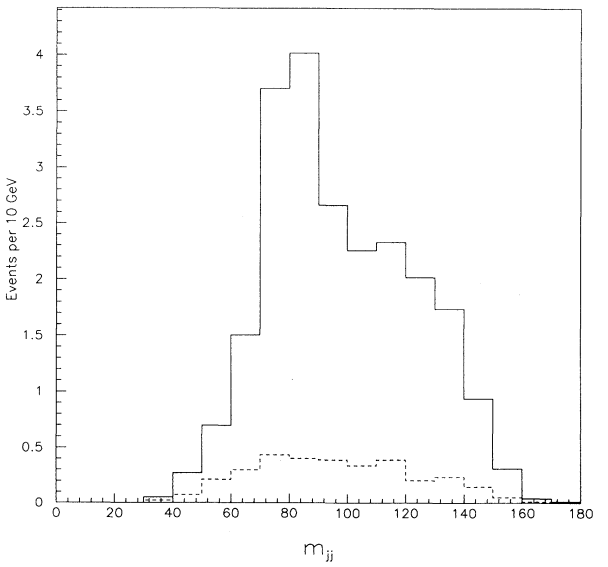

FIG. 3. Mass distribution (three combinations per event) for dijets formed from hadronically decaying tops ($m_t = 175\,\mathrm{GeV}/c^2$) identified using the top mass reconstruction algorithm [cuts (a) and (b), $|m_{j\ell\nu} - m_{\mathrm{trial}}| < 20\,\mathrm{GeV}/c^2$ with $m_{\mathrm{trial}} = 175\,\mathrm{GeV}/c^2$, and $|m_{jjj} - m_t| < 15\,\mathrm{GeV}/c^2$] at $\int \mathcal{L}dt = 100\ \mathrm{pb}^{-1}$. The W boson mass peak, which is *not* used in the analysis, shows up clearly. The dashed curve, representing the background alone, is shown for comparison.

[2]At much higher energies, such as at the CERN Large Hadron Collider (LHC), there is no clear way to choose the correct neutrino solution in order to evaluate $m_{j\ell\nu}$. For such a case, one can avoid choosing by instead using $m^*_{j\ell\nu}$ which is defined by minimizing $m_{j\ell\nu}$ with respect to η_ν. To an excellent degree of approximation, that is equivalent to assigning η_ν to the p_T-weighted average $\eta_\nu^* = (p_T^\ell\eta_\ell + p_T^j\eta_j)/(p_T^\ell + p_T^j)$. The "Jacobian peak" in the amount of phase space near the minimum causes a sharp peak in the probability distribution for $m^*_{j\ell\nu}$ at a value only slightly lower than the true peak in $m_{j\ell\nu}$. The quantity $m^*_{j\ell\nu}$ is analogous to the "transverse mass" variable used in measuring m_W.

that our method, by relying totally on $m_{j\ell\nu}$ for the trijet selection, avoids introducing bias into the trijet mass. Of course, we have not simulated the effect of initial and final state radiation, and leptonic decay of the b quarks — the latter especially would likely cause a systematic lowering of the trijet mass. Further detailed study is required for a maximally realistic assessment of the methods proposed here.

A nice cross-check of a top quark peak found using this method is shown in Fig. 3, where for each trio of jets in the peak, each of the three dijet mass combinations is plotted (with weight 1/3 each). A clear peak at m_W appears, which will provide a unique calibration for the hadronic calorimetry and the jet-finding algorithm. The combinatoric background under the W peak is substantial, which shows the wisdom of *not* trying to recognize $W \to jj$ as part of the $t\bar{t}$ event selection. We also note that the method proposed here makes no use of b tagging, and could be used in tandem with the b-tagging cuts imposed by the CDF Collaboration to raise signal acceptance rate, by accepting (in addition to single or double b-tagged events) any *untagged* events that pass the cuts described above.

In conclusion, we have demonstrated the usefulness of the binary decision tree technique in separating signal and background events for top quark production at the Tevatron. We showed why the new observables m_{jj}^6, m_{jjj}^4, and $p_T^4(j)$ strongly enhance the signal. We introduced an algorithm to determine the top quark mass, which yields a directly observable $t \to bq\bar{q}$ peak in a certain m_{jjj} distribution. We further showed that $S/B \approx 4$ is achievable (for $m_t = 175\,\mathrm{GeV}/c^2$), with only about 50% loss of the signal beyond typical minimal experimental acceptance cuts.

We wish to emphasize here that the explicit cuts given in Table I are meant only to illustrate the kind of cuts that should be used. The actual parameters in the cuts will change somewhat when the parton-level simulation is upgraded to a full QCD Monte Carlo calculation, and detector effects are included in more detail.

Finally, we point out that the methods derived here for $t\bar{t} \to \ell + 4$ jets could be used in an analogous fashion to observe the total hadronic signature $t\bar{t} \to 6$ jets [13]. We expect a similar substantial increase of S/B through HASTAC-derived cuts. Given the higher event rate, further background suppression by requiring one b tagged jet would greatly reduce B but leave sufficient signal events. In analogy to the tag on $t \to b\ell\nu$, we would pick from the five other jets the pair that (1) reconstructs a W boson and (2) best reconstructs $t \to jjj$ with the tagged b. Then the invariant mass of the *other* three jets should have an unbiased peak at m_t. Work in this direction, as well as refinements of our method, such as inclusion of b-tagging information for $t\bar{t} \to \ell + 4$ jets and consideration of events with initial and final state radiation, is in progress.

We thank the other HASTAC collaborators, including R. Hatcher, J. Linnemann, and in particular J. Hughes for much assistance in implementing the current algorithm. S. Chao provided crucial advice on the optimization algorithm. We also had useful discussions with H. Miettinen, H. Weerts, and C.-P. Yuan. P.A. was supported in part by NSF Grant No. PHY-9396022. D.B.-C. was supported in part by NSF Grant No. PHY-9307980.

[1] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995).
[2] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74**, 2632 (1995).
[3] R. P. Brent, "Fast Training Algorithms for Multi-layer Neural Nets," Australian National Laboratory report (unpublished).
[4] For an example of such a study see V. Barger, J. Ohnemus, and R. Phillips, Phys. Rev. D **48,** R3953 (1993); V. Barger, E. Mirkes, J. Ohnemus, and R. J. N. Phillips, Phys. Lett. B **344**, 329 (1995).
[5] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, Phys. Rev. Lett. **65,** 1321 (1990).
[6] H. Baer, D. D. Karatas, and G. F. Giudice, Phys. Rev. D **46,** 4901 1992.
[7] P. C. Bhat, in *The Albuquerque Meeting,* Meeting of the Division of Particles and Fields by the APS, Albuquerque, New Mexico, 1994, edited by S. Seidel (World Scientific, Singapore, 1995).
[8] D. Bowser-Chao and D. L. Dzialo, Phys. Rev. D **47,** 1900 (1993), and references therein.
[9] D. Bowser-Chao, R. Hatcher, J. Hughes, and J. Linnemann (unpublished).
[10] D. A. Wismer and R. Chattergy, *Introduction to Nonlinear Optimization* (North-Holland, New York, 1978).
[11] F. A. Berends, W. T. Giele, and H. Kuijf, Nucl. Phys. **B321**, 39 (1989); F. A. Berends, W. T. Giele, H. Kuijf, and B. Tausk, *ibid.* **B357**, 32 (1991).
[12] CTEQ Collaboration, J. Botts *et al.*, Phys. Lett. B **304**, 59 (1993).
[13] W. T. Giele *et al.*, Phys. Rev. D **48**, 5226 (1993).