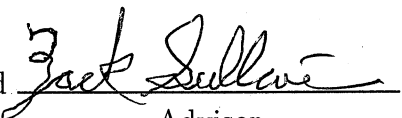


EXPANDING THE HEP FRONTIER WITH  
BOOSTED  $B$ -TAGS AND THE QCD POWER SPECTRUM

BY  
KEITH PEDERSEN

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor or Philosophy in Physics  
in the Graduate College of the  
Illinois Institute of Technology

Approved   
Advisor

Chicago, Illinois  
May 2018



## ACKNOWLEDGMENT

This thesis could not have been written without Zack Sullivan, my adviser, mentor and friend. By guiding me through my first few projects, Zack helped me develop the tools of a particle theorist. By cultivating my rather vague ideas about a thesis project, then giving me the freedom to develop them, he permitted me to grow into a self-reliant scientist. Additionally, by editing my writing with care, Zack taught me the invaluable lesson that a good scientific paper cannot stand on its equations; it must “tell a story” and constantly remind the reader “where we’ve been and where we’re going”.

I also owe an immeasurable debt to Ariana, my wife and partner, whose tremendous patience and encouragement helped me stay the course. But even more importantly, Ariana gave me the push I needed to pull myself out my rut and go back to grad school. Her motivation was essential to this work.

Tracing back further I thank my parents, Neil and Diane, who always encouraged me to explore the world. From music lessons to Boy Scouts, summer camps and church groups — the lessons and experiences of my youth formed me into the person I am today. These experiences were not my doing, they were the guiding hand of my parents, an embodiment of their love and support.

Additional thanks to: Andrew Webster for lowering the activation energy to formalize the robust inversion method in Appendix A.1; Daniel Kaplan, for reading my thesis with great attention and teaching me that “which” and “that” are not synonymous; and Carlos Wagner for his help with the SUSY  $\Delta m_b$  correction in Chapter 3.2.2.2.

This work was supported by the U.S. Department of Energy under award No. DE-SC0008347.

# TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENT . . . . .	iii
LIST OF TABLES . . . . .	vi
LIST OF FIGURES . . . . .	xiii
NATURAL UNITS . . . . .	xiv
LIST OF SYMBOLS . . . . .	xv
ABSTRACT . . . . .	xvi
CHAPTER	
1. INTRODUCTION . . . . .	1
1.1. Answering the challenge . . . . .	2
1.2. The collider . . . . .	7
1.3. The detector . . . . .	14
1.4. Event reconstruction . . . . .	23
2. THE $\mu_X$ BOOSTED-BOTTOM-JET TAG . . . . .	28
2.1. Tagging a heavy-flavored jet . . . . .	32
2.2. The $\mu_x$ boosted- $b$ tag . . . . .	34
2.3. $\mu_x$ tagging results . . . . .	42
3. SEARCHES FOR NEW PHYSICS VIA HEAVY-FLAVOR JETS	50
3.1. Leptophobic $Z'$ . . . . .	50
3.2. Two Higgs doublet models . . . . .	55
4. THE SHAPE OF QCD . . . . .	69
4.1. The pioneering work of Fox and Wolfram . . . . .	74
4.2. QCD power spectra $H_l$ . . . . .	83
4.3. Particle multiplicity and sampling noise . . . . .	99
4.4. Smearing out the sampling noise . . . . .	111
5. POWER JETS . . . . .	126
5.1. Jet shape . . . . .	129
5.2. The missing pieces . . . . .	136
5.3. Fitting power jets and pileup . . . . .	146



5.4. The future of the QCD power spectrum . . . . .	160
6. THESIS SUMMARY . . . . .	163
APPENDIX . . . . .	165
A. NUMERICAL STABILITY . . . . .	165
A.1. A robust inversion method . . . . .	166
A.2. Relativistic kinematics . . . . .	181
A.3. A precise vector class . . . . .	184
A.4. Stable recursion . . . . .	191
B. SMALL DETAILS . . . . .	194
B.1. A finite, isotropic sample . . . . .	195
B.2. Drawing uneven uniform variates . . . . .	201
BIBLIOGRAPHY . . . . .	205

## LIST OF TABLES

Table		Page
1.1	Jet constituents, sorted by electric charge and primary detection mode.	16
2.1	A summary of parameters chosen for $\mu_x$ boosted bottom jet tagging.	42
5.1	Reconstructed 3-jet kinematics for the 2-jet-like event (Fig. 5.8a). .	153
5.2	Reconstructed 3-jet kinematics for the 3-jet-like event (Fig. 5.8b). .	154

## LIST OF FIGURES

Figure		Page
1.1	Pairs of particles are only correlated on the same-side ( $\Delta\phi = 0$ ) of the beam in <i>high-multiplicity</i> proton-proton collision (i.e., a large number of reconstructed charged particles $N_{\text{ch}}^{\text{reco}}$ ). . . . .	6
1.2	A 3-jet event at PETRA, circa 1979 (19 GeV $e^+e^-$ collisions), as seen in the transverse plane. Solid curves are charged particle tracks; dotted lines are neutral particles (calorimeter towers not struck by tracks). . . . .	12
1.3	A multijet event at the LHC, circa 2015 (13 TeV $pp$ collisions). Left view is the transverse plane, right view is rotated $90^\circ$ . Cyan curves are reconstructed tracks; dots are tracker hits. There are at least 17 events (each emanating from its own vertex), but only one is useful for new physics (i.e., 16 pileup events). . . . .	12
1.4	The reconstructed tracks (lines) and pileup collision vertices (dots) during a special high-pileup run at CMS, producing $\mu = \mathcal{O}(100)$ pileup collisions. . . . .	14
1.5	A cutaway of the CMS detector. Most of its volume is the muon chamber. . . . .	15
1.6	A transverse slice of ATLAS, depicting the propagation of various particles through the tracker, calorimeter, and muon system. . . . .	16
1.7	A jet inside CMS, with constituent tracks (solid curves) and towers (blocks, with dotted line depicting reconstructed neutral $\mathbf{p}$ ). The energy of towers is depicted via their radial height, with ECal towers (brown) stacked before HCal towers (light green). The cone of the reconstructed jet is shaded. . . . .	18
1.8	The probability distribution of calorimeter boost for a square tower ( $a = 1$ ), using 65,000 samples of $N$ random particles uniformly distributed across the square, with energies drawn uniformly from the interval [1 GeV, 100 GeV]. . . . .	20
1.9	Two “equal area” calorimeters; (a) the “Crystal Ball” at SPEAR and (b) a hemisphere from the square-cell pseudo-detector used by the author for basic phenomenological studies. . . . .	22

1.10	Two jet definitions applied to the same set of massless protojets (charged towers from a simulated square-celled calorimeter). The “Lego” plot unrolls each protojet’s 3D direction $(y, \phi)$ into a plane, depicting energy $(p_T)$ as height. The colored regions depict the jet areas, with conical jets appearing elliptical. . . . .	27
2.1	The non- $b$ jet efficiency of the “track counting high purity” algorithm, as a function of $p_T$ , in CMS simulation. . . . .	30
2.2	The secondary vertex of a $B$ hadron decay. . . . .	32
2.3	Nomenclature for the center-of-momentum frame and boosted lab frame. . . . .	35
2.4	A muon’s lab frame $x$ as a function of its CM frame $z'$ , for several choices of $\kappa$ . For all $\kappa > 1$ , there are two $z'$ which map to a given $x$ . . . . .	37
2.5	Theoretical muon distribution $\frac{d\Gamma}{\Gamma dx}$ in the lab frame (with $\beta_B \rightarrow 1$ ) for various muon boosts $\gamma'_\mu$ in the CM frame of the decay. . . . .	38
2.6	$x(\xi)$ for a subjet with $\gamma_{\text{core}} = 250$ for (a) a harder muon ( $\lambda = 1/7$ ) and (b) a softer muon ( $\lambda = 1/13$ ). . . . .	41
2.7	$\mu_x$ tagging efficiency vs. (a) jet $p_T$ and (b) $\eta_{\text{jet}}$ . . . . .	45
2.8	Density of reconstructed candidate tags with $\mu = 40$ pileup events as a function of $f_{\text{subjet}}$ vs. $x$ (summing over all $p_T$ and $\eta_{\text{jet}}$ bins). . . . .	47
3.1	Events per bin expected for $5\sigma$ discovery of a $M_{Z'_B} = 2.5$ TeV signal, and backgrounds, in the (a) 2-tag and (b) 1-tag analyses using $100 \text{ fb}^{-1}$ of integrated luminosity at Run II of the LHC. . . . .	53
3.2	Estimated Run II ( $100 \text{ fb}^{-1}$ ) $5\sigma$ discovery potential and 95% confidence level exclusion limits for $g_B$ vs. $M_{Z'_B}$ for the 2-tag and 1-tag analyses. 2-tag 95% C.L. exclusion reach (not shown) is comparable to the 1-tag discovery reach. . . . .	54
3.3	The estimates of Figure 3.2 overlaid on existing experimental experimental exclusions for leptophobic $Z'$ at 95% C.L.. . . . .	54
3.4	A Feynman diagram for the leading order associated production of a charged Higgs, where each proton contributes a gluon. . . . .	56
3.5	A CMS search for $H^\pm$ found no charged Higgs in the gray regions of parameter space, excluding 2HDM with those $\tan \beta$ and $m_{H^\pm}$ . CMS is not sensitive to 2HDM in the white “wedge” region that opens up at large charged Higgs mass (as well as to $m_{H^\pm} \approx m_t$ , which bisects the figures). . . . .	58

3.6	The theoretical reach for charged Higgs mass near 1 TeV; (a) at the LHC and (b) at the LHC (hashed) and a 100 TeV future collider (light/salmon). These predictions are inconsistent with each other.	58
3.7	Predicted exclusion regime, at a 95% confidence level, for a generic 2HDM at a 14 TeV LHC in terms of (a) the effective Yukawa coupling $y_{tb}$ , and (b) the corresponding $\tan\beta$ .	64
3.8	Predicted exclusion regime, at a 95% confidence level, for a generic 2HDM at a 100 TeV FCC in terms of (a) the effective Yukawa coupling $y_{tb}$ , and (b) the corresponding $\tan\beta$ .	65
3.9	Predicted exclusion regime, at a 95% confidence level, for the MSSM at a 14 TeV LHC, with the sign of $\mu$ (a) positive, or (b) negative.	66
3.10	Predicted exclusion regime, at a 95% confidence level, for the MSSM at a 100 TeV FCC, taking the sign of $\mu$ to be (a) positive, or (b) negative.	67
4.1	The discovery of quark jets at SLAC, using different $E_{\text{cm}} \equiv \sqrt{S}$ : (a) The sphericity of events, with the predictions of (solid) the jet model and (dotted) isotropic phase space. (b) The azimuthal angle of the dijet axis.	72
4.2	The angular power spectrum of the cosmic microwave background, but annotated as if it were the power spectrum of a QCD event. $\mathcal{D}_l$ indicates what fraction of the CMB is composed of spherical harmonics of degree $l$ .	73
4.3	The spherical harmonics $Y_l^m$ to degree $l = 3$ . At each $l$ , the order $m$ runs from $-l$ to $l$ , producing $2l + 1$ orthogonal orientations. The complex phase is rounded to $\pm 1$ (white/black). A quadrupole ( $l = 2$ ) has the same phase on opposite sides of the origin.	73
4.4	The $\Upsilon$ resonance in independent detectors at DORIS (1978).	76
4.5	The original $H_l$ predictions for (left) free partons and (right) measurable particles.	77
4.6	The $\Upsilon(4S)$ resonance at CLEO (1980). A cut on $H_2$ dramatically reduces background.	79
4.7	The power spectrum of a 2-particle final state in its CM frame.	91

4.8	The power spectrum of a $q\bar{q}g$ final state with a soft gluon (2-jet-like). (a) The low- $l$ behavior and (b) the asymptotic behavior, with kinematic depiction. $H_l$ is depicted for (circle) partons and (square) $\mathcal{O}(100)$ measurable particles. Odd powers are hollow, red, and thin. Even powers are blue, filled and thick. . . . .	92
4.9	The power spectrum of a $q\bar{q}g$ final state with a hard gluon (3-jet-like). (a) The low- $l$ behavior and (b) the asymptotic behavior, with kinematic depiction. $H_l$ is depicted for (circle) partons and (square) $\mathcal{O}(100)$ measurable particles. Odd powers are hollow, red, and thin. Even powers are blue, filled and thick. . . . .	93
4.10	The phase space for a $q\bar{q}g$ event. . . . .	95
4.11	The probability $f(H_2)$ and $f(H_3)$ for $q\bar{q}g$ and $ggg$ final states, given phase space points $(x_1, x_2)$ where $H_2 \leq 0.95$ . . . . .	97
4.12	The probability $f(H_8)$ and $f(H_9)$ for $q\bar{q}g$ and $ggg$ final states, given phase space points $(x_1, x_2)$ where $H_2 \leq 0.95$ . . . . .	98
4.13	The probability $f(H_8)$ and $f(H_9)$ for $q\bar{q}g$ and $ggg$ final states, given phase space points $(x_1, x_2)$ where $0.45 \leq H_2 \leq 0.5$ . . . . .	98
4.14	Charged particle multiplicity $n_{\text{track}}$ vs. jet $p_T$ at the LHC. . . . .	101
4.15	The ratio $\sum p_T^{\text{track}}/p_T^{\text{jet}}$ is a proxy for charged multiplicity to total particle multiplicity. “The momentum ratio of charged particles to all particles is nearly 2/3 due to the number of pion species ...”. . . . .	102
4.16	The power spectrum for three random, isotropic samples ( $N = 64$ , $N = 1,024$ , $N = 16,384$ ). For each sample, the $\langle f f \rangle$ prediction of Equation 4.47 is depicted with the dotted line. The circle shows where $l = \pi/(2 \arcsin(1/\sqrt{N}))$ . . . . .	106
4.17	The power spectrum for a uniformly filled calorimeter in a spherical pseudo-detector with (left) $\Delta\theta = 20^\circ$ towers and (right) $\Delta\theta = 5^\circ$ towers on a log-scale. The dashed line shows $1/N$ and the arrows indicates $l = 2\pi/\Delta\theta$ . . . . .	108
4.18	A unit square filled with (a) uniformly random points and (b) a uniform grid, with their respective discrete Fourier transforms. . . . .	110
4.19	A Dirac comb is an infinite series of $\delta$ functions with period T. . . . .	111
4.20	The raw angular correlation function for the measurable particles in a 2-jet-like event, truncating the series to (a) $l_{\text{max}} = 16$ and (b) $l_{\text{max}} = 256$ . . . . .	113

4.21	A particle shape function $h(\hat{r})$ that is pseudo-Gaussian in polar angle $\theta$ (relative to the particle's nominal direction of travel $\hat{p}_i$ ). The radial position of the red line indicates the fraction of the particle that is found at each $\hat{r}$ . . . . .	114
4.22	A pair of shape functions which have been rotated such that $h_{(i)}(\hat{r})$ is parallel to the $z$ -axis (on-axis), maintaining the interior angle with $h_{(j)}(\hat{r})$ (off-axis). . . . .	117
4.23	The squared, on-axis coefficient $\bar{h}_l^2$ for the Gaussian shape function, depicting several smearing angles $\lambda$ on a (a) linear and (b) log-linear scale. . . . .	120
4.24	The angular correlation function for the measurable final-state particles in a $q\bar{q}g$ final state with a soft gluon (2-jet-like), with (inset) a kinematic depiction of the original partons. $H_l$ for this event was shown in Figure 4.8. (top) Small and medium smearing angle $\lambda$ ; (bottom) medium and large smearing angle $\lambda$ . . . . .	121
4.25	The angular correlation function for the measurable final-state particles in a $q\bar{q}g$ final state with a hard gluon (3-jet-like), inset with a kinematic depiction of the original partons. $H_l$ for this event was shown in Figure 4.9. (top) Small and medium smearing angle $\lambda$ ; (bottom) medium and large smearing angle $\lambda$ . . . . .	123
5.1	A 3-jet-like, $q\bar{q}g$ event with (solid) $H_l^{\text{obs}}$ and (dot-dashed) $H_l^{\text{jet}}$ , fit using $\delta$ -function jets in (a) a 3-jet model and (b) a 7-jet model. The "Fit" line depicts $l_{\text{max}}$ . . . . .	128
5.2	For anti- $k_T$ jets of radius $R = 0.7$ at CMS, the fraction $\rho$ of jet $p_T$ residing in an annulus of radius $r$ from the jet's centroid (with inner and outer radius denoted by the horizontal error bar). Results are shown for jets with (a) low energy and (b) high energy. . . . .	130
5.3	Particles splitting into jets. (a) A single $a \rightarrow b+c$ splitting, depicting the rotation $\phi$ of the splitting plane. (b) A binary splitting tree with three nodes and four "leaves", one for each jet. . . . .	140
5.4	A next-to-leading order depiction of (a) 2-jet and (b) 3-jet events. In power jets, hard QCD radiation (partons splitting) is modeled by distinct prongs, which are given extensive shape by soft QCD radiation. . . . .	147
5.5	The power jet fit for the 2-jet-like $q\bar{q}g$ event previously seen in Figure 4.8, with (red, thin) $H_l^{\text{obs}}$ and (blue, thick) $H_l^{\text{jet}}$ ; (a) the "refined" 3-prong fit ( $l_{\text{max}} = 10$ ) and (b) the final 4-prong fit ( $l_{\text{max}} = 36$ ). . . . .	149

5.6	The power jet fit for the 3-jet-like $q\bar{q}g$ event previously seen in Figure 4.9, with (red, thin) $H_l^{\text{obs}}$ and (blue, thick) $H_l^{\text{jet}}$ ; (a) the “refined” 3-prong fit ( $l_{\text{max}} = 10$ ) and (b) the final 6-prong fit ( $l_{\text{max}} = 36$ ). . . . .	150
5.7	The boundaries of (a) $k_T$ and (b) anti- $k_T$ jets for the same original particles. The irregular boundaries of $k_T$ jets depend on the fine details of soft QCD radiation, while the “conical” boundaries of anti- $k_T$ jets do not. . . . .	151
5.8	The best power jets fit of a $q\bar{q}g$ jet from twenty trials, each with a different, random isotropic pileup ( $f_{\text{PU}} = 1\%$ ), with (red, thin) $H_l^{\text{obs}}$ and (blue, thick) $H_l^{\text{jet}}$ ; (a) 2-jet-like and (b) 3-jet-like event. . . . .	153
5.9	The $H_l$ fit of the 2-jet-like event, with (red, thin) $H_l^{\text{obs}}$ and (blue, thick) $H_l^{\text{jet}}$ , and a kinematic depiction of the fit. Random pileup is added until (a) $S/N = 1$ and (b) $S/N = 1/5$ . . . . .	158
5.10	The reconstructed kinematics (blue, solid) of the 2-jet-like event, versus (red, dashed) the three original partons, for power jets and anti- $k_T$ with random pileup; (a) $S/N = 1$ and (b) $S/N = 1/5$ . . . . .	158
5.11	The $H_l$ fit of the 3-jet-like event, with (red, thin) $H_l^{\text{obs}}$ and (blue, thick) $H_l^{\text{jet}}$ , and a kinematic depiction of the fit. Random pileup is added until (a) $S/N = 1$ and (b) $S/N = 1/5$ . . . . .	159
5.12	The reconstructed kinematics (blue, solid) of the 3-jet-like event, (red, dashed) versus the three original partons, for power jets and anti- $k_T$ with random pileup; (a) $S/N = 1$ and (b) $S/N = 1/5$ . . . . .	159
A.1	The $\lambda = 1$ exponential distribution $f(x) = e^{-x}$ ; (a) the quantile function $Q_1$ (solid) and its condition number (dashed) and (b) the “quantile flip-flop” — in the domain $0 < u \leq 1/2$ , each $Q$ maps out half of $f$ ’s sample space while remaining well-conditioned. . . . .	170
A.2	A visual depiction (using floats with $P = 4$ for clarity) of each possible $u$ for (E) even $\{U_{\text{E}}[0, 1]\}$ and (N) uneven $\{U_{\text{N}}[0, 1]\}$ . The height of each tic indicates its relative probability, which is proportional to the width of the number-line segment which rounds to it. . . . .	173
A.3	The bits of precision <i>lost</i> ( $D_{\text{KL}}$ ) when sampling the $\lambda = 1$ exponential distribution via (▼) GNU, (▲) GNU + <code>log1p</code> , and (○) our robust inversion method. The median ( $u = 1/2$ ) bisects the sample-space into two tails, with improbable values near the edges. The sampled variate $x = Q(u)$ is shown on the top axis. . . . .	178



A.4	Recursive $\bar{h}_l$ for the Gaussian shape functions (Eq. 4.63). (a) $\bar{h}_l$ for three smearing angles, connected by lines. (b) $R_l$ (as discrete points) for the same angles, with the prediction of Equation A.81 plotted with lines. . . . .	192
B.1	A Voronoi diagram for a flat square filled with seven points. . . .	195
B.2	The $h(\tilde{f})$ distribution for an $N = 16,384$ sample of isotropic Voronoi vectors. The shape is fit via Equation B.18 to determine the empirical Voronoi constant. . . . .	201

## NATURAL UNITS

This thesis uses the natural unit convention of particle physics; energy, momentum, mass, length and time can all be written with units of *energy* (specifically gigaelectronvolts, or GeV). This convention keeps constants from cluttering up equations. To convert to a non-energy unit, use dimensional analysis to figure out how many powers of  $c$  or  $\hbar c$  are needed (e.g., distance [in meters] =  $\hbar c/\text{energy}$ ).

## LIST OF SYMBOLS

Symbol	Definition
1.234(56)	Uncertainty notation: The parenthesis denote the uncertainty in the last few digits. In this case, $1.234(56) \mapsto 1.234 \pm 0.056$ .
eV	An electronvolt: the amount of kinetic energy an electron acquires after accelerating through a 1 volt potential; $1 \text{ MeV} = 10^6 \text{ eV}$ ; $1 \text{ GeV} = 1000 \text{ MeV}$ ; $1 \text{ TeV} = 1000 \text{ GeV}$ .
$\hbar c$	Planck's reduced constant times the speed of light in vacuum: $\hbar c = 197.326\,9718(12) \text{ MeV fm}$ [1]; $c = 299,792,458 \text{ m/s}$ [1].
$\vec{v}$ , $v$ , $\hat{v}$ , $v^i$	A 3-vector, its length, unit direction, and $i^{\text{th}}$ component ( $v =  \vec{v} $ and $\hat{v} = \vec{v}/v$ ). Roman indices run over $i = 1-3$ .
$\mathbf{M}$ , $M_{ij}$	A matrix (upper-case) and one of its elements.
$a_i b_i \mapsto \sum_i a_i b_i$	Einstein summation notation: repeated indices indicate implicit summation (unless otherwise stated).
$E$ , $\vec{p}$	A particle's energy and 3-momentum.
$p_L$ , $p_T$	The length of $\vec{p}$ along the longitudinal (L) axis (parallel to the colliding beams) and in the transverse (T) plane.
$\not{p}_T$	Missing transverse momentum: $\not{p}_T = - \sum \vec{p}_T $ .
$\vec{\beta}$ , $\beta$	Dimensionless 3-velocity: $\vec{\beta} = \vec{p}/E$ and $\beta =  \vec{\beta}  = v/c$ .
$\gamma$	Lorentz boost factor: $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ and $E = \gamma m$ .
$\mathbf{p}$ , $p^\mu$	A 4-vector (lower-case) and its $\mu^{\text{th}}$ component. Greek indices run over $\mu = 0-3$ . $\mathbf{x} = [ct, \vec{x}]$ ; $\mathbf{p} = [E, \vec{p}] = \gamma m[1, \vec{\beta}]$ .
$g_{\mu\nu}$	The metric of flat space: $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ .
$\mathbf{p}_1 \cdot \mathbf{p}_2$	4-vector contraction: $\mathbf{p}_1 \cdot \mathbf{p}_2 \equiv g_{\mu\nu} p_1^\mu p_2^\nu = E_1 E_2 - \vec{p}_1 \cdot \vec{p}_2$ .
$\mathbf{p}^2$	Self-contraction (mass squared): $\mathbf{p}^2 \equiv \mathbf{p} \cdot \mathbf{p} = m^2$ .

## ABSTRACT

As particle physics continues to expand into the high-energy and high-luminosity frontiers, it is encountering event topologies with extreme boosts and intense pileup. This creates unique challenges that limit our ability to use QCD jets to find new physics and conduct precision tests of the standard model. In this thesis, I present two tools that greatly expand our ability to use jets for these important purposes: (i) The  $\mu_x$  boosted-bottom jet tag, whose  $\mathcal{O}(100)$  signal to background ratio does not falter as jet  $p_T$  exceeds 1 TeV, and which is robust to pileup due to its foundation in boosted kinematics. (ii) Power jets, the first stage in a larger program to harness the power spectrum of QCD radiation to better utilize the vast amount of information collected about each collider event. Using the full power spectrum of a detected event, the power jets framework not only provides an accurate and precise recovery of jet kinematics, but also naturally facilitates a *global fit* to pileup intensity (rather than a local subtraction of pileup energy, which inadvertently strips soft QCD that belongs to the hard scatter).

## CHAPTER 1

### INTRODUCTION

2016 was the first year the Large Hadron Collider (LHC) achieved its design luminosity:  $\sim 0.5 \text{ fb}^{-1}/\text{day}$  at a collision energy of 13 TeV (92% of its design energy).<sup>1</sup> Over the course of June and July, ATLAS and CMS (the LHC’s general purpose detectors) each collected  $18 \text{ fb}^{-1}$  of data, five times more data than during the whole of 2015. This was more than enough to answer an important question: *Is there a neutral particle with a mass of about 750 GeV?* “No,” said the data, “the particle-like excess in the 2015 data, the one that generated all that excitement, was merely a statistical fluctuation [2, 3].” There was widespread disappointment, but little surprise.

This was a known problem; a truly exhaustive search for new physics is bound to uncover statistical anomalies (in fact, we should get worried if it doesn’t). Unfortunately, statistical fluctuations are merely one of the hurdles that high energy physics

---

<sup>1</sup>A brief primer on collider jargon. The “femtobarn” fb is a very tiny area

$$1 \text{ fb} = \underbrace{10^{-15}}_{\text{femto}} \underbrace{10^{-28} \text{ m}^2}_{\text{barn}} = 10^{-43} \text{ m}^2 \quad (1.1)$$

The historical development of scattering theory means we still express the probability of observing some phenomenon in terms of its “cross section”  $\sigma$  (an area).

This makes sense if we imagine a collider consisting of two shotguns aimed at each other. Pulling the triggers simultaneously, the probability that two of the pellets collide depends on their intrinsic cross section  $\sigma$ , but also on the collider’s “luminosity”  $L$  (how good it is at manufacturing collisions).  $L$  depends on the number of pellets per shell, their density/spread upon leaving the barrel, and the number of shells per second. The integrated luminosity  $L_{\text{int}} = \int L dt$  tabulates the amount of data collected in terms of an *inverse* area, permitting a trivial calculation of the number of collisions expected in the data set

$$\text{Ex (events)} = \sigma(\text{fb})L_{\text{int}}(\text{fb}^{-1}). \quad (1.2)$$

will face in the next few decades. The search for new phenomena inevitably forces us into unexplored corners of parameter space — more energy per collision, more collisions per bunch crossing, and more bunch crossings per second. This environment is already taxing our ability to determine when *jets* (conical sprays of particles) originate from top or bottom quarks, and may even hinder our ability to see jets at all. We will need better tools, but we may also need an entirely new language for high energy physics; today’s jet may be tomorrow’s plum pudding. The future is exciting!

### 1.1 Answering the challenge

The challenge of particle physics is to answer the question: **“What is matter, and how does it work?”** The answer is not quite so laconic, and has led to the construction of (arguably) the most complicated machine that humanity has ever built, the Large Hadron Collider (LHC). Experiments at the LHC use the standard particle physics formula for studying matter; manufacture an “event” by smashing two well-known particles together (two protons at the LHC), creating some interesting intermediate state  $X$ , then study that intermediate state via the debris it decays into

$$pp \rightarrow X \rightarrow \text{particles} . \tag{1.3}$$

With enough energy and enough events, it should be possible to determine all the constituents of matter and map out all the forces with which they interact.

This formula has taken us from plum pudding, at the atomic length scale of one ångström ( $\text{\AA} = 10^{-10} \text{ m}$ ), to the unprecedented length scales probed by LHC collisions. In fact, if we convert the maximum usable energy of the LHC to a distance ( $\sim 4 \text{ TeV} \mapsto 5 \times 10^{-10} \text{\AA}$ ) we find that particle physics has drilled down to an *atom’s* ångström. Yet while coming this far has required answering “the question” in great detail, the answer remains unsatisfactory. We cannot explain why a top quark is so heavy, whereas an electron is quite light, nor why neutrinos have any mass at all (and

especially such an infinitesimal one). Nor can we explain why 94% of the universe appears to be made of stuff we’ve never seen before.

The particle physics approach to such conundrums is to assume that our best approximation of Nature’s laws, the Standard Model (SM), is incomplete. There is some new force which operates at some higher energy scale, and the SM is simply its ground state. To suss out this new force, particle physics uses our time-tested, two part strategy: (i) When you think you understand what you’re seeing, crank up the energy scale of the interactions to unlock new phenomena. (ii) Take data at this new energy long enough to get good statistics and resolve the fine details of what you’re seeing. The LHC has completed the first step by maxing out its energy, and is now prepared to take data for the next 20+ years. To expedite this process, in 2023 the LHC will undertake its high luminosity upgrade (HL-LHC), enhancing the collisions per second by an order of magnitude. Hopefully, this data will contain new physics beyond the Standard Model (BSM), so that we can obtain the hints we need to develop the next model. However, extracting information from this high-energy and high-luminosity frontier will not be easy.

**1.1.1 The high-energy frontier.** Since the LHC smashes protons together, which are built from colored particles (quarks and gluons), most of the physics at the LHC is produced by colored particles interacting via the strong force. And since these interactions are quite energetic, we can describe them via the perturbative theory of the strong force, quantum chromodynamics (QCD). Yet these interacting colored particles are never actually seen. Instead, we see jets — shotgun blasts of long-lived particles from the “decay” of a colored particle.<sup>2</sup>

---

Jets are one of the most important ways to search for new physics, and one of

<sup>2</sup> This over-simplified description of jets will be made more rigorous in Section 1.2.1.

the most promising channels in which to look for BSM phenomena is the decay of a new, massive particle into top and bottom quark jets. The benefit of this channel is that heavy-flavored jets ( $t$  and  $b$ ) have *relatively* minimal sources of background — SM processes that mimic the BSM signal — compared to light-flavored jets (down, up, strange, charm or gluon, or *du<sub>s</sub>c<sub>g</sub>*). Additionally, their heavy quark masses give them unique decay signatures. And since top quarks decay to bottom quarks nearly 100% of the time ( $t \rightarrow bW^\pm$ ), reliably identifying (or “tagging”) jets that originated from bottom quarks is an absolutely crucial component of searches for new physics.

Unfortunately, current methods for tagging  $b$  jets lose efficiency and experience a large rise in fake rate (light jets posing as  $b$  jets) as jet energies approach 1 TeV. This causes a massive decline in the signal-to-noise ratio, which severely fogs an important lens for discovery on the high-energy frontier. This led our group to develop a new technique for tagging heavy-flavor jets with  $p_T > 500$  GeV, called the “ $\mu_x$ -boosted-bottom jet tag” [4], whose development we will cover in Chapter 2. The  $\mu_x$  tag probes the angular and energy substructure of a jet to look for muons coming from boosted  $B$  hadron decays. The main virtue of the  $\mu_x$  tag is that its efficiency  $\epsilon_b = 14\%$  and fake rate  $\epsilon_{\text{light}} = 0.1\%$  are effectively insensitive to jet  $p_T$ , leading to a relatively constant signal-to-noise ratio across the TeV scale [4].

In Chapter 3, we explore how the  $\mu_x$  tag can be used to extend the LHC’s reach to find or exclude two hypothetical extensions to the Standard Model, each of which manifests as a very heavy  $m = \mathcal{O}(\text{TeV})$  particle — a leptophobic  $Z'$  boson or a charged Higgs boson  $H^\pm$ . Each of these BSM theories is an important candidate for new physics, and both vector bosons are frequently the subject of new physics searches at the LHC. Yet the LHC experiments have trouble drawing firm conclusions for  $m \gtrsim \mathcal{O}(1 \text{ TeV})$  because their  $b$  tags lose their power. Our results clearly demonstrate that judicious application of the  $\mu_x$  boosted  $b$  tag will help dismantle this arbitrary



wall [4, 5], proving that the  $\mu_x$  tag is an important, general tool sorely needed at the high energy frontier.

**1.1.2 The high-luminosity frontier.** One of the major sources of systematic error at the LHC is pileup: dozens of boring, low-energy events superimposed upon the one interesting, high-energy event. Unfortunately, the pileup problem is only going to get worse at the HL-LHC (dozens  $\rightarrow$  hundred), complicating the observations of subtle, low-energy phenomena. However, a detector filled with pileup — with so many particles that they are difficult to disentangle, forming a nearly continuous surface of event activity — also creates a bounty of correlated observables. A global analysis of these correlations might be the solution to the pileup problem.

Global analyses of extremely busy collider events are already the status quo for heavy-ion collisions, where the final state often resembles a shapeless blob. Studying inter-particle correlations within the blob is one of the only ways to extract useful information from it. In Pb-Pb collisions at the LHC, pairs of final-state particles are seen correlated at opposite azimuthal angles  $\phi$  ( $\Delta\phi = \pi$ , indicating a  $p_T$ -preserving common origin), but also on the *same side* of the beam ( $\Delta\phi = 0$ ). This same-side correlation is often attributed to a quark-gluon plasma that forms when the two heavy nuclei collide.

Curiously, the same-side correlation is also evident in proton-proton collisions [6], in which there is simply no time for quark-gluon plasma to form, suggesting an alternative mechanism. The data also requires that this mechanism is somehow related to event multiplicity (the number of particles in the final state), since the same-side correlation is only evident in high-multiplicity  $pp$  events. The emergence of the bump is clearly evident in Figure 1.1a–d, where the number of charged particles increases from around 20 to more than 120. Is it possible that the same-side correlations seen in heavy-ion collisions is merely the sum of many low-temperature, jet-like

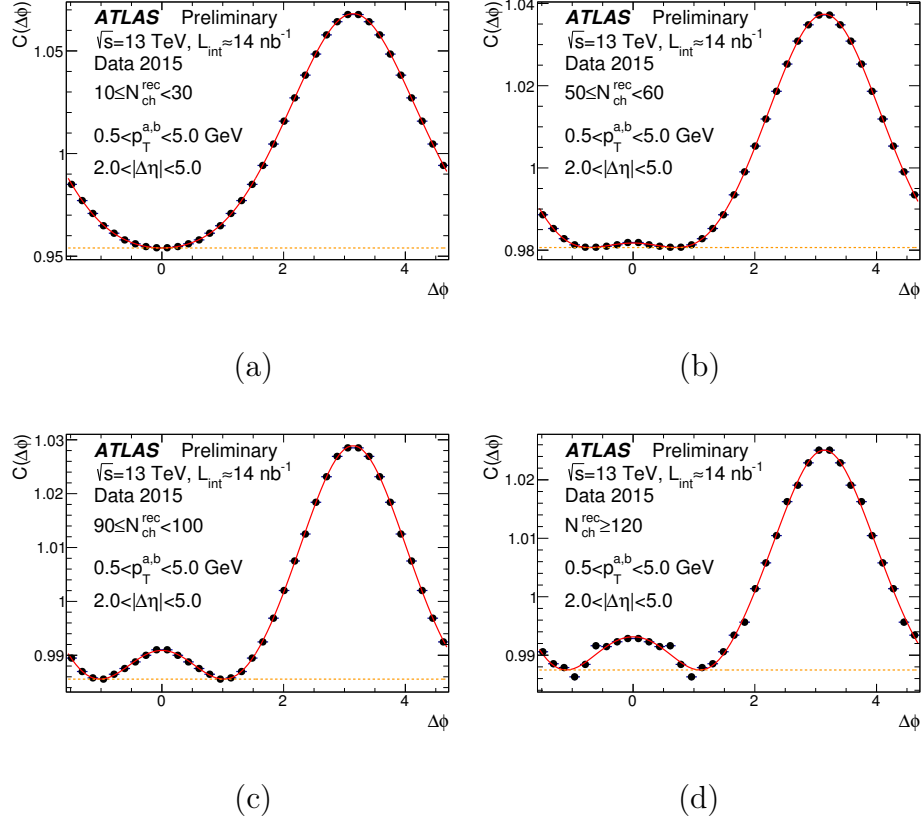


Figure 1.1. Pairs of particles are only correlated on the same-side ( $\Delta\phi = 0$ ) of the beam in *high-multiplicity* proton-proton collision (i.e., a large number of reconstructed charged particles  $N_{\text{ch}}^{\text{reco}}$ ) [6].

correlations? Or perhaps its multiplicity dependence indicates some meso-scale in between the collision of two nucleons ( $pp$ ) and hundreds of nucleons (PbPb); if this were the case, then characterizing this meso-scale would undoubtedly unlock new and important truths about nature. In either case, understanding why we see this effect in  $pp$  events will clearly require an approach that better utilizes the many inter-particle correlations.

This point is further emphasized in in Section 1.4, where we introduce the standard methodology for reconstructing  $pp$  collisions; given a detector brimming with correlated information, only one correlation is used at any given time. Why not use all simultaneously? This is the approach we adopt in Chapter 4, where we

begin to analyze QCD radiation using its angular power spectrum (a tool familiar to cosmologists in their study of the cosmic microwave background). This harks back to some of the original methods that theorists used to study QCD; methods which fell by the wayside partially because event multiplicity was too low.

As we venture down this path of using global correlations, we will see that QCD radiation is dominated by jets and their substructure. In Chapter 5, we will introduce power jets — a novel jet definition which uses all the information in the event to define jets without borders. Not only do power jets offer a new and exciting way to study QCD, they provide a natural framework to separate pileup from signal, even when the signal-to-noise ratio is as low as  $S/N = 1/5$ . With these preliminary successes, I anticipate that the angular power spectrum has a lot to offer to high energy proton physics, and I hope to take a leading role in these investigations.

The remainder of this chapter will lay the foundation for my research, introducing the basic theoretical and experimental underpinnings of high energy particle physics.

## 1.2 The collider

The frontier of high energy physics is probed at a particle collider, where two beams of particles are accelerated to ultrarelativistic speeds and smashed together. When a particle from the forward beam ( $p_1^\mu = E [1, 0, 0, +1]$ ) collides with a particle from the backward beam ( $p_2^\mu = E [1, 0, 0, -1]$ ), the resulting “event” has a very large invariant mass<sup>3</sup>

$$\sqrt{s} \equiv \sqrt{(\mathbf{p}_1 + \mathbf{p}_2)^2} = 2E. \quad (1.4)$$

This unlocks a wide range of physics in the event’s final state, with each particular

---

<sup>3</sup> While the beam particles have mass, they are ultrarelativistic, and therefore effectively massless. We depict them here as identically massless ( $\beta = 1$ ).

option possessing a finite cross section  $\sigma$ . New and exotic events are necessarily rare, with tiny cross section  $\sigma$ , so most events seen by the collider are not interesting; this is the *background* — formerly exciting physics that is now well understood (and thus mundane). One of the main tasks of phenomenology is to design clever tricks to reduce this background.

To “see” the final state, a nearly hermetic particle detector (trapping  $4\pi$  of solid angle, minus two holes for the beams) is wrapped around the collision point. A spherical-polar coordinate system is a natural choice for such a detector, with the colliding beams defining the longitudinal axis:

$$\vec{r} = [r_T, r_L] = r [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta]. \quad (1.5)$$

The polar angle  $\theta$  is relative to the forward beam, and the azimuthal angle  $\phi$  parameterizes the transverse plane. This is a natural system because, if the beams are unpolarized, all cross sections must have azimuthal symmetry. Furthermore, in the center-of-momentum (CM) frame of the collision, the simplest class of event (a  $2 \rightarrow 2$  scattering, for which the final state particles are back-to-back) has a cross section that depends only on the scattering angle  $\theta$  and  $\sqrt{s}$ . Given the head-on nature of the beams, the collision has no transverse momentum (to first approximation); hence,  $\sum \vec{p}_T = 0$  for the final state. This allows “missing” transverse momentum,

$$\cancel{p}_T = - \left| \sum \vec{p}_T \right|, \quad (1.6)$$

to serve as a proxy for particles that the detector cannot see (such as neutrinos).

At lepton colliders, an event’s initial state is simply the two beam particles (e.g., an electron and a positron annihilating to a quark-antiquark pair,  $e^-e^+ \rightarrow q\bar{q}$ ). If the lepton beams have approximately symmetric energies, the CM frame is very close to the lab frame. But the current high energy frontier occurs at the LHC, a

proton collider. Here, it's not  $pp \rightarrow q\bar{q}$ , but  $u\bar{u} \rightarrow q\bar{q}$  or  $gg \rightarrow q\bar{q}$ .<sup>4</sup> The initial state is two *partons*, two constituents of the proton (a quark or gluon). And even though the momentum  $\mathbf{P}$  of each proton is well known, the parton carries only some fraction  $x$ :

$$\mathbf{p}_1 = x_1 \mathbf{P}_1 \quad \text{for} \quad 0 \leq x_1 \leq 1. \quad (1.7)$$

We can therefore calculate

$$\mathbf{p}_{\text{cm}} = \mathbf{p}_1 + \mathbf{p}_2 = E [x_1 + x_2, 0, 0, x_1 - x_2], \quad (1.8)$$

$$\sqrt{s} = 2E\sqrt{x_1 x_2}. \quad (1.9)$$

These equations tell us that each individual proton-proton event has a unique invariant mass  $\sqrt{s}$ . For this reason, we will use  $S = (\mathbf{P}_1 + \mathbf{P}_2)^2$  to denote the steady invariant mass of two beam particles and  $s = (\mathbf{p}_1 + \mathbf{p}_2)^2$  to denote the unique invariant mass of each hard scatter.<sup>5</sup> A side effect of this unique  $\sqrt{s}$  is that the CM frame of each event is longitudinally boosted relative to the lab frame by factor of

$$\gamma_L = \frac{x_1 + x_2}{2\sqrt{x_1 x_2}}. \quad (1.10)$$

While  $\gamma_L$  is observable in principle, this requires detecting *every* final-state particle from the hard scatter; if any escape through the beam holes, or there are any neutrinos, or any other signals overlapping the hard scatter, then  $\gamma_L$  is lost.

Without a means to reliably boost the lab frame into the CM frame at a proton collider, spherical-polar coordinates cannot reliably map experiment onto theory. Of

---

<sup>4</sup> When colliding protons ( $uud$ ), how can the initial state contain an anti-up quark? Because if you probe a proton with more than enough energy to create a  $u\bar{u}$  pair inside it, you will find  $\bar{u}$ . It's Heisenberg's microscope with second quantization. Similarly, the proton contains lots of gluons, especially low-energy ones.

<sup>5</sup> In HEP slang, "hard" means energetic and "soft" means "not energetic relative to the hardest observed scale". Colliders cannot accelerate single particles, only "bunches" containing billions, so there can be dozens of collisions per "bunch crossing" (two bunches meeting head-on). Most collisions in a bunch crossing are soft, with the "hard scatter" being the hardest (and therefore most interesting).

course, transverse momentum and azimuthal symmetry are unaffected by  $\gamma_L$ , so  $p_T$  and  $\phi$  are still good CM variables. But  $\theta$  is heavily transformed by  $\gamma_L$ , and requires replacement. The natural choice is rapidity

$$y = \frac{1}{2} \log \left( \frac{E + p_L}{E - p_L} \right) = \operatorname{arctanh}(\beta_L), \quad (1.11)$$

since the  $\Delta y$  between two particles is invariant to longitudinal boosts. Thus, as long as one can express final-state observables in terms of *pairs* of particles, those observables can still be measured even when  $\gamma_L$  is unknown.

Unfortunately, rapidity is not fully geometric; it depends on mass. As such, components in the detector cannot be assigned a unique rapidity. Luckily, ultrarelativistic particles ( $\beta \gtrsim 0.99$ ,  $\gamma \gtrsim 7$ ) are effectively massless ( $E \rightarrow |\vec{p}|$ ), and when  $\beta \rightarrow 1$ , rapidity becomes pseudorapidity, which has a geometric interpretation

$$\eta = \frac{1}{2} \log \left( \frac{|\vec{p}| + p_L}{|\vec{p}| - p_L} \right) = -\log \left[ \tan \left( \frac{\theta}{2} \right) \right]. \quad (1.12)$$

Since most final-state particles are indeed ultrarelativistic, individual elements inside the detector can be mapped out in  $\eta$ .

**1.2.1 What is a jet?** QCD is the quantum gauge theory of the strong force, which exhibits SU(3) symmetry. The strong force acts between particles carrying SU(3) “color” charge, of which there are three classes: singlets (no color charge, such as photons, electrons, and neutrinos), triplets (quarks, the building blocks of nucleons) and octets (gluons, the mediator of the strong force). All stable/observable particles are color singlets, because the strong force is *so strong* that naked color charge will polarize the vacuum until it neutralizes itself. As such, colored particles are found only in bound states like  $\pi^0$ ,  $K^-$ , and protons. This self-neutralizing property is called color confinement.

Because of color confinement, physicists cannot study QCD directly through colored particles; it can only be examined by mapping the behavior of the un-colored

matter which emerges from particle collisions, then inferring the interactions which occurred within the black box of confinement. Yet color confinement doesn't happen immediately; an energetic quark is born free and naked. But as it propagates, its color charge creates a cloud of gluons (which can then split to quark/anti-quark pairs). This shows up in electromagnetism as well; a cloud of virtual photons surrounds an electron, and the virtual  $e^+e^-$  pairs screen/diminish the apparent charge of the electron.

But photons aren't charged under the force they mediate, whereas gluons are. Thus, the color charge of the gluon cloud *increases* the color charge of the bare quark; the cloud precipitates its own growth. Boosting into the lab frame, this growing cloud wrapped around a speeding quark resembles the original quark radiating colored particles at relatively shallow angles. This process is called QCD showering. Eventually, confinement takes hold, and the colored particles begin to organize into color singlets (hadrons built from quarks and bound by gluons). This is called fragmentation/hadronization. What starts as a naked quark is detected as a collimated cone of hadrons and their decay products — this is called a jet. In this work, jet formation via QCD showering and fragmentation is modeled by PYTHIA 8 [7, 8], a Monte Carlo program that stitches the perturbative and non-perturbative aspects of jet formation into a very useful tool.

Jets are important because, in this grossly simplified picture of how a quark converts into a jet, the four-momentum  $\mathbf{p}$  of the jet is a proxy for the  $\mathbf{p}$  of the quark. This is called jet-parton duality. If some previously unseen heavy  $Z'$  particle decays into a bottom and anti-bottom quark ( $pp \rightarrow Z' \rightarrow b\bar{b}$ ), detecting both bottom jets and measuring their  $\mathbf{p}$  permits the determination of the mass of the  $Z'$

$$m_{Z'} = \sqrt{(\mathbf{p}_3 + \mathbf{p}_4)^2} = \sqrt{E_3 E_4 - \vec{p}_3 \cdot \vec{p}_4}. \quad (1.13)$$

Jets are also the most direct way to study QCD interactions above the confinement

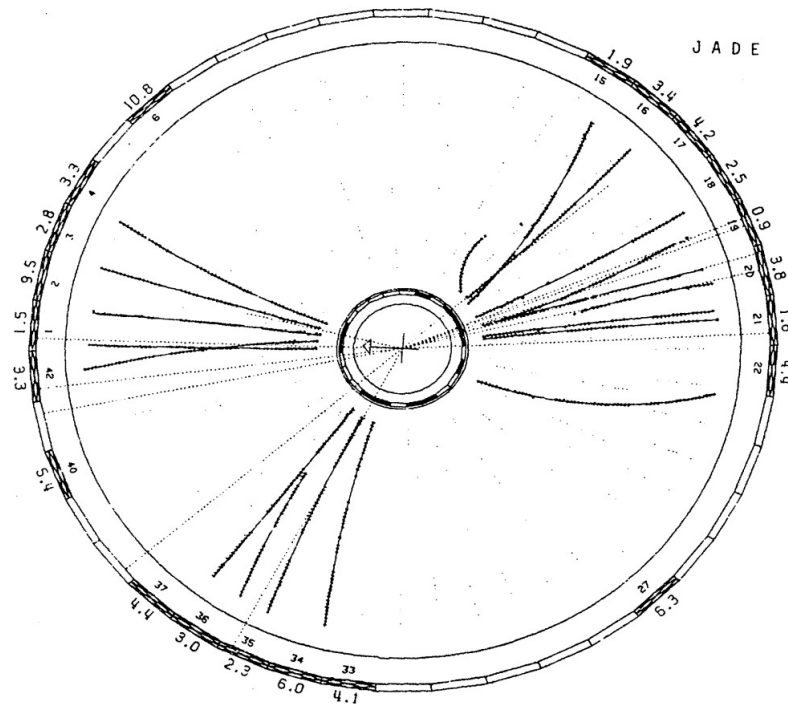


Figure 1.2. A 3-jet event at PETRA, circa 1979 (19 GeV  $e^+e^-$  collisions), as seen in the transverse plane. Solid curves are charged particle tracks; dotted lines are neutral particles (calorimeter towers not struck by tracks) [9].

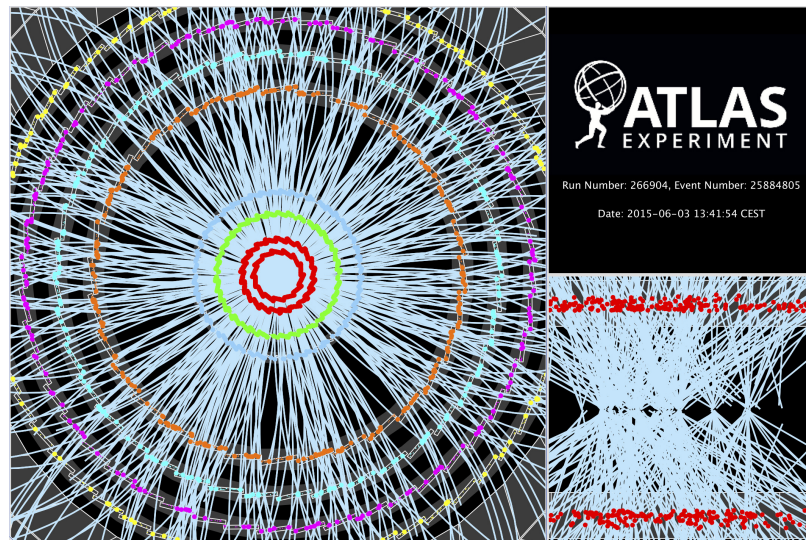


Figure 1.3. A multijet event at the LHC, circa 2015 (13 TeV  $pp$  collisions). Left view is the transverse plane, right view is rotated  $90^\circ$ . Cyan curves are reconstructed tracks; dots are tracker hits. There are at least 17 events (each emanating from its own vertex), but only one is useful for new physics (i.e., 16 pileup events) [10].



scale, since all energetic colored particles (including gluons) will form jets. In the early days of jet physics, jets could often be reconstructed by eye; Figure 1.2 was one of the first events that showed evidence of a gluon jet ( $e^+e^- \rightarrow q\bar{q}g$ , where the gluon was radiated by one of the quarks). Crank up the collision energy by 3 orders of magnitude and use a colored initial state (see Fig. 1.3), and the hard scatter becomes shrouded in the underlying event (which we will now define).

**1.2.2 The underlying event.** Creating an event requires a collider, but the violent process of smashing two beams together creates a number of side-effects that contaminate the event. These processes are collectively called the underlying event (UE).

One source of UE comes from the *charge* of the beam particles, which can radiate before the hard interaction. For lepton colliders, this initial-state radiation (ISR) is photons. For proton colliders, where the hard scatter steals one colored parton from each proton, the ISR *can* be photons but is more likely a colored parton (such as gluons). A second source of UE arises because interesting final states have an extremely small cross section. This requires a high-luminosity collider with tightly focused beams. This tight focusing creates soft, elastic scattering of beam particles called pileup (which at a proton collider also includes inelastic scattering). At the LHC’s design luminosity, each interesting event is accompanied by  $\mu \approx 40$  pileup collisions per beam crossing.<sup>6</sup> A final source of UE occurs only a proton collider: the hard interaction breaks two protons asunder, whose fragments then scatter (albeit mostly through the beam hole).

These three effects spray particles all over the detector, but the beam momenta tend to focus UE particles in the forward ( $+\hat{z}$ ) and backward ( $-\hat{z}$ ) directions (creating

---

<sup>6</sup>The pileup collisions are random and uncorrelated, but occur at a constant average rate that depends on the collider’s luminosity. Hence, the number of pileup collisions per bunch crossing *roughly* follows a Poisson distribution with mean  $\mu$ .

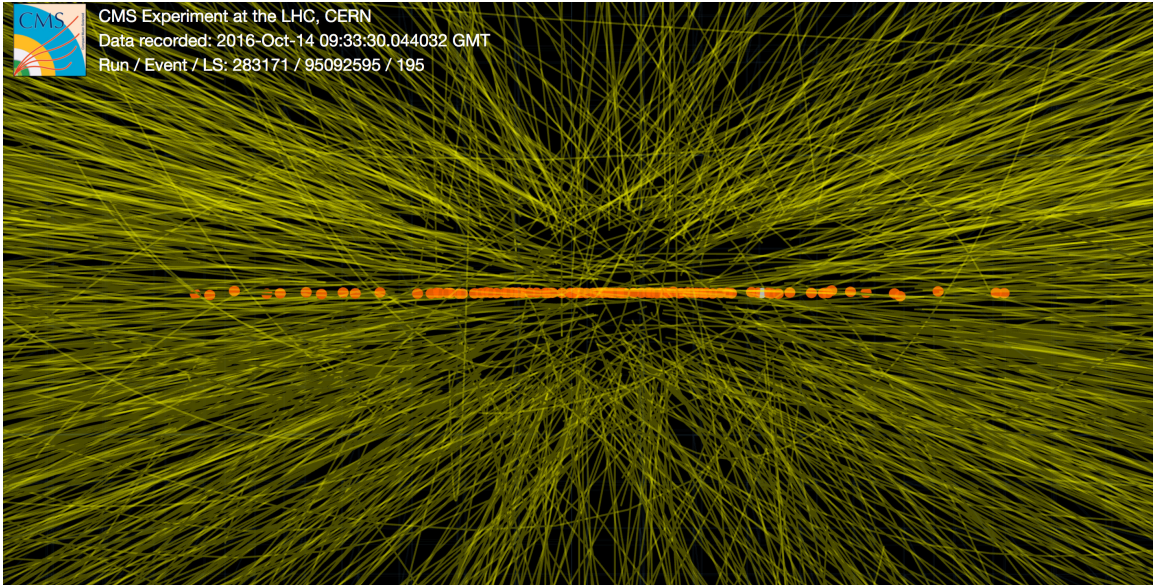


Figure 1.4. The reconstructed tracks (lines) and pileup collision vertices (dots) during a special high-pileup run at CMS, producing  $\mu = \mathcal{O}(100)$  pileup collisions [11].

what are known as the “beam” jets, which can surround the beams like a halo). However, a significant fraction of UE particles are created at wide angles to the beam, bathing the central regions of the detector with diffuse particles. This overlaps the signal of the hard event, adding a significant amount of noise to the hard jets, but also randomly collecting to create concentrated spikes resembling soft QCD jets (which the UE therefore fakes). Of these three effects, pileup is often the most problematic, and will be especially strong at the HL-LHC, where the number of pileup events per bunch crossing is expected to be  $\mu = \mathcal{O}(100)$ . Each bunch crossing will therefore resemble the tangled mess depicted in Figure 1.4 (which shows only the event’s charged components).

### 1.3 The detector

To study the Standard Model and the physics beyond, a collider must be wrapped in a detector that can “see” an entire event. This requires being able measure the energy and angles of the two known forms of matter — quarks (via jets) and

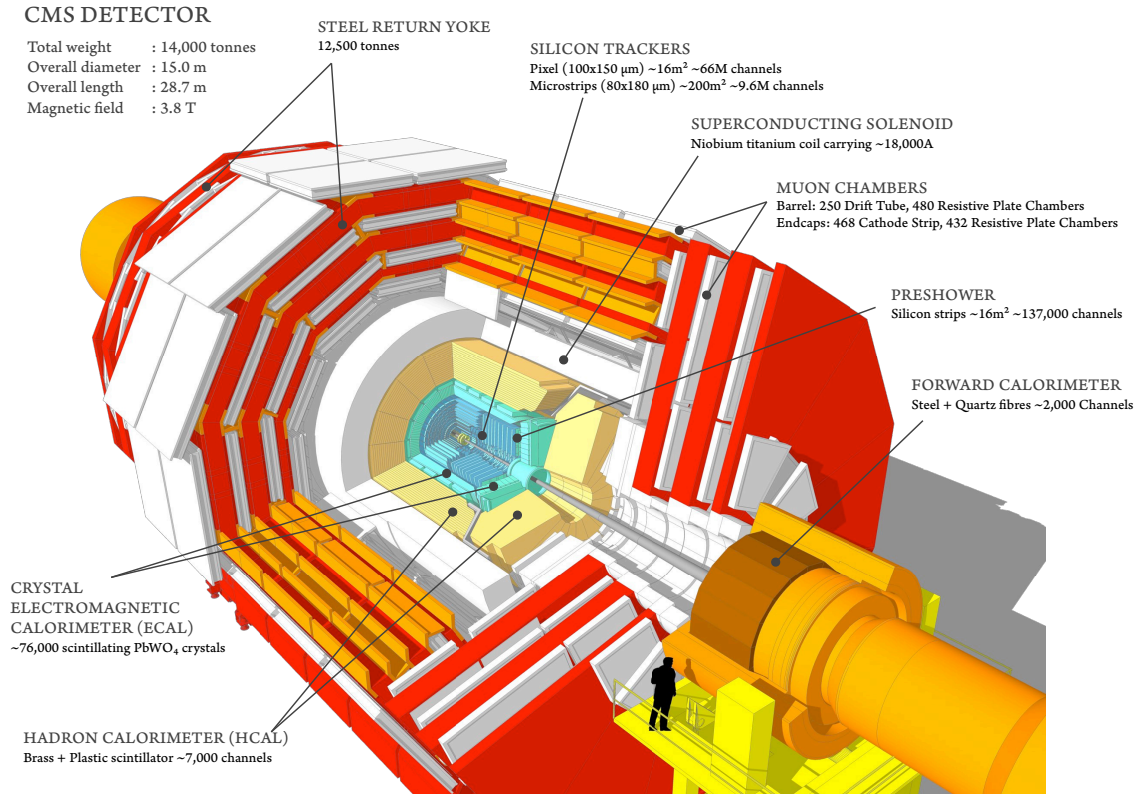


Figure 1.5. A cutaway of the CMS detector [12]. Most of its volume is the muon chamber.

leptons — and the one stable force carrier, the photon  $\gamma$ . The primary components of jets are: pions ( $\pi^\pm$ ,  $\pi^0$ ), kaons ( $K^\pm$ ,  $K^0$ ), and nucleons ( $p$ ,  $n$ ). Neutral pions are terribly unstable, and immediately decay to two photons ( $\pi^0 \rightarrow 2\gamma$ ) inside the jet. Various decays of unstable hadrons also sprinkle jets with charged leptons ( $e^\pm$ ,  $\mu^\pm$ ).

Figure 1.5 depicts one of the major detectors/experiments at the LHC, the Compact Muon Solenoid (CMS). CMS's sister detector, ATLAS, has essentially the same design, with some differences in implementation. Both detectors output two basic classes of physics objects — tracks and towers. It is generally not possible to detect the mass of either object; the best one can do is assume they are ultrarelativistic (which they generally are), and therefore effectively massless. As such, the reconstructed energy  $E = |\vec{p}|$  and direction of travel  $\hat{p}$  of each objects can be used to

Table 1.1. Jet constituents, sorted by electric charge and primary detection mode.

	Electromagnetic (ECal)	Hadronic (HCal)	Muon chamber
Charged	$e^\pm$	$\pi^\pm, K^\pm, p$	$\mu^\pm$
Neutral	$\gamma$	$K^0, n$	

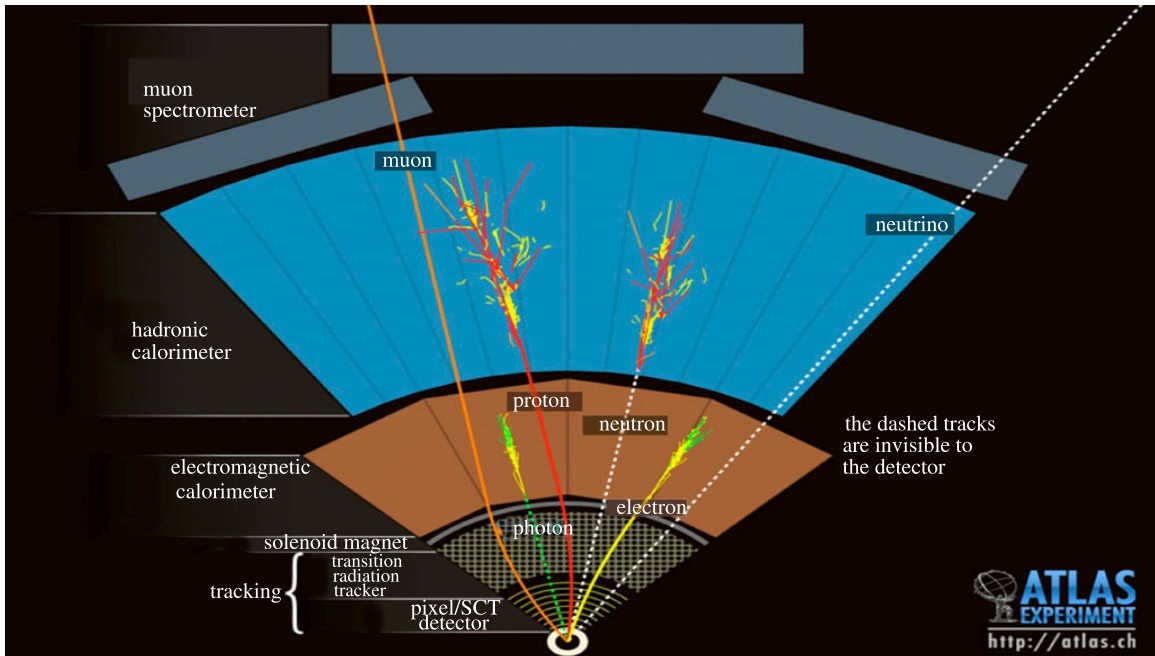


Figure 1.6. A transverse slice of ATLAS, depicting the propagation of various particles through the tracker, calorimeter, and muon system [13].

accurately approximate its four-momentum  $\mathbf{p} = E[1, \hat{p}]$ .

The first layer of the detector is the tracker — concentric shells of 2D pixels inside a uniform, longitudinal magnetic field (with strength  $B$ , measured in tesla). The field deflects charged particles in the transverse plane, so that they propagate away from the collision along helical paths, ionizing the pixels and leaving a helical trail of “hits” (in Fig. 1.6, the  $\vec{B}$  field is coming out of the page). Pattern finding software groups the hits into reconstructed tracks. Given the track’s charge  $q$  (in units of elementary charge) and radius of curvature  $R$  (in meters), its transverse

momentum (in GeV) is

$$p_T = 0.3 |q| R B \quad (1.14)$$

(where the numerical factor is the conversion from eV to GeV, times  $c$ ). Since “stable” particles have been extensively cataloged, and all have  $|q| = 1$ , the magnitude of  $q$  can be assumed, and its sign determined from the direction of the curvature. Given the  $p_T$  and the path, the track’s original four-momentum  $\mathbf{p}$  can be determined.

Neutral particles are not deflected by the magnetic field, nor do they ionize the tracker pixels; they pass straight through to the calorimeter. Both charged and neutral particles dump their kinetic energy into the calorimeter via particle showers (as depicted in Fig. 1.6). These showers activate scintillating material proportional to the energy deposited. The calorimeter is segmented into cells stacked into a deep tower radiating from the collision point (radial lines define the towers in Fig. 1.6).<sup>7</sup> The inner layers (ECal) are designed to capture particles interacting electromagnetically, while the outer layers (HCal) capture hadrons via nuclear interactions (see Table 1.1). Because particles shed their energy over multiple layers, only the total energy of the tower  $E_{\text{total}}$  contains a meaningful energy (although this can be split into its ECal and HCal components).

Since a tower’s angular sub-structure is indeterminate, the best one can do is orient its four-momentum at its geometric center. So unlike tracks, whose high-granularity pixel hits allow them to be reconstructed with nearly any angular position, tower centers can exist only at discrete points (often on a semi-regular grid). Figure 1.7 highlights this distinction, showing how tracks and towers are gathered into a single jet at CMS. Care must be taken when using tracks and towers simultaneously,

---

<sup>7</sup>In reality, a tower with clean radial boundaries is a crude description of the calorimeter; there are many overlapping absorbing plates and sensing units. A tower is a composite object that stitches these volumes together into a particle shower that cannot be resolved at any smaller scale.



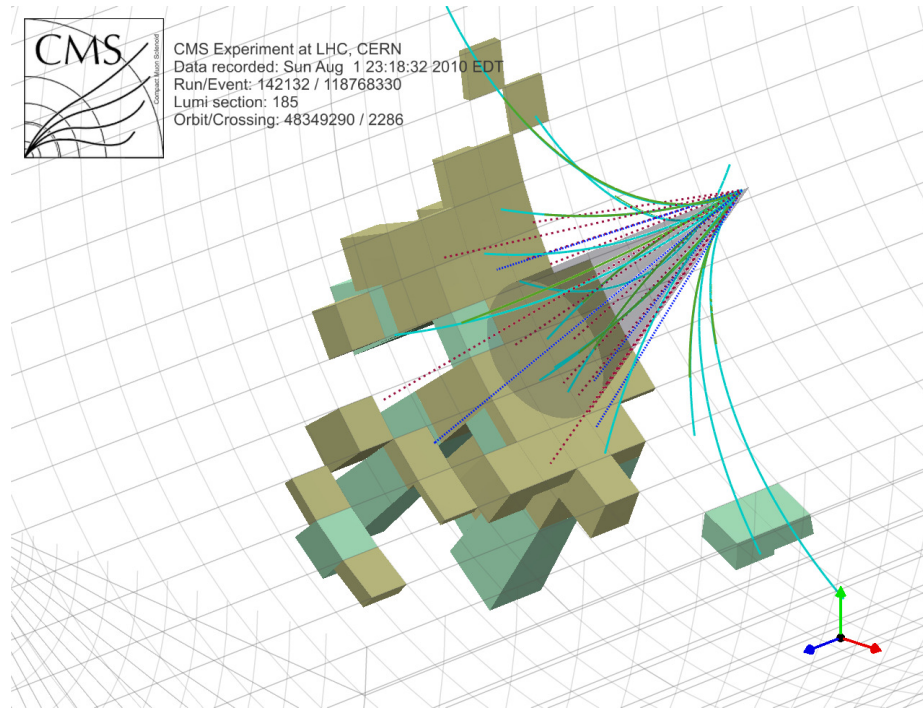


Figure 1.7. A jet inside CMS, with constituent tracks (solid curves) and towers (blocks, with dotted line depicting reconstructed neutral  $\mathbf{p}$ ). The energy of towers is depicted via their radial height, with ECal towers (brown) stacked before HCal towers (light green). The cone of the reconstructed jet is shaded [14].

since track energy is double counted (charged particles deposit their energy into the calorimeter; note the four outlying tracks in Figure 1.7). This can be rectified by extrapolating each track to the tower it ostensibly struck, then subtracting the track’s well-measured energy (via its  $p_T$ ) from the tower’s  $E_{\text{total}}$ . This “track-subtraction” technique creates a neutral tower, which should contain only the energy of neutral particles. This technique is a key component of CMS’s particle-flow algorithm [15], and significantly increases the multiplicity of physics objects *and* the angular resolution of each event’s energy.

The last layer of the detector system is the muon tracker. Muons interact with matter minimally, at least until their energy surpasses a few hundred GeV. Thus, most muons leave only small traces in the inner tracker and the calorimeter, and fully constraining their  $p_T$  requires a large secondary tracker after the calorimeter. This

annoyance comes with a fringe benefit; muons are about the only particle that make it through the calorimeter to the muon tracker, so their mere presence identifies them as muons. Conversely, electrons must be selected from the full set of inner tracks (which also includes protons,  $\pi^\pm$ , and  $K^\pm$ ), based upon certain electron-like properties that can be faked or obscured (either by photons or  $\pi^\pm/K^\pm$  that begin showering in the ECal). This makes electrons more difficult to reliably identify, especially inside a very energetic jet with dozens of overlapping tracks.

**1.3.1 Are towers effectively massless?** Calorimeter towers have a *much* coarser granularity than tracker pixels. From the perspective of the particles leaving the collision, towers are approximately rectangles of solid angle  $\Omega = (w) \times (aw)$ , with angular dimension  $w$  and aspect ratio  $0 < a \leq 1$ . It is therefore common for towers to collect multiple particles. Even if these constituent particles are each effectively massless, the tower acquires a mass from their angular spread. It is therefore important to verify that a tower's boost is large enough (i.e.,  $\gamma \gtrsim 7$ ) that its unobservable mass is indeed negligible. Setting  $m \rightarrow 0$  for the tower's constituents, and requiring  $w \ll 1$ ,

$$\gamma_{\text{tower}} \approx \frac{E_{\text{total}}}{\sqrt{(\sum \mathbf{p}_i)^2}} \approx \frac{E_{\text{total}}}{\sqrt{2E_i(\delta_{ij} - \cos \xi_{ij})E_j}} \approx \sqrt{\frac{2}{f_i \xi_{ij}^2 f_j}}, \quad (1.15)$$

where  $\xi_{ij}$  are the opening angles between constituents and  $f_i = E_i/E_{\text{total}}$  is each constituent's energy fraction. A pathologically small boost arises when both the opening angle  $\xi_{ij}^2$  and energy product  $f_i f_j$  are maximized, and the absolute worst case is two particles of equal energy at opposite corners, for which  $\gamma_{\text{tower}} \approx \frac{1}{w} \sqrt{\frac{4}{1+a^2}}$ . But this worst case is obviously rare, so what is  $\text{Ex}(\gamma_{\text{tower}})$ ?

Since Equation 1.15 already uses the small angle approximation, it is safe to treat the tower's spherical cap as a flat rectangle. We can then run a Monte Carlo simulation for  $N$  random particles using a toy model where (i) particles are distributed

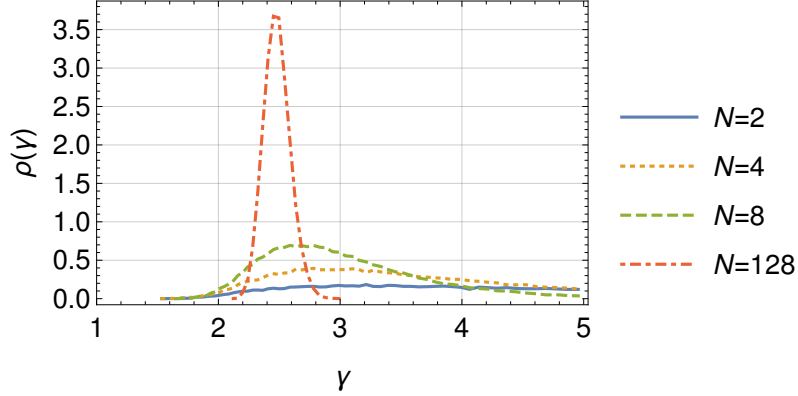


Figure 1.8. The probability distribution of calorimeter boost for a square tower ( $a = 1$ ), using 65,000 samples of  $N$  random particles uniformly distributed across the square, with energies drawn uniformly from the interval [1 GeV, 100 GeV].

uniformly across the tower and (ii) their energy distribution is uniform in *log*-space,

$$\rho(E) = \frac{1}{\log(E_{\max}) - \log(E_{\min})} \frac{1}{E}. \quad (1.16)$$

This choice of energy distribution prefers soft particles, with a bit of a hard tail.

Figure 1.8 depicts the results for a square tower. For small  $N$ , the distributions stretch far beyond the right edge of the graph. As  $N$  becomes large, the distribution peaks around the expected value  $\text{Ex}(\gamma_{\text{tower}}) = \frac{2}{w} \sqrt{\frac{3}{1+a^2}} \Big|_{a \rightarrow 1} \approx 2.45/w$  (using  $\text{Ex}(f) = 1/N$  and  $\text{Ex}(\xi^2) = \frac{w^2}{6}(1+a^2)$ ). Since the mean boost for samples with lower particle multiplicity is much larger than this large- $N$  expectation, it is useful for determining the maximum width of calorimeter towers which are effectively massless ( $\beta \geq 0.99, \gamma > 7$ ):

$$w_{\max} = \frac{2}{7} \sqrt{\frac{3}{1+a^2}} \Big|_{a \rightarrow 1} \approx 0.35. \quad (1.17)$$

The calorimeter towers at ATLAS and CMS fall well within this limit (ranging from  $w = 0.02$ – $0.2$ ), and it is therefore safe to treat them as effectively massless.

**1.3.2 Fast detector simulations.** To understand how their extremely complex detectors function as whole unit, the ATLAS and CMS collaborations conduct extensive simulation of individual components in their detector using the Geant4 toolkit [16–18].



This requires accurately simulating the passage of particles through  $\mathcal{O}(10^8)$  volume elements (one for each of the detector’s readout channels), which is often the only way to understand the detector’s systematic errors. To make theoretical predictions about what these LHC experiments should measure, a theorist must simulate their detectors. Not only is the Geant4 level of granularity computationally expensive, it is totally unnecessary for phenomenological studies; a theorist’s predictions should have a sound physical basis, and should not be sensitive to the small details of detector construction (which are not publicly available anyway).

The “fast detector” model takes the average efficiencies and errors published by the detector experiments, converts them into a simplified model (e.g., approximate the error curve with a Gaussian), then roughly simulates the various components of the detector. Our studies of the  $\mu_x$  boosted  $b$  tag in Chapters 2 and 3 use DELPHES 3.2 [19], a fast detector simulation developed by scientists with the CMS experiment. This simulation is fast and versatile, allowing for easy modification of the detector’s design, and is a common choice for phenomenological studies of the LHC. As a member of the DELPHES user community, I helped improve its code by discovering and reporting several bugs introduced when the main developers implemented new features. I additionally published the  $\mu_x$  tag as a DELPHES module [20], so that others scientists can easily implement  $\mu_x$  tagging in their own analyses.

The exploration of QCD power spectra in Chapters 4 and 5 is still in the proof of concept phase, where it is important not to conflate fundamental physics with detector effects. Hence, these studies either utilize no detector (using the exact  $\mathbf{p}$  of all visible particles, discarding neutrinos), or they use a bare-bones “pseudo-detector.” This pseudo-detector is designed to extract the maximum observable information from a collider event, and is constructed as follows: Tracks with  $p_T > 300$  MeV are perfectly detected; this  $p_T$  floor accounts for “loopers”, whose tiny radius of curvature

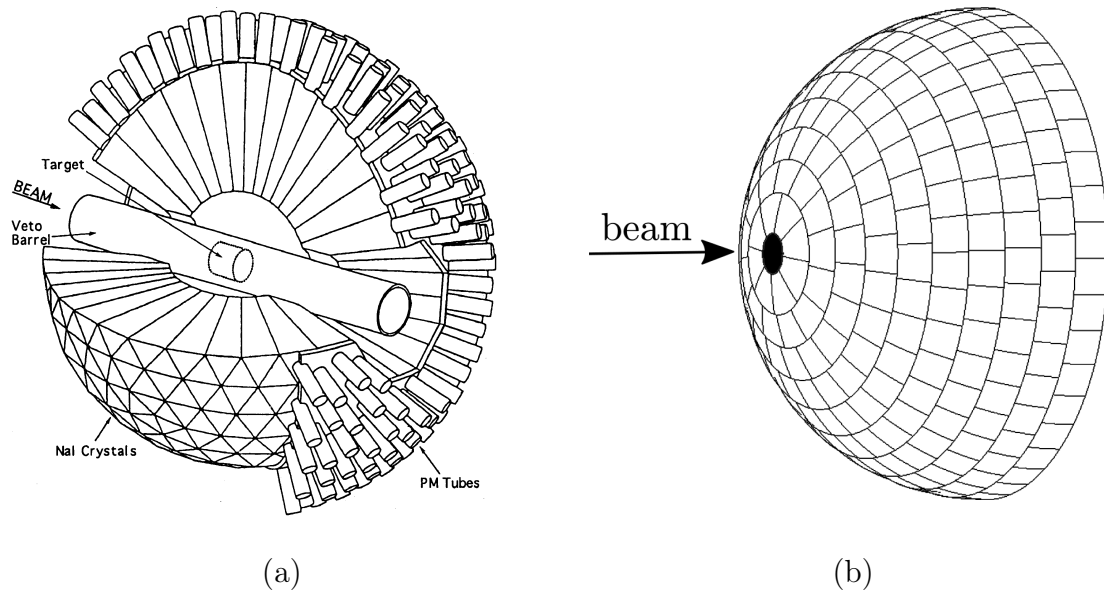


Figure 1.9. Two “equal area” calorimeters; (a) the “Crystal Ball” at SPEAR [21] and (b) a hemisphere resembling the square-cell pseudo-detector used by the author for basic phenomenological studies (figure modified from Ref. [22]).

$R$  ensures they can’t be seen in real life. The calorimeter is composed of square towers of approximately equal solid angle  $\Omega = w^2$ , formed into azimuthal belts as depicted in Figure 1.9b.<sup>8</sup> The calorimeter detects all neutral energy perfectly (except neutrinos), but none of the charged energy, simulating perfect track-subtracted neutral towers.

The pseudo-calorimeter does not resemble ATLAS or CMS; these experiments combine  $B$  field tracking with a cylindrical calorimeter, so that forward/backward tracks have more room to bend, permitting a more consistent measurement of track  $|\vec{p}|$  across the detector. To reconcile these facts with the spherical shape of the pseudo-calorimeter, one can imagine that the pseudo-calorimeter merely represents the solid angle subtended by each tower. Its towers’ physical locations can then be projected from the unit sphere onto a cylindrical shape. However, the exact geometry

---

<sup>8</sup>There are an even number of towers per belt and two identical hemispheres, each with an inactive beam hole. Each band has a width  $\Delta\theta = w$ , and to maintain the equal-area property the azimuthal pitch  $\Delta\phi \approx w/\sin\theta$  gets larger as the belts approach the beam hole (fewer towers per belt).

of the calorimeter towers is not terribly important; the main purpose of the pseudo-calorimeter is to smear the angular position of neutral energy into an extensive shape, since our ability to resolve this angular information is fundamentally limited by the inability of neutral matter to directly ionize components within a physical detector. Hence, the pseudo-detector fulfills its goal of approximating the maximum information observable in a collider event.

## 1.4 Event reconstruction

As we just saw, a detector can resolve a jet only via its constituents (massless tracks and neutral towers) which must then be reconstructed into the original jet. The recipe which governs this reconstruction is the jet definition. Given the potential complexity of a QCD final state, jets need a definition that is (i) useful for physics, (ii) theoretically sound (e.g., stable under small perturbations) and (iii) tractably automated. In this section we will explore one of the most common jet definitions.

**1.4.1 Singlet jets.** A singlet jet is clustered from protojets. At the start of clustering, each massless physics object seen by the detector becomes a protojet (the tracks and neutral towers seen at PETRA in Figure 1.2 and CMS in Figure 1.7). Clustering is the process of bundling these protojets into massive, extensive jets, with each protojet belonging to only one jet. Figure 1.7 depicts all the constituents of one jet, with the jet's conical shape overlaid. Since protojets are color singlets, jets clustered from protojets must also be color singlets. How can a singlet be a proxy for a colored quark (triplet) or gluon (octet)?

The problem with mapping singlet jets onto partons arises from charge conservation; a color triplet (quark) cannot disintegrate into a collection of color singlets without absorbing an anti-triplet. Thus, a quark cannot become a jet without some color connections with the rest of the final state (e.g., if the final state is  $q\bar{q}$ , the quark

and antiquark must exchange some relatively soft gluons). A singlet jet necessarily conflates the four-momentum  $\mathbf{p}$  of the original parton with the  $\mathbf{p}$  of these color connections. In spite of this failure, singlet jets tend to be good approximations for the hardest jets in the event (generally the ones being studied), and are also relatively simple to implement. There are two classes of singlet jet definitions: top-down and bottom-up.

A top-down jet definition tries to define jets via some final criteria (fitting jets to some specified shape [23, 24], finding some global extrema [25, 26], or gathering protojets inside cones which don't appreciably shift when each constituent is added/removed [27, 28]). Not only do these algorithms involve a lot of trial and error (since they are searching for a stationary state), they can be unstable under small perturbations. These definitions are not widely used at the LHC.

A bottom-up jet definition builds jets piece by piece [29–33]. The most common is the binary scheme, which looks at the  $\mathcal{O}(N^2)$  unique pairs of protojets and ranks them according to how much they resemble two particles from the same jet. The highest-ranking pair is merged (summing the two  $\mathbf{p}$  into a new, composite protojet), the pairs are re-ranked (given the merged pair) and the process is repeated. When one of the aggregated protojets begins to look like a jet, it is removed from the protojet pool and promoted to a jet. Eventually, there are no more protojets.

**1.4.2 The  $k_T$ -family jet definition.** Currently, the most popular jet definitions belong to the  $k_T$ -family — a binary, bottom-up scheme parameterized by its clustering power  $p$  and radius  $R$ . As of 2010, both general-purpose LHC experiments (ATLAS and CMS) use anti- $k_T$  ( $p = -1$ ) [33] as their primary jet definition (and in 2016 agreed that  $R = 0.4$  was the best choice of radius). Cambridge-Aachen ( $p = 0$ ) and  $k_T$  ( $p = 1$ ) clustering remain popular in niche applications, but suffer from greater sensitivity to soft QCD and the underlying event [33].

$k_T$ -family clustering revolves around two “distances” based upon each protojet’s transverse momentum ( $k_T \equiv p_T$ ). These “distances” should not be interpreted as lengths, but as a metric that gauges how much a protojet resembles some phenomenon, with short distances corresponding to close resemblance. A protojet’s “beam” distance gauges how much the protojet looks like isolated radiation:

$$d_i = k_{T,i}^p. \quad (1.18)$$

This beam distance is an intrinsic property of each protojet. The inter-protojet distance gauges how much each *pair* of protojets looks like QCD radiation of a common origin, via a geometric separation  $\Delta R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}$  (which is invariant to longitudinal boosts)

$$d_{ij} = \min(k_{T,i}^p, k_{T,j}^p) \frac{\Delta R_{ij}}{R} \quad (j \neq i). \quad (1.19)$$

At each clustering step, the smallest distance is found, and an action is taken. The nature of the final clustered jets depends on the power  $p$  (which controls the flow of energy) and the radius  $R$  (which sets the angular scale of the jet “cones”).

Given  $N$  starting protojets, clustering continues until no protojets remain (exactly  $N$  iterations). The canonical algorithm is:<sup>9</sup>

1. Calculate  $d_{ij}$  for all  $j > i$  and all beam distances  $d_i$ .
2. Find the smallest distance.
  - (a) When the smallest distance is a beam distance ( $d_i$ ), remove the protojet from the list of protojets and promote it to a finalized jet.
  - (b) When the smallest distance is between two protojets ( $d_{ij}$ ), merge them.

---

<sup>9</sup> A naïve implementation (recalculate every  $d_{ij}$  at every step) has a computational complexity of  $\mathcal{O}(N^3)$ . One that caches the smallest distances can approach  $\mathcal{O}(N^2)$ . For  $N > 10,000$ , it becomes advantageous to implement a nearest-neighbor approach, which is even faster ( $\mathcal{O}(N \log N)$ , see Ref. [34]).

3. Recalculate any distance affected by step 2, then repeat step 2.

Step 2a has important implications, and it will help us understand  $k_T$  clustering. Imagine two ways to convert a set of original protojets  $P$  into a set of final jets  $J$  with radius  $R$ : **(a)** cluster  $P$  into jets of radius  $R$  directly or **(b)** first cluster  $P$  into jets with a smaller radius  $R^* < R$ , then use the  $R^*$  jets as the input protojets for a second round of clustering.

$$\mathbf{a} : \quad P \xrightarrow{R} J_a \quad (1.20)$$

$$\mathbf{b} : \quad P \xrightarrow{R^*} J_b^* \xrightarrow{R} J_b \quad (1.21)$$

During **b**'s first round of clustering, step 2a will remove radius  $R^*$  jets from the clustering pool before the first round is complete. These jets will return for the second round of clustering, but their temporary removal will ensure that **a** and **b** lack identical sets of partially-clustered protojets at every step. This forks the clustering histories, so that  $J_b \neq J_a$ .<sup>10</sup> If one then selects a radius  $R$  jet from  $J_a$  and reclusters its constituents using the smaller radius  $R^*$ , they will not map back onto the same jets found in  $J_b^*$ .

One reason for anti- $k_T$ 's ( $p = -1$ ) popularity at the LHC is its simple nature; it builds perfectly conical jets, starting with the highest- $p_T$  (or hardest) jets in the event. This causes a harder jet's cone to eat into a softer jet (see Fig. 1.10 near  $y = 2$  and  $\phi = 5$ ), which makes anti- $k_T$  relatively insensitive to soft QCD radiation or the underlying event. Another of anti- $k_T$ 's strengths is its ability to run efficiently on the large number of protojets that the LHC supplies. The SIS cone definition (a top-down algorithm used at the Tevatron) is much slower, and produces more ragged jet boundaries, which can be sensitive to soft radiation.

---

<sup>10</sup>A notable exception is Cambridge-Aachen (CA) clustering, since  $p = 0$  makes every beam distance *larger* than any valid merger. This forces *all* protojet mergers to occur before *any* jet can leave, making  $J_b^*$  an exact intermediate state for  $J_a$ .

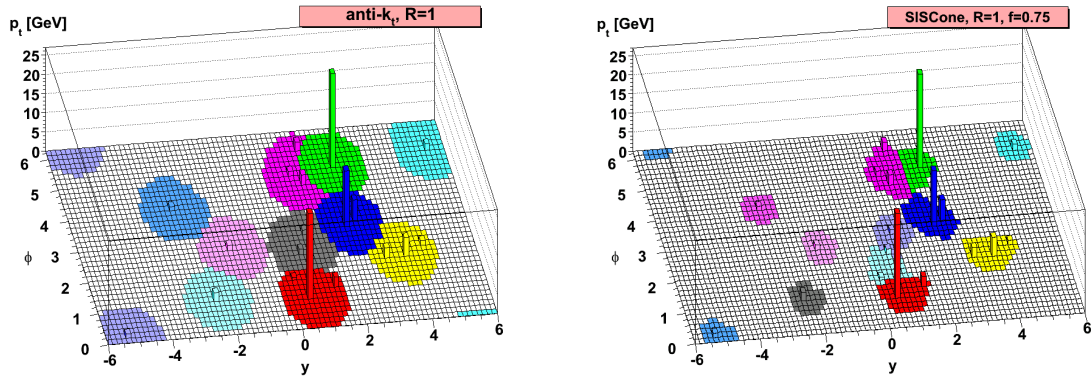


Figure 1.10. Two jet definitions applied to the same set of massless protojets (charged towers from a simulated square-celled calorimeter) [33]. The “Lego” plot unrolls each protojet’s 3D direction ( $y, \phi$ ) into a plane, depicting energy ( $p_T$ ) as height. The colored regions depict the jet areas, with conical jets appearing elliptical.

With these basic definitions, we can move to Chapter 2, where we will discuss why determining the flavor of clustered jets (which of the six quarks they originated from) is a vital tool for modern particle physics. We will then find that the existing tools for heavy-flavor identification are not very effective in the LHC’s most energetic events, limiting our ability to resolve new physics. In response to this problem, we develop the  $\mu_x$  boosted-bottom jet tag, a new tool that shows great promise in extending the LHC’s reach into uncharted territories.

## CHAPTER 2

THE  $\mu_x$  BOOSTED-BOTTOM-JET TAG

*Note: To first approximation, this chapter should be considered a verbatim reprint of the first few sections of my journal article [4], but fleshed out where more clarity was needed.*

The driving force behind searches for new physics is the widely held belief that the Standard Model is incomplete. For example, the Standard Model cannot predict the mass of the Higgs boson, nor of any of the quarks and leptons, nor can it explain the large hierarchy in masses among the three generations. Furthermore, the Standard Model (SM) was formulated with massless neutrinos, whereas neutrino oscillations and cosmological observations provide compelling evidence that neutrinos have masses of  $\mathcal{O}(0.1 \text{ eV})$ , a million times lighter than the electron — neutrino mass creates a *second* hierarchy problem. Additionally, all observations indicate that QCD does not violate CP symmetry, which either requires fine tuning the QCD Lagrangian *or* some physics *beyond* the Standard Model [35].

Natural law provides a much more sensible reason to believe that the Standard Model is incomplete. The intrinsic nature of science is to constantly overturn or extend the previous model, resolving the tension in the new, high-precision measurements. Human beings have never had the full picture before the Standard Model, so why should we expect that we have it now? Unfortunately, we are in the unenviable position that the new, high-precision measurements require the most complex machine ever built, and there is only one that we all must share. In spite of these hurdles, the Large Hadron Collider (LHC) is committed to the search for new physics, and specifically the search for new, massive resonances which are beyond the Standard Model (BSM).



Prime candidates for new physics are narrow vector current particles, generally called  $Z'$  or  $W'$  bosons, which are a main focus of the exotics groups in experiments at the LHC. These particles arise in many extensions of the Standard Model, such as the sequential Standard Model [36], broken  $SU(2)_L \times SU(2)_R$  symmetry [37–39], grand unified models [40–42], Kaluza-Klein excitations in models of extra dimensions [43,44], non-commuting extended technicolor [45], general extended symmetries [46,47], and more. Using 8 TeV LHC data, the ATLAS [48] and CMS [49] collaborations set bounds on many types of  $Z'$  bosons that decay to dileptons ( $Z' \rightarrow l^+l^-$ ) below around 2.9 TeV.

A more challenging search is hadronically decaying resonances like leptophobic gauge bosons (those which do not decay to leptons), such as a top-color  $Z'$  boson, which is excluded up to 2.4 TeV [50–52], or a right-handed  $W'$  boson, which is excluded for SM-like couplings up to 1.9 TeV [53, 54]. This latter boson is most strongly constrained by the  $W' \rightarrow tb$  final state [55, 56]. The primary challenge with these hadronic final states is that the SM “multijet” background  $pp \rightarrow jj(j)$  (where  $j$  is a light quark or gluon) is many orders of magnitude larger than for  $pp \rightarrow l^+l^-$ , so that hadronically decaying resonances are a tiny bump on a huge background.

A primary tool to cut through the background is to isolate the decay to heavy-flavored quarks (e.g.,  $Z' \rightarrow b\bar{b}$  or  $W' \rightarrow tb$ ), since the SM background to produce heavy-flavored jets (e.g.,  $pp \rightarrow b\bar{b}$ ) is many orders of magnitude lower than the total multijet background. Isolating the heavy-flavored final states requires “flavor tagging” the heavy-flavored jets via the unique signatures of the bottom and top quark. Unfortunately, standard flavor tagging techniques lose much of their power in searches for vector boson resonances above 1.5 TeV, since the large invariant mass of the resonance boosts the decay products of each quark, breaking the standard tagging algorithms and requiring special techniques to tag boosted-top jets [57–65] and boosted-bottom

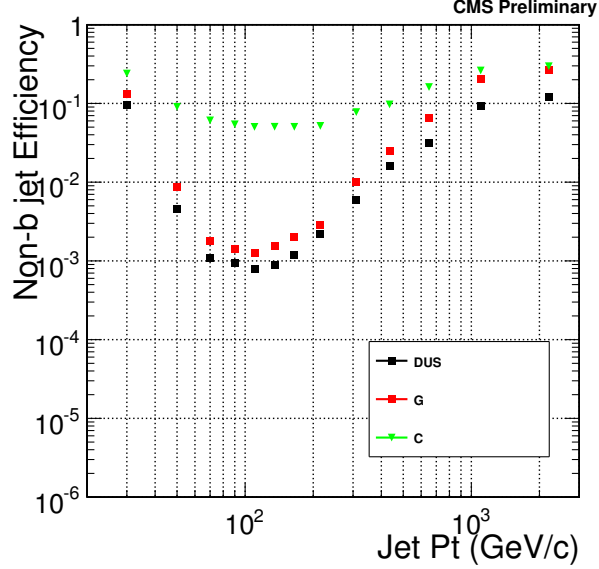


Figure 2.1. The non- $b$  jet efficiency of the “track counting high purity” algorithm, as a function of  $p_T$ , in CMS simulation [68].

jets [56, 66, 67].

The breakdown of  $b$  tagging for boosted- $b$  jets is quite clear in Figure 2.1. The tagging parameters in each bin are tuned so that the  $b$ -tag efficiency  $\epsilon_b$  (the probability to tag a  $b$  jet as a  $b$  jet) is maintained at 50%. The resulting fake rate  $\epsilon_{\text{non-}b}$  (the probability to falsely tag a non- $b$  jet a  $b$  jet) rises dramatically as jet  $p_T \rightarrow \mathcal{O}(\text{TeV})$  [67]. The large fake rate for charm quarks is somewhat irreducible, as will be discussed in the next section. Much more distressing is the dramatic rise in the light jet fake rate (for light quarks  $dus$  or gluons  $g$ ), which dramatically increases the magnitude of multijet background. Note especially the disproportionate increase in the gluon-jet fake rate, which is primarily caused by the presence of real  $B$  hadrons from gluon splitting  $g \rightarrow b\bar{b}$  after the hard scatter. As an example, one CMS search for exotic resonances above 1.2 TeV encountered fake rates above 10% per jet [69]. Ideally, one prefers a much purer  $b$  tag that can maintain  $\epsilon_b/\epsilon_{\text{light}} = \mathcal{O}(100)$ .

The high- $p_T$  degradation of  $b$ -tagging also manifests in the systematic uncer-

tainties in  $b$ -tagging efficiency and fake rates, which dominate the current  $W' \rightarrow tb$  limits, and have so far closed the  $Z' \rightarrow b\bar{b}$  searches from consideration. A recent ATLAS  $W'$  search [53, 70] found a 35% uncertainty in the  $b$ -jet tagging efficiency for jets with  $p_T$  above 500 GeV (i.e.,  $M_{W'} \gtrsim 1$  TeV). This is mainly driven by a lack of clean samples of high- $p_T$   $b$  jets tagged with a complementary method, which are necessary to cross-check the signal/background efficiencies of the  $b$  tags [71–73].

This chapter presents a new technique for tagging boosted, heavy-flavor jets with  $p_T > 500$  GeV called “ $\mu_x$  tagging,” which I originally published in [4]. This technique is an improvement to the *boosted-bottom-jet tag* first proposed in Ref. [56]. Here, the focus is on  $b$  quarks which are themselves highly boosted, instead of boosted topologies which contain bottom quarks (e.g., boosted  $t \rightarrow Wb$  or  $H \rightarrow b\bar{b}$ ). In Section 2.1 we explain why existing tagging methods are insufficient at high energies, and then derive from first principles the muon-based  $\mu_x$  tag in Section 2.2. In Section 2.3 we present the  $\mu_x$  tagging efficiencies for bottom and charm flavored jets, along with small light-jet fake rates, using a detailed simulation based on the ATLAS detector.

The  $\mu_x$  tag combines angular information and jet substructure to tag  $b$  jets,  $c$  jets, light jets, and “light-heavy” jets (a light jet containing  $B$  hadrons from gluon splitting) with efficiencies  $\epsilon_b = 14\%$ ,  $\epsilon_c = 6.5\%$ ,  $\epsilon_{\text{light-light}} = 0.14\%$ , and  $\epsilon_{\text{light-heavy}} = 0.5\%$ , respectively. These efficiencies are nearly independent of transverse momentum at high energy (i.e., a constant light-jet fake rate), providing  $\mathcal{O}(100)$  signal purity. Not only does the  $\mu_x$  tag solve a major problem with existing  $b$  tags, it also provides a complementary tagging scheme that can be used to dramatically reduce systematic uncertainties in tagging efficiencies for high- $p_T$  jets. In the following chapter, we demonstrate the efficacy of the  $\mu_x$  tag via a full signal and background study for several interesting beyond the Standard Model (BSM) theories: a leptophobic  $Z'$  boson and a charged Higgs boson  $H^\pm$  from a two-Higgs doublet.

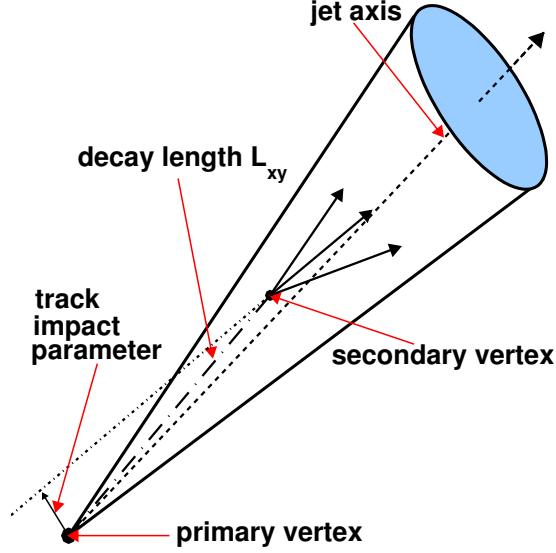


Figure 2.2. The secondary vertex of a  $B$  hadron decay [67].

## 2.1 Tagging a heavy-flavored jet

Heavy-quark ( $b$  or  $c$ ) initiated jets shower and hadronize in a manner that is distinct from light parton ( $d$ ,  $u$ ,  $s$ , or  $g$ ) initiated jets. The large masses of the heavy quarks ( $m \gtrsim \Lambda_{\text{QCD}}$ ) cause their fragmentation functions to peak near  $z = 1$ , (where  $z$  is the fraction of energy retained by the original quark when it hadronizes). This means that  $b$  and  $c$  quarks tend to retain their momentum during fragmentation [74], spawning heavy hadrons which carry a large fraction of their jet's momentum. These hadrons have long lifetimes ( $c\tau(B/D) \approx \mathcal{O}(10^{-4} \text{ m})$ ), so if the  $b/c$  quark is even moderately boosted ( $\gamma \gtrsim 10$ ), they will travel millimeters from the collider's collision point (the primary vertex) before decaying (see Fig. 2.2). The charged decay daughters will then point back to a secondary vertex (SV) which is far enough from the primary vertex to be resolved, but close enough to distinguish them from other meta-stable particles (e.g.,  $c\tau(K_S^0) = 3 \times 10^{-1} \text{ m}$ ). Additionally, the significant rate of semi-leptonic decay of  $b/c$  hadrons ( $\mathcal{B}(X_{b/c} \rightarrow l \nu_l Y) \approx 0.1$  for each  $l \in \{e, \mu\}$ )

enriches these jets with energetic leptons. Since bottom hadrons decay primarily to charm hadrons,  $b$  jets have twice the probability of  $c$  jets to contain leptons (20% versus 10%, respectively).

**2.1.1 Challenges for existing  $b$  tags.** Modern  $b$ -tagging algorithms are essentially track-based tags that search for evidence of a secondary vertex using the charged particles seen by the detector [75,76]. While they frequently use neural nets and multiple inputs, their efficiencies are predominantly determined by the impact parameter of a jet’s tracks and the mass of its reconstructed SV. Although light jets also contain secondary vertices (e.g.,  $K_S^0/\Lambda$  decay or material interaction [77]), this background is largely reducible for jets with  $p_T < 300$  GeV, giving track tags high  $b$  jet efficiency (50-80%) and light jet fake rates of  $\mathcal{O}(1\%)$ . Above  $p_T = 300$  GeV, the increasingly boosted nature of the jet makes track-tagging difficult. Boosted tracks bend less, and are thus harder to constrain and more sensitive to tracker resolution and alignment.

These problems are exacerbated in boosted heavy-flavor jets, where the primary hadron can decay *after* traversing one or more pixel layers, leading to fewer hits inside the tracker and making it difficult or impossible for its daughters to produce the “high purity” tracks needed by most SV tagging algorithms. Additionally, if these collimated daughters strike *adjacent* pixels, they can create a “merged cluster” which also hinders reconstruction [76,78]. These problems are well exemplified by Figure 2.1, where the light-jet fake rate increases 100-fold as jet  $p_T$  increases from 100 GeV to 1 TeV.

Another component of current  $b$ -tagging algorithms is  $p_T^{\text{rel}}$  tagging, which measures the momentum of leptons transverse to the centroid of their jet. Compared to light hadrons, heavy hadrons have a larger mass and carry a larger fraction of their jet’s momentum; thus, leptons produced by heavy hadrons will have more energy and will arrive at wider angles inside the jet. These effects conspire to produce larger

values of  $p_T^{\text{rel}}$  [68, 79]. Since electrons are difficult to identify inside jets,  $p_T^{\text{rel}}$  tagging generally utilizes only muons. In ATLAS and CMS, muon  $p_T^{\text{rel}}$  tags give  $\sim 10\%$   $b$  jet efficiency and a light jet rejection (inverse tagging efficiency) of about 300 [79]. However, once jet  $p_T$  exceeds about 140 GeV, the underlying boost makes  $p_T^{\text{rel}}$  distributions for heavy and light jets nearly indistinguishable [71], precluding the tag.

## 2.2 The $\mu_x$ boosted- $b$ tag

The failure of existing tagging methods to adequately reject high- $p_T$  light jets is a problem. For track tagging, it is essentially a problem of detector resolution, so any improvements will likely involve novel utilization of the hardware and track observables. For  $p_T^{\text{rel}}$  tagging it is potentially a problem of definition;  $p_T^{\text{rel}}$  dilutes a well measured muon angle with a more poorly measured muon energy. This drove the development of the “boosted-bottom tag” [56], a purely angular tag on jets containing muons within  $\Delta R = 0.1$  of their centroid. This tag achieves nearly ideal signal efficiency (given the muonic branching fraction), but suffers from a continuous rise with energy in light jet fake rate. Since the centroid of an *entire* jet is not necessarily aligned with the  $B$  hadron’s decay, and the boost cone of muon emission should tighten as the jet’s boost increases,  $b$  jet decay should be reexamined in the context of jet substructure. This will provide the basis for a new,  $\mu_x$  boosted- $b$  tag.

**2.2.1 Theory of the  $\mu_x$  tag.** Since the boost of a high-energy  $b$  jet is unavoidable, a good tag should try to use the boost to its advantage. In this section we examine the kinematics of boosted decay to define an observable  $x$ . In the subsequent sections, we will explore how  $x$  is observed and why it is a strong handle for identifying boosted  $b$  jets.

Consider a jet containing a  $B$  meson that decays semi-muonically, as depicted in Figure 2.3. In the decay’s center-of-momentum (CM) frame (denoted with a prime),

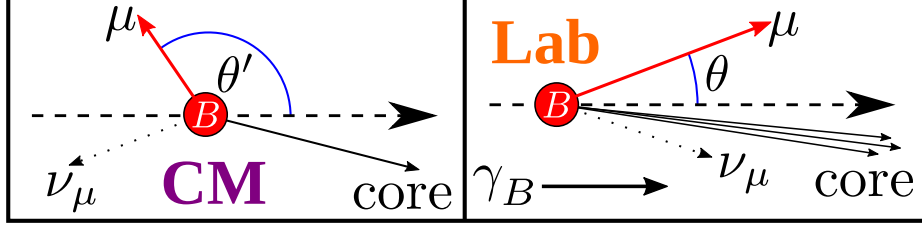


Figure 2.3. Nomenclature for the center-of-momentum frame and boosted lab frame.

the muon is emitted with speed  $\beta'_\mu$  and angle  $\theta'$  with respect to the boost axis. In the lab frame, the  $B$  meson's decay products are boosted by  $\gamma_B$  into a *subject* with a hadronic “core” (typically a charm hadron that later decays). The four-momentum of this  $B$  hadron subject is

$$\mathbf{p}_{\text{subject}} = \mathbf{p}_{\text{core}} + \mathbf{p}_\mu + \mathbf{p}_{\nu_\mu}, \quad (2.1)$$

and the muon now makes the angle  $\theta$  with the  $B$  meson's direction of travel.

To determine the relationship between  $\theta$  and  $\theta'$ , we set the  $z$ -axis parallel to the motion of the  $B$  hadron. Placing the muon in the  $yz$ -plane, its CM four-momentum is

$$\mathbf{p}'_\mu = E' [1, 0, \beta'_\mu \sin(\theta'), \beta'_\mu \cos(\theta')]. \quad (2.2)$$

We then use the Lorentz boost matrix of the  $B$  hadron

$$\mathbf{\Lambda} = \begin{pmatrix} \gamma_B & 0 & 0 & \beta_B \gamma_B \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta_B \gamma_B & 0 & 0 & \gamma_B \end{pmatrix} \quad (2.3)$$

to boost the muon into the lab frame

$$\mathbf{p}_\mu = \mathbf{\Lambda} \cdot \mathbf{p}'_\mu = E' [\gamma_B (1 + \beta_B) \cos(\theta'), 0, \beta'_\mu \sin(\theta'), \gamma_B (\beta_B + \beta'_\mu \cos(\theta'))]. \quad (2.4)$$

Using the muon's lab frame momentum, we now define a lab frame observable

$$x \equiv \gamma_B \tan(\theta) = \frac{\sin(\theta')}{\kappa + \cos(\theta')}. \quad (2.5)$$

This observable  $x$  depends on  $\kappa \equiv \beta_B/\beta'_\mu$ , which itself depends on the boost of the muon  $\gamma'_\mu$  in the CM frame. This boost is generally not measurable, but when the system is highly boosted,  $\kappa \rightarrow 1$  and  $x$  loses its  $\kappa$ -dependence. This makes  $x$  nearly scale-free for highly boosted systems, with  $x \approx 0$  corresponding to a very central muon (close to the centroid of the subjet). Fortunately, the kinematics of  $B$  meson decay ensure that the muon is usually quite boosted in the CM frame ( $\gamma'_\mu \gtrsim 3$ ), while the  $b$  jets of interest (jet  $p_T > 300$  GeV) have an even larger boost  $\gamma_B \gtrsim \mathcal{O}(100)$ . This sends  $\kappa \rightarrow 1$ , giving muons in boosted  $b$  jets a nearly universal  $x$  distribution, and allowing muon  $x$  to become a robust handle for identifying boosted  $b$  jets.

To determine this universal  $x$  distribution, we can first assume that  $\kappa > 1$  (the  $b$  jet's lab frame boost is greater than the muon's CM frame boost). This will ensure that  $x > 0$  (i.e., no muon is detected more than  $90^\circ$  from the subjet's centroid). We now need to determine the angular distributions of muons in the CM frame. Luckily,  $B$  mesons are scalar particles ( $J = 0$ ), and lacking vector degrees of freedom to orient their decay, emitted muons can only be distributed isotropically in the CM frame. Giving an equal helping of decay width  $\Gamma$  to each patch of solid angle  $d\Omega$ , we find

$$\frac{d\Gamma}{d\Omega} = \frac{\Gamma}{4\pi} \quad \implies \quad \frac{d\Gamma}{\Gamma d\Omega} = \frac{1}{4\pi}. \quad (2.6)$$

Since  $d\Omega = \sin(\theta')d\theta'd\phi'$ , and integrating over  $\phi'$ ,

$$\frac{d\Gamma}{\Gamma d\theta'} = \frac{1}{2} \sin(\theta'). \quad (2.7)$$

This differential decay width is more easily manipulated via the change of variable  $z' \equiv \cos(\theta')$ . Multiplying by the absolute Jacobian  $\left| \frac{d\theta'}{dz'} \right| = \frac{1}{\sin(\theta')}$ , we find

$$\frac{d\Gamma}{\Gamma dz'} = \frac{1}{2}. \quad (2.8)$$

To boost this distribution into the lab frame, we need to study the mapping

$$x = \frac{\sqrt{1 - z'^2}}{\kappa + z'}. \quad (2.9)$$



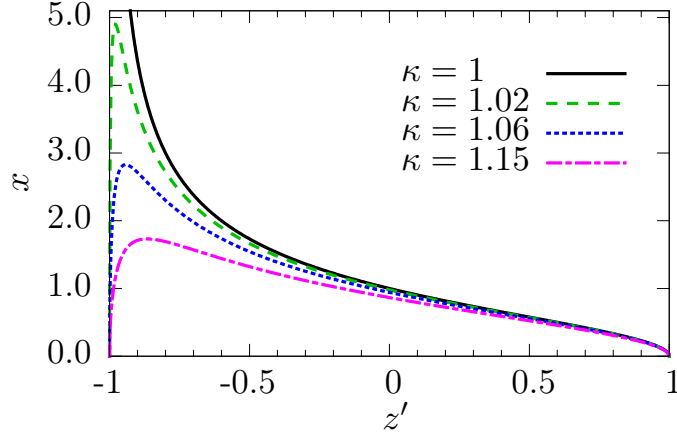


Figure 2.4. A muon's lab frame  $x$  as a function of its CM frame  $z'$ , for several choices of  $\kappa$ . For all  $\kappa > 1$ , there are two  $z'$  which map to a given  $x$ .

We plot this mapping in Figure 2.4 for several values of  $\kappa > 1$ . A muon emitted exactly backward in the CM frame ( $z' = -1$ ) is flipped forward by the boost and shows up exactly centrally in the lab frame ( $x = 0$ ). As muons become less backward in the CM frame ( $z'$  increases), they become less central in the lab frame, moving to larger  $x$ . But muons are restricted to a boost cone  $0 \leq x \leq 1/\sqrt{\kappa^2 - 1}$ , so once they achieve this maximum  $x$ , increasing  $z'$  causes them to become central again ( $x \rightarrow 0$ ).

This behavior is important as we implement the change of variable

$$\frac{d\Gamma}{\Gamma dx} = \underbrace{\frac{d\Gamma}{\Gamma dz'}}_{z' \mapsto x} \times \left| \frac{dz'}{dx} \right|. \quad (2.10)$$

This requires solving Equation 2.9 for  $z'$ , to first covert  $z' \mapsto x$  in Equation 2.8 (which happens to have no  $z'$  dependence) and then to find  $|dz'/dx|$ . But as we just determined in the last paragraph, two  $z'$  map to the same  $x$ , so there are two solutions

$$z'_{L/R} = \frac{-\kappa x^2 \mp \sqrt{1 + x^2 - \kappa^2 x^2}}{1 + x^2}, \quad (2.11)$$

one each for to the left and right sides of the peaks shown in Figure 2.4. Hence, the Jacobian also has two solutions;

$$\frac{dz'_{L/R}}{dx} = \pm \frac{x}{(1 + x^2)^2} \frac{(1 + \kappa^2) + x^2(1 - \kappa^2) \mp 2\kappa\sqrt{1 + x^2 - \kappa^2 x^2}}{\sqrt{1 + x^2(1 - \kappa^2)}}. \quad (2.12)$$

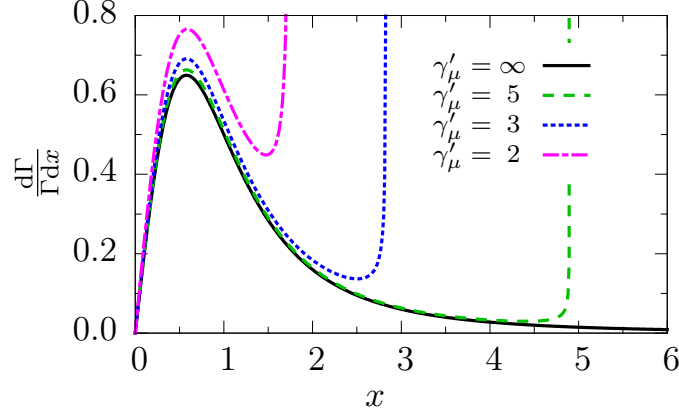


Figure 2.5. Theoretical muon distribution  $\frac{d\Gamma}{\Gamma dx}$  in the lab frame (with  $\beta_B \rightarrow 1$ ) for various muon boosts  $\gamma'_\mu$  in the CM frame of the decay.

To build the absolute Jacobian for  $\kappa \geq 1$ , we notice that  $\frac{dz'_L}{dx} > 0$  and  $\frac{dz'_R}{dx} < 0$ , so

$$\left| \frac{dz'}{dx} \right| = \left( \frac{dz'_L}{dx} - \frac{dz'_R}{dx} \right) = \frac{2x}{(1+x^2)^2} \frac{(1+\kappa^2) + x^2(1-\kappa^2)}{\sqrt{1+x^2(1-\kappa^2)}}. \quad (2.13)$$

We can now calculate the muons' differential decay width in the lab frame

$$\frac{d\Gamma}{\Gamma dx} = \frac{2x}{(x^2+1)^2} K(x, \kappa), \quad (2.14)$$

where all  $\kappa$  dependence (i.e., scaling) has been shifted into the factor

$$K(x, \kappa) = \begin{cases} \frac{(1+\kappa^2)+x^2(1-\kappa^2)}{2\sqrt{1+x^2(1-\kappa^2)}} & 0 \leq x \leq 1/\sqrt{\kappa^2-1} \\ 0 & \text{everywhere else} \end{cases}. \quad (2.15)$$

The piece-wise definition of  $K(x, \kappa)$  enforces the boundary of the  $B$ 's boost cone; when  $\gamma_B \gg \gamma'_\mu$ , the maximum value of  $x$  for a lab frame muon is  $x \approx \sqrt{\gamma_{\mu, \text{cm}}^2 - 1}$ .

We now have an  $x$  distribution for lab-frame muons which is split into a universal shape  $2x/(x^2+1)^2$  and a  $\kappa$  dependent factor  $K(x, \kappa)$ . But as  $\kappa \rightarrow 1$  in the boosted limit,  $K(x, \kappa) \rightarrow 1$  also, so that only the universal shape survives. This can be seen in Figure 2.5, where the  $x$  distribution is shown for several muon boosts  $\gamma'_\mu$ . Since the typical muon boost in the CM frame is  $\gamma'_\mu > 3$ , for which the  $x$  distribution

only slightly deviates from the universal shape, the  $d\Gamma/(\Gamma dx)$  of a typical muon is well approximated by the  $\kappa = 1$  limit. This makes the universal  $x$ -shape useful for identifying muons from a boosted decay.

Assuming that all muons follow the universal shape, we can calculate the fraction  $\rho$  of muons which fall within some  $x_\rho$  of the subjet's centroid

$$\rho = \int_0^{x_\rho} \frac{2x}{(x^2 + 1)^2} dx = \frac{x_\rho^2}{1 + x_\rho^2}. \quad (2.16)$$

We can then invert this equation to find the value of  $x_\rho$  which confines a fraction  $\rho$  of lab frame muons

$$x_\rho = \sqrt{\frac{\rho}{1 - \rho}}. \quad (2.17)$$

Since slow muons are distributed at smaller  $x$  than ultrarelativistic muons,  $x_\rho$  is valid for all muons, not just those that are ultrarelativistic in the CM frame of the decay.

We can now define a cut  $x_{\max} = x_{90\%}$  which will accept 90% of muons which are compatible with boosted  $B$  hadron decay. In addition, we use the hard fragmentation of  $b$  quarks (i.e., they should retain the majority of their jet's momentum) to motivate a cut on the  $p_T$  fraction of the  $B$  hadron subjet to the total jet  $p_T$

$$f_{\text{subjet}} \equiv \frac{p_{T,\text{subjet}}}{p_{T,\text{jet}}} \geq 0.5. \quad (2.18)$$

These two cuts ( $x \leq x_{\max}$  and  $f_{\text{subjet}} \geq f_{\text{subjet}}^{\min}$ ) define the  $\mu_x$  boosted-bottom-jet tag.

**2.2.2 Reconstructing  $\mathbf{p}_{\text{subjet}}$  and measuring  $x$ .** Using the  $\mu_x$  tag requires measuring  $x$ , which we originally defined in terms of an isolated decay of a bottom hadron. In reality,  $\mathbf{p}_{\text{subjet}}$  will overlap QCD radiation within the jet. Furthermore, half of a  $b$  jet's semi-leptonic decays come from charm hadrons. Therefore, it is not possible to measure  $\gamma_B$  — the boost of the primary  $B$  hadron — only  $\gamma_{\text{subjet}}$ , the boost of the muon-core system. In spite of this limitation, we will see that it is still possible to reconstruct a meaningful  $x$ . To arrive at this conclusion, we must first work through the technical details of reconstructing the subjet.

We cluster jets using the anti- $k_T$  algorithm and a radius parameter  $R = 0.4$ . Muons participate in jet clustering, which lets hard muons seed jet formation. Candidate jets for  $\mu_x$  tagging must contain a *taggable* muon ( $p_{T,\mu} \geq 10$  GeV) to ensure good muon reconstruction. While a taggable muon’s associated neutrino is inevitably lost, most of the muon and neutrino momentum comes from their shared boost, making the muon an acceptable neutrino analog. We use the simplest choice:  $\mathbf{p}_{\nu_\mu} = \mathbf{p}_\mu$ .

To find the core (the hadronically decaying charm hadron), we compose a list of core candidates by reclustering the jet with the anti- $k_T$  algorithm using  $R_{\text{core}} = 0.04$ ; this radius is designed to localize the core to a  $3 \times 3$  grid, based on the fixed width  $w$  of the calorimeter towers ( $\sqrt{2}w < R_{\text{core}} < 2w$ ). All jet constituents are used during reclustering (allowing taggable muons to seed core formation) *except* towers failing a cut on jet  $p_T$  fraction (we choose  $f_{\text{tower}}^{\text{min}} = 0.05$ ); this reduces the core’s sensitivity to pileup, the underlying event, and soft QCD. Since the calorimeter granularity produces an ill-measured core mass, we fix the mass of each core candidate to a charm hadron mass  $m_{\text{core}} = 2$  GeV. The “correct” core is the candidate which brings  $\sqrt{\mathbf{p}_{\text{subjet}}^2}$  closest to  $m_B$ , the nominal mass of the  $b$  hadron admixture (we choose  $m_B = 5.3$  GeV).

Given our neutrino strategy ( $\mathbf{p}_{\nu_\mu} = \mathbf{p}_\mu$ ), we can study the value of  $x$  that will be observed for an *arbitrary* muon-subjet system, which *could* be the remnants of a  $B$  hadron, but could also be a random association of jet constituents. Such a system can be fully described using three lab frame observables:  $\gamma_{\text{core}}$  (the energy of the core),  $\lambda = 2E_\mu/E_{\text{core}}$  (the energy of the muon plus neutrino, relative to the core), and  $\xi$  (the lab-frame angle between the muon and the *core*). Assuming that both the muon and the core are ultra-relativistic in the lab frame (i.e.,  $\beta \rightarrow 1$ ), the  $x$  of this

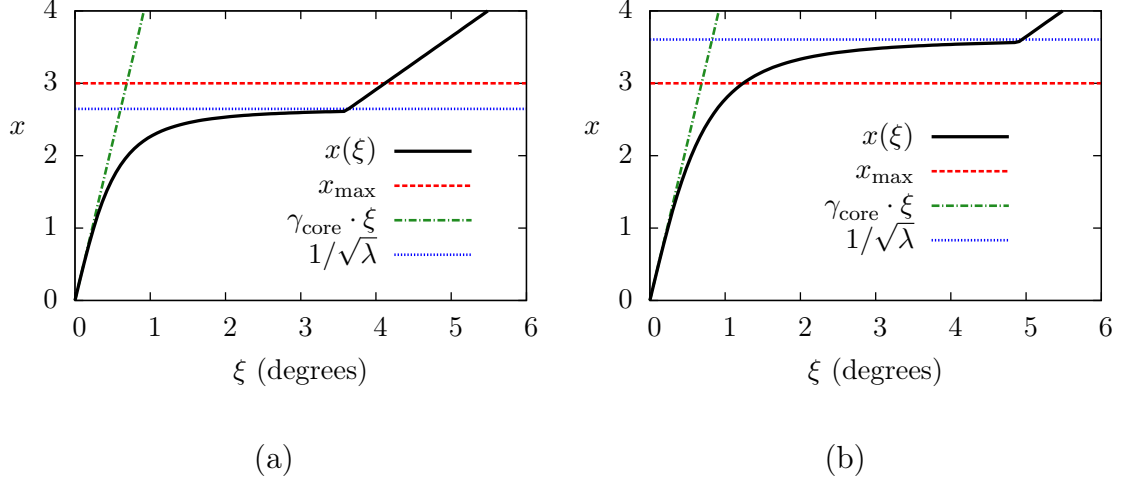


Figure 2.6.  $x(\xi)$  for a subject with  $\gamma_{\text{core}} = 250$  for (a) a harder muon ( $\lambda = 1/7$ ) and (b) a softer muon ( $\lambda = 1/13$ ).

arbitrary system is a function of the muon-core angle

$$x(\xi) \approx \underbrace{\gamma_{\text{core}} \frac{1 + \lambda}{\sqrt{1 + 2\lambda \gamma_{\text{core}}^2 [1 - \cos(\xi)]}}}_{\gamma_{\text{subject}}} \underbrace{\frac{\sin(\xi)}{\cos(\xi) + \lambda}}_{\tan(\theta)}. \quad (2.19)$$

This  $x(\xi)$  function has two distinct  $\xi$  regimes, visible in Figure 2.6. When the muon and the core nearly overlap (so that  $\xi$  is vanishingly small),  $x \approx \gamma_{\text{core}} \xi$ . As the muon gets farther from the core,  $x$  flattens into a plateau at  $x \approx 1/\sqrt{\lambda}$ . This  $x$  plateau exists because, as  $\xi$  rises, every increase in  $\tan(\theta)$  is compensated by an increase in  $m_{\text{subject}}$  that drives down  $x$ . Eventually,  $m_{\text{subject}} \gg m_B$  and the reconstructed subject is no longer consistent with a decaying  $B$  hadron. This requires limiting  $m_{\text{subject}}$  (we choose  $m_{\text{subject}}^{\text{max}} = 12$  GeV), which forces the  $x$  of poorly reconstructed (or fake) subjects to abruptly return to a nearly linear  $\xi$  dependence, a discontinuity clearly visible in Figure 2.6.

The angle where  $x(\xi)$  crosses  $x_{\text{max}}$  in Figure 2.6 defines a cone of radius  $\xi_{\text{max}}$ : if the muon falls within this cone, it triggers a  $b$  tag, otherwise it does not. For the harder muon in Figure 2.6a, its  $x$  plateau is below  $x_{\text{max}}$ , so the crossing occurs at a

large angle

$$\xi_{\max}^{\text{hard}} \approx \frac{1}{\gamma_{\text{core}}} \underbrace{\left( x_{\max} \frac{m_{\text{subject}}^{\max}}{m_{\text{core}}} \right)}_{\text{tag parameters}}. \quad (2.20)$$

For *hard* muons ( $\lambda \geq x_{\max}^{-2}$ ),  $x_{\max}$  is a purely angular cut which scales inversely proportional to the energy of the core, with no additional dependence on the energy of the muon. For the softer muon in Figure 2.6b, its  $x$  plateau is above  $x_{\max}$ , so the crossing occurs at a small angle

$$\xi_{\max}^{\text{soft}} \approx \frac{1}{\gamma_{\text{core}}} \left( \frac{x_{\max}}{\sqrt{1 - \lambda x_{\max}^2}} \right). \quad (2.21)$$

For *soft* muons ( $\lambda < x_{\max}^{-2}$ ),  $x_{\max}$  is a much tighter angular cut which scales with the energy of both the core and the muon. But unless  $\lambda$  is near  $x_{\max}^{-2}$  (the soft/hard muon boundary),  $\xi_{\max}^{\text{soft}}$  is only mildly sensitive to  $\lambda$ , the energy of the muon.

Combining these two limits,  $x_{\max}$  is effectively a dual angular cut: a very tight cut for soft muons, a looser cut for hard muons, and a quick transition region (as a function of muon energy) between these two regimes. As such, the  $\mu_x$  tag depends primarily on well measured angles. In the next section, we will see that this robustness makes it nearly insensitive energy mis-measurements, such as those created by pileup. For convenience, we summarize the parameters chosen for  $\mu_x$  tagging in the following table.

Table 2.1. A summary of parameters chosen for  $\mu_x$  boosted bottom jet tagging.

$R$	0.4	$m_{\text{core}}$	2 GeV	$p_{T,\mu}^{\min}$	10 GeV
$R_{\text{core}}$	0.04	$m_B$	5.3 GeV	$x_{\max}$	3 ( $x_{90\%}$ )
$f_{\text{tower}}^{\min}$	0.05	$m_{\text{subject}}^{\max}$	12 GeV	$f_{\text{subject}}^{\min}$	0.5

### 2.3 $\mu_x$ tagging results

We extract the  $\mu_x$  tagging efficiency for individual jets by simulating detector reconstruction for samples of flavored dijets. We generate all samples at  $\sqrt{S} = 13$  TeV

using MADGRAPH5 v2.2.3 [80] with CT14llo PDFs [81]. We use PYTHIA 8.210 [7, 8] for all fragmentation, hadronization, and decay, using the default PYTHIA tune and PDF set for everything except pileup, for which we use the settings described in Table 7 of Ref. [82]. To allow in-flight muon production, we activate  $K_L^0$ ,  $K^\pm$  and  $\pi^\pm$  decays inside PYTHIA.

We use FastJet 3.1.2 [83] to reconstruct jets, and a modified version of DELPHES 3.2 [19] to simulate the ATLAS detector at the LHC. Since the  $\mu_x$  tag relies heavily on muon angle, with in-flight  $\pi^+/K^+$  decays being a large source of muon background, we developed a custom module `AllParticlePropagator` to properly handle such decays. The module which implements  $\mu_x$  tagging `MuXboostedBTagging` (available on GitHub [20]) can be used in conjunction with the default  $b$  tagging module `BTagging` in DELPHES. It is important to note that, until the more recent version of DELPHES (3.3), the default DELPHES cards define `BTagging` efficiencies which are *not accurate* at high  $p_T$  (e.g. light-jet fake rates are constant everywhere, and  $b/c$  jet efficiencies are constant for jets with  $p_T \gtrsim 150$  GeV). The DELPHES 3.3 efficiencies for 1–2 TeV jets are now 14–28% for  $b$ -tags and 1–2% for light jet fake rates. Our goal is to provide similar  $b$ -tagging efficiency with a factor of 10 improvement in fake rates.

Muon reconstruction efficiencies and  $p_T$  resolutions are taken from public ATLAS plots [84, 85] for *standalone* muons (muons seen in the Muon Spectrometer [MS], but not necessarily the main tracker). Because the MS experiences limited punch-through from non-muons, it can reconstruct muons with  $p_T \geq 10$  GeV with high efficiency (95–99%), even inside boosted jets. Because we focus on the ATLAS MS, our results reflect the holes for detector services and support feet, which cause (i) a dip in muon reconstruction efficiency at  $\eta = 0$  [86], precisely where the dijet  $dN/d\eta$  distribution peaks, and (ii) 80% geometric acceptance of the Level-1 muon trigger in

the barrel [87]. This latter restriction can be resolved by relying on jet triggers (jet  $p_T$  or event  $H_T$ ) to select pertinent events, since  $\mu_x$  tagging only works for high- $p_T$  jets.

There are several sources of standalone muon background which we are unable to simulate: (i) cosmic muons, (ii) decay muons from particles produced in the calorimeter shower, (iii) fake muons from punch-through, and (iv) fake muons from noise. Nonetheless, since the  $\mu_x$  tag is effectively a tight angular cut with a reasonably high  $p_{T,\mu}$  threshold, we expect these backgrounds to be negligible compared to the light jet background which we simulate.

The direction of the core is extremely important in  $\mu_x$  tagging, and tracks would provide the best information. However, the core’s intrinsic collimation hampers track reconstruction in a manner difficult to model in a fast detector simulation. As such, we build jets (and cores) solely from calorimeter towers and muons. The coarse granularity of the hadronic calorimeter (HCal) is mitigated by using the finer granularity of the EM calorimeter (ECal) to orient the combined tower (“ECal pointing”). This is implemented in DELPHES’ `Calorimeter` module by giving both ECal and HCal the segmentation of ATLAS’s ECal Layer-2 ( $\Delta\phi \times \Delta\eta = 0.025 \times 0.025$  in the barrel). To ensure that we are not overly sensitive to this resolution, we also test a granularity twice as coarse ( $0.05 \times 0.05$ ), finding negligible degradation in the heavy jet tagging efficiency, with only a slight rise in light jet fake rate (1.2 times larger at  $p_T = 600$  GeV, but dropping to no increase at 2.1 TeV).

**2.3.1 Tagging efficiencies.** To test the  $\mu_x$  tag, we create samples of  $b\bar{b}$ ,  $c\bar{c}$ , and  $j\bar{j}$  ( $j \in \{u, d, s, g\}$ ) spanning  $p_T = 0.1$ –2.1 TeV. We then find the efficiency to tag the top two jets (ranked by  $p_T$ ) in each event. Since heavy hadrons from gluon splitting ( $g \rightarrow b\bar{b}/c\bar{c}$ ) are an inevitable component of our light-jet sample, especially at high  $p_T$ , it is important to determine the extent to which this background can be reduced.



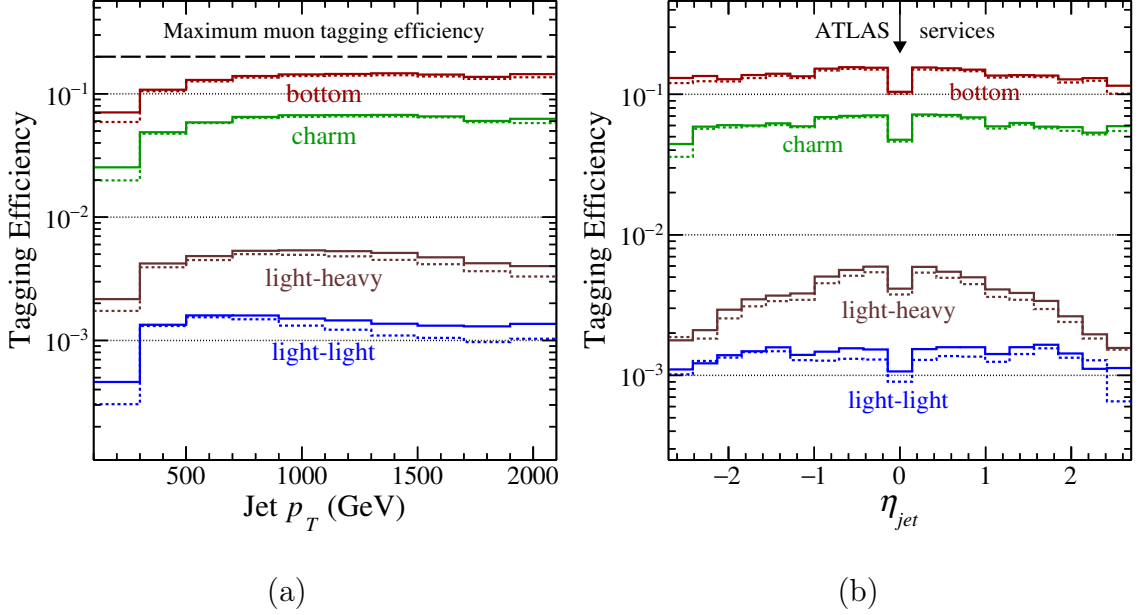


Figure 2.7.  $\mu_x$  tagging efficiency vs. (a) jet  $p_T$  and (b)  $\eta_{jet}$ .

We sort the light jet sample via the truth-level flavor of a taggable muon’s primary hadronic precursor. This classifies each attempted tag as *light-heavy* (where the muon descends from a  $b/c$  hadron inside a jet initiated by a light parton) or *light-light* (where the muon’s lineage is purely light-flavored).

In Figure 2.7 we show our predicted efficiencies for the four classes of  $\mu_x$  tags. The solid lines represent the efficiencies without pileup, while the dotted lines show the efficiencies when a random number of pileup events (drawn from a Poisson distribution with  $\mu = 40$ ) are added to each hard event. Since we do not utilize non-muon tracking, and are working with TeV-scale jets, we do not attempt any pileup subtraction.

Each  $p_T$  bin in Figure 2.7a sums over all available  $\eta_{jet}$ . When the boosted approximations are valid (jet  $p_T \geq 300$  GeV), the efficiency to tag heavy jets is nearly flat versus  $p_T$ , while the efficiency to tag light jets decreases slightly. We find asymptotic tagging efficiencies of  $\epsilon_b = 14\%$ ,  $\epsilon_c = 6.5\%$ ,  $\epsilon_{light-light} = 0.14\%$ , and

$\epsilon_{\text{light-heavy}} = 0.5\%$ , respectively. This light-light rejection provides us the full factor of 10 improvement over existing algorithms. At low- $p_T$  (where  $B$  hadrons are no longer strongly boosted and track tagging is superior) all  $\mu_x$  efficiencies plummet, although the relative rates remain approximately the same. Notice that pileup actually *improves* the performance of  $\mu_x$  tagging above 1 TeV, causing almost no degradation in heavy-jet efficiencies, but a significant drop in light-jet efficiency. This is a consequence of the increased probability for light jets to reconstruct a subjet with a low fraction of total jet energy, thereby failing the cut on  $f_{\text{subjet}}$ .

Since the  $\mu_x$  tag is not effective at low  $p_T$ , each  $\eta_{\text{jet}}$  bin in Figure 2.7b requires  $p_T \geq 300$  GeV. We can see that both heavy and light-light jet efficiencies are flat with  $\eta_{\text{jet}}$ . The light-heavy efficiency decreases significantly with  $|\eta_{\text{jet}}|$ , indicating a rising rejection of heavy hadron background from gluon splitting. This suggests the intriguing possibility that the  $g \rightarrow b\bar{b}$  contribution to  $b$  jets could be extracted from data, and used to calibrate the Monte Carlo event generators for highly boosted jets.

The validity of the cuts made by the  $\mu_x$  tag is evident in Figure 2.8, where we plot a 2D histogram of the  $x$  and  $f_{\text{subjet}}$  for each jet. The  $x$  distribution for *bottom* jets peaks at  $x \approx 0.8$ , versus the theoretical prediction of  $x \approx 0.6$  from Equation 2.14. This discrepancy occurs because half the muons in  $b$  jets are from secondary-charm decay, which have a larger  $x$  on average (the charm jets peak around  $x = 1$  in Figure 2.8b). Nonetheless, most values of  $x$  are smaller than  $x_{\text{max}}$  and, in both heavy-jet classes, the  $f_{\text{subjet}}$  distributions favor subjets carrying nearly all of their jet's momentum.

The  $x$  distribution for *light-light* jets with sufficiently hard subjets passing  $f_{\text{subjet}}^{\text{min}}$  peaks to the right of  $x_{\text{max}}$ , whereas muons with taggable  $x$  tend to be clustered into overly soft subjets. Since *light-heavy* jets actually contain heavy hadrons, their high- $f_{\text{subjet}}$  muons should (and do) have  $b$ -like values of  $x$ . However, since the initial jet momentum must be shared between a pair of heavy hadrons, many light-heavy muons



While we derive the  $\mu_x$  boosted-bottom-jet tag from basic kinematics in Section 2.1, in this chapter we examined its effectiveness at the LHC in the context of the ATLAS detector. This choice is driven by the public ATLAS standalone/non-isolated muon reconstruction capabilities as a function of  $p_T$  and  $\eta$  [85]. We ensure that our  $b$  tag is robust in a realistic detector environment by simulating ATLAS detector subsystems in DELPHES, and establishing an insensitivity to the detector details. Given that the  $\mu_x$  boosted- $b$  tag is driven by physical principles, and not detector idiosyncrasies, we are confident it will work just as well with the CMS detector provided they can reconstruct the non-isolated muons.

Naturally, the  $\mu_x$  tag will require experimental validation using heavy-flavor enriched and deficient control samples from CMS and ATLAS. A comparison of the  $\mu_x$  tag to existing  $b$  tags around 500 GeV (the lowest energy with good efficiency overlap) will permit the extension of the  $\mu_x$  tag to the highly boosted regime, where smaller uncertainties are sorely needed. The  $b$  jet efficiency could be extracted from  $t\bar{t}$  events ( $\sim 36\%$  of which should contain a semi-muonic  $B$  decay). To calibrate the light-jet fake rate we suggest looking in a light-jet enriched dijet sample: where one jet lacks a muon and fails a “loose” track-based  $b$  tag, and the other jet contains a muon.

It is possible that additional improvements to the  $\mu_x$  tag can be made using capabilities specific to a given experiment. For example, the final layer of the ATLAS inner detector has very fine  $\phi$  resolution, while the first layer of the ATLAS ECal has excellent  $\eta$  resolution. Since the direction of the “core” subjet is more important than the properties of its charged constituents (track quantity, impact parameters, opening angles), it may be possible to interrogate the *global* nature of the core without attempting to reconstruct its individual tracks. Given enough angular resolution, a direct measurement of  $m_{\text{core}}$  could replace its manual constraint. This procedure is

essentially an extension of CMS's particle flow algorithm to very boosted hadronic substructure.

Having verified the effectiveness of the  $\mu_x$  tag, which fills a hole in LHC tagging performance, we can move to the next chapter, where we will study the application of  $\mu_x$  to searches for new, heavy resonances. These results show that the implementation of the  $\mu_x$  tag at the LHC will greatly improve the LHC's sensitivity to important new physics.

## CHAPTER 3

## SEARCHES FOR NEW PHYSICS VIA HEAVY-FLAVOR JETS

The development of the  $\mu_x$  tag was motivated by the need for a better  $b$  tagging scheme for TeV jets, so that the LHC may continue the search for heavy resonances. To test the efficacy of our new method, we perform a full signal and background study for two very different resonances. The first study is the search for a leptophobic  $Z'$ , whose  $Z' \rightarrow b\bar{b}$  decay was the direct motivation for the  $\mu_x$  tag. As such, Section 3.1 can be treated as an immediate continuation of Chapter 2. The second study is a search for the charged Higgs boson, which arises when nature has *two* Higgs fields; this study will use an improved implementation of the  $\mu_x$  tag.

### 3.1 Leptophobic $Z'$

*Note: To first approximation, this section should be considered a verbatim reprint of the last few sections of my journal article [4], slightly expanded for greater clarity.*

A simple extension of the Standard Model involves the addition of a broken  $U(1)'$  symmetry mediated by a heavy neutral  $Z'$  boson. If the new symmetry is associated with baryon number  $B$ , one would not expect to see a dilepton signal, since only SM quarks would be charged under the  $U(1)'$ . To cancel anomalous couplings (currents arising from nowhere), this  $U(1)'_B$  should couple to vector-like quarks, and come with at least one scalar field whose vacuum expectation value breaks the symmetry [46, 47]. If the vector-like quarks are kinematically inaccessible at the LHC (i.e., they are too heavy to produce), a flavor-independent  $Z'_B$  gauge coupling to SM quarks [46]

$$\mathcal{L} \subseteq \frac{g_B}{6} Z'_{B\mu} \bar{q} \gamma^\mu q \tag{3.1}$$

would mean that the dijet decay of the  $Z'$  is the only detectable signature of this new physics at the LHC.

We would expect the purity of a dijet  $Z'$  signal to be very low, since QCD production of dijets has an enormous cross section. This is where  $\mu_x$  tagging is useful, as the rejection of light-jet fakes seen in Section 2.3.1 is  $\mathcal{O}(10^3)$ . We simulate a search for a narrow  $Z'_B$  peak above the dijet background at Run II of the LHC (i.e., looking for an excess in the  $d\sigma/dm_{jj}$ ). We examine the experimental reach in two dijet samples: 2-tag and 1-tag inclusive (where  $N$ -tag requires at least  $N$  of the top two  $p_T$ -ranked jets to be  $\mu_x$ -tagged).

We model  $Z'_B$  production for a variety of  $M_{Z'_B}$  spanning 1–4 TeV, using the production mode  $pp \rightarrow Z'_B \rightarrow b\bar{b}/c\bar{c}(j)$  at  $\sqrt{S} = 13$  TeV. The cross section involving the radiation of one additional jet slightly enhances the overall  $Z'_B$  rate, but is mostly useful to improve the differential jet distribution to more closely approximate the next-to-leading-order result. This is accomplished via MLM jet matching [88] in both MADGRAPH and PYTHIA (in “shower-kt” mode [89], using a matching scale of  $M_{Z'_B}/20$ ). We reconstruct  $R = 0.4$  anti- $k_T$  jets using **FastJet**, and  $\mu_x$  tag them via our modified DELPHES code that simulates  $\mu_x$  tagging at the ATLAS detector (as described in Sections 2.2.2 and 2.3).

The relevant background is pure QCD, as no other SM processes have competing cross-sections. Both 2-tag and 1-tag backgrounds include  $pp \rightarrow b\bar{b}/c\bar{c}/j\bar{j}(j)$ . The 1-tag background also includes a large contribution from  $pq_h \rightarrow jq_h(j)$  (a heavy quark from one proton scattering off any parton from the other proton). To obtain good tagging statistics, multiple background sets are generated, using identical MLM matching parameters as their corresponding signal set.

The minuscule light-jet tagging efficiency forces us to estimate the *second* tag

for the 2-tag light-dijet background sample. First we bin the light-jet sample from Section 2.3.1 into a two-dimensional  $p_T \times \eta_{\text{jet}}$  histogram of tagging efficiency. We then fit the surface of this histogram with a continuous function  $\epsilon_l(p_T, \eta_{\text{jet}})$ . When exactly one leading jet is tagged (the “tagged” jet), we use this surface to estimate the probability  $\epsilon_l$  to tag the other jet (the “estimated” jet), we then use Bayes’ theorem to estimate the conditional probability of a 2-tag event, given the observation of only one tag (with the number of tags represented by an integer)

$$\Pr(2|1) = \frac{\Pr(1|2) \times \Pr(2)}{\Pr(1)} \approx \frac{\epsilon_l^2}{2\epsilon_l(1 - \epsilon_l)} = \frac{\epsilon_l}{2(1 - \epsilon_l)}. \quad (3.2)$$

This probability  $\Pr(2|1)$  is used to re-weight the 1-tag event. Its denominator accounts for the fact that when there is only one tag, either jet could have been the “tagged” jet, which effectively doubles the 1-tag rate of a two-jet event (since  $1 - \epsilon_l \approx 1$ ). When both leading jets are tagged, the event must be discarded, otherwise it would be double-counted by this method.

Additional cuts for our analysis include a requirement that the pseudorapidity interval between jets is small,  $|\Delta\eta_{jj}| \leq 1.5$ , in order to suppress much of the  $t$ -channel dijet background. We also require  $|\eta_{\text{jet}}| \leq 2.7$  to ensure that both jets fall within the muon spectrometer. While we considered including the effects of higher order final state radiation in our mass reconstruction, we find that adding a hard third jet to the dijet system causes an unacceptable hardening of the QCD continuum. Not including this radiation, combined with the estimation of hard neutrino momenta inherent to  $\mu_x$  tagging, degrades the mass resolution of the intrinsically narrow  $Z'_B$  bosons of this model ( $\Gamma_{Z'} = \frac{1}{6}\alpha_B(1 + \alpha_s/\pi)M_{Z'}$ ). Hence, we require a rather wide mass window ( $[0.85, 1.25] \times M_{Z'_B}$ ) to capture most of the signal.

The signal and backgrounds for a  $5\sigma$  discovery of a  $M_{Z'_B} = 2.5$  TeV  $Z'_B$  boson, using our cuts for the 2-tag and 1-tag analyses, can be seen in Figure 3.1 for  $100 \text{ fb}^{-1}$  of integrated luminosity at the 13 TeV LHC. The signal to background ratio



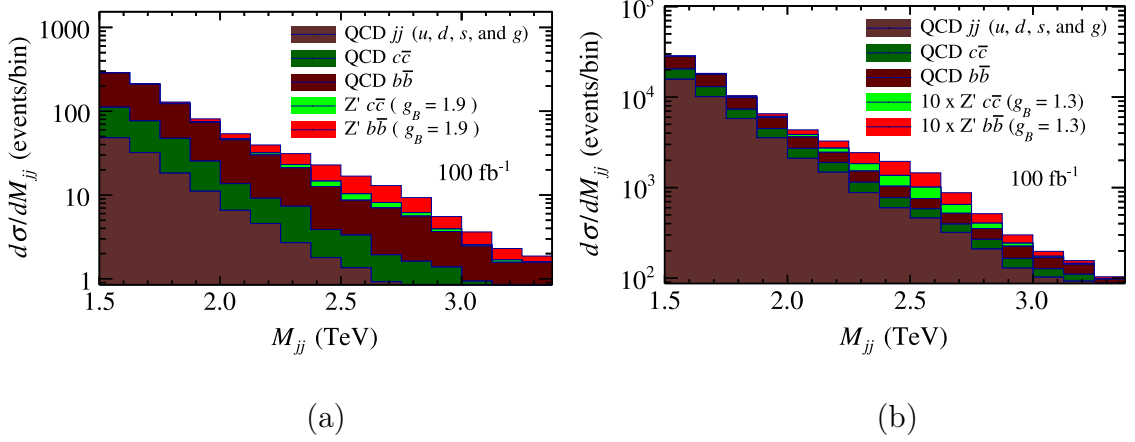


Figure 3.1. Events per bin expected for  $5\sigma$  discovery of a  $M_{Z'_B} = 2.5$  TeV signal, and backgrounds, in the (a) 2-tag and (b) 1-tag analyses using  $100 \text{ fb}^{-1}$  of integrated luminosity at Run II of the LHC.

$S/B = 1/2$  for the 2-tag sample, indicating an excellent purity. The 1-tag sample has  $S/B = 1/12$ , still acceptable given the factor of 12 more signal events that would appear in the sample. The peak in the 1-tag sample is slightly narrower than that in the 2-tag sample because only one neutrino is estimated via the muon proxy.

In Figure 3.2 we depict the estimated discovery potential for the 2- and 1-tag analyses, along with the 1-tag 95% confidence level (C.L.) exclusion limits, for the LHC Run II with the scheduled luminosity of  $100 \text{ fb}^{-1}$ . In Figure 3.3, we overlay these predictions on the existing exclusions obtained from Ref. [46]. Our 2-tag discovery reach is about 500 GeV higher in mass for large coupling constant  $g_B$ , and is right at the limit for smaller  $g_B$ . Not shown in the figure is the 95% C.L. exclusion limit for the 2-tag search, which is slightly better than the  $5\sigma$  discovery reach in the 1-tag search. The 1-tag search dramatically improves the mass reach by  $\sim 1.5$  TeV beyond the current limits at large  $g_B$  and, more importantly, can attain  $g_B < 1$  below 2(3) TeV for discovery(exclusion). The  $\mu_x$  boosted-bottom tag opens a new window into leptophobic  $Z'$  boson physics.

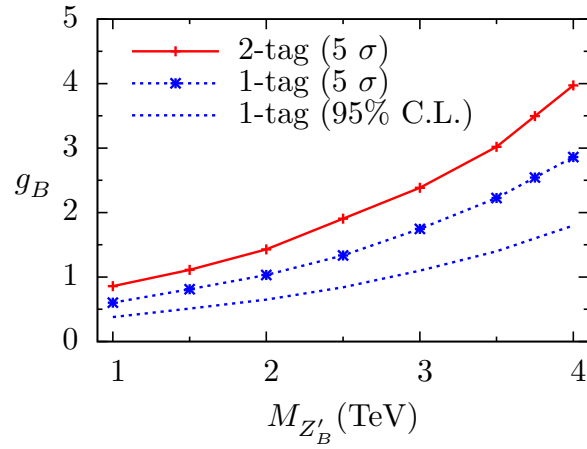


Figure 3.2. Estimated Run II ( $100 \text{ fb}^{-1}$ )  $5\sigma$  discovery potential and 95% confidence level exclusion limits for  $g_B$  vs.  $M_{Z'_B}$  for the 2-tag and 1-tag analyses. 2-tag 95% C.L. exclusion reach (not shown) is comparable to the 1-tag discovery reach.

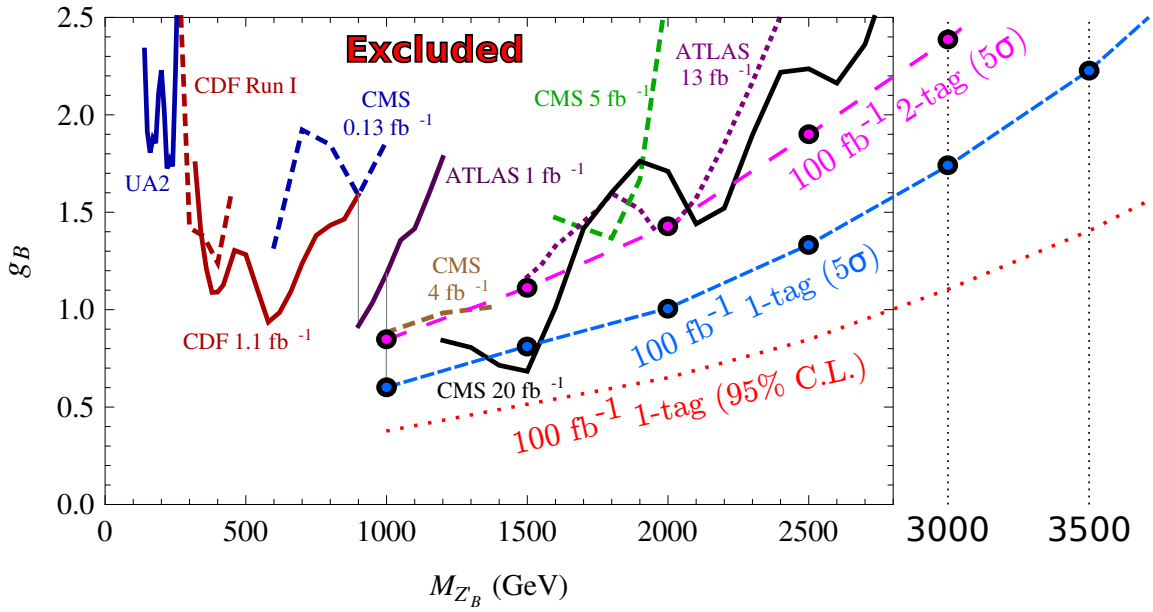


Figure 3.3. The estimates of Figure 3.2 overlaid on existing experimental exclusions for leptophobic  $Z'$  at 95% C.L. [46].

### 3.2 Two Higgs doublet models

*Note: To first approximation, this section should be considered a verbatim reprint of my journal article [5], but slightly expanded for clarity.*

With the discovery of a 125 GeV boson at the CERN Large Hadron Collider (LHC) [90], one which behaves uncannily like the massive scalar of the Standard Model’s (SM) singular SU(2) doublet, the question turns to whether an additional scale of physics can be found in a collider environment. A generic way to accommodate another scale of symmetry breaking is to add an additional scalar field, creating a two Higgs doublet model (2HDM) [91]. 2HDMs are commonly associated with supersymmetry (SUSY) [92–96], but they also appear in axion models suppressing CP violation in QCD [35, 97] and to explain the matter/antimatter imbalance [98–100].

2HDM are primarily characterized by  $\tan \beta$  (the ratio of the two vacuum expectation values of the Higgs doublets) and  $\beta - \alpha$  (the doublet mixing angle). Symmetry breaking produces four scalar Higgs bosons ( $h, H, H^\pm$ ) and a pseudo-scalar boson ( $A$ ). If the fine tuning of the various parameters is minimal, then  $h$  is the lightest physical particle [101]. Given that a wide range of measurements have effectively ruled out flavor changing neutral currents at tree-level, realistic 2HDM are restricted to four general models [101], of which two are worth noting here: type-I, where all quarks couple to only one of the doublets, and type-II, where  $u_R^i$  and  $d_R^i$  couple to opposite doublets (a requirement of SUSY). We restrict our attention to type-II Higgs theories.

The SM-like nature of the recently discovered scalar boson (especially in its per-channel signal strength [102]) constrains many type-II 2HDM rather tightly to the limit where the mixing angle “aligns”:  $\cos(\beta - \alpha) \rightarrow 0$  [103, 104]. This forces the

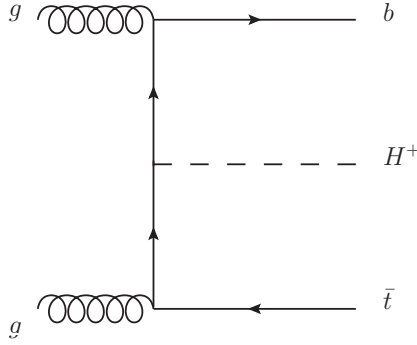


Figure 3.4. A Feynman diagram for the leading order associated production of a charged Higgs, where each proton contributes a gluon [108].

2HDM Higgs  $h$  to look like a Standard Model Higgs  $H_{\text{SM}}^0$ .<sup>11</sup> If there is also a near-degeneracy in the masses of the remaining 2HDM Higgs  $H$ ,  $A$ , and  $H^\pm$  — a natural consequence of SUSY when it does not couple strongly to normal matter [105, 106] (as all experimental evidence suggests) — then these heavy Higgs are kinematically forbidden from decaying to each other. This mass degeneracy also occurs in more generic 2HDM models which favor natural SM alignment without decoupling (e.g., softly broken  $\text{SO}(5)$  [107]). For these reasons, we explore the degenerate mass sector, where the coupling of the heavy charged Higgs boson to the Standard Model is dominated by the heavy third generation.

Detecting  $pp \rightarrow H/A$  is difficult as both the signal and background have identical initial and final states ( $gg \rightarrow q\bar{q}$ ), and the resulting interference gives  $H/A$  resonances an unusual dip-then-bump shape that is more easily mimicked by statistical fluctuations in QCD background than a traditional symmetric bump [109, 110]. Measuring  $H/A$  in association with an additional heavy quark pair eliminates this interference. For a charged Higgs boson, associated production is the leading order production mode ( $pp \rightarrow H^\pm t(b)$ ) (as depicted in Fig. 3.4), where the associated  $b$  is sometimes resummed into the beam fragments, and thus is not always in the

<sup>11</sup>For  $\tan\beta > 10$ , non-aligned 2HDM are still allowed on a thin trajectory.

measurable final state. This study focuses on the charged Higgs production channel.

Assuming quasi-degeneracy of the heavy Higgs masses, one finds [101]

$$\mathcal{L}_{\text{eff}} = -H^+ \bar{t} (y_t P_L + y_b P_R) b + \text{h.c.}, \quad (3.3)$$

where  $P_{L/R}$  are the chiral projection operators and the Yukawa couplings  $y$  are similar to the Standard Model Higgs; proportional to the running quark mass over the electroweak vacuum expectation value  $v$ :  $y_t = \sqrt{2} m_t \cot \beta / v$  and  $y_b = \sqrt{2} m_b \tan \beta / v$ . Appealing to naturalness ( $\tan \beta = \mathcal{O}(1)$ ), and keeping  $y_{tb} = \sqrt{y_t^2 + y_b^2}$  sensibly perturbative ( $y_{tb} \lesssim 1$ ), leads to the expectation that

$$\tan \beta \geq \frac{\sqrt{2} m_t}{v} \quad \text{and} \quad \tan \beta \leq \frac{v}{\sqrt{2} m_b}, \quad (3.4)$$

which corresponds to  $\tan \beta \in [0.83, 73]$  at  $Q_{2\text{HDM}} = 2 \text{ TeV}$ . At the center of this region ( $\tan \beta = \sqrt{m_t/m_b}$ ) lies a “wedge” of low production cross section, where the coupling transitions from top-dominated at low  $\tan \beta$  to bottom-dominated at large  $\tan \beta$ . The wedge is quite visible in recent experimental searches for  $H^\pm$  using 8 TeV LHC run-I data [108, 111], as shown in Figure 3.5. That the wedge obfuscates a large swath of interesting parameter space is quite evident from the dotted line, which shows the  $\tan \beta$  predicted by the Minimally Supersymmetric Standard Model (MSSM, the simplest version of a functional supersymmetry). This prediction cuts squarely through the center of the wedge.

The wedge should shrink in the LHC’s run-II, with a larger collision energy and more data, but the sensitivity predictions range from rather pessimistic for  $m_{H^\pm} = 0.5\text{--}1 \text{ TeV}$  [104] — as seen in Figure 3.6a, where the wedge does not close — to quite optimistic for  $m_{H^\pm} = 0.5\text{--}2 \text{ TeV}$  [112, 113] — as seen in Figure 3.6b where the wedge just barely closes at the LHC, and is completely closed at a 100 TeV collider for  $m_{H^\pm} \lesssim 10 \text{ TeV}$ . It is our assessment that the variations in previous estimates are primarily due to choices made when simulating a standard “track-vertex”  $b$  tag to

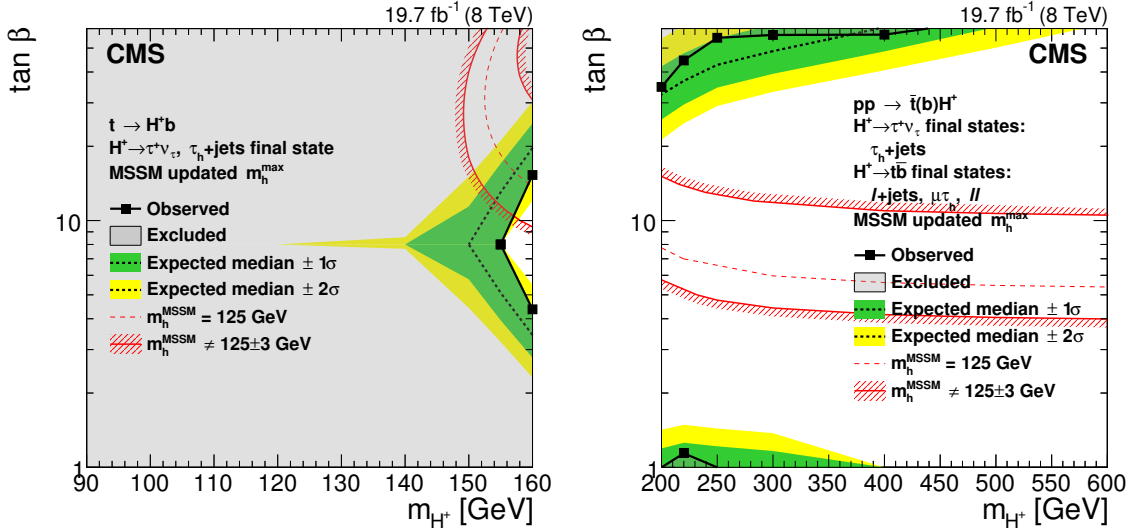


Figure 3.5. A CMS search for  $H^\pm$  found no charged Higgs in the gray regions of parameter space, excluding 2HDM with those  $\tan\beta$  and  $m_{H^\pm}$ . CMS is not sensitive to 2HDM in the white “wedge” region that opens up at large charged Higgs mass (as well as to  $m_{H^\pm} \approx m_t$ , which bisects the figures) [111].

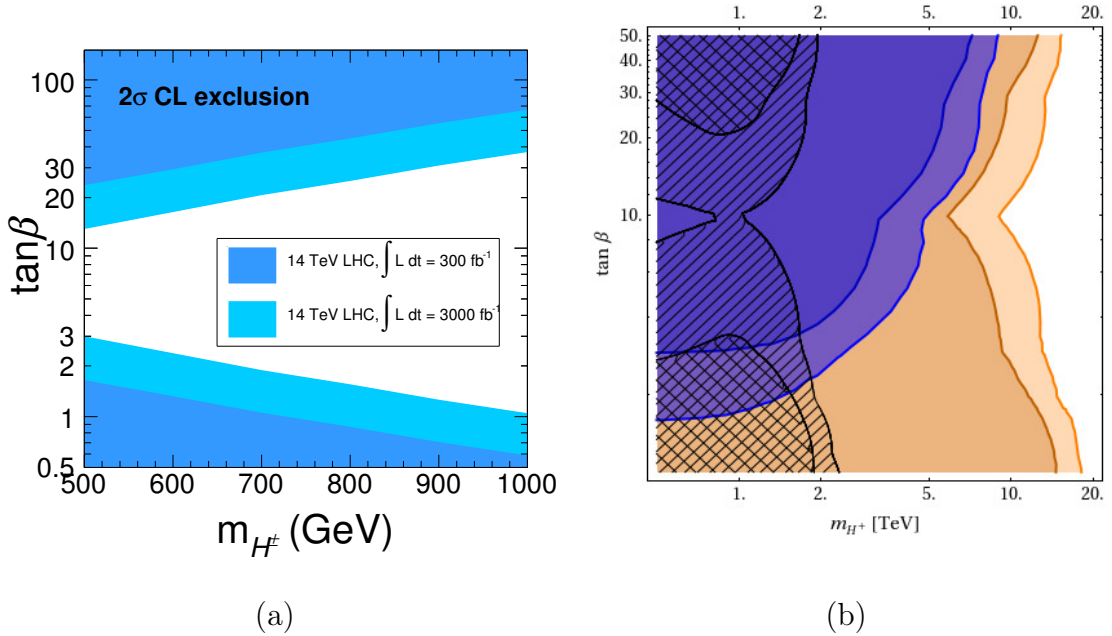


Figure 3.6. The theoretical reach for charged Higgs mass near 1 TeV; (a) at the LHC [104] and (b) at the LHC (hashed) and a 100 TeV future collider (light/salmon) [112]. These predictions are inconsistent with each other.

suppress QCD background. This becomes more difficult as the mass of the charged Higgs moves above one TeV, as the bottom quark become significantly boosted, making theoretical predictions sensitive to careful modeling of real-world degradation of  $b$ -tagging efficiencies for boosted bottom jets.

In this section, we predict the experimental reach for  $m_{H^\pm} > 1$  TeV through its associated production with a top quark, and its decay to boosted top and boosted bottom jets, in both a generic two Higgs doublet model and in SUSY. In Section 3.2.1 we describe our selection cuts and tagging efficiencies in the boosted regime. In Section 3.2.2 we present our numerical results for the LHC at  $\sqrt{S} = 14$  TeV. We find that the LHC has limited reach to observe a charged Higgs boson, and so extend our examination to show the reach of a 100 TeV future circular collider (FCC).

**3.2.1 Methods.** Given the disparity between previous predictions for the reconstruction of  $tH^\pm \rightarrow ttb$  at large charged Higgs boson mass, this study concentrates on careful modeling of boosted bottom jets at the LHC and at a FCC. In this section, we address improvements to our existing  $\mu_x$  boosted-bottom-jet tag, and detail improvements to signal selection over previous studies.

**3.2.1.1 Bottom-jet tagging.** In the previous sections, we implemented the  $\mu_x$  tag by utilizing the resolution of the electromagnetic calorimeter and avoided using tracks [4]. In the present study, we improve upon our prior implementation by allowing  $\mu_x$  to access high-resolution angular information in tracks to locate the jet core. While we find that combining tracking with normal-resolution calorimetry does not change the tagging efficiency at 14 TeV, tracking becomes absolutely essential at 100 TeV. The large radius (6 m) and strong magnetic field (6 T) of the hypothetical FCC tracking system [19] smears the charged constituents in  $\phi$ , reducing the correlation between charged tracks and the towers they strike, severely diminishing the

usefulness of tower-only jets.

**3.2.1.2 Signal selection.** There are two major production modes for  $tH^+$  at a proton collider: the “4b” final state  $gg \rightarrow [H^+ \rightarrow \bar{t}b]t\bar{b}$  (with  $t \rightarrow bW^+$ ), and the “3b” final state  $gb \rightarrow (H^+ \rightarrow \bar{t}b)t$ . Since the 3b final state is the dominant mode, accounting for at least 60% of the total cross section for all masses, the inclusive (3b + 4b) final state is a natural starting point. This requires tagging a boosted bottom jet and two tops: a boosted top jet from the  $H^\pm$  decay, and a much softer, resolvable, associated top.

Using the  $\mu_x$  tag to identify the boosted- $b$  jet unavoidably selects events containing hard neutrinos from semi-leptonic  $B$  hadron decay. This smears the missing  $\not{p}_T$  of any leptonically decaying tops, reducing the effectiveness of  $\not{p}_T$  for top identification or reconstruction, and limiting  $H^\pm$  mass resolution if the boosted top decays leptonically. These limitations are easily side-stepped by using only the fully hadronic decay of the boosted top, tagging the unique shape of  $t \rightarrow W^+b$  merged into a single “fat” jet [58]. Conversely, the associated top is slow enough to be resolved into isolated daughters, so its fully hadronic final state is quite susceptible to QCD background. It is safer to resolve the associated top into an isolated lepton ( $e/\mu$ ) and a  $b$  jet (which is soft enough that high-efficiency track tags remain robust).

The  $t\bar{t}$  portion of the inclusive final state provides multiple handles to suppress pure multijet background, leaving  $ttj+X$  the dominant background (where  $j = guds$ ). Here, the light flavored jet is both hard and “mis-tagged” as a primary boosted- $b$  jet. This usually occurs when the jet showers  $g \rightarrow b\bar{b}$ , creating a real  $B$ -hadron inside a jet of light-flavor origin. The sub-dominant background is  $tt(bb/cc)$  — effectively the same final state, but with the gluon splitting at a much higher scale. Other final states (e.g.,  $tjj + X$  and  $ttbj$ ) are found to be negligible.



Event reconstruction begins with jet reconstruction. First, “narrow” jets are clustered using an anti- $k_T$  algorithm with  $R = 0.4$  [33], and “fat” jets are clustered using a Cambridge-Aachen algorithm with  $R = 0.8$  [32]. Both boosted jets must have  $p_{Tj} \geq 350$  GeV, and all jets must have  $p_{Tj} \geq 20(40)$  GeV for 14(100) TeV collisions. Additionally, all jets must have  $|\eta_j| < 2.1(3.0)$ , so that the edge of the tracker lies outside the clustering radius of narrow jets. We require exactly one isolated lepton with  $p_T^{\text{lepton}} > 15(25)$  GeV. The lepton is considered isolated if  $p_T^{\text{lepton}} / \sum_i p_T^i < 5\%$  for all tracks and towers within a cone of  $\Delta R < 10 \text{ GeV} / p_T^{\text{lepton}}$ , as prescribed in a recent experimental search [114]. Additionally, the lepton cannot fall within a  $\Delta R = R_{\text{cluster}}$  cone surrounding any of the candidate jets.

Narrow jets are sorted by  $p_T$  (high to low), and the first narrow jet which is  $\mu_x$  tagged becomes the boosted  $b$  candidate. To exclude the situation where the boosted top decays leptonically (and the associated top hadronically), we require that the boosted- $b$  plus lepton system has a mass inconsistent with a top quark ( $m_{bl} > 172$  GeV). This cut is primarily used to properly model the  $ttj + X$  background, but is redundant in other systems because it effectively overlaps the requirement that the lepton reside outside of the boosted- $b$  jet.

Next, fat jets are sorted by  $p_T$ , and the first one which has a boosted hadronic top tag is the boosted top candidate. We then require that  $\Delta R_{bt} \geq 2$  and  $|\Delta\eta_{bt}| \leq 2$  for the two boosted candidates. The latter cut is used to restrict  $t$ -channel background from hardening the tail of the  $m_{bt}$  distribution, although it removes about a fifth of all  $H^\pm$  (whose isotropic decay is minimally boosted in the transverse direction, due to its large mass). We do not impose any constraints on the mass of the boosted top jet, as these are already built into the boosted top tag.

We then attempt to reconstruct the associated top by finding a  $b$  jet compatible with the isolated lepton. From the set of narrow jets whose  $p_T$  is smaller than that

of the boosted  $b$ , we take at most two jets which are  $b$ -tagged and reside outside an  $R = 1.2$  cone around the boosted top (which should contain its own  $b$  jet). We then attempt to find a  $b$ -lepton system with  $p_T$  less than that of the boosted top, and an invariant mass consistent with a top quark missing its neutrino ( $70 \text{ GeV} < m_{bl} < 180 \text{ GeV}$ , where the slightly elevated ceiling permits detector smearing). If two  $b$  candidates pass these cuts, the one whose  $m_{bl}$  is above 110 GeV is selected; if both are above 110 GeV, the one which is closer to 110 GeV is selected.

The total branching ratio of the hadronic/leptonic  $t\bar{t}$  decay (14%), combined with the efficiency of the two boosted flavor tags ( $\epsilon_b \approx 0.14$  and  $\epsilon_t \approx 0.45$ ) and the event shape cuts for the inclusive final state, produce an overall  $H^\pm$  acceptance of  $\mathcal{O}(0.1\%)$ . The QCD background acceptance is an order of magnitude lower, though a more important consideration is the ratio of  $ttj+X$  to  $tt(bb/cc)$ . For the inclusive cuts, the ratio is consistently about 5:1, which is small enough that there is no clear benefit to independently reconstructing the  $4b$  final state, as was previously done [104, 112], since the process is already signal constrained at the LHC.

**3.2.2 Results.** We calculate all cross sections using a generic 2HDM from FeynRules [115–118] with MADGRAPH 5 v2.3.3 [80] and the CT14llo parton distribution functions [81]. Events are showered and hadronized using PYTHIA 8.210 [7, 8], and reconstructed using FastJet 3.1.3 [83] and the DELPHES 3 [19] detector simulation. For the 14 TeV analysis, we modify the ATLAS card supplied with DELPHES to simulate the  $\mu_x$  boosted  $b$  tag (using the MuXBoostedBTag module available on GitHub [20]). Both the track-based  $b$  tag and the boosted top tag are applied using an efficiency function  $\epsilon_f(p_T)$  based upon jet  $p_T$  and truth-level jet flavor  $f$ . For the track-vertex  $b$  tag, we use the run 2 efficiency from the ATLAS card (based upon Ref. [119]), and for the top tag, we use the efficiencies depicted in Ref. [58], which closely match those given in more recent publications [62, 120]. At 100 TeV, we use the FCC card supplied

with DELPHES (again modified to simulate  $\mu_x$ ), with two major changes: (i) we use the same track-vertex  $b$  tagging efficiency formula used for 14 TeV and (ii) we use a more conservative tracking domain ( $|\eta| \leq 3.5$ ).

At both collider energies, we use DELPHES’ “EFlow” objects (which subtracts track energy from the calorimeter towers they strike, after both tracks and towers have their energy smeared). We then cluster jets from tracks (minus isolated leptons) and track-subtracted towers. To estimate the neutrino  $\cancel{p}_T$  inherent to the  $\mu_x$  tag, we simply double the momentum of the tagging muon [4]. This does a reasonably effective job of reconstructing the  $H^\pm$  peak, allowing us to use a mass window of  $[0.9, 1.15] \times M_{H^\pm}$  at both 14 and 100 TeV to capture the majority of the signal. Without neutrino estimation, the  $H^\pm$  peak has a noticeably longer low-mass tail.

**3.2.2.1  $tH^\pm \rightarrow ttb$  in a generic 2HDM.** We first explore the reach for a charged Higgs boson produced in association with a top quark for a generic 2HDM. We convert the leading order  $y_{tb}$  used by MadGraph to a next-to-leading order  $y_{tb}$  by using the running quark masses at one-loop in QCD [121], which shifts the center of the  $\tan\beta$  wedge upwards. In Figure 3.7, we show the 95% confidence level (C.L.) limit for  $H^\pm$  exclusion at a 14 TeV LHC with 300 or 3000  $\text{fb}^{-1}$  of data. In order to compare directly with Refs. [104,112], we show (a) the limit obtainable on the effective Yukawa coupling  $y_{tb}$ , and (b) the corresponding region of  $\tan\beta$  probed. It turns out that the only Yukawa couplings  $y_{tb}$  or values of  $\tan\beta$  that can be probed at the LHC are on the border of the non-perturbative regions of parameter space.

The accessible region of parameter space at the LHC is entirely limited by the production cross section, as  $S/B = \mathcal{O}(1/2)$  across the entire mass range. Because the  $tH^\pm$  cross section at 14 TeV is quite small, the reach in  $\tan\beta$  is poor at the LHC. Once  $M_{H^\pm}$  surpasses 2 TeV, the  $H^\pm$  begins to grow noticeably off-shell, which weakens the narrow width approximation we use to extrapolate from our working

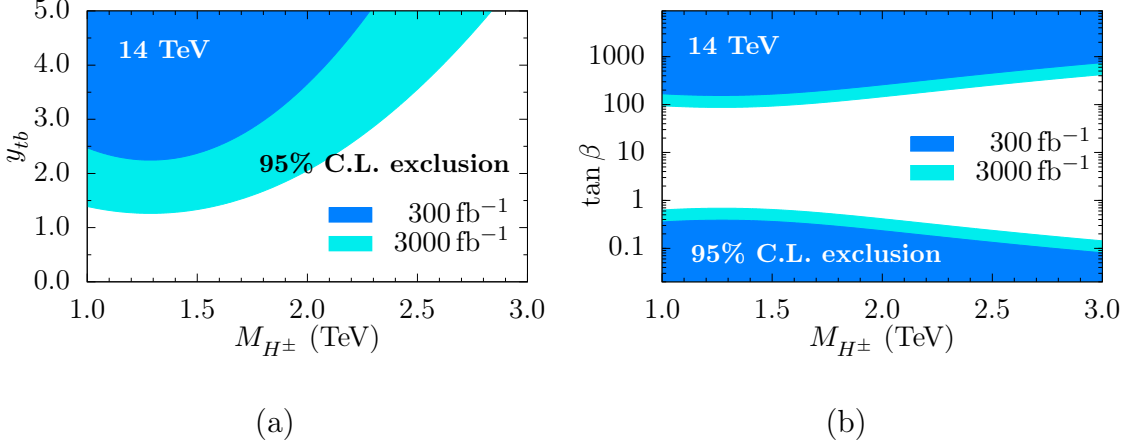


Figure 3.7. Predicted exclusion regime, at a 95% confidence level, for a generic 2HDM at a 14 TeV LHC in terms of (a) the effective Yukawa coupling  $y_{tb}$ , and (b) the corresponding  $\tan \beta$ .

value of  $\tan \beta$  to the 95% limit. The loss of reach approaching 1 TeV is due to signal/background attenuation, a combination of the 350 GeV minimum  $p_T$  cut imposed on both boosted jets and the swiftly diminishing efficiency of both boosted flavor tags below 500 GeV. Given this feature, our results are consistent with extending the predictions of Ref. [104] into the TeV regime. Charged Higgs bosons are unlikely to be observed at the LHC.

The  $tH^\pm$  cross section is strongly dependent on collider energy. A 100 TeV collider, such as a FCC, promises significantly more reach for charged Higgs bosons. At 100 TeV, the reach becomes background limited, with  $S/B$  rising from  $\sim 1\%$  at 1 TeV to  $\sim 5\%$  at 6 TeV. In Figure 3.8 we observe that the reach in effective Yukawa coupling is an order-of-magnitude better than at the LHC. This allows the wedge region to close as the integrated luminosity rises above  $3 \text{ ab}^{-1}$  up to a charged Higgs mass of 2 TeV. While this analysis is robust, more sophisticated techniques — boosted decision trees (BDT) or neural nets (NN) — might improve the reach. However, since BDT/NN techniques are highly dependent on the quality of the observables with which they train, it is difficult to make accurate predictions this far from a realized

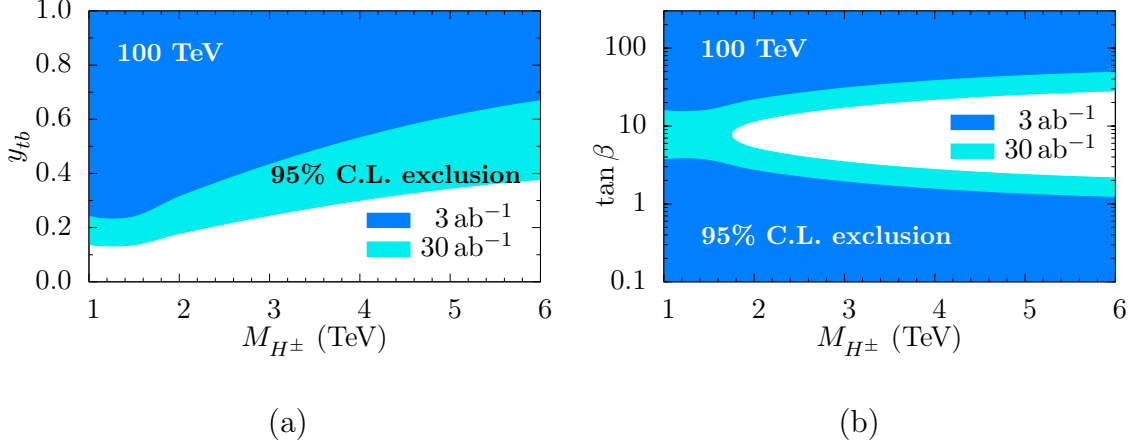


Figure 3.8. Predicted exclusion regime, at a 95% confidence level, for a generic 2HDM at a 100 TeV FCC in terms of (a) the effective Yukawa coupling  $y_{tb}$ , and (b) the corresponding  $\tan \beta$ .

100 TeV detector system, especially using a fast detector simulator. Regardless, our results suggest that search for TeV-scale charged Higgs bosons is the domain of future colliders.

**3.2.2.2  $tH^\pm \rightarrow ttb$  in a supersymmetric model.** One-loop corrections in the minimal supersymmetric Standard Model (MSSM) modify the fermionic couplings to  $H^\pm$  bosons. The effect is most significant for the bottom quark [121,122], and can be absorbed into the Yukawa coupling as

$$y_b^{\text{SQCD}} = y_b \frac{1}{1 + \Delta m_b} \quad (3.5)$$

(here we ignore supersymmetric electroweak corrections, using only those from supersymmetric QCD).  $\Delta m_b$  explicitly depends on the gluino mass, the mass of the two bottom squark eigenstates and  $\mu$ , the mass parameter coefficient of the  $\epsilon_{ij} H_i^1 H_j^2$  term in the superpotential. In the quasi-degenerate limit, where all these mass parameters are of equal size, only the sign of  $\mu$  survives [121]. At large  $\tan \beta$  ( $\sin \beta \approx 1$ )

$$\Delta m_b \approx \text{sign}(\mu) \frac{\alpha_s(Q_{\text{SUSY}})}{3\pi} \tan \beta, \quad (3.6)$$

where  $Q_{\text{SUSY}}$  is the heavy SUSY scale (which we take to be 10 TeV, although the

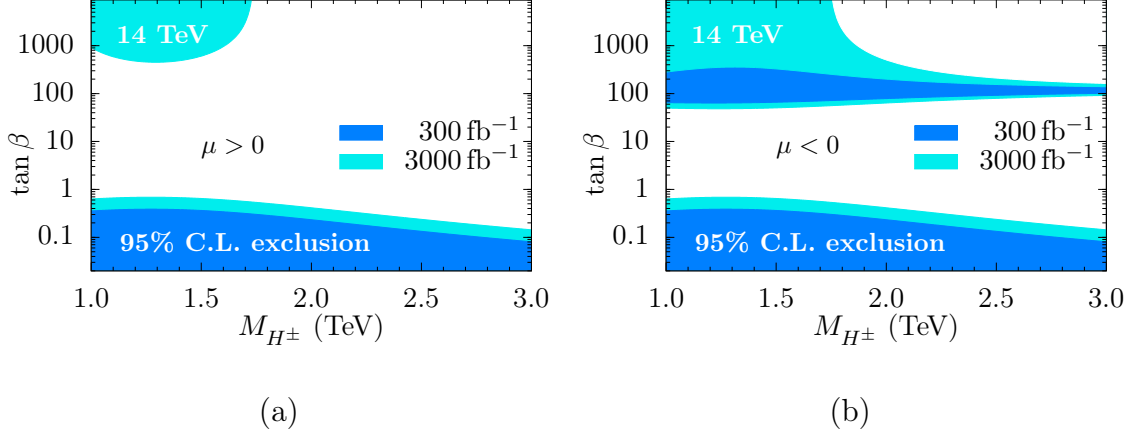


Figure 3.9. Predicted exclusion regime, at a 95% confidence level, for the MSSM at a 14 TeV LHC, with the sign of  $\mu$  (a) positive, or (b) negative.

result is not heavily dependent upon the choice of  $Q_{\text{SUSY}}$ , since  $\alpha_s$  runs slowly above a few TeV). These results are shown in Figure 3.9.

Comparing Figure 3.9 to Figure 3.7, it is readily apparent that the  $\Delta m_b$  correction has a significant impact on the reach at 14 TeV, where the production cross section is so small that only very large  $\tan\beta$  are accessible. For a positive  $\mu$ , the  $\Delta m_b$  correction counteracts the cross section enhancement of large  $\tan\beta$ , shifting high  $\tan\beta$  parameter space completely out of reach. Conversely, the negative  $\mu$  correction enhances the cross section beyond the generic 2HDM in a small region of  $\tan\beta \sim 100$ , but decreases it at larger values of  $\tan\beta$ . At small values of  $\tan\beta < 0.5$ , the top-quark Yukawa coupling becomes so large the theory is non-perturbative. If charged Higgs boson searches are difficult at the LHC in a generic 2HDM, in SUSY they are nearly impossible.

In stark contrast, Figure 3.10 shows that the effect of  $\Delta m_b$  is noticeable at a 100 TeV collider, but it manifests only as a moderate shift in the upper bound of the wedge, without a dramatic change in shape. This serves to underline the nature of the  $\Delta m_b$  effect; for a *signal limited* search (14 TeV), it is very important, while for a *background limited* search (100 TeV) it is more-or-less negligible. The lack of

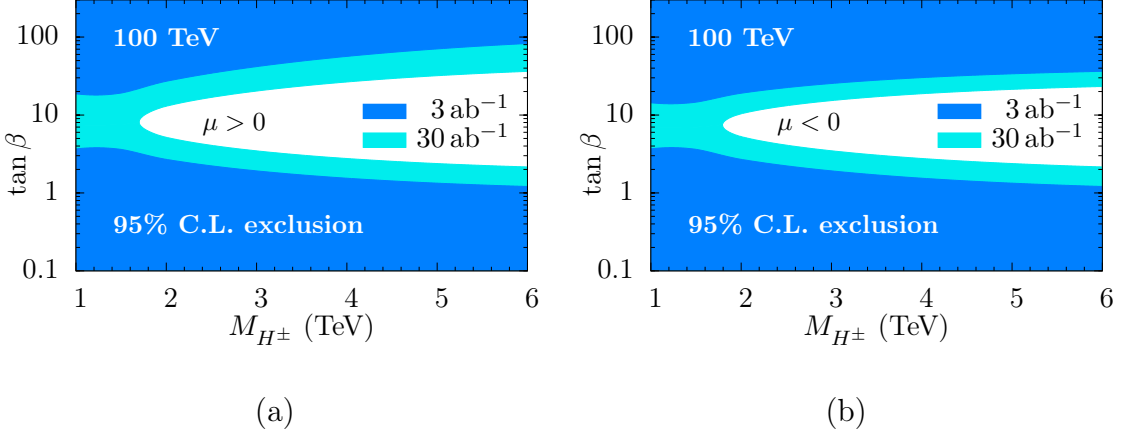


Figure 3.10. Predicted exclusion regime, at a 95% confidence level, for the MSSM at a 100 TeV FCC, taking the sign of  $\mu$  to be (a) positive, or (b) negative.

sensitivity to SUSY corrections at 100 TeV demonstrates the low model dependence in the reach for charged Higgs bosons at a future collider.

**3.2.3 Conclusions.** We have examined the predicted experimental reach for charged Higgs bosons in  $tH^\pm \rightarrow ttb$  at both the LHC and at a 100 TeV future collider, using a type-II two Higgs doublet model with mass degenerate heavy Higgs bosons. In the limit where  $H^\pm$  couples mostly to  $tb$ , we find that the LHC has access only to relatively large effective Yukawa couplings  $y_{tb}$  when  $m_{H^\pm} > 1$  TeV — confirming and extending expectations from Ref. [104]. Additionally, we find that supersymmetric corrections to the bottom Yukawa coupling are large, and further reduce sensitivity to a MSSM charged Higgs boson at the LHC. These findings indicate that a next-generation collider will probably be necessary to examine TeV-scale charged Higgs bosons that couple strongly to the third generation of quarks. In comparison to more optimistic predictions [112], we stress the importance of using realistic  $b$ -tagging efficiencies [4, 20] in phenomenological predictions covering TeV-scale physics.

Our particular choice of 2HDM (type-II with degenerate masses) ensures that  $H^\pm tb$  is the only pertinent coupling. A less restrictive model (e.g., where  $H^\pm$  couples to charm [123]), or one with alternate decay channels, such as  $H^\pm \rightarrow W^\pm H$ , may still

be visible at the LHC given sufficient integrated luminosity. In those cases, one can convert our limit on  $y_{tb}$  to a limit on cross-section times branching fraction for the channel  $tH^\pm \rightarrow ttb$  in those models.

Finally, we find a 100 TeV proton collider has the potential to close the moderate  $\tan\beta$  “wedge” region below 2 TeV. While the charged Higgs-top associated channel will be background limited at such a machine, charged Higgs bosons with masses up to 6 TeV can be probed with very little dependence on model parameters (such as the sign of the  $\mu$ -parameter in SUSY). Hence, a future circular collider shows great promise in shedding light on the structure of multiplets in the Higgs boson sector.

In this chapter, I have examined the application of my  $\mu_x$  boosted  $b$  tag to searches for  $Z'$  and  $H^\pm$ . The  $\mu_x$  tag’s enhanced light-jet rejection unlocks previously inaccessible parameter space to searches for new physics, permitting  $Z'$  sensitivity comparable to that of  $Z' \rightarrow l^+l^-$  searches for non-leptophobic models. And compared to some previous predictions, our more realistic  $b$  tagging efficiencies for high- $p_T$  jets demonstrate that an  $H^\pm$  from a 2HDM (with strong  $tb$  coupling) is likely only visible at a next generation, 100 TeV future circular collider.



## CHAPTER 4

### THE SHAPE OF QCD

In this chapter we will undertake a more global analysis of QCD events, via the fully-correlated QCD power spectrum. Unlike sequential jet clustering (e.g., anti- $k_T$ ), this global approach will *simultaneously* utilize correlated information from across the detector, much like the event shape variables popular in the early days of QCD. We will find challenges inherent to this global approach, but upon rigorous investigation will uncover enough clues to determine appropriate solutions, steadily accumulating the tools necessary to define the power jets model of Chapter 5. This novel method will significantly expand our ability to characterize QCD.

For 40 years, jets have been one of the primary tools for studying QCD, providing a rigorous (and quite successful) test of perturbative QCD theory. Some of the strongest measurements to date have been via a differential cross section ( $\frac{1}{\sigma} \frac{d\sigma}{d\chi}$ ), where  $\chi$  is some inclusive, kinematic observable built from jets (energy, momentum, mass, angles between the leading jets, momentum fraction of charged tracks, etc.). Each of these distributions is determined by filtering the predictions of particle-level QCD through the jet definition.

There is a good reason that jets are the primary tool for studying QCD; soft (low-energy) and collinear QCD radiation create a pole in the cross section, since these emissions have an effectively infinite probability of occurring. The reason for this pole is relatively simple: perturbation theory only works for a finite number of transient interactions. But a quark can radiate an *infinite* number of zero-energy gluons, and when it radiates a collinear gluon, the quark-gluon pair can continue to interact *indefinitely*.

A jet definition solves this problem by coarse-graining the phase space, so that in singular regions the radiation is “integrated out” (i.e., it cannot be separated from the radiating particle). This procedure maps directly onto experimental systems, where a parton which emits soft or collinear radiation will look no different in a detector than a non-radiating parton.<sup>12</sup> Grouping detected particles into bundles of adjacent  $\mathbf{p}$ , the jet becomes a proxy for the original parton, since the  $\sum \mathbf{p}$  of all its descendants is the parton’s original  $\mathbf{p}$ . This jet-parton duality allows jets to absorb the infinite regions of partonic phase space (i.e., the pathological and unobservable soft/collinear radiation resides within the jet), permitting the calculation of finite cross sections.

Nonetheless, modern jet definitions (like anti- $k_T$ ) have their limitations. For one, the choice of radius parameter  $R$  is somewhat arbitrary for individual events. A detector collaboration will generally pick one radius  $R$  to use for all jets, regardless of the nature of the event. Yet in the process  $pp \rightarrow b\bar{b}$  at the LHC, the boost of each  $b$ -jet is  $\gamma = Q/(2m_b)$ , which depends on the interaction scale  $Q$ , which varies from event to event. As the jets’ boosts increase with  $Q$ , they become more collimated, giving them a smaller effective radius. Clearly, it would make more sense if  $R$  were a dynamic quantity. Additionally, a constant  $R$  introduces a sharp cutoff in the angular information. To see angular correlations below the scale of  $R$ , jets must be reclustered with a smaller radius, which inevitable changes the “shape” of the event (there are more jets, each with less energy, after reclustering).

Second, modern jets are infested with energy from the pileup of many soft interactions per beam crossing, which must be removed. One method is to subtract all soft energy below a given threshold before clustering [124]. Alternatively, one can

---

<sup>12</sup>Here we use *parton* as a generic label for any quark or gluon. It originated as the name for particles which are *part* of the proton (its constituents).

build the jets, then subtract pileup proportional to a jet’s area [125], or recluster each jet using a smaller radius and remove the softest sub-jets [126–128]. While these procedures generally improve the jet’s reconstructed energy and mass, they inevitably throw away real information about the hard scatter — true jet substructure — as well as any information about long-range QCD correlations (such as the same-side bump in high-multiplicity  $pp$  events shown in Figure 1.1, which cannot be readily explained with existing models).

Finally, the evolution of sequential jet clustering is guided at each iteration by one “special” correlation. In the case of anti- $k_T$  jets, it is the minimum distance  $d_{ij}^2$  which determines which two particles should merge, or which jet should be finalized. But given  $N$  physics objects there are  $\mathcal{O}(N^2)$  energy and angular correlations, and a more complete reconstruction would simultaneously use *all* the information from the whole detector. In fact, clustering particles into definite, mutually exclusive jets is not the only way to characterize a QCD event. Some of the most important early discoveries used an event shape variable to classify the distribution of particles in individual events using one (or a handful) of scalars.

“Sphericity”  $S$  measures the isotropy of an event, with  $S = 0$  corresponding to two back-to-back particles and  $S \rightarrow 1$  indicating a completely isotropic distribution [129]. Measuring  $S$  requires using the detected physics objects to define the sphericity axis, which in the case of  $e^+e^- \rightarrow q\bar{q}$  corresponds to the dijet axis. Sphericity was used quite successfully to demonstrate the existence of quark jets in 1975 at SLAC [130]. The smoking gun is shown in Figure 4.1a; calculating  $S$  for each event and summing many thousands, their collective sphericity distribution heavily favors the jet model over isotropy. Another important support comes from the azimuthal angle of each event’s sphericity/dijet axis, which was predicted to follow a  $1 + b \sin^2(\theta) \cos(2\phi)$  distribution (where  $b$  is proportional to the beam polariza-

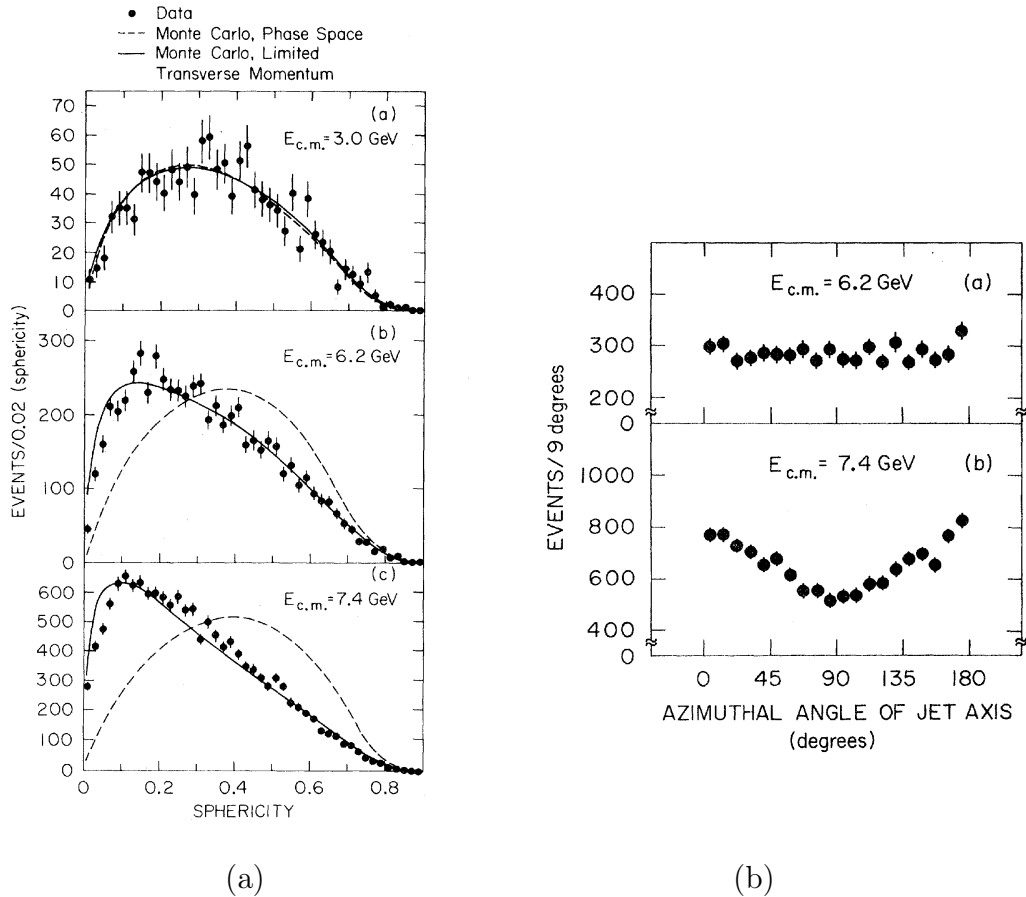


Figure 4.1. The discovery of quark jets at SLAC [130], using different  $E_{cm} \equiv \sqrt{S}$ : (a) The sphericity of events, with the predictions of (solid) the jet model and (dotted) isotropic phase space. (b) The azimuthal angle of the dijet axis.

tions). Figure 4.1 shows the azimuthal distribution of the jet axis, summing over  $|\cos\theta| \leq 0.6$  (the geometric acceptance of the detector); Figure 4.1a is from collisions with an unpolarized beam, and thus exhibit the flat curve of azimuthal symmetry, whereas Figure 4.1b is from collisions with about 50% polarization, which clearly reflects the predicted  $\cos(2\phi)$  distribution.

Progress was swift in the late 1970s. Four years after the discovery of dijets at SLAC, four separate collaborations at PETRA were able to demonstrate the existence of gluon jets using shape variables like “oblateness” and “thrust” [129]. However, with the contemporaneous invention of jet clustering algorithms — providing accurate

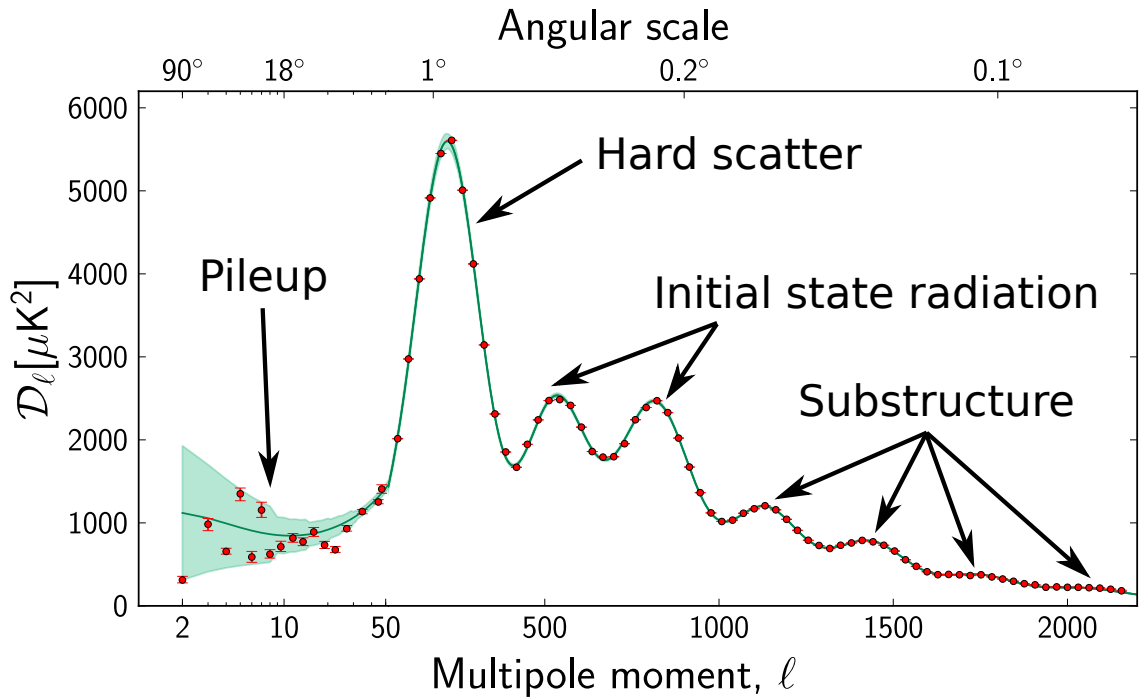


Figure 4.2. The angular power spectrum of the cosmic microwave background [131], but annotated as if it were the power spectrum of a QCD event.  $\mathcal{D}_l$  indicates what fraction of the CMB is composed of spherical harmonics of degree  $l$ .

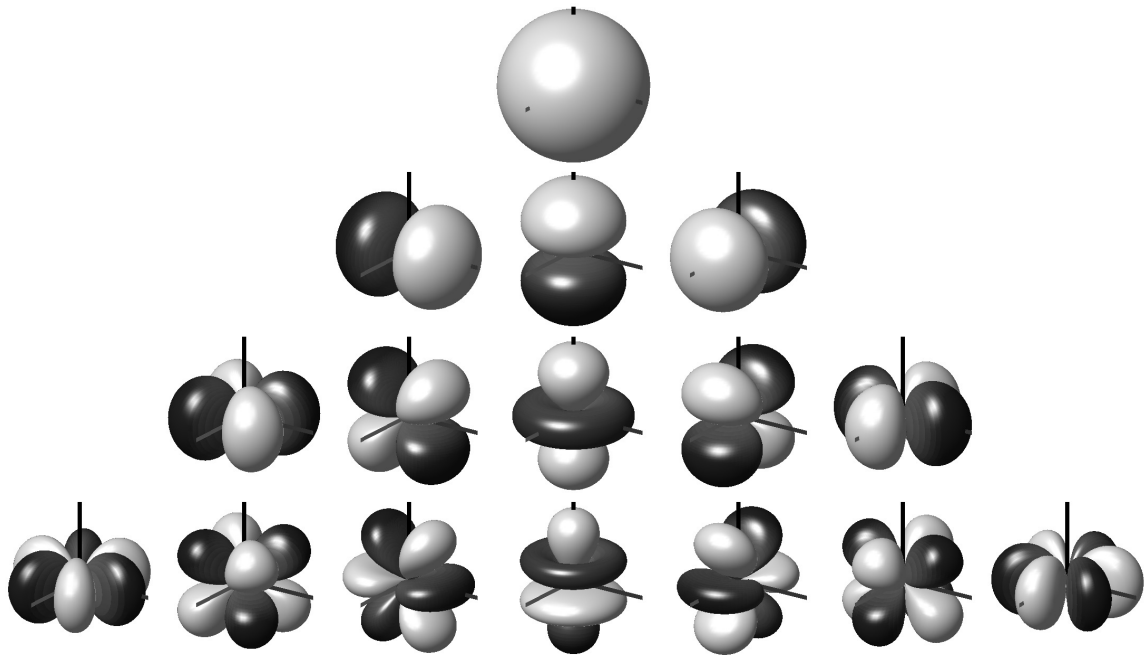


Figure 4.3. The spherical harmonics  $Y_l^m$  to degree  $l = 3$  [132]. At each  $l$ , the order  $m$  runs from  $-l$  to  $l$ , producing  $2l + 1$  orthogonal orientations. The complex phase is rounded to  $\pm 1$  (white/black). A quadrupole ( $l = 2$ ) has the same phase on opposite sides of the origin.

jet  $\mapsto$  parton proxies for final-state quarks and gluons — event shape variables found a place of diminishing importance. One reason for this transition is rather obvious; event shape variables tend to condense the event down to one, or a handful, of scalar observables. This is not tremendously useful for discovering hadronically decaying resonances, which requires calculating the invariant mass of the decaying system. But as we just saw, clustering throws out a tremendous amount of information by using only one correlation at a time. Is there a way to combine the holistic nature of event shape variables with the partonic mapping of spatially resolved jets?

For example, what if we could analyze QCD events like the cosmic microwave background, as characterized in Figure 4.2? The angular power spectrum of a QCD event decomposes the detected particles into spherical harmonics  $Y_l^m$  (see Fig. 4.3), seeking out the multipole structure of particle associations. This power spectrum could then be fit to known phenomena. Both pileup and long-distance QCD would appear at large angles (small  $l$ ), the hard scatter would dominate the power at large to medium angles (medium  $l$ ), and jet substructure would show up at small angles (large  $l$ ). Compare this to traditional event shape variables (such as sphericity), where each event produces *one* scalar observable, and *many thousands* of events must be summed together to produce a shape curve (such as Figure 4.1a). The QCD power spectrum could build a shape curve from a *single event!*

#### 4.1 The pioneering work of Fox and Wolfram

The idea of using spherical harmonics to characterize the shape of QCD events is not new; it was first proposed by Fox and Wolfram in 1978 [133, 134]. Their dimensionless power spectrum  $H_l$  was designed to work at an electron-positron collider of collision energy  $\sqrt{S}$ , and is constructed from the energy  $E$  and direction of travel

$\hat{p}$  of all the particles seen in the detector:

$$H_l = \frac{1}{S} E_i P_l(\hat{p}_i \cdot \hat{p}_j) E_j. \quad (4.1)$$

The Legendre polynomial  $P_l$  operates individually upon each inter-particle angle  $\cos \xi = \hat{p}_i \cdot \hat{p}_j$ , creating a matrix that is sandwiched between the vector of particle energies. The contraction over  $i$  and  $j$  gives the power spectrum  $H_l$ , the total magnitude of the  $l$ -pole correlation in the event. For example, one naïvely expects a dijet event to have a large 2-prong power  $H_2$ .

As a proof of concept, Fox and Wolfram used  $H_l$  to distinguish between  $q\bar{q}$ ,  $q\bar{q}g$  and  $ggg$  final states, which have very different origins at an electron-positron collider. The basic process is an electron and positron annihilating to a virtual photon, which then splits to  $q\bar{q}$ . One of the quarks will sometimes radiate a gluon, with a probability  $\mathcal{O}(\alpha_s)$ :

$$e^+e^- \rightarrow \gamma \rightarrow q\bar{q}(g). \quad (4.2)$$

Alternatively, the electron-positron annihilation can create a heavy resonance  $X$ , which then decays to three gluons:

$$e^+e^- \rightarrow X \rightarrow ggg. \quad (4.3)$$

**4.1.1 The  $\Upsilon$  resonance.** Studying the gluon was one of the major focuses of the late 1970s, and a trio of gluons decaying from a resonance of known mass was an attractive laboratory in which to do so. It is generally difficult to determine whether a jet originates from a quark or gluon,<sup>13</sup> but one place to look is the decay of a vector ( $J = 1$ ) quarkonium meson ( $q\bar{q}$ ). In its ground state, a vector quarkonium built from heavy quarks ( $Q\bar{Q}$ ) would likely decay to three gluons, since its mass

---

<sup>13</sup>The question itself is ill-conditioned, since an appropriate choice of gauge will rotate  $g \mapsto q\bar{q}$ . Color structure (singlet, triplet, octet) is gauge invariant.

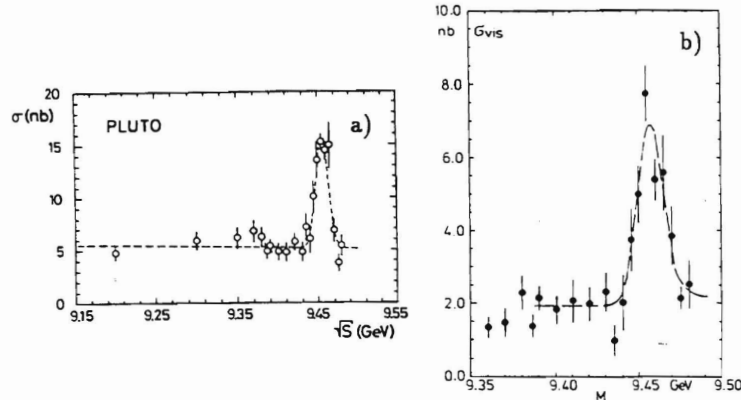


Figure 12:  $e^+e^-$  annihilation cross section in the  $\Upsilon$  region observed at the DORIS storage ring by a) PLUTO and b) DASP-2 collaborations [41].

Figure 4.4. The  $\Upsilon$  resonance in independent detectors at DORIS (1978) [136].

would be just below threshold for decaying to a pair of light-heavy mesons ( $Q\bar{q}'$ ). The  $\Upsilon$  meson (a  $J = 1$ ,  $b\bar{b}$  ground state) had just been discovered by the Fermilab E-288 experiment in 1977, with a rough mass of  $m \approx 9.5$  GeV [135]. This discovery occurred in the  $\Upsilon \rightarrow \mu^+\mu^-$  channel — the only discovery channel clean enough for a fixed-target experiment [136], even though only  $\sim 2.5\%$  of  $\Upsilon$ 's decay to two muons. QCD predicted that  $\mathcal{O}(90\%)$  of decays would use the  $\Upsilon \rightarrow ggg$  channel, which was accessible only at a collider.

Within a year, physicists at the DORIS collider had rammed through a beam upgrade [136], confirming a hadronically decaying resonance with  $m_{\text{had}} \approx 9.46(1)$  GeV [137, 138]. This result was perfectly consistent with the updated E-288 measurement of  $m_{\mu\mu} \approx 9.45(5)$  GeV [139], which used three times more data after several more months of data taking [136]. Figure 4.4 shows the  $\Upsilon \rightarrow ggg$  resonance seen at DORIS, perched atop the  $\gamma \rightarrow q\bar{q}g$  background. Distinguishing these two final states was unnecessary for the purposes of discovering the  $\Upsilon$ , because the background cross section is flat enough, and the  $S/N$  ratio high enough, that the resonant structure is obvious. However, QCD predicts very different shapes for these final states, and



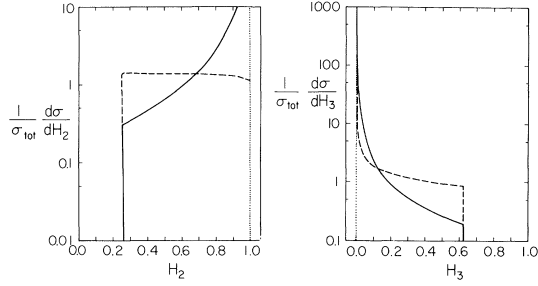


FIG. 1. The distributions in  $H_2$  and  $H_3$  for the processes  $e^+e^- \rightarrow q\bar{q}$  (dotted lines),  $e^+e^- \rightarrow q\bar{q}G$  (full lines), and  $e^+e^- \rightarrow \text{heavy resonance} \rightarrow GGG$  (dashed lines). The process  $e^+e^- \rightarrow q\bar{q}G$  alone yields an infinite total cross section, but when added to  $e^+e^- \rightarrow q\bar{q}$  calculated through  $O(g^2)$  the combination of processes [denoted by  $e^+e^- \rightarrow q\bar{q}(G)$ ] gives a finite cross section. We have taken  $\alpha_s = 0.25$  for the  $e^+e^- \rightarrow q\bar{q}G$  distribution.

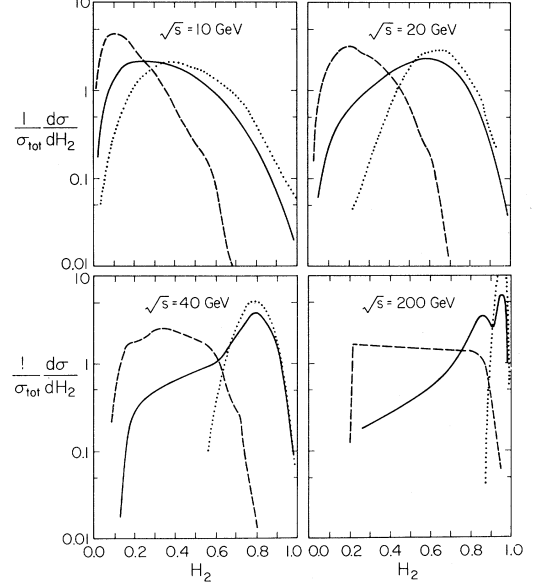


FIG. 2. The  $H_2$  distributions predicted for hadronic events resulting from the processes  $e^+e^- \rightarrow q\bar{q}$  (dotted lines),  $e^+e^- \rightarrow q\bar{q}(G)$  (full lines), and  $e^+e^- \rightarrow \text{heavy resonance} \rightarrow GGG$  (dashed lines), at various center-of-mass energies  $\sqrt{s}$ .

Figure 4.5. The original  $H_l$  predictions for (left) free partons and (right) measurable particles [133].

measuring this difference offers a precision test of QCD.

Fox and Wolfram's approach was to calculate power spectrum probability distributions  $f(H_l)$  for each of the three final states ( $\gamma \rightarrow q\bar{q}$ ,  $\gamma \rightarrow q\bar{q}g$  and  $X \rightarrow ggg$ ). If these  $f(H_l)$  distributions were distinct enough, then they could be used to take the power spectrum for an observed event and construct the likelihood that it originated from each of the three final states. The left pane of Figure 4.5 shows  $H_2$  (2-prong) and  $H_3$  (3-prong) distributions (on a log-scale), assuming *free* quarks and gluons (no showering, no jet structure). Since  $q\bar{q}$  events are always back-to-back in the CM frame (which is the lab frame at an  $e^+e^-$  collider), they always have  $H_2 = 1$  and  $H_3 = 0$ . Similarly,  $q\bar{q}g$  events tend to have large  $H_2$  and small  $H_3$  because their quarks are mostly back-to-back (the gluon preferring to radiate at small angles). In  $ggg$  events, the partons lack intense correlations, and thus have more moderate values of  $H_2$  and

$H_3$ . Apparently, the power spectrum can discriminate these topologies quite well.

Unfortunately, free quarks and gluons are not observable, so parton level predictions are not necessarily useful. The right pane of Figure 4.5 shows more realistic event shapes, at various collider energies, using an early, phenomenological showering/fragmentation model by Feynman and Field [140]. These non-perturbative effects are clearly important at  $\sqrt{S} = 10$  GeV, because only at much higher energies ( $\sqrt{S} \rightarrow 200$  GeV) do the final states begin to resemble the free parton predictions. Nonetheless, even at low energy, the average three-gluon event has a much lower value of  $H_2$ . This was put to good use by CLEO in 1980, as seen in Figure 4.6. The  $\Upsilon(4s)$  resonance (an excited state of the  $\Upsilon$ ) can clearly be seen in the original data (Fig. 4.6a), but requiring  $H_2/H_0 < 0.3$  (Fig. 4.6b) suppressed a significant amount of the  $q\bar{q}(g)$  background by rejecting its more two-pronged shape.

**4.1.2 What about jets?** After the initial success at CLEO, the 2-prong power  $H_2$  became a common tool in suppressing 2-prong background in the measurement of particle decays and branching fractions [142]. Suppressing  $n$ -prong background is especially useful at “ $B$  factories” (experiments devoted to producing and measuring  $b$ -flavored hadrons), and the power spectrum  $H_l$  has been modified to assist these discoveries [143–145]. However, in these particle-decay applications,  $H_l$  are used to distinguish a handful of decay products from the hadronic background debris; they are not used to study the shape of QCD events containing distinct, extensive jets. While  $H_l$  have occasionally been used for jet physics [146], this original purpose has never been fully realized. Why not?

Because Fox-Wolfram distributions  $f(H_l)$  are the wrong tool for harnessing the QCD power spectrum. In the Fox-Wolfram approach, one integrates over the differential cross section for some process, encoding the kinematic configurations of its final-state partons into  $f(H_l)$ . To find the likelihood that an observed event

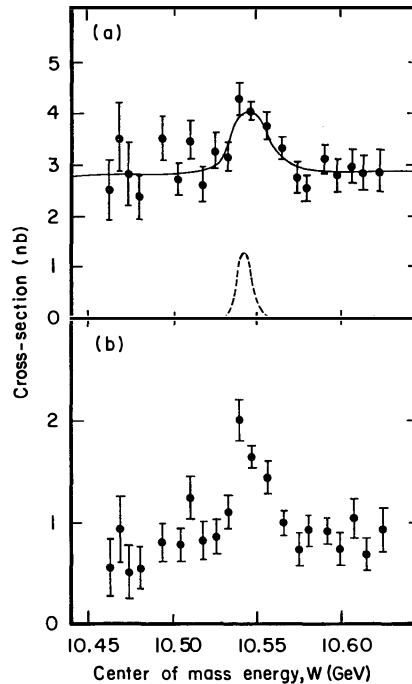


Fig. 8. Hadronic cross sections corrected for acceptance for the  $\Upsilon(4S)$  region. In addition to the statistical errors shown, there is an overall systematic error of  $\pm 15\%$ . (a) The total hadronic cross section. The curve shows the fit described in the text. The dashed curve indicates the beam energy resolution. (b) Partial cross section for events with  $H_2/H_0 < 0.3$ .

Figure 4.6. The  $\Upsilon(4S)$  resonance at CLEO (1980). A cut on  $H_2$  dramatically reduces background [141].

originated from a given process, the observed power spectrum is compared to each process'  $f(H_l)$  distributions. In a way, one is using these  $f(H_l)$  distributions to indirectly reconstruct the kinematic configuration of the observed event.

In this thesis, I present power jets, a more direct method to map the observed  $H_l$  onto the kinematic configuration of its original partons. In order to understand how power jets should work, we need to understand the two very practical reasons why Fox-Wolfram distributions do not.

**4.1.2.1 Jet-parton duality requires  $Q \gg 1$  GeV.** At the time of Fox and Wolfram's original work, the state of the art in  $e^+e^-$  collider energies was  $\sqrt{S} \approx 10$  GeV, the lowest scale in their original predictions (Fig. 4.5). Comparing  $H_2$  for free par-

ton (left) and hadronized particles (right) at this scale, it is clear that jet formation (showering and fragmentation) strongly alters the event shapes. It is not intense enough to eliminate the distinction between  $\gamma \rightarrow q\bar{q}g$  and  $X \rightarrow ggg$  (indeed,  $H_2$  was used to discriminate these final states at CLEO, as we say in Fig. 4.6), but it does heavily distort the original QCD spectrum.

This emphasizes the importance of the interaction scale  $Q$  on jet-parton duality; only very energetic partons have anything approaching freedom from their siblings ( $\alpha_s \rightarrow 0$  as  $Q \rightarrow \infty$ ). Confinement requires hadronization into color singlets, and the non-perturbative (NP) effects of this requirement occur at the scale  $\Lambda_{\text{QCD}} \sim 1 \text{ GeV}$ . Perturbative QCD is not very good at making predictions in this regime, so that predictions for the distribution of hadrons in the final states depend heavily on the *phenomenological* jet formation model — an approximation of the NP physics of hadronization with many free parameters to tune to data. When an observable is heavily altered by a phenomenological model, it becomes sensitive to all the model's assumptions about NP physics, and begins to lose objectivity.

Given the clear strength of non-perturbative effects at  $Q = 10 \text{ GeV}$ , it is not surprising that experiments in the late 1970s observed the resonant  $\Upsilon \rightarrow ggg$  decay *not* as three distinct jets, but as two hadrons and a third, jet-like object [129]. While the  $H_2$  of this decay is generally useful for background discrimination (or perhaps tuning the parameters in the jet formation model), it cannot provide strong insight into perturbative QCD. Yet Figure 4.5 also demonstrates that at much higher scales such as  $Q = 200 \text{ GeV}$ , the hadronized  $H_2$  distributions look much more like the free parton distributions. Jet-parton duality requires a sufficiently high scale. This limitation was extremely important in 1978, since  $e^+e^-$  colliders would not be able to reach  $Q > \mathcal{O}(100 \text{ GeV})$  for 10–20 years, by which time  $H_l$  were largely forgotten outside of  $B$ -factories. Of course, hadron colliders (like CERN's SPS) were already

beginning to probe 100 GeV scales in 1978 — so why weren't  $f(H_l)$  used there?

**4.1.2.2 At hadron colliders, neither the CM frame nor  $Q$  is fixed.** The power spectrum is invariant to rotations of the final state, but  $H_l$  is not invariant to boosts (we will prove these assertions in the next section). At an  $e^+e^-$  collider, the lab frame and CM frame coincide quite well. But as we saw in Section 1.2, at hadron colliders the CM frame of the hard scatter has some longitudinal boost  $y_{\text{cm}}$ . This creates two problems which require a wholesale shift in the way the  $H_l$  are used. We will outline the problems here, and present a solution in the next section.

A detector will never perfectly reconstruct a hard scatter, which precludes a perfect reconstruction of its longitudinal boost  $y_{\text{cm}}$ . In addition to detector mis-measurement, only *transverse* missing momentum  $\cancel{p}_T$  can be observed; additionally, initial state radiation (ISR) mimics the radiation of jets from the hard scatter. These effects collectively shift an event's reconstructed CM frame away from its true value. Without  $y_{\text{cm}}$ , one cannot accurately compare the  $H_l$  observed in the lab frame to  $f(H_l)$  calculated in the CM frame. It is possible to *approximate*  $y_{\text{cm}}$  using the hardest few jets, then boost into that frame to measure  $H_l$ , but there will be error.

To account for this effect, Fox-Wolfram distributions  $f(H_l)$  at a hadron collider should convolve  $H_l$  predictions from the CM frame with boosts from the probability distribution of boost errors  $f(\Delta y_{\text{cm}})$ , smearing  $f(H_l)$  so that different processes look more similar. This prevents theorists from making universal predictions for a process of interest with the Fox-Wolfram approach, because the  $y$  uncertainty depends on the detector (energy resolution, active area, response, etc.) *and* the triggers used.

But even if one were committed to Fox-Wolfram approach, and calculated  $f(H_l)$  distributions dedicated to a given detector, it is still impossible to make universal predictions because the interaction scale  $Q$  is not fixed. The CM frame is

longitudinally boosted because only one parton from each hadron interacts, taking only a fraction  $x$  of its mother's energy/momentum; the scale of each interaction is therefore  $Q \sim \sqrt{s} = \sqrt{x_1 x_2} \sqrt{S}$ , where  $\sqrt{S}$  is the invariant mass of the hadron-hadron collision. Since the two  $x_i$  for each collision are randomly drawn from a parton distribution, the  $Q$  of all observed collisions spans many orders of magnitude.

Given the detector-dependent error in  $y_{\text{cm}}$  and the random  $Q$ , predictions about a specific process at a hadron collider would require a large library of  $f(H_l)$  curves. This gives theorists an almost impossible task: for each process of interest, supply experimentalists with predictions for various detectors (the full details of which may not be public), a finite number of scales to run the simulations (with intermediate scales accessed via interpolation), and different jet formation models (since different groups prefer different models).

**4.1.3 Rebooting the power spectrum.** It is now clear that in 1978, lepton colliders were too weak to use power spectra to accurately study perturbative QCD, and Fox-Wolfram distributions  $f(H_l)$  were unsuitable for hadron colliders. In 2018, the collider situation is not much better. There *was* a lepton collider capable of reaching  $\sqrt{S} = 200 \text{ GeV}$  (LEP II), but it was dismantled in 2001 to make way for the collider of the future: the only collider operating today which is capable of studying perturbative, non-nuclear QCD at sufficiently high  $Q$  — the Large *Hadron* Collider. A few next generation lepton colliders are on the drawing board, but their future is uncertain. So if all we have is a hadron collider, why am I talking about  $H_l$ ?

Because the difficulties at a hadron collider are not with the power spectrum itself, but with the probability distributions  $f(H_l)$ . These distributions are problematic even at a lepton collider. As we will see in the next section, QCD events exhibit rather complicated  $H_l$  patterns which in no way resemble the smooth power spectrum of the CMB; this gives  $f(H_l)$  rather non-trivial shapes. And consider how  $f(H_l)$  are

used: to determine the kinematic configuration of the original partons through the convoluted filter of probability distributions. Why not remove this intermediate filter?

This is the “power jets” approach; it takes an observed  $H_l^{\text{obs}}$  and works out the partons that produced it. It does this by calculating the power spectrum  $H_l^{\text{jet}}$  for a handful of jet-like objects, then *fits* the jets by minimizing the difference between  $H_l^{\text{jet}}$  and  $H_l^{\text{obs}}$ . The final fit corresponds to a point in partonic phase space, from which one can determine the likelihood that each candidate process produced it. And by starting with  $H_l^{\text{obs}}$  and working backward, power jets automatically account for different scales and different detectors. Thus, theorists need only generate differential cross sections for the processes of interest.

In order to soundly understand how and why power jets work, we must take the scenic route (this is a thesis, after all). For the remainder of this chapter, we will study  $H_l$  in the CM frame (i.e., at a lepton collider). Even this is complicated enough that it becomes quite prudent to ignore a boost until the next chapter. First we will derive  $H_l$  from first principles and observe the power spectrum for a few simple QCD events. This will make it abundantly clear why Fox and Wolfram’s  $f(H_l)$  distributions are the wrong approach. It will also reveal the problem of sampling noise, which manifests through particle multiplicity (the number of observed particles  $N$ ), which severely limits the amount of information that can be extracted from  $H_l$ . Finally, we will develop a scheme to control sampling noise; this will set the stage for the next chapter, where we develop the power jet model.

## 4.2 QCD power spectra $H_l$

Deep inside the heart of a particle collider, an “event” occurs. An electron and a positron annihilate, producing a quark-antiquark pair. These fly away from the collision in opposite directions, and within  $\mathcal{O}(\text{fm})$  they shower and fragment into

long-lived particles — these are seen by the detector as two collimated jets. Neutral particles travel in straight lines; their four momentum  $\mathbf{p}$  is approximated by the calorimeter cell they strike. Charged particles are deflected by the magnetic field, which determines the sign of their charge and the magnitude of their momentum. Extrapolating tracks back to the interaction point gives their original  $\mathbf{p}$ . Combining towers and tracks, the detector outputs a set of  $N$  four-momenta emerging from the interaction point. From this set of  $N$  objects, we would like to characterize the event's dominant 2-jet shape, as well the higher-order shapes of QCD interactions.

**4.2.1 Defining the power spectrum  $H_l$ .** The spherical harmonics  $Y_l^m$  are an obvious choice for this shape characterization, since they provide a complete, orthonormal basis which has been extensively studied (so there are many identities we can draw upon). The spherical harmonics  $Y_l^m(\hat{r})$  are a set of complex functions on the surface of the unit sphere, each of which encodes an  $l$ -prong shape. Any complex scalar function  $f(\hat{r})$  on the unit sphere can be decomposed into  $Y_l^m(\hat{r})$  of degree  $l$  and order  $m$ :

$$f(\hat{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_l^m Y_l^m(\hat{r}). \quad (4.4)$$

To apply spherical harmonics to collider events, we must decide what  $f(\hat{r})$  should encode. Since jets are localized bundles of energy/momentum, with a 2-jet event having most of its energy/momentum back-to-back, the obvious choice is for  $f(\hat{r})$  to encode the energy (or momentum) of the event projected on the unit sphere. Looking at the first few spherical harmonics in Figure 4.3, we naïvely expect our 2-jet event to look more 2-prong ( $l = 2$ ) than 3-prong ( $l = 3$ ).

Of course, spherical harmonics operate in three dimensions of space, and encode 3 degrees of freedom (d.o.f.) ( $2 \times$  position + intensity), whereas particles emanating from an event have 4 d.o.f. ( $3 \times$  space + time). To create  $f(\hat{r})$  from collider events, we need to make a few approximations that constrain the fourth degree of



freedom:

- Placing the origin at the interaction point, all QCD occurs within an  $\mathcal{O}(\text{fm})$  radius, outside of which the particles no longer interact with each other. Charged particle tracks bend in the magnetic field, but this can be rewound back to the interaction point.<sup>14</sup> Hence, if we ignore in-flight particle decays for the time being, we can approximate all  $N$  physics objects as moving radially outward from the interaction point:  $\vec{p} \parallel \hat{r}$ .
- Each physics object is effectively massless ( $\beta \approx 1$ ), which constrains its fourth d.o.f. ( $E = |\vec{p}|$ ). This condition is forced upon us by the detector, which cannot reliably measure the mass of tracks or towers. Yet as we saw in Section 1.3, massless physics objects are a good approximation of reality. If each particle is massless, then it travels at the speed of light, and its position at any given time is simply its radial direction of travel times the light-distance since the collision

$$\mathbf{x} = ct [1, \hat{p}]. \quad (4.5)$$

That being so, all  $N$  physics objects move outward from the collision on the surface of a sphere. Since they move exactly radially from the interaction point, this surface is unchanging, and contains all observable information about the event's shape.

Using the radial, massless approximation, the shape of a QCD event the moment after hadronization is fully encoded in its energy distribution  $E(\hat{r})$  projected onto the unit sphere (very much like how the cosmic microwave background is a snapshot of the universe the moment after recombination).

---

<sup>14</sup>There are also unwanted interactions with the detector material. Some can be corrected, but many cannot, and irreversibly smear the information we seek.

Since events are composed of a finite number of discrete particles, an event's energy distribution is also discrete. It sums over  $N$  physics objects with finite energy, each of which exist only at discrete radial positions (via delta functions):

$$E(\hat{r}) = \sum_{i=1}^N E_i \delta^3(\hat{x} - \hat{p}_i) = \sum_{i=1}^N E_i \frac{\delta^2(\hat{r} - \hat{p}_i)}{\sin \theta_i}. \quad (4.6)$$

Changing the delta function from Cartesian to spherical coordinates is convenient, since it removes an unnecessary degree of freedom, but this operation requires dividing by the determinant of the Jacobian  $\mathbf{J} = \partial x^i / \partial r^j$ :

$$\delta^3(\vec{x} - \vec{x}_0) = \frac{\delta^3(\vec{r} - \vec{r}_0)}{\det(\mathbf{J})}. \quad (4.7)$$

This is required so that  $E(\hat{r})$  is properly normalized (since  $d\Omega = \sin \theta d\theta d\phi$ ):

$$E_{\text{tot}} = \sum_{i=1}^N E_i = \int_{\Omega} d\Omega E(\hat{r}). \quad (4.8)$$

The energy distribution is decomposed into spherical harmonics by integrating over the surface of a sphere:

$$E_l^m = \int_{\Omega} d\Omega Y_l^{m*}(\hat{r}) E(\hat{r}). \quad (4.9)$$

These coefficients indicate “how much” of  $E(\hat{r})$  is 2-prong ( $l = 2$ ), 3-prong ( $l = 3$ ), etc, probing the angular scale of these  $l$ -prong shapes

$$\xi = 2\pi/l. \quad (4.10)$$

The resulting  $Y_l^m$  decomposition is exact, provided that the total squared power

$$T = \int_{\Omega} d\Omega |E(\hat{r})|^2 \quad (4.11)$$

is finite (which seems reasonable, since a physical collider/detector system must have a physical energy). However, the exactness of the decomposition also relies on the orthonormality of the basis,

$$\int_{\Omega} d\Omega Y_l^{m*}(\hat{r}) Y_{l'}^{m'}(\hat{r}) = \delta_{ll'} \delta_{mm'}. \quad (4.12)$$

In practice, a detector's beam holes leave an unobservable solid angle which break the orthogonality of  $Y_l^m$ , causing spectral leakage among  $E_l^m$ . This problem predominantly manifests at high  $l$ , so we leave its exploration until Section 5.4.2.

Note that the spherical harmonics used by various disciplines can differ in their normalization, which can lead to confusion in prefactors. For completeness, we now define the conventions used here. We choose the orthonormal spherical harmonics commonly employed in quantum mechanics,

$$Y_l^m(\theta, \phi) = (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) e^{im\phi}, \quad (4.13)$$

which are built from the *associated* Legendre polynomials

$$P_l^m(x) = (-1)^m (1-x^2)^{(m/2)} \frac{d^m}{dx^m} [P_l(x)], \quad (4.14)$$

which themselves derive from the ordinary Legendre polynomials

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} [(x^2 - 1)^n]. \quad (4.15)$$

$Y_l^m$  are basis functions for SO(3) (the group of rotations in 3-space), but the arbitrary definition of the longitudinal axis makes individual moments only *partially* invariant to rotations;  $E_l^m$  of the same degree  $l$  will mix among orders  $m$ . This mixing confounds the characterization of QCD event shapes via unique “fingerprints” of  $E_l^m$  moments. Yet since  $Y_l^m$  are orthogonal, summing over  $m$  produces a rotationally invariant power spectrum

$$S_l \equiv \sum_{m=-l}^l |E_l^m|^2 = \sum_{m=-l}^l \int_{\Omega} d\Omega Y_l^{m*}(\hat{r}) E(\hat{r}) \int_{\Omega'} d\Omega' Y_l^m(\hat{r}') E(\hat{r}'). \quad (4.16)$$

Bringing the sum inside the integral allows us to use the addition theorem

$$P_l(\hat{r} \cdot \hat{r}') = \frac{4\pi}{2l+1} \sum_{m=-l}^l Y_l^{m*}(\hat{r}) Y_l^m(\hat{r}'), \quad (4.17)$$

so that  $S_l$  can be rewritten as the non-separable integral

$$S_l \equiv \sum_{m=-l}^l |E_l^m|^2 = \frac{2l+1}{4\pi} \int_{\Omega} d\Omega \int_{\Omega'} d\Omega' E(\hat{r})E(\hat{r}')P_l(\hat{r} \cdot \hat{r}'). \quad (4.18)$$

This form makes the absolute rotational invariance of the power spectrum manifest, as the  $l$ -dependent piece of the integral depends only on  $\hat{r} \cdot \hat{r}' = \cos \xi$ , the rotationally invariant *interior* angle between two parts of the distribution.

Combining Equations 4.4, 4.11, and 4.16, the total squared power  $T$  can be related to the absolute power spectrum  $S_l$  by

$$T = \sum_{l=0}^{\infty} S_l. \quad (4.19)$$

Because the absolute scale of  $S_l$  depends on  $T$  (and therefore  $E_{\text{tot}}$ ), normalizing  $S_l$  to the isotropic power  $S_0$  gives a more meaningful descriptor of the pure *shape* of the event. Furthermore, the near-linear factor of  $2l+1$  is a nuisance to plot (and very easily recovered when desired), so we choose to discard it. Henceforth, the “power spectrum” refers to the dimensionless, flattened

$$H_l \equiv \frac{1}{2l+1} \frac{S_l}{S_0} \quad (4.20)$$

(so that  $H_0$  is always unity). Since  $P_0(x) = 1$  for any  $x$ , the  $S_0$  integral in Equation 4.16 is separable, so that  $S_0 = E_{\text{tot}}^2/(4\pi)$ . If we then define the normalized event shape

$$\rho(\hat{r}) \equiv \frac{E(\hat{r})}{E_{\text{tot}}}, \quad (4.21)$$

we can replace  $E(\hat{r})$  in Equation 4.16 with  $\rho(\hat{r})$ , leading to

$$H_l \equiv \frac{4\pi}{2l+1} \sum_{m=-l}^l |\rho_l^m|^2 = \int_{\Omega} d\Omega \int_{\Omega'} d\Omega' \rho(\hat{r})\rho(\hat{r}')P_l(\hat{r} \cdot \hat{r}'). \quad (4.22)$$

**4.2.2  $H_l$  for discrete samples.** The definition of  $H_l$  in Equation 4.22 is agnostic to the content of  $\rho(\hat{r})$ , but as we previously defined, a physical detector will generally build its  $E(\hat{r})$  from discrete particles localized to spatial delta functions (see

Eq. 4.6). Since  $\rho(\hat{r}) = E(\hat{r})/E_{\text{tot}}$ , we can replace particle energy  $E_i$  in Equation 4.6 with energy fraction

$$f_i \equiv \frac{E_i}{E_{\text{tot}}}. \quad (4.23)$$

Evaluating Equation 4.22 for this discrete  $\rho(\hat{r})$ , the delta functions collapse the integral to the sum over all inter-particle terms, with the Legendre polynomial  $P_l$  acting individually upon each element of the angular correlation matrix  $\Xi$ ,

$$H_l = f_i P_l(\hat{p}_i \cdot \hat{p}_j) f_j = \langle f | P_l(\Xi) | f \rangle \quad (4.24)$$

$$\Xi \equiv |\hat{p}\rangle \cdot \langle \hat{p}| = \hat{p}_i \cdot \hat{p}_j = \cos \xi_{ij}. \quad (4.25)$$

Here, we use bra-ket notation to reduce the clutter of indices, where

$$\langle a | = (a_1 \quad \dots \quad a_N) \quad (\text{a row-vector}) \quad (4.26)$$

$$|b\rangle = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} \quad (\text{a column-vector}) \quad (4.27)$$

$$\langle a | b \rangle = \sum_i a_i b_i \quad (\text{inner product}) \quad (4.28)$$

$$|b\rangle \cdot \langle a| = f_i \cdot f_j \quad (\text{outer product}). \quad (4.29)$$

With a complete definition of the power spectrum, we can constrain  $H_l$ . We know that  $H_l \geq 0$  by construction (Eq. 4.22), since each term sums the total squared magnitude over all orders  $m$ . To find the maximum possible  $H_l$ , we can design an angular correlation matrix  $\Xi$  which maximizes the sum. Since the energy fractions in  $|f\rangle$  are manifestly positive, and  $|f\rangle$  is contracted by  $\Xi$ , it is relatively straightforward to deduce that the maximizing  $\Xi$  should be filled with large, positive values. Since  $-1 \leq P_n(z) \leq 1$  and  $P_n(1) = 1$ , the best option is  $\Xi = |1\rangle \langle 1|$  (a matrix of ones). This gives

$$H_l^{\text{max}} = \langle f | 1 \rangle \langle 1 | f \rangle = \left( \sum f_i \right)^2 = 1. \quad (4.30)$$

While this unitary  $H_l^{\text{max}}$  is mathematically possible, is it also physical?

Given that  $P_n(z) < 1$  for  $|z| < 1$  and  $P_n(\pm 1) = (\pm 1)^n$ , the  $\Xi$  which maximizes  $H_l$  can only be formed by  $z = \cos \xi = \pm 1$ . For odd  $l$ , this requires only collinear particles. For even  $l$ , this requires all particles to lie along one axis, so that they are either parallel or antiparallel to every other particle. Since collinear, massless particles are indistinguishable from a single composite particle, the configuration which causes  $H_{\text{odd}} = 1$  is a lone particle, and the configuration which causes  $H_{\text{even}} = 1$  is two back-to-back particles (albeit without the equal energy constraint of the CM frame). Both options are physical, so we can fully constrain

$$0 \leq H_l \leq 1. \quad (4.31)$$

This exemplifies why it is useful to scale  $E_{\text{tot}}$  and  $(2l + 1)$  out of  $H_l$ ; it puts the entire power spectrum into the very manageable unit interval.

One useful property that  $H_l$  does not possess is boost invariance. Calculating  $H_1$  via Equation 4.24, we note that  $P_1(x) = x$ , which means that  $H_1$  can be expressed as the dimensionless squared norm of the total momentum vector  $\vec{p}_{\text{tot}}$ :

$$H_1 = \langle f | \hat{p} \rangle \cdot \langle \hat{p} | f \rangle = |(f_1 \hat{p}_1 + f_2 \hat{p}_2 + \dots)|^2 = \frac{|\vec{p}_{\text{tot}}|^2}{E_{\text{tot}}^2} = \beta_{\text{tot}}^2. \quad (4.32)$$

Hence,  $H_1$  is the speed of the CM frame of *detected* particles (which is not necessarily the CM frame of the hard scatter). Since a boost alters angles, all other  $H_l$  are also frame dependent.

**4.2.3 The power spectrum for simple events.** We have thus far experienced  $H_l$  in a very rigorous mathematical framework. But what does  $H_l$  look like for an actual event?

The simplest QCD process we can construct at a lepton collider is  $e^+e^- \rightarrow q\bar{q}$  for massless quarks. In the CM frame of the final state, each quark has energy fraction  $f = 0.5$  and they are always back-to-back. In fact, this is the configuration for any

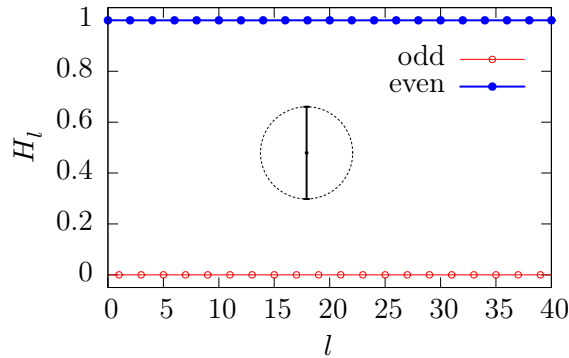


Figure 4.7. The power spectrum of a 2-particle final state in its CM frame.

system of two massless particles in its CM frame — and since  $H_l$  are rotationally invariant, all 2-quark power spectra are identical:  $H_{\text{even}} = 1$  and  $H_{\text{odd}} = 0$ , as depicted in Figure 4.7, where the momenta of the two partons is depicted in the inset.

It may seem odd that a 2-parton final state is not simply  $H_2 = 1$ , with zeroes in all other powers. But recall what an  $l = 2$  quadrupole actually looks like (see Fig. 4.3); each of the two poles is an extensive lobe, almost spherical. They look nothing like the  $\delta$  functions that are used to describe each parton’s spatial location in  $\rho(\hat{r})$ . The power spectrum of these back-to-back  $\delta$  functions tells us that the only way to reconstruct them from spherical harmonics is to add up every *even* multipole, and none of the odd ones.

**4.2.3.1  $H_l$  does *not* look like the CMB!** The next most complicated QCD process radiates a gluon from one of the quarks. This 3-parton final state is strikingly different from the 2-parton one. First we will examine two specific  $q\bar{q}g$  final states, noting some interesting features in each, then we will discuss their implications.

In Figure 4.8 we show the power spectrum for an event where the gluon is soft, so that we expect the event’s shape to be very two-jet-like. The three unobservable partons shower and hadronize into measurable final-state particles (the “measurable”

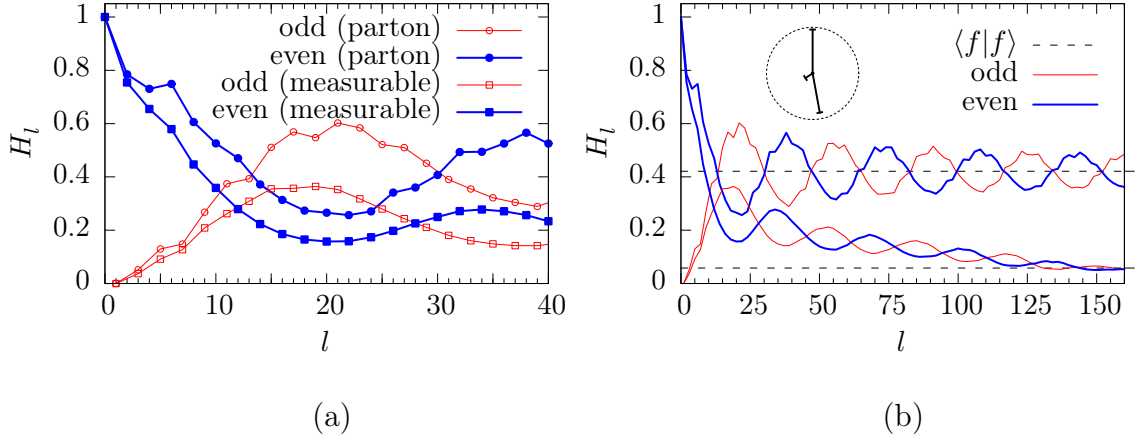


Figure 4.8. The power spectrum of a  $q\bar{q}g$  final state with a soft gluon (2-jet-like). (a) The low- $l$  behavior and (b) the asymptotic behavior, with kinematic depiction.  $H_l$  is depicted for (circle) partons and (square)  $\mathcal{O}(100)$  measurable particles. Odd powers are hollow, red, and thin. Even powers are blue, filled and thick.

final state), and  $H_l$  is shown for both. A depiction of the energy fraction and relative orientation of the three partons is shown in the inset of Figure 4.8b (the soft gluon having the soft momentum). Note that Figure 4.8b shows the connecting lines, but not the points at integer  $l$ . This will be the default depiction of power spectra in subsequent sections; even though  $l$  is always an integer, lines are simply easier to see.

If we naïvely expect a two-jet-like power spectrum to look like a two-parton power spectrum, then the event shown in Figure 4.8a meets our expectations at low- $l$ ;  $H_2$  is large and  $H_3$  is small, and the same large/small trend holds for the first few even/odd powers. But as  $l$  increases, the even powers meet the odds, then lock into a strange oscillatory dance. Another interesting feature is the  $H_l$  of the many-particle, measurable final state. It initially follows the 3-parton power spectrum quite closely, but quickly locks into a different oscillation. It then slowly attenuates towards  $H_l \approx \langle f|f \rangle$ . In fact, it is clear in Figure 4.8b that both  $H_l$  approach this same asymptotic limit, as depicted by the dashed line (using their respective energy fraction vectors  $|f\rangle$ ).



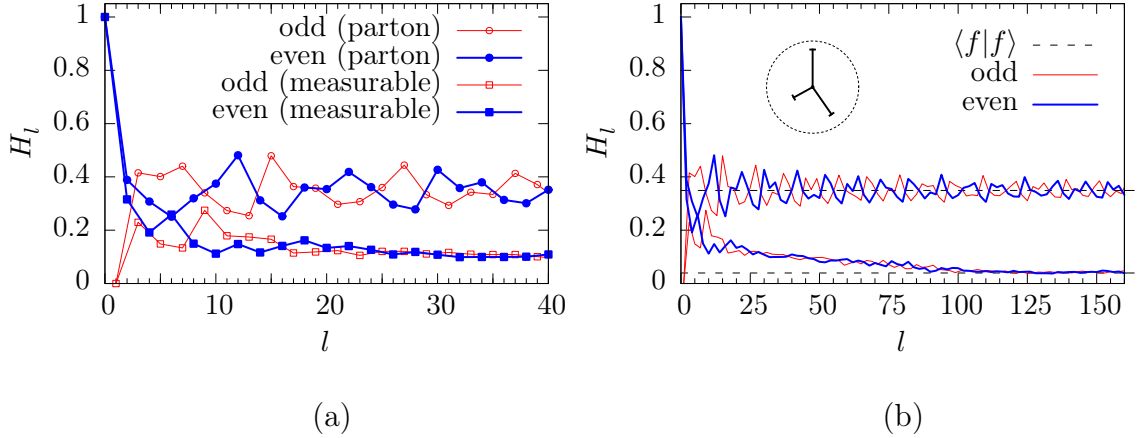


Figure 4.9. The power spectrum of a  $q\bar{q}g$  final state with a hard gluon (3-jet-like). (a) The low- $l$  behavior and (b) the asymptotic behavior, with kinematic depiction.  $H_l$  is depicted for (circle) partons and (square)  $\mathcal{O}(100)$  measurable particles. Odd powers are hollow, red, and thin. Even powers are blue, filled and thick.

Turning our attention to Figure 4.9, where the gluon is significantly *harder*, we find additional strange behavior. The kinematic depiction shows that this event is truly three-jet-like. This is reflected in the power spectrum, where  $H_3 > H_2$  for the original partons. But while the measurable particles'  $H_3$  is still large (relative to their  $H_2$ ), there is a significant overall mismatch between the power spectra of the measurable particles and their partonic forebearers. Additionally, the oscillations in the power spectra are far less regular and at a higher frequency. In fact, the vast differences in the shape of the power spectra between the 2-jet-like and 3-jet-like events indicate that  $H_l$  is quite sensitive to an event's coarse, jet-like structure. Finally, just as with the 2-jet-like event, the power spectra for partons and measurable particles asymptotically approach a mysterious plateau at  $H_l \approx \langle f|f \rangle$ .

We have examined only two events, but the same behavior appears in *each* of the hundreds of jet-like events studied. A major takeaway is that the power spectrum seems dominated by the coarse, jet-like structure of the event, a result consistent (at least at low- $l$ ) with the angular resolution  $\xi = 2\pi/l$  of the spherical harmonics. But we have yet to explain some other, general features:

- **$H_l$  oscillates:** Unlike the CMB, there are no broad shapes to fit.
- **$H_l$  is unending:** As  $l \rightarrow \infty$ ,  $H_l$  oscillates around some power spectrum plateau at  $H_l \sim \langle f|f \rangle$ . Therefore, its total power  $T = \sum H_l$  is infinite.
- **$N \neq n$ :** The peaks and valleys of the measurable power spectrum with  $N$  particles does not match the  $H_l$  for their  $n$  originating partons. This is not a problem with jet-parton duality at low  $Q$ , as these simulations were conducted at  $\sqrt{S} = 400$  GeV. The shape of QCD radiation inside the jets matters, and seems to manifest as a slow attenuation of  $H_l$  to its lower asymptotic value.

We will spend the rest of this chapter understanding these effects, so that we can gain meaningful access to the massive amount of correlated information embedded in the power spectrum.

The remainder of this section will examine the side-effects of oscillation, further demonstrating that calculating Fox-Wolfram distributions  $f(H_l)$  is indeed the wrong approach. Section 4.3 will show that the unending  $H_l$  plateau is due to sampling noise in a discrete sample. Finally, Section 4.4 will show that sampling noise can be mitigated by giving tracks and towers physical extent (versus infinitesimally thin  $\delta$  functions). This will help explain the slow attenuation of the  $H_l$  for measurable particles, which power jets must be able to accurately model. This will put us in a position to begin developing the power jet model in Chapter 5.

**4.2.3.2  $f(H_l)$  distributions for three-parton events.** We will now replicate the  $f(H_l)$  distributions originally calculated by Fox and Wolfram (Figure 4.5). This requires integrating over each process' differential cross section in a 3-parton phase space, which will get rather technical. But when we are done, we will understand quite precisely why calculating  $f(H_l)$  is not the right way to use the power spectrum, solidly motivating the basic concept of the power jets model.

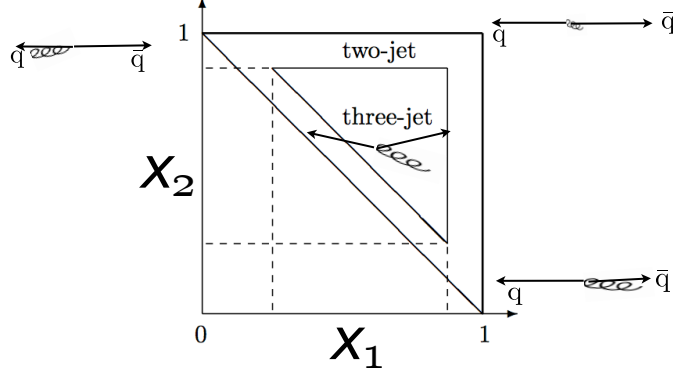


Figure 4.10. The phase space for a  $q\bar{q}g$  event.

For  $e^+e^- \rightarrow \gamma \rightarrow j_1 j_2 j_3$ , where  $j$  are massless partons (a quark or gluon) that sum to the four-momentum  $\mathbf{p}_0$  of the photon  $\gamma$  (with  $\mathbf{p}_i^2 = 0$  and  $\mathbf{p}_0^2 = s$ ),

$$\mathbf{p}_0 = \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3. \quad (4.33)$$

It is convenient to define a dimensionless variable

$$x_i \equiv \frac{2\mathbf{p}_i \cdot \mathbf{p}_0}{s} = \frac{2E_i}{\sqrt{s}}, \quad (4.34)$$

where  $E_i$  is the particle's final state energy in the CM frame. Total energy conservation then requires  $x_3 = 2 - x_1 - x_2$ .

Using  $x$  is useful for this 3-parton system because we can move one particle to the LHS of the equation and square it, calculating (where  $i \neq j \neq k$ )

$$\frac{1}{s}(\mathbf{p}_0 - \mathbf{p}_i)^2 = (1 - x_i) = \frac{2}{s}\mathbf{p}_j \cdot \mathbf{p}_k = \frac{1}{2}x_j x_k (1 - \cos(\xi_{jk})). \quad (4.35)$$

Hence, from  $\mathbf{p}$  conservation alone we constrain all inter-particle angles

$$\cos(\xi_{jk}) = 1 - \frac{2(1 - x_i)}{x_j x_k}. \quad (4.36)$$

Since  $H_i$  is invariant to the absolute orientation of the event, the two remaining d.o.f. ( $x_1$  and  $x_2$ ) constrain the entire power spectrum.

Since  $E_i \geq 0$ , we know that  $x_i \geq 0$ . This means that the last term in Equation 4.35 can never be negative, which tells us that  $x_i \leq 1$ . Applying this constraint to  $x_3$ , we find that  $x_1 + x_2 \geq 1$ . This constrains the phase space of  $x_1$  and  $x_2$  to lie in the triangle depicted in Figure 4.10, which depicts the 3-parton configurations for the differential cross section of  $e^+e^- \rightarrow q\bar{q}g$  [147], where  $x_1$  is for  $q$  and  $x_2$  is for  $\bar{q}$ ,

$$\frac{d\sigma}{\sigma dx_1 dx_2} = \frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}. \quad (4.37)$$

In the central “three-jet” region,  $\frac{d\sigma}{\sigma dx_1 dx_2}$  is well-behaved. But along the top and right edges, the gluon is either soft or collinear, and  $\frac{d\sigma}{\sigma dx_1 dx_2}$  becomes singular. A jet definition avoids this singularity by defining some phase space boundary, beyond which partonic configurations are experimentally indistinguishable from each other (they look like two jets). This creates a “two-jet” region which contains all the singularities, and in which the gluon is integrated out.

Normally, one still needs a regularization scheme to integrate over the two-jet region and obtain the total cross section. But for the sake of generating  $f(H_l)$ , the total cross section doesn’t matter; we simply normalize to the area of the 3-jet region over which we integrate. However, we will need to define the two-jet boundary using some jet definition, so for the sake of this section we temporarily choose  $H_2 \geq 0.95$  as our two-jet definition (this is not how power jets will be defined).

The  $X \rightarrow ggg$  process shares the same kinematics and phase space cuts, it just has a different differential cross section. Because the three gluons are indistinguishable, the distribution is symmetric [129]

$$\frac{d\sigma}{\sigma dx_1 dx_2} = \frac{6}{(\pi^2 - 9)x_1^2 x_2^2 x_3^2} (x_1^2(1-x_1)^2 + x_2^2(1-x_2)^2 + x_3^2(1-x_3)^2). \quad (4.38)$$

This differential cross section is totally finite, so no additional regularization is needed. We compose each Fox-Wolfram distribution  $f(H_l)$  via Monte Carlo integration; we draw  $x_1$  randomly from the uniform distribution  $U(0,1)$  and  $x_2$  randomly from

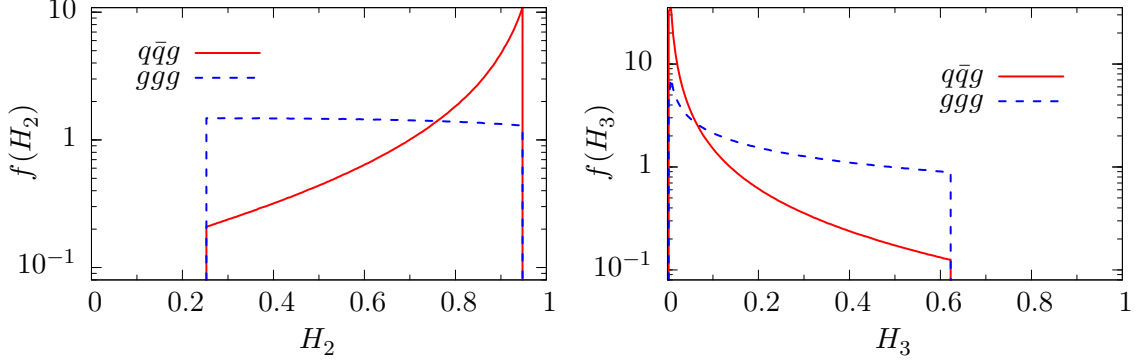


Figure 4.11. The probability  $f(H_2)$  and  $f(H_3)$  for  $q\bar{q}g$  and  $ggg$  final states, given phase space points  $(x_1, x_2)$  where  $H_2 \leq 0.95$ .

$U(1 - x_1, 1)$ , then bin the resulting  $H_l$  into separate histograms for each  $l$ , weighting each phase space point by  $w = x_1 \times \frac{d\sigma}{\sigma dx_1 dx_2}$  (the weight of the phase space point divided by the probability of drawing it). Monte Carlo integration works best here because mapping phase space to  $H_l$  is a topological nightmare for  $l > 2$ .

In Figure 4.11 we replicate the predictions of Fox and Wolfram (see Fig. 4.5, which has a different aspect ratio, but the same shapes). Just like their predictions, the  $q\bar{q}g$  process has a much higher preference for a 2-jet-like event, and exhibits much higher values of  $H_2$ . In Figure 4.12 we examine  $H_8$  and  $H_9$  for the same process. The smooth shapes of  $H_2$  and  $H_3$  are replaced with quite “peaky” ones (reminiscent of attractors in nonlinear systems, such as the bifurcation diagram of the logistic map). Apparently, the periodic oscillations in the asymptotic  $H_l$  of massless partons (as seen in Fig. 4.9b and 4.9b) have a non-trivial preference for certain values at high  $l$  (the peaks), and these values depend on the underlying physics process.

To probe this behavior further, in Figure 4.13 we plot the  $H_8$  and  $H_9$  distributions for the portions of phase space in which  $H_2$  has a very specific value:  $0.45 \leq H_2 \leq 0.5$ . This results in more “peaky”  $H_8$  and  $H_9$  curves, but this time the two physics processes are nearly indistinguishable. This further supports the idea

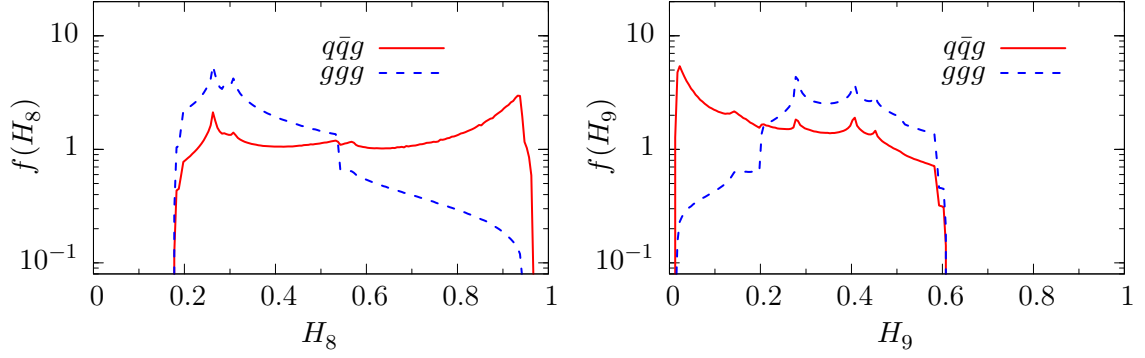


Figure 4.12. The probability  $f(H_8)$  and  $f(H_9)$  for  $q\bar{q}g$  and  $ggg$  final states, given phase space points  $(x_1, x_2)$  where  $H_2 \leq 0.95$ .

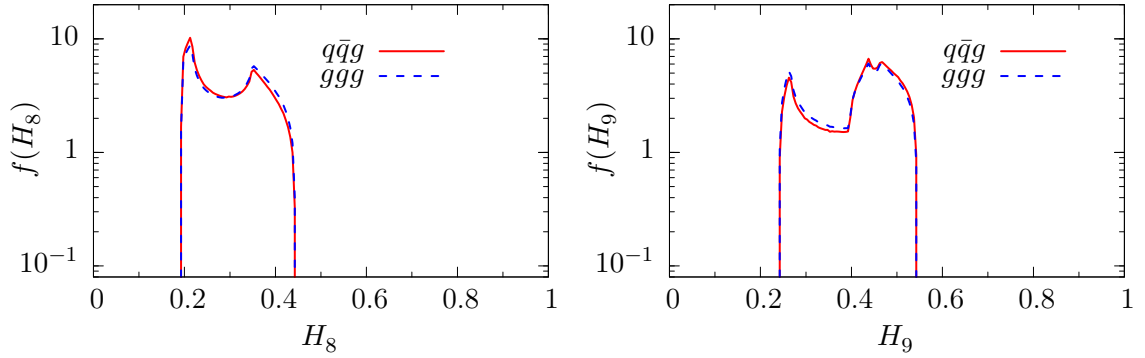


Figure 4.13. The probability  $f(H_8)$  and  $f(H_9)$  for  $q\bar{q}g$  and  $ggg$  final states, given phase space points  $(x_1, x_2)$  where  $0.45 \leq H_2 \leq 0.5$ .

that periodic  $H_l$  oscillations are a manifestation of intense correlations between the power spectrum at high and low  $l$ , because now the shape of the distributions depends much more on the value of  $H_2$  than it does on the originating process.

These results demonstrate the main problem with  $f(H_l)$  distributions — it not possible to construct a meaningful likelihood from a few independent  $f(H_l)$ , because  $H_l$  is highly inter-correlated. Accounting for these correlations *a priori* (when theory predictions are made), then accounting for experimental errors and  $Q$ -dependent jet formation is an extremely complicated beast. This is why the power jets method skips

$f(H_l)$  entirely, and directly fits the jet-like model to the observed power spectrum.

### 4.3 Particle multiplicity and sampling noise

For the power jet method to be successful, we must understand one of the primary observations of the previous section. Why does  $H_l$  oscillate around some asymptotic value  $H_l \sim \langle f|f \rangle$ ? And for the measurable particles in jets, what causes the gradual attenuation to this value?

We will find that the asymptotic power spectrum is inversely proportional to particle multiplicity  $N$ ; this defines a noise floor. The larger the multiplicity, the lower the floor, and the more useful information can be extracted from higher degree  $l$ . This is easy to understand physically: *more* particles give a better sample of the underlying event shape  $\rho(\hat{r})$ . Understanding the mathematical mechanism by which sampling noise manifests will allow us to define a scheme that mitigates the noise, while simultaneously replicating the slow attenuation to the noise floor. This will give us the final tools we need to begin designing power jets.

**4.3.1 Multiplicity limits information.** The primary limiting factor when extracting information from  $H_l$  is particle multiplicity. We can understand this from a relatively simple question. **Q:** Given a power spectrum, but not the particles which created it, how much information is needed to exactly reproduce  $H_l$ ? **A:** The information necessary to describe the originating particles.

For two massless particles in their CM frame, we need *no* information. This is because we know that such a system must have  $f_1 = f_2 = 0.5$ , and  $H_l$  is invariant to the axis which they define. Adding a third massless particle, conservation of momentum,

$$\sum_i \vec{p}_i = 0, \quad (4.39)$$

constrains it:

$$\vec{p}_3 = -(\vec{p}_1 + \vec{p}_2). \quad (4.40)$$

In fact, we have already solved this system when we derived 3-parton phase space in the previous section (Section 4.2.3.2) — only two d.o.f. are needed to describe the  $H_l$  of three massless particles.

Via induction, the minimum amount of information needed to exactly reproduce an  $N$ -particle power spectrum is proportional to particle multiplicity  $N$ :

$$\text{d.o.f.} = 3(N - 2) - 1. \quad (4.41)$$

We start with 3 d.o.f. for every particle, then subtract 3 d.o.f. for conservation of momentum in the CM frame, 3 d.o.f. for the three Euler angles of  $H_l$ 's rotational invariance, and 1 d.o.f. for scaling out  $E_{\text{tot}}$ . That there exists a minimum amount of information needed to reproduce an  $N$ -particle power spectrum  $H_l$  is highly suggestive that Equation 4.41 is also the *maximum* amount of information contained in  $H_l$ . This implies that using  $H_l$  to arbitrarily high  $l$  will not be meaningful; there must be some  $l_{\text{max}}$  above which no new information lies (although we cannot simply assume that  $l_{\text{max}} = \# \text{d.o.f.}$ , since the  $H_l$  are correlated).<sup>15</sup> This conclusion will become increasingly obvious as we continue to explore event multiplicity.

**4.3.1.1 Multiplicity scales with  $Q$ .** If the information content of  $H_l$  is limited by particle multiplicity, there is an immediate experimental constraint, because particle multiplicity within a jet scales with jet energy (which itself scales with the energy scale  $Q$  of the hard scatter).

In addition to being more free from its siblings, a high energy parton has more phase space available for showering because it is farther from the hadronization scale.

---

<sup>15</sup>Note that Equation 4.41 is far fewer d.o.f. than the naïve estimate of  $N(N - 1)/2$  energy correlations and  $N(N - 1)/2$  angular correlations.



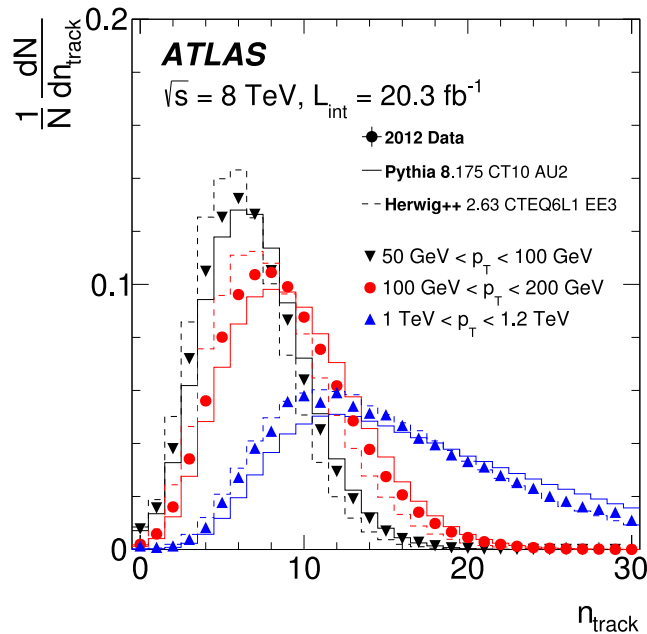


Figure 1: The distribution of the number of reconstructed tracks associated with a jet (not unfolded) in three example jet  $p_T$  ranges:  $50 \text{ GeV} < p_T < 100 \text{ GeV}$ ,  $100 \text{ GeV} < p_T < 200 \text{ GeV}$ , and  $1 \text{ TeV} < p_T < 1.2 \text{ TeV}$  for data and for Pythia 8 and Herwig++ predictions. The simulated samples are described in Sect. 3. The data points have statistical uncertainties which in all bins are smaller than the marker size. There is one entry per jet.

Figure 4.14. Charged particle multiplicity  $n_{\text{track}}$  vs. jet  $p_T$  at the LHC [148].

This effect is clearly visible in Figure 4.14, where track multiplicity increase with jet energy. This study compares observed data (points) to the predictions of two different showering Monte Carlos (histograms) [148]. While it is difficult to determine a jet's total multiplicity (since neutral particles are not individually observable), the ratio of a jet's track energy to its total energy ( $f_{\text{track}} = \sum p_T^{\text{track}}/p_T^{\text{jet}}$ ) gives a rough estimate of its total particle multiplicity. In Figure 4.15, theory predicts  $f_{\text{track}} \approx 60\%$ , but ATLAS measures  $f_{\text{track}} \approx 50\%$  (with a rather large energy dependence from detector effects [149]). Thus, a jet's total particle multiplicity is roughly double that of its track multiplicity, so that Figure 4.14 reveals that low- $Q$  collisions have a more discrete sampling of event shape, making them more susceptible to random variations.

**4.3.2 The multiplicity plateau.** Looking again at the power spectra of the 3-jet-like event (Fig. 4.9b), it is relatively easy to see the loss of new information as

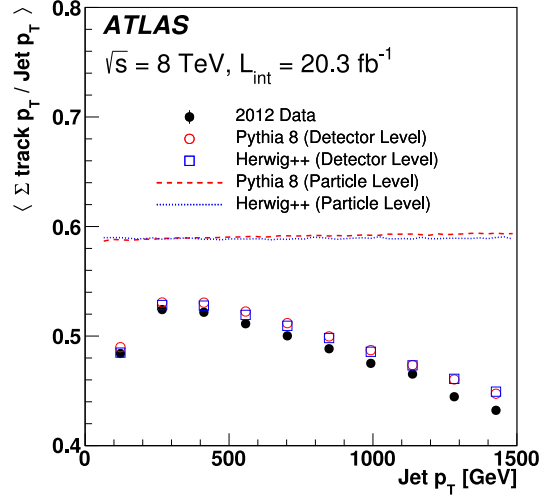


Figure 4.15. The ratio  $\sum p_T^{\text{track}}/p_T^{\text{jet}}$  is a proxy for charged multiplicity to total particle multiplicity [149]. “The momentum ratio of charged particles to all particles is nearly 2/3 due to the number of pion species . . .” [149].

$l \rightarrow \infty$ . When  $H_l$  flattens to the plateau at  $H_l \sim \langle f|f \rangle$ , the oscillations around this value also diminish, making it ever easier to predict the next  $H_{l+1}$ . As discussed in Appendix A.1.3 (and Eq. A.11), high information content requires *un*-predictability.

To understand how to deal with the  $H_l$  plateau, we need to understand its root mathematical cause. Looking back at Equation 4.24, and recalling that  $P_n(1) = 1$ , one will find that the angular correlation matrix  $\Xi = |\hat{p}\rangle \cdot \langle \hat{p}| = \cos \xi_{ij}$  is always unitary along the diagonal (because the angle between each particle and itself is always zero). This is each particle’s self-correlation, which can be separated from the terms describing the inter-particle correlations:

$$H_l = \underbrace{\langle f|f \rangle}_{\text{self}} + \underbrace{\langle f| (P_l(|\hat{p}\rangle \cdot \langle \hat{p}|) - \mathbf{1}) |f \rangle}_{\text{inter-particle}}. \quad (4.42)$$

The angular scale of the power spectrum is  $\xi = 2\pi/l$  (see Eq. 4.10), and the inter-particle term will tend to vanish when  $l$  surpasses the scale of the smallest inter-particle angles in the event. This is because there are no correlations to resolve except for sampling noise, which appears increasingly random at high  $l$ . Once the

inter-particle correlations begin to interfere destructively, the constant self-correlation term  $\langle f|f \rangle$  will dominate  $H_l$ , leading to the asymptotic plateau. We will now relate  $\langle f|f \rangle$  to particle multiplicity by calculating its expected value.

**4.3.2.1 We expect  $\langle f|f \rangle \propto N^{-1}$ .** To study the expected value  $\text{Ex}(\langle f|f \rangle)$ , we can assume that there is some generic physics process that consistently produces a many-particle final state with variable multiplicity  $N$ . Yet the underlying physics is consistent, so one can write down a smooth probability distribution  $h(E)$  for the energy of the produced particles.

Converting this  $h(E)$  distribution to energy fraction  $h(f)$ , we find the undesirable property that, by construction, the mean of  $h(f)$  must be inversely proportional to  $N$  ( $\text{Ex}(f) = N^{-1}$ , since  $\sum_i f_i = 1$  and there are  $N$  particles). It is therefore useful to define the scale-free energy fraction

$$\tilde{f} \equiv \frac{f}{\text{Ex}(f)} = Nf, \quad (4.43)$$

since the probability distribution  $h(\tilde{f})$  has a shape that depends only on the physics process, but not the multiplicity of any given event (i.e.,  $\text{Ex}(\tilde{f}) = 1$ ). We then require that  $h(\tilde{f})$  has a finite variance

$$\sigma^2 = \text{Ex}(\tilde{f}^2) - \text{Ex}(\tilde{f})^2 = \text{Ex}(\tilde{f}^2) - 1. \quad (4.44)$$

Let us now study a random instance of this physics process. The energy fraction vector  $|f \rangle$  for this instance draws  $N$  energy fractions  $\tilde{f}$  from  $h(\tilde{f})$ , then normalizes to their sum:

$$|f \rangle = \frac{\{\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_N\}}{\tilde{f}_1 + \tilde{f}_2 + \dots + \tilde{f}_N}. \quad (4.45)$$

The expectation value of  $\langle f|f \rangle$  is therefore

$$\text{Ex}(\langle f|f \rangle) = \int d\tilde{f}_1 d\tilde{f}_2 \dots d\tilde{f}_N \frac{\tilde{f}_1^2 + \tilde{f}_2^2 + \dots + \tilde{f}_N^2}{(\tilde{f}_1 + \tilde{f}_2 + \dots + \tilde{f}_N)^2} h(\tilde{f}_1) h(\tilde{f}_2) \dots h(\tilde{f}_N). \quad (4.46)$$

Assuming each  $\tilde{f}_i$  is independent (i.e., all correlations are built into the shape of  $h(\tilde{f})$ ), we can treat each  $\tilde{f}_i$  separately, and use the linearity of expectation to obtain

$$\text{Ex}(\langle f|f \rangle) = \frac{N \text{Ex}(\tilde{f}^2)}{(N \text{Ex}(\tilde{f}))^2} = \frac{1}{N} \text{Ex}(\tilde{f}^2) = \frac{1}{N}(1 + \sigma^2). \quad (4.47)$$

This tells us that the height of the  $H_l$  plateau (Eq. 4.42) is inversely proportional to the particle multiplicity. And since  $\sigma^2$  must be non-negative, this multiplicity plateau is somewhat greater than  $1/N$ .

**4.3.2.2 The smallest possible  $\langle f|f \rangle_{\min} = N^{-1}$ .** As a sanity check, we can also find the absolute minimum possible value of  $\langle f|f \rangle$  *without* assuming a some well-behaved physics process with a limiting distribution  $h(\tilde{f})$ . We define a normalized energy fraction vector for some arbitrary set of particles

$$|f\rangle = \{f_1, \dots, f_{N-1}, f_N\}, \quad \text{where} \quad f_N \equiv \left(1 - \sum_{i=1}^{N-1} f_i\right) \geq 0. \quad (4.48)$$

Evaluating the gradient  $\vec{\nabla} \langle f|f \rangle$ , each  $f_i$  minimizes  $\langle f|f \rangle$  when

$$\partial_i \langle f|f \rangle = \partial_i (f_i^2 + f_N^2) = 2(f_i - f_N) = 0 \quad (4.49)$$

(since the only terms in the inner product which depend on  $f_i$  are  $f_i^2$  and  $f_N^2$ ). This tells us that  $\langle f|f \rangle$  is minimized when  $f_i = f_N$ . If every energy fraction is equal to the final energy fraction, then all  $f_i$  must be the same. The total normalization then requires  $f_i = N^{-1}$ , so that one finds

$$\langle f|f \rangle_{\min} = N(N^{-2}) = \frac{1}{N}. \quad (4.50)$$

This result corresponds exactly to  $\sigma = 0$  for Equation 4.47, and once again we find that the multiplicity plateau is intrinsically linked to particle multiplicity.

**4.3.3 Studying sampling noise with isotropy.** We have just predicted a connection between the asymptotic noise plateau of  $H_l$  and the multiplicity of its sample. This is a strong prediction, which we need to test with an actual event shape. For

simplicity, we choose a toy model which is guaranteed to be featureless *except* for sampling noise.

The most trivial event shape  $\rho(\hat{r})$  is a perfectly isotropic, homogeneous event shape; each patch of solid angle is given the same helping of cross section

$$\frac{d\sigma}{d\Omega} = \frac{\sigma}{4\pi} \quad \Longrightarrow \quad \rho(\hat{r}) = \frac{1}{4\pi} = \sqrt{\frac{1}{4\pi}} Y_0^0(\hat{r}). \quad (4.51)$$

Since this event shape *is itself* a spherical harmonic, its power spectrum is trivial;  $H_l = 0$  for  $l > 0$  and  $H_0 = 1$  (while  $H_0 = 1$  for all  $\rho(\hat{r})$ , it is especially true here).

However, this trivial power spectrum is for a continuous event shape, which a particle detector can only sample discretely. In any discrete sample of a continuous distribution, there always will be fluctuations from the expectation. A QCD event is especially discrete because: (i) The fields are quantized, so particle multiplicity is intrinsically finite, giving rise to random fluctuations. (ii) The angular positions of neutral particles are constrained to a calorimeter lattice, creating spatial quantization artifacts. Both effects create sampling noise in a QCD power spectrum, and our isotropic event shape provides the perfect vehicle to understand them.

**4.3.3.1 The sampling noise of a finite (“track-only”) sample.** First we will study the noise of *finite* sampling, assuming perfect angular resolution. This is very much like a power spectrum built only from charged particle tracks (i.e., a “track-only”  $H_l$ ). We randomly sample isotropic massless particles via the algorithm outlined in Appendix B.1 and depict their power spectra in Figure 4.16.

A log-scale makes the full behavior quite visible. At low  $l$ , the power spectra match the expectations of an isotropic distribution; there is little power in anything but the  $l = 0$  mode (not depicted). This indicates that the isotropic sampling algorithm is quite successful, at least at low angular resolution. As the angular resolution increases with  $l$ , the power spectrum begins to detect the random correlations inside

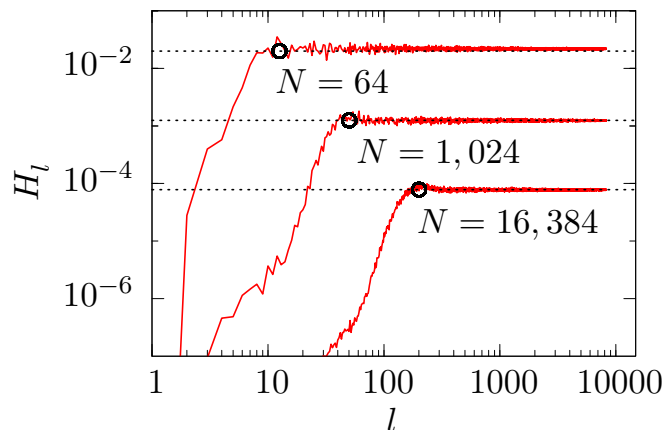


Figure 4.16. The power spectrum for three random, isotropic samples ( $N = 64$ ,  $N = 1,024$ ,  $N = 16,384$ ). For each sample, the  $\langle f|f \rangle$  prediction of Equation 4.47 is depicted with the dotted line. The circle shows where  $l = \pi/(2 \arcsin(1/\sqrt{N}))$ .

the discrete sample, and the power steadily rises until it flattens to each sample's multiplicity plateau.

Recall that Equation 4.42 explained the multiplicity plateau by assuming that inter-particle correlations destructively interfere at sufficiently large  $l$ , so that  $H_l$  is dominated by the sum of self-correlations  $\langle f|f \rangle$ . A dotted line for each sample size  $N$  shows the  $\text{Ex}(\langle f|f \rangle)$  predicted by Equation 4.47 (using  $\text{Ex}(\tilde{f}^2) \approx 1.28$ , a value determined in Appendix B.1.1). It is clear that this prediction is accurate for all three samples, even though each power spectrum stabilizes to its actual  $\langle f|f \rangle$ , which contains statistical fluctuations that displace it from the expectation value of its limiting distribution  $h(\tilde{f})$ .

We now have a consistent model to explain how finite sampling creates a noisy asymptotic power spectrum — once  $H_l$  hits the multiplicity plateau, it is difficult to imagine extracting useful information from the flat, shapeless power spectrum. Having verified that the *height* of the multiplicity plateau is  $\langle f|f \rangle$ , one can simply calculate  $\langle f|f \rangle$  for a given event, find the  $l$  where  $H_l$  approaches that height, and use no higher  $l$ . But it is also possible to predict where this will occur, because we expect

inter-particle correlations to become totally random and destructively interfere when the power spectrum's angular resolution  $\xi = 2\pi/l$  drops below the mean inter-particle angle in the sample. This is the point where  $H_l$  begins probing the smallest angular scale in the event.

To find this maximum useful  $l$ , we can calculate each particle's expected Voronoi area  $A$  — the locus of points for which no other particle is closer, it's unique turf on the unit sphere (see Fig. B.1 for a visual depiction). The average isotropic particle has  $\text{Ex}(A) = 4\pi/N$ , and this area can be approximated by a circular cap of angular radius  $\theta_r$ ;

$$A = \int_0^{\theta_r} 2\pi \sin(\theta) d\theta = 4\pi \sin^2\left(\frac{1}{2}\theta_r\right) \implies \theta_r = 2 \arcsin\left(\sqrt{\frac{A}{4\pi}}\right) \quad (4.52)$$

This gives the average adjacent particles an angular separation of  $\xi = 2 \text{Ex}(\theta_r)$ . The multiplicity plateau should begin at the  $l$  which achieves this angular resolution;

$$l_{\max} = \frac{\pi}{2 \arcsin(1/\sqrt{N})} \underset{N \gg 1}{\approx} \frac{\pi}{2} \sqrt{N}. \quad (4.53)$$

This prediction is depicted for each sample in Figure 4.16 with a black circle, and it quite successfully predicts the transition to noise. Note that this is a lower limit on  $l_{\max}$ , derived by assuming that particles are distributed as sparsely as possible; events with dense clusters will have a smaller minimum angular scale.

**4.3.3.2 The sampling noise of a calorimeter.** In addition to noise from an intrinsically discrete sample, we also expect artifacts when space is quantized, as in a calorimeter lattice.

We study the  $e^+e^-$  pseudo-detector described in Section 1.3 — an “equal area” spherical calorimeter covering  $4\pi$  of solid angle as uniformly as possible, with only two holes for the colliding beams. Each calorimeter tower covers approximately the same solid angle  $\Omega \approx (\Delta\theta)^2$  (requiring larger  $\Delta\phi$  in the more forward/backward azimuthal

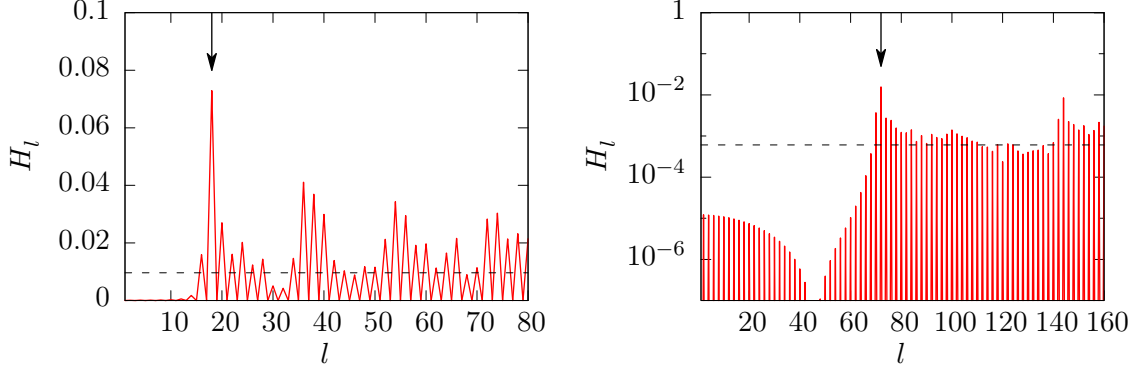


Figure 4.17. The power spectrum for a uniformly filled calorimeter in a spherical pseudo-detector with (left)  $\Delta\theta = 20^\circ$  towers and (right)  $\Delta\theta = 5^\circ$  towers on a log-scale. The dashed line shows  $1/N$  and the arrows indicates  $l = 2\pi/\Delta\theta$ .

belts). To focus on the systematic effects of the angular scale  $\Delta\theta$ , we can bathe this calorimeter in isotropic energy long enough that random fluctuations vanish. This creates a “particle” at the center of each tower with an energy fraction  $f_i = \Omega_i/(4\pi)$  proportional to its exact solid angle. The sum of all towers in such an “event” is nearly isotropic *and* homogeneous.

Figure 4.17 shows the power spectrum for (left)  $5^\circ$  and (right)  $20^\circ$  towers (on a log-scale). As with the random, isotropic particles in Figure 4.16,  $H_l$  is initially consistent with isotropy (albeit with some small beam-hole artifacts, visible in Fig. 4.17b). But when  $l$  becomes large enough to probe the towers’ angular dimension  $\Delta\theta$ , the power rises dramatically to reveal the artifacts of the calorimeter lattice. Instead of a plateau at  $\sim 1/N$ , it is a bed of nails at approximately the same height; this highly-oscillatory shape occurs because every “particle”  $\vec{p}_i$  has an equal and opposite partner on the other side of the calorimeter (with  $\vec{p}_j = -\vec{p}_i$ ).  $P_n(\pm 1) = (\pm 1)^n$ , so for odd/even  $l$ , the pair’s self-correlations are exactly canceled/doubled by their inter-particle correlation

$$f_i^2 P_l(1) + f_j^2 P_l(1) + 2f_i f_j P_l(-1) = \begin{cases} 4f_i^2 & \text{even} \\ 0 & \text{odd} \end{cases}. \quad (4.54)$$



Given a calorimeter’s intrinsically limited angular scale and lattice artifacts, one might be tempted to use a track-only power spectrum. However, we saw in the previous section that the angular resolution of a track-only sample was limited by its multiplicity. Similarly, while the onset of detector artifacts in Figure 4.17 is primarily governed by the angular scale of the towers, multiplicity governs the asymptotic magnitude of  $H_l$  as  $l \rightarrow \infty$  (which is why the artifacts are relatively small). The multiplicity sacrifice of a “track-only” power spectrum is just too high. Furthermore, we will soon find a better way to build limited angular resolution into an observation.

**4.3.4 The side effects of  $\delta$  distributions.** The sampling noise effects seen in the previous two sections are not unique to  $H_l$ ; they are a general feature of a discrete sample decomposed into an infinite basis. Understanding this fact will give us the clues we need to design a tool for mitigating sampling noise in the next section.

Figure 4.18 shows two squares filled with  $N = 256$  points; one with uniformly distributed random points and the second with a uniform grid. Taking the discrete Fourier transform of these points, we can calculate their spectral power by plotting the norm of these Fourier coefficients. For random points, the power is spread evenly across all the frequencies (with an appreciable amount of random jitter), while for a uniform grid, the power is concentrated into spikes corresponding to the grid’s period.

This behavior results from describing each point as a  $\delta$  function. Consider the 1D Fourier transform for a set of  $N$  points  $\{x_i\}$  randomly distributed in some interval

$$f(k) = \int_{-\infty}^{\infty} \frac{1}{N} \left( \sum_i \delta(x - x_i) \right) \exp(-2\pi i x k) = \frac{1}{N} \sum_i \exp(-2\pi i x_i k). \quad (4.55)$$

This  $f(k)$  is the sum of uncorrelated complex sinusoids of many different frequencies, so that its norm is flat and jittery because there is no feature to encourage destructive or constructive interference. Hence, the power spectrum  $|f(k)|$  goes as  $1/N$ , and looks

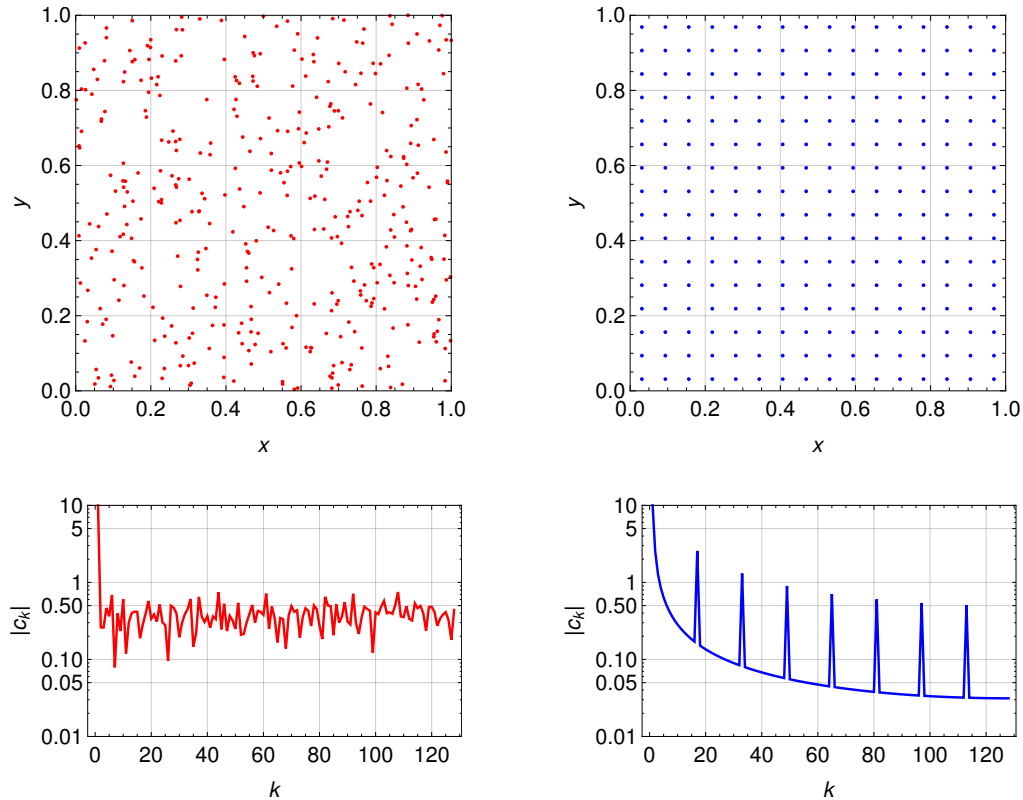


Figure 4.18. A unit square filled with (a) uniformly random points and (b) a uniform grid, with their respective discrete Fourier transforms.

like flat white noise.<sup>16</sup> Conversely, the Dirac comb  $f(x)$  depicted in Figure 4.19 (an infinite train of periodic delta functions) has a Fourier transform  $f(k)$  that is also a Dirac comb. A Dirac comb  $f(x)$  of *finite* length has an  $f(k)$  which is a “softened” Dirac comb — its spikes have thickness, and are connected by a  $\sim 1/N$  baseline.

This explains the equivalent shapes seen in the asymptotic power spectra of the track-only and calorimeter samples; they use an event shape built from  $\delta$  functions:

$$\rho(\hat{r}) = \sum_i f_i \frac{\delta^2(\hat{r} - \hat{p}_i)}{\sin \theta_i}. \quad (4.56)$$

The total square power  $T = \sum_l H_l$  of this shape must be infinite because objects with infinitesimal spatial extent ( $\delta$  functions) contain *all* frequencies. That  $T$  is infinite

---

<sup>16</sup> White noise has equal power in each frequency.

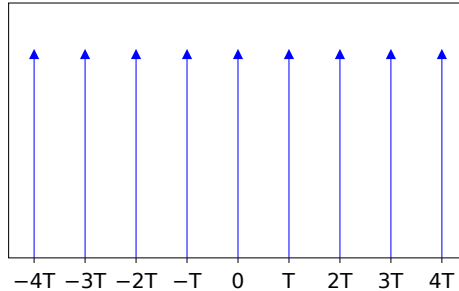


Figure 4.19. A Dirac comb is an infinite series of  $\delta$  functions with period  $T$  [150].

indicates that the  $H_l$  decomposition is inexact (see Eq. 4.11), and this inexactness manifests as the multiplicity plateau. This suggests that if we were to describe our event shape as the sum of objects with *spatial extent*,  $H_l$  would asymptotically vanish and the total square power  $T$  would be finite. This is the final piece we need to define power jets.

#### 4.4 Smearing out the sampling noise

In the previous section we showed that the asymptotic plateau of a QCD power spectrum is the inevitable result of discrete sampling noise. We also determined that its mathematical cause is an event shape composed of spatial  $\delta$  functions. The ultimate problem with  $\delta$  functions is that they assert impossibly perfect knowledge of a particle's angular position. In this section we will find that a sensible solution is to give each measured particle spatial extent — to smear it around its measured position using some shape function  $h(\hat{r})$ . This permits a fully continuous event shape

$$\rho(\hat{r}) = \sum_i f_i h_i(\hat{r}), \quad (4.57)$$

where each shape function  $h_i(\hat{r})$  surrounds the nominal/measured position  $\hat{p}_i$  of the  $i^{\text{th}}$  particle. If each shape function has a finite total square power  $T = \int_{\Omega} |h_i(\hat{r})|^2 d\Omega$ , then so will the whole power spectrum.

It is possible to interpret a shape function  $h(\hat{r})$  as either (i) a detector's mea-

surement uncertainty or (ii) an angular filter which accepts only meaningful, large-angle correlations, suppressing the small-angle correlations that are sensitive to discrete sampling noise. Both interpretations are critical in the definition of power jets, which will fit extensive jets (which have intrinsic shape) to the  $H_l$  calculated from a discrete set of point-like tracks and extensive towers (i.e., towers have some uncertainty about the angular location of the particle(s) initiating the shower). Additionally, the power jets fit should not be overly sensitive to random fluctuations in the detector's discrete sample. Because a shape function smears out small-angle information, it automatically gives more weight to low- $l$  information. In this section we will dive into the technical details of calculating  $H_l$  for extensive physics objects described by shape functions.

But before we dive into the weeds, let us develop a more qualitative understanding of why a shape function is needed. A far more intuitive way to look at a power spectrum is via its angular correlation function  $A(\xi)$ , which uses  $H_l$  as the coefficients of a Legendre series [133]

$$A(\xi) = \sum_{l=0}^{\infty} (2l+1) H_l P_l(\cos(\xi)). \quad (4.58)$$

The height of  $A(\xi)$  indicates how much of the event's energy is distributed at inter-particle angle  $\xi$ .<sup>17</sup> For example, a standard 2-jet-like event should be large at  $\xi = 0$  (jets have lots of nearby particles), large at  $\xi \approx \pi$  (there are many particles that are nearly back-to-back), and much smaller in between. But if  $H_l$  never goes to zero,  $A(\xi)$ 's series never converges. A naïve solution is to truncate the series to some  $l_{\max}$ .

In Figure 4.20 we show  $A(\xi)$  for a two-jet-like event, truncating its  $A(\xi)$  to two different values of  $l_{\max}$ . Figure 4.20a uses  $l_{\max} = 16$  to obtain a course look at  $A(\xi)$ , and it is possible to see the peaks and valleys are exactly where we expect them to be

---

<sup>17</sup>For  $A(\xi)$  to be fully integrable, it should be multiplied by  $\sin(\xi)$ . We leave out this complication since we use  $A(\xi)$  only for illustration.

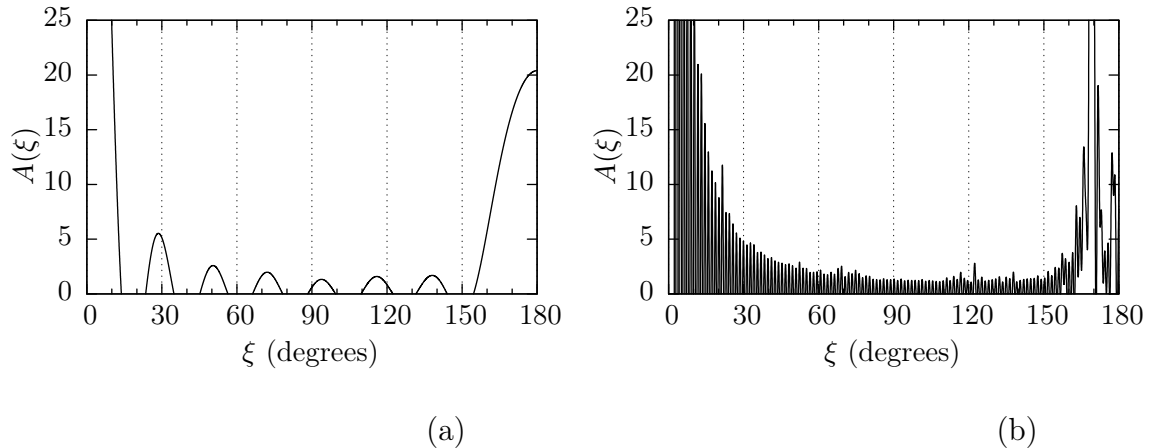


Figure 4.20. The raw angular correlation function for the measurable particles in a 2-jet-like event, truncating the series to (a)  $l_{\max} = 16$  and (b)  $l_{\max} = 256$ .

for a 2-jet-like event. Increasing the resolution in Figure 4.20b, we begin to resolve the structure of the two jets in the inter-jet peak at  $\xi \approx 170^\circ$ . But it also becomes clear that an arbitrary cut on  $l$  creates “ringing” artifacts that make the angular correlation function rather unusable. As we study shape functions in the remainder of this section, we will find that their main role is to force  $H_l \rightarrow 0$  asymptotically. Thus, when calculating  $A(\xi)$  the series will naturally converge, without an arbitrary  $l_{\max}$ . When we reexamine  $A(\xi)$  at the end of this section, we will find much more meaningful shapes.

**4.4.1 The extensive particle.** If we *need* a particle shape function, what should it look like? Placing the polar  $\hat{z}$ -axis parallel to the particle’s nominal/detected position  $\hat{p}$ , there is no reason the shape function should lack azimuthal symmetry. And to distribute the particle in polar angle  $\theta$ , why not use a Gaussian? Not only does this shape appear everywhere in nature, it describes multiple scattering at small angles [1] (e.g., propagating through a particle detector). A Gaussian prototype is

$$h(\theta) \approx C \exp\left(-\frac{\theta^2}{2\lambda^2}\right), \quad (4.59)$$

which smears the particle by some angle  $\lambda$  (with normalization  $C$ ). Yet this prototype

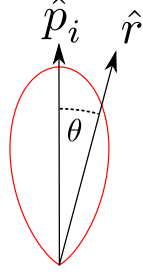


Figure 4.21. A particle shape function  $h(\hat{r})$  that is pseudo-Gaussian in polar angle  $\theta$  (relative to the particle's nominal direction of travel  $\hat{p}_i$ ). The radial position of the red line indicates the fraction of the particle that is found at each  $\hat{r}$ .

is immediately problematic because its slope  $h'(\theta)$  is discontinuous at  $\theta = \pi$  (i.e., the non-vanishing gradient always points *away* from the south pole).

As a second attempt, we can choose a function which approaches Equation 4.59 for  $\theta \ll 1$ , but where the  $\mathcal{O}(\theta^2)$  function inside the exponential goes smoothly to some constant as  $\theta \rightarrow \pi$ . A natural choice is

$$h(\theta) = C \exp\left(-\frac{2 \sin^2(\theta/2)}{\lambda^2}\right), \quad (4.60)$$

This shape function is vastly simplified by the change of variable  $z \equiv \cos(\theta)$

$$h(z) = C \exp\left(-\frac{1-z}{\lambda^2}\right), \quad (4.61)$$

because then  $d\Omega = dz d\phi$ , and we can easily solve for  $C$  by normalizing to unity

$$\int h(z) dz d\phi = 2\pi \lambda^2 h(z)|_{-1}^1 = 1 \quad \implies \quad C = \frac{1}{2\pi \lambda^2 (1 - e^{-2/\lambda^2})}. \quad (4.62)$$

To make this distribution generic for any particle  $\vec{p}$  and variable of integration  $\hat{r}$ , we simply replace  $z = \cos(\theta)$  with the inter-particle angle  $\cos \xi = \hat{r} \cdot \hat{p}$ :

$$h(\hat{r}) = \frac{1}{2\pi \lambda^2 (1 - e^{-2/\lambda^2})} \exp\left(-\frac{1 - \hat{r} \cdot \hat{p}_i}{\lambda^2}\right). \quad (4.63)$$

This shape function is roughly depicted in Figure 4.21.

Because this distribution is only *pseudo*-Gaussian, we need to establish a firm connection between the smearing angle  $\lambda$  and the particle's spatial extent. One

method is to calculate the radius  $\theta_r$  (in polar angle) of the circular cap which bounds some fraction  $u$  of the distribution's probability. Using  $z_r \equiv \cos \theta_r$ , we can calculate the Gaussian shape function's cumulative distribution function

$$\text{CDF}(z_r) = 2\pi \int_{z_r}^1 h(z) dz = \frac{1 - e^{-(1-z_r)/\lambda^2}}{1 - e^{-2/\lambda^2}}. \quad (4.64)$$

Setting the probability fraction  $u = \text{CDF}(\cos \theta_r)$ , we can solve for the cap's radius;

$$\cos \theta_r = 1 + \lambda^2 \log(1 - u(1 - e^{-2/\lambda^2})). \quad (4.65)$$

Using the substitution  $1 - \cos(\theta) = 2 \sin^2(\theta/2)$  helps prevent a floating point cancellation (see Sec A.2.1) when solving for  $\lambda$

$$\lambda = \sin\left(\frac{\theta_r}{2}\right) \sqrt{\frac{-2}{\log(1 - u(1 - e^{-2/\lambda^2}))}}. \quad (4.66)$$

For example, to get a shape function where a  $\theta_r = 5^\circ$  circular cap contains  $u = 90\%$  of the particle, one can use  $\lambda \approx 2.33^\circ$ .

**4.4.2 Calculating  $H_l$  for continuous distributions.** Smearing particles into extensive objects solves one problem but creates another; we can no longer use Equation 4.24 to calculate  $H_l$  via the simple contraction over  $N^2$  inter-particle angles and energy correlations:  $H_l = \langle f | P_l(\Xi) | f \rangle$ . This expression is valid only for a discrete event shape  $\rho(\hat{r})$ , built solely from  $\delta$  functions which collapse the  $H_l$  integral to some trivial linear algebra. Now that  $\rho(\hat{r})$  is continuous, we must use Equation 4.22 to sum over its  $Y_l^m$  coefficients;

$$H_l = \frac{4\pi}{2l+1} \sum_{m=-l}^l |\rho_l^m|^2. \quad (4.67)$$

With  $\rho(\hat{r})$  built from  $N$  shape functions pointing in different directions, this task becomes non-trivial. Yet with a few pages of some admittedly dense calculations, it only requires a small amendment to the previous form of  $H_l$ .

We compose the event shape from  $N$  extensive particles with individual shapes

$$\begin{aligned}\rho(\hat{r}) &= \rho_{(1)}(\hat{r}) + \rho_{(2)}(\hat{r}) + \dots + \rho_{(N)}(\hat{r}), \\ &= f_1 h_{(1)}(\hat{r}) + f_2 h_{(2)}(\hat{r}) + \dots + f_N h_{(N)}(\hat{r}).\end{aligned}\quad (4.68)$$

The  $\rho_l^m$  integral is separable for each shape function, so we can expand Equation 4.67 into a series of diagonal and cross terms for each shape function's  $Y_l^m$  coefficient  $h_{(1)l}^m$  (recalling that  $a^*$  indicates the complex conjugate)

$$H_l = \frac{4\pi}{2l+1} \sum_{m=-l}^{+l} \left( f_1^2 h_{(1)l}^m h_{(1)l}^{m*} + 2f_1 f_2 h_{(1)l}^m h_{(2)l}^{m*} + \dots \right). \quad (4.69)$$

Due to the addition theorem (Eq. 4.17) and the rotational invariance of the scalar product, each of the terms in this series is rotationally invariant (just like  $H_l$  itself)

$$\sum_{m=-l}^l h_{(1)l}^m h_{(2)l}^{m*} = \sum_{m=-l}^l \int_{\Omega} d\Omega \int_{\Omega'} d\Omega' Y_l^{m*}(\hat{r}) Y_l^m(\hat{r}') h_{(1)}(\hat{r}) h_{(2)}(\hat{r}') \quad (4.70)$$

$$= \int_{\Omega} d\Omega \int_{\Omega'} d\Omega' h_{(1)}(\hat{r}) h_{(2)}(\hat{r}') P_l(\hat{r} \cdot \hat{r}'). \quad (4.71)$$

The rotational invariance of each term in Equation 4.69 is extremely useful, because we can compute each term separately, choosing whatever orientation simplifies the calculation, so long as each pair of shape functions maintains its interior angle  $\xi_{ij}$ .

With the freedom to choose orientation, we can recall that the pseudo-Gaussian shape function which forced us down this road (Eq. 4.63) is azimuthally symmetric about its central axis. Defining an arbitrary cross term  $h_{(i)l}^m h_{(j)l}^{m*}$  between two particles smeared by the pseudo-Gaussian (with implicit summation over  $m$ ), a natural choice of orientation is to rotate the distributions such that  $h_{(i)}$  is parallel to the polar axis  $\hat{z}$ , with  $h_{(j)}$  sticking out off-axis at some arbitrary azimuthal angle  $\phi_j$  (see Fig. 4.22). Because  $h_{(i)}$  is azimuthally symmetric *and* aligned with the polar axis, the  $\phi$  integral of its coefficient  $h_{(i)l}^m$  will always be  $\int_0^{2\pi} e^{im\phi} d\phi = 0$ , so that  $h_{(i)l}^m = 0$  for any order  $m \neq 0$ .



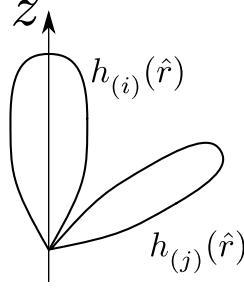


Figure 4.22. A pair of shape functions which have been rotated such that  $h_{(i)}(\hat{r})$  is parallel to the  $z$ -axis (on-axis), maintaining the interior angle with  $h_{(j)}(\hat{r})$  (off-axis).

Even though  $h_{(j)}$  is not parallel to the  $z$ -axis, which breaks the azimuthal symmetry of its integral and creates non-zero coefficients  $h_{(j)l}^{m*}$ , the inner product over  $m$  with  $h_{(i)l}^m$  ensures that only the  $m = 0$  coefficients are used. Thus, for both shape functions, we only need to calculate  $h_{(i)l}^0$ , the simplest of the  $Y_l^m$  coefficients

$$h_{(i)l}^0 = \sqrt{\frac{2l+1}{4\pi}} \int_{\Omega} P_l(\cos\theta) h_{(i)}(\hat{r}) d\Omega. \quad (4.72)$$

To simplify this calculation a bit more, note that Equation 4.69 (which accumulates all the diagonal and cross terms) has a  $4\pi/(2l+1)$  prefactor. This can be factored through to Equation 4.72 (as  $\sqrt{4\pi/(2l+1)}$ , splitting the full factor between the two coefficients in each term). This results in a more useful  $m = 0$  coefficient, denoted by a half-circle overline, which is simply the Legendre integral:

$$\check{h}_{(i)l} = \sqrt{\frac{4\pi}{2l+1}} h_{(i)l}^0 = \int_{\Omega} P_l(\cos\theta) h_{(i)}(\hat{r}) d\Omega. \quad (4.73)$$

Expanding upon this nomenclature, when the overline is straight (i.e., it becomes a bar), it denotes that the  $m = 0$  coefficient is calculated “on-axis”:

$$\bar{h}_{(i)l} = \int_0^{2\pi} d\phi \int_{-1}^1 dz P_l(z) h(z). \quad (4.74)$$

Defining these  $m = 0$  shape coefficients vastly simplifies the  $H_l$  for a total event shape  $\rho(\hat{r})$  composed of azimuthally symmetric shape functions:

$$H_l = \left( f_1^2 \bar{h}_{(1)l}^2 + 2f_1 f_2 \bar{h}_{(1)l} \check{h}_{(2)l} + 2f_1 f_3 \bar{h}_{(1)l} \check{h}_{(3)l} + \dots \right). \quad (4.75)$$

But note the imbalance in the definition of Equation 4.75: why is it always the first shape function which is placed on-axis? In our arbitrary cross term  $h_{(i)l}^m h_{(j)l}^{m*}$ , we could have originally rotated the pair so that  $h_{(j)}$  was on-axis and utilized the same  $m = 0$  trick. Thus, there are two equivalent ways to calculate every cross term:

$$\bar{h}_{(j)l} \check{h}_{(i)l} = \bar{h}_{(i)l} \check{h}_{(j)l}. \quad (4.76)$$

Since this equation was derived assuming only that both particle shapes are azimuthally symmetric about their central axis  $\hat{p}$  (i.e., they do not have to be the pseudo-Gaussian shape function of Eq. 4.63), we can craft a more useful statement. Let  $h_{(i)}$  be *some* azimuthally symmetric distribution, then choose  $h_{(j)}$  to be a very special azimuthally symmetric distribution, the delta function  $h_{(j)}(\hat{r}) = \frac{\delta^2(\hat{r}-\hat{p}_j)}{\sin\theta_j}$ , which has trivial coefficients

$$\bar{h}_{(j)l} = P_l(1) = 1, \quad (4.77)$$

$$\check{h}_{(j)l} = P_l(\xi_{ij}) \quad (4.78)$$

(the latter coefficient relies on maintaining the inter-particle angle  $\xi_{ij}$ ). Plugging these coefficients into Equation 4.76,

$$\check{h}_{(i)l} = \bar{h}_{(i)l} P_l(\xi_{ij}). \quad (4.79)$$

It doesn't matter that we forced  $h_{(j)}$  to be some special shape function, because  $h_{(i)}$  was kept as general as possible, and  $h_{(i)}$  is what this equation talks about.

This is an extremely important result! If extensive particles have azimuthally symmetric shape functions, we can combine Equation 4.75 with Equation 4.79 to calculate their “smeared” power spectrum

$$H_l = \left( f_1^2 \bar{h}_{(1)l}^2 + 2f_1 f_2 P_l(\cos \xi_{12}) \bar{h}_{(1)l} \bar{h}_{(2)l} + \dots \right). \quad (4.80)$$

The only remaining task is to calculate the on-axis coefficients  $\bar{h}_{(i)l}$ , an integral vastly simplified by azimuthal symmetry. The topological nightmare of moving  $h_{(j)}$  off-axis

to calculate  $\check{h}_{(j)_l}$  is avoided. Luckily, the objects we plan on describing with shape functions (tracks, towers and jet-like objects) are all well-approximated by azimuthal symmetry about their central axis, so this is the only power spectrum recipe we need.

**4.4.2.1 Calculating  $\bar{h}_l$  for extensive particles.** We now demonstrate the calculation of an on-axis coefficient  $\bar{h}_l$  for the pseudo-Gaussian shape function (Eq. 4.63). Setting  $\hat{p} \rightarrow \hat{z}$  allows us to use its  $z$ -version (Eq. 4.61), so that

$$\bar{h}_l = \int_0^{2\pi} d\phi \int_{-1}^1 dz C e^{-(1-z)/\lambda^2} = 2\pi C \int_{-1}^1 dz P_l(z) e^{-(1-z)/\lambda^2}. \quad (4.81)$$

To compute this integral, we can use a Legendre identity

$$(2l+1)P_l(z) = \frac{d}{dz} [P_{l+1}(z) - P_{l-1}(z)]. \quad (4.82)$$

We can then define an integral  $A_l = \bar{h}_l/(2\pi C)$  and solve it with integration by parts:

$$\begin{aligned} A_l &= \int_{-1}^1 dz e^{-(1-z)/\lambda^2} P_l(z) \\ &= \int_{-1}^1 dz e^{-(1-z)/\lambda^2} \frac{1}{2l+1} \frac{d}{dz} [P_{l+1}(z) - P_{l-1}(z)] \\ &= \frac{1}{2l+1} \left[ \left( e^{-(1-z)/\lambda^2} (P_{l+1}(z) - P_{l-1}(z)) \right) \Big|_{-1}^1 \right. \\ &\quad \left. - \int_{-1}^1 dz \frac{e^{-(1-z)/\lambda^2}}{\lambda^2} (P_{l+1}(z) - P_{l-1}(z)) \right] \\ &= \frac{1}{\lambda^2(2l+1)} (A_{l-1} - A_{l+1}). \end{aligned} \quad (4.83)$$

We were able to kill off the boundary term because the two Legendre polynomials will either both be odd or both be even, and thus at  $z = \pm 1$  will have identical values.

We can rearrange this result into a useful recurrence relation:

$$A_{l+1} = -(2l+1)\lambda^2 A_l + A_{l-1}. \quad (4.84)$$

And since  $\bar{h}_l$  is proportional to  $A_l$ , it has the same recurrence relation:

$$\bar{h}_{l+1} = -(2l+1)\lambda^2 \bar{h}_l + \bar{h}_{l-1}. \quad (4.85)$$

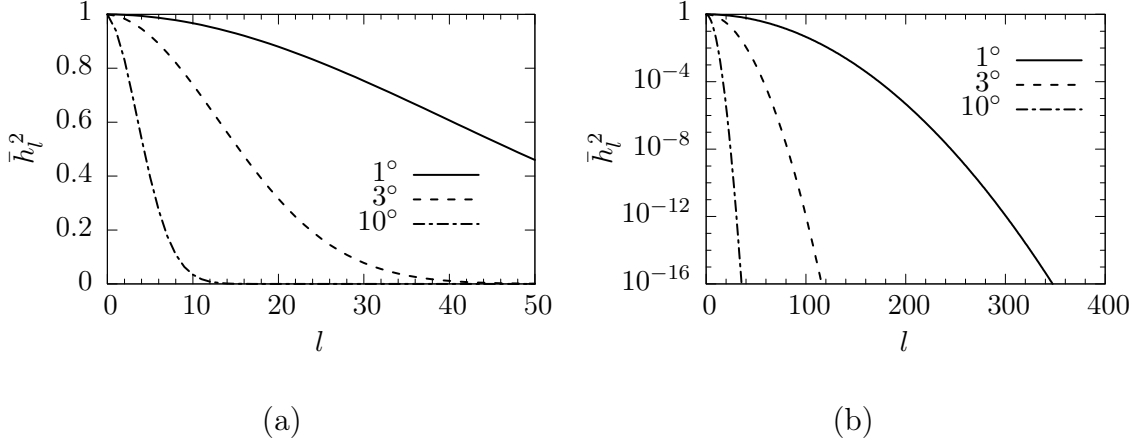


Figure 4.23. The squared, on-axis coefficient  $\bar{h}_l^2$  for the Gaussian shape function, depicting several smearing angles  $\lambda$  on a (a) linear and (b) log-linear scale.

So long as we can calculate the  $l = 0$  and  $l = 1$  terms by hand, the rest can be worked out recursively. The  $l = 0$  integral is equivalent to the normalization condition;  $\bar{h}_0 = 1$ , regardless of the distribution. For the Gaussian shape of Equation 4.63,

$$\begin{aligned}
 \bar{h}_1 &= \frac{1}{\lambda^2(1 - e^{-2/\lambda^2})} \int_{-1}^1 dz e^{-(1-z)/\lambda^2} z \\
 &= \frac{1}{\lambda^2(1 - e^{-2/\lambda^2})} \lambda^2(1 - \lambda^2 + e^{-2/\lambda^2}(1 + \lambda^2)) \\
 &= \frac{1 + e^{-2/\lambda^2}}{1 - e^{-2/\lambda^2}} - \lambda^2 = \frac{1}{\tanh(\lambda^{-2})} - \lambda^2
 \end{aligned} \tag{4.86}$$

Unfortunately, this recursion becomes unstable as  $\bar{h}_l \rightarrow 0$ , and needs special handling to behave sensibly. This procedure is explained in Appendix A.4.

**4.4.3 Shape functions as low-pass filters.** We now have all the pieces we need to calculate the power spectrum of extensive particles. For our test run, we choose to smear every particle in the event using the same pseudo-Gaussian smearing angle  $\lambda$ . In this case, all the identical on-axis  $\bar{h}_l$  in Equation 4.87 can be factored out, and the smeared power spectrum is simply the original multiplied by an  $l$ -dependent prefactor:

$$H_l = \bar{h}_l^2 \langle f | P_l(\Xi) | f \rangle . \tag{4.87}$$

The values of  $\bar{h}_l^2$  for several smearing angles  $\lambda$  are shown in Figure 4.23. Since the

smearing power spectrum is produced by multiplying the raw  $H_l$  by  $\bar{h}_l^2$ , the near-exponential fall-off of  $\bar{h}_l^2$  at large  $l$  will successfully kill the multiplicity plateau  $\langle f|f \rangle$ . Hence, giving particles physical extent clearly suppresses high- $l$  (small-angle) information, and a larger smearing angle  $\lambda$  suppresses more high- $l$  information — the shape function is a *low-pass filter*. This is exactly what we were looking for, and it has immediate applications.

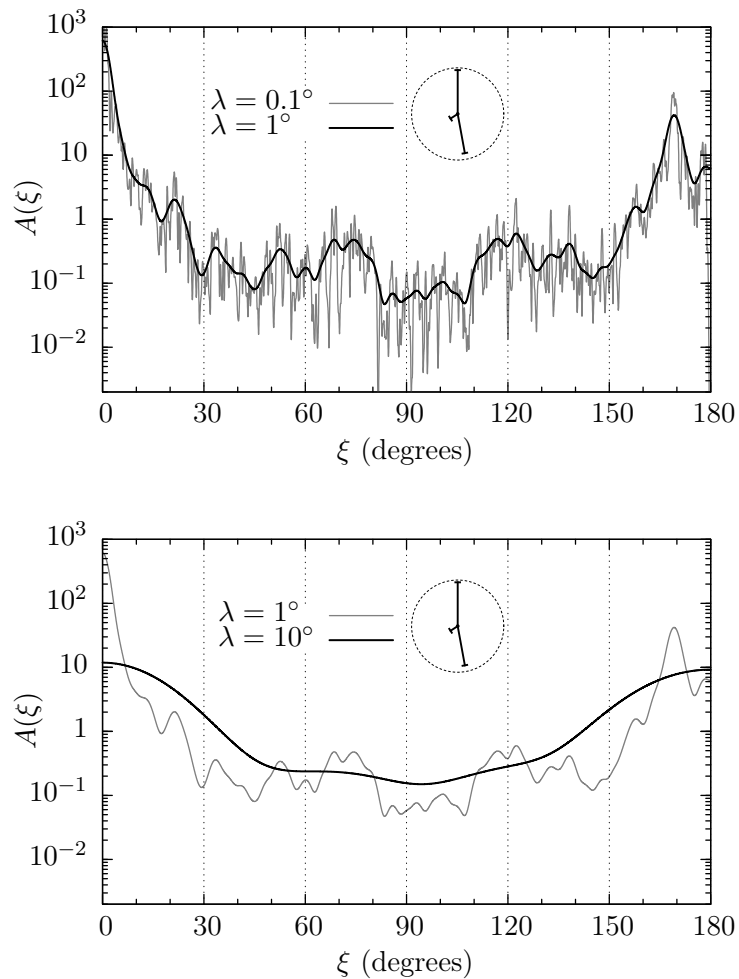


Figure 4.24. The angular correlation function for the measurable final-state particles in a  $q\bar{q}g$  final state with a soft gluon (2-jet-like), with (inset) a kinematic depiction of the original partons.  $H_l$  for this event was shown in Figure 4.8. (top) Small and medium smearing angle  $\lambda$ ; (bottom) medium and large smearing angle  $\lambda$ .

Recall the angular correlation function  $A(\xi)$  for the 2-jet-like event that was

shown at the beginning of this section (Fig. 4.20). To calculate a finite  $A(\xi)$ , we were forced to truncate its Legendre series at arbitrary  $l_{\max}$ , with rather messy results. Now that we have the ability to calculate a smeared power spectrum where  $H_l \rightarrow 0$  asymptotically, no arbitrary  $l_{\max}$  is needed. In Figure 4.24, we show the angular correlation function for the same event as before, for three choices of smearing angle  $\lambda$ . Since the values of  $A(\xi)$  span multiple scales, we use a log-scale.

In the top sub-figure, we show  $A(\xi)$  using two different smearing angles. With almost no smearing at all ( $\lambda = 0.1^\circ$ ; gray, thin), very little high- $l$  information is discarded, and there is a peak for nearly every inter-particle correlation. This is not much better than Fig. 4.20, but note two key differences:  $A(\xi) > 0$  everywhere and the high-frequency behavior looks a lot less like systematic “ringing”. The lack of ringing occurs because the smooth attenuation of the Gaussian smearing function is a better low-pass filter than some simple cut on  $l$  (i.e., fully trust  $H_l$  for  $l < l_{\max}$ , but reject any higher  $l$ ).

We can also see a medium smearing angle in the top sub-figure ( $\lambda = 1^\circ$ ; black, thick), which exemplifies the virtue of extensive particles. Throwing out the very-high- $l$  information gives a much better sense of the correlations among clusters of particles, without focusing too finely on individuals. Furthermore, the correlations are in the same overall location, just with coarser angular resolution. This makes it abundantly clear that this event has two large jets, about  $170^\circ$  apart, with a bunch of soft QCD in between. As expected, there is no unequivocally jet-like peak for the soft gluon. This becomes more obvious when we switch to the bottom sub-figure, where a *large* amount of smearing gives us the quintessential 2-jet-like event: major correlations at  $\xi = 0$  and  $\xi = \pi$ , with minimal correlations in between.

Examining the angular correlation of a 3-jet-like event in Figure 4.25 (this time without the log-scale), we notice similar behavior. Minimal smearing ( $\lambda = 0.1^\circ$ )

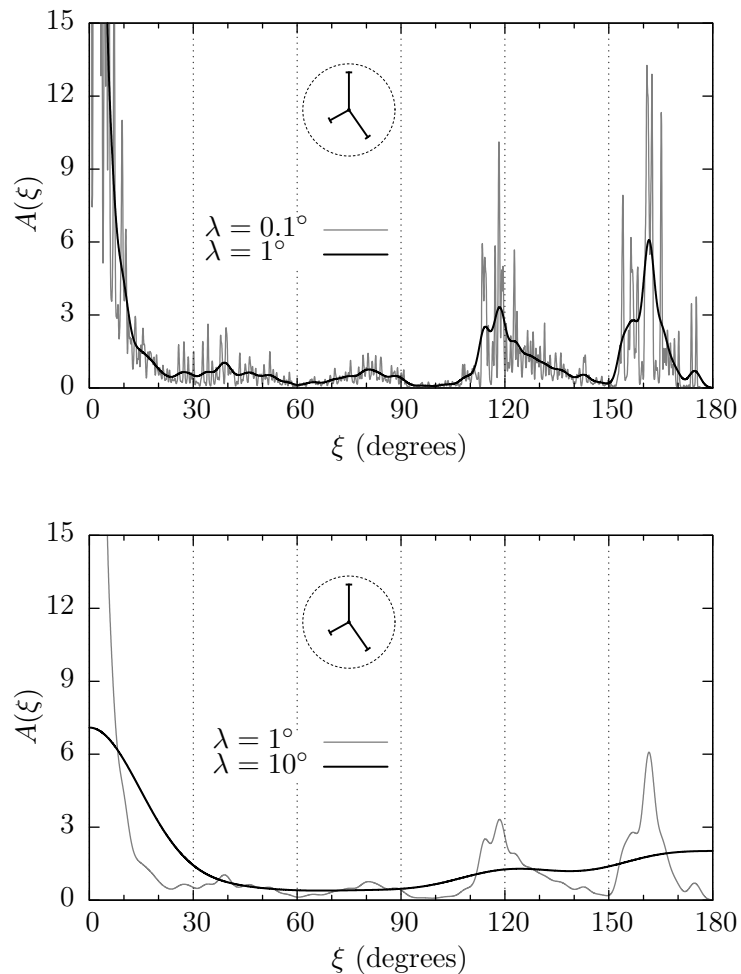


Figure 4.25. The angular correlation function for the measurable final-state particles in a  $q\bar{q}g$  final state with a hard gluon (3-jet-like), inset with a kinematic depiction of the original partons.  $H_l$  for this event was shown in Figure 4.9. (top) Small and medium smearing angle  $\lambda$ ; (bottom) medium and large smearing angle  $\lambda$ .

leaves too much fine detail, while medium smearing ( $\lambda = 1^\circ$ ) reveals the main picture. There is clearly a lot of energy correlated at  $\xi = 120^\circ$  (the quark-gluon correlation) and  $\xi = 160^\circ$  (the quark-antiquark correlation). This is strong evidence that there are three distinct jets. Smearing at the large angle in the bottom sub-figure smothers the event's fine features, but still reveals that this event *is not* two back-to-back jets. This is further evidence that, at low  $l$ , the power spectrum is dominated by the event's

coarse, jet-like structure.

**4.4.3.1 Infrared and collinear safety.** The angular correlation function  $A(\xi)$  shown in Figures 4.24 and 4.25 provided a good demonstration that giving particles spatial extent helps isolate the useful information in an event shape and suppress sampling noise. A shape function is indeed a low-pass filter. These figures also demonstrate a useful corollary:  $H_l$  is infrared and collinear (IRC) safe at low  $l$ .

The probability for QCD to create soft and collinear radiation is so large that perturbation theory breaks down, and the region of phase space describing such effects must be absorbed into a jet definition. A useful QCD observable must therefore be insensitive (safe) to infrared and collinear radiation. That  $H_l$  is IRC safe is demonstrated qualitatively in Figures 4.24 and 4.25, since the angular correlation function for  $H_l$  with  $\lambda = 1^\circ$  smearing looks *exactly* like a smoothed version of  $H_l$  with  $\lambda = 0.1^\circ$  smearing. This can be the case only if the low- $l$  power spectra are insensitive to the small-angle splittings and soft radiations that characterize the fine-grained event ( $\lambda = 0.1^\circ$ ). This follows naturally from the coarse angular resolution of  $Y_l^m$  at low  $l$ .

For a more quantitative guarantee [151], note that the weight  $w_{ij} = f_i f_j$  of each angular contribution in  $H_l$  makes it insensitive to the addition of a handful of soft particles (which have  $f \ll 1$  by definition). Similarly, a particle splitting to two nearly parallel particles ( $a \rightarrow bc$ ) minimally alters  $H_l$ . First we examine particle  $a$ 's total contribution to the power spectrum *before* the splitting (where  $z_{aj} \equiv \hat{p}_a \cdot \hat{p}_j$ )

$$H_{l,a} = f_a \sum_j f_j P_l(z_{aj}). \quad (4.88)$$

When particle  $a$  splits,  $\hat{p}_{b/c} = \hat{p}_a + \vec{\delta}_{b/c}$  (for tiny  $\vec{\delta}_{b/c}$ ), and this contribution becomes

$$H_{l,a} = f_b \sum_j f_j P_l(z_{aj} + \delta z_{bj}) + f_c \sum_j f_j P_l(z_{aj} + \delta z_{cj}).$$

Because  $f_a = f_b + f_c$ , the total contribution  $H_{l,a}$  is perturbed only in the  $P_l$  terms. Only when  $l$  becomes large — and  $P_l(z + \delta z)$  highly oscillatory — can a small  $\delta z$  give



rise to significant changes in  $H_l$ . Hence, for reasonably coarse event shapes (low to moderate  $l$ ), the power spectrum is infrared and collinear safe.

**4.4.4 Summary.** In this chapter we explored the power spectrum  $H_l$  in depth, and determined some very important properties (such as infrared and collinear safety). One of the major takeaways is that the power spectrum is dominated by the coarse, jet-like shape of the event. Furthermore, we now understand two of the three features of QCD events we identified in Section 4.2.3.1:

- **$H_l$  oscillates:** In Section 4.2.3.2, we showed that  $H_l$  is highly correlated between  $l$ , so that  $f(H_l)$  distributions at different  $l$  are not independent. This, combined with the complications of longitudinal boost and scale, makes Fox-Wolfram distributions  $f(H_l)$  relatively useless for characterizing QCD events at a hadron collider.
- **$H_l$  is unending:** We showed in Section 4.3 that the multiplicity plateau is a manifestation of discrete sampling noise, and dominates the power spectrum at high  $l$ , where  $H_l$  runs out of new information. In Section 4.4, we showed that particle shape functions create a low-pass filter that can be used to suppress the meaningless high- $l$  information.

Yet a final feature remains unexplained:

- **$N \neq n$ :** The peaks and valleys of the measurable power spectrum with  $N$  particles does not match  $H_l$  for their  $n$  originating partons. The shape of QCD radiation inside the jets matters, and seems to manifest as a slow attenuation of  $H_l$  to its asymptotic value.

It turns out that this feature is a manifestation of *jet shape*, which we will explore in the next chapter as we define the power jet model.

## CHAPTER 5

### POWER JETS

In Chapter 4 we saw that the QCD power spectrum  $H_l$  can characterize an event's shape.  $H_3$  is large for a 3-jet-like event (see Fig. 4.9) and small for a 2-jet-like event (see Fig. 4.8). Additionally, the angular correlation function (Figures 4.24 and 4.25) matches the kinematics of the 2-jet-like and 3-jet-like event, *if* a shape function is used to smear the small-angle information and minimize the effects of discrete sampling noise. This spatial smearing demonstrated that the power spectrum is IRC safe at low  $l$ , where  $H_l$  is dominated by the jet-like structure of the event — the hard QCD radiation.

Chapter 4 also revealed that the highly oscillatory nature of the power spectrum makes the Fox-Wolfram distributions  $f(H_l)$  rather unusable. In order to extract useful information from the fully correlated QCD power spectrum, we must be able to determine how jets produce these oscillations. This chapter proposes a scheme called power jets, whose basic concept is relatively straightforward: use the highly inter-correlated  $H_l$  oscillations to *fit* the phase space of a small number of jet-like objects, extracting the hard QCD signature from  $H_l$ . This combines a global event shape with the kinematic reconstruction of sequential jet clustering.

As a first implementation of power jets, suppose we detect a QCD event via  $N$  tracks and towers, whose power spectrum  $H_l^{\text{obs}}$  we calculate. We would then like to replicate this observed power spectrum with an  $n$ -jet model, where  $n \ll N$  (otherwise we will simply reproduce the tracks and towers). Our prototype jet model will start with  $n$  isolated partons:

$$\rho(\hat{r})_{\text{jet}} = \sum_{j=1}^n f_j \frac{\delta^2(\hat{r} - \hat{p}_j)}{\sin \theta_j}. \quad (5.1)$$

Calculating the  $n$ -jet power spectrum  $H_l^{\text{jet}}$ , we can find the square of the residuals between the two power spectra (where  $l_{\text{max}}$  is the cutoff of the fit)

$$\chi_l \equiv H_l^{\text{jet}} - H_l^{\text{obs}}, \quad (5.2)$$

$$\chi^2 = \sum_{l=0}^{l_{\text{max}}} \chi_l^2. \quad (5.3)$$

The best fit minimizes  $\chi^2$ , finding the  $n$ -jet event shape  $\rho(\hat{r})_{\text{jet}}$  which most closely matches the particles observed in the detector.

To perform a fit, we need a vector of parameters  $\vec{b}$  that define the event shape  $\rho(\hat{r})_{\text{jet}}$ ; the best-fit  $\vec{b}$  will minimize  $\chi^2$ ;

$$\frac{\partial \chi^2}{\partial b_k} = 2 \sum_l \chi_l \frac{\partial \chi_l}{\partial b_k} = 0. \quad (5.4)$$

Because  $H_l$  is nonlinear, the downward gradient  $J_{lk} \equiv -\frac{\partial \chi_l}{\partial b_k}$  is also nonlinear; this precludes a closed-form solution to the  $k$  equations embedded in Equation 5.4, *even if* the analytic calculation of  $J_{lk}$  were tractable.<sup>18</sup> Given this enormous restriction, a nonlinear least squared (NLLS) algorithm is used to find a local minimum via ever-better guesses for  $\vec{b}_{i+1}$ . This is accomplished by finding  $J_{lk}$  at the current  $\vec{b}_i$ , then stepping some  $\delta \vec{b}$  based upon this gradient. This scheme imposes a massive constraint on NLLS fits: they are only guaranteed to find a local minimum. If one wants a global minimum, the initial guess  $\vec{b}_0$  must reside within the  $\chi^2$  basin of the global minimum (i.e., starting at  $\vec{b}_0$  and following  $J_{lk}$  leads directly to the global minimum).

Skipping the implementation details for the time being, we can fit a  $q\bar{q}g$  event using the power jet procedure described thus far. In Figure 5.1, we fit the measurable final state using (a) a 3-jet model and (b) a 7-jet model. This highlights an immediate problem with the prototype event shape defined in Equation 5.1: to fit to

---

<sup>18</sup>For an  $n$ -jet model with arbitrary  $l_{\text{max}}$  and  $n$ ,  $J_{lk}$  is impractical to calculate analytically, and must instead be estimated numerically, via tiny shifts in  $b_k$ .

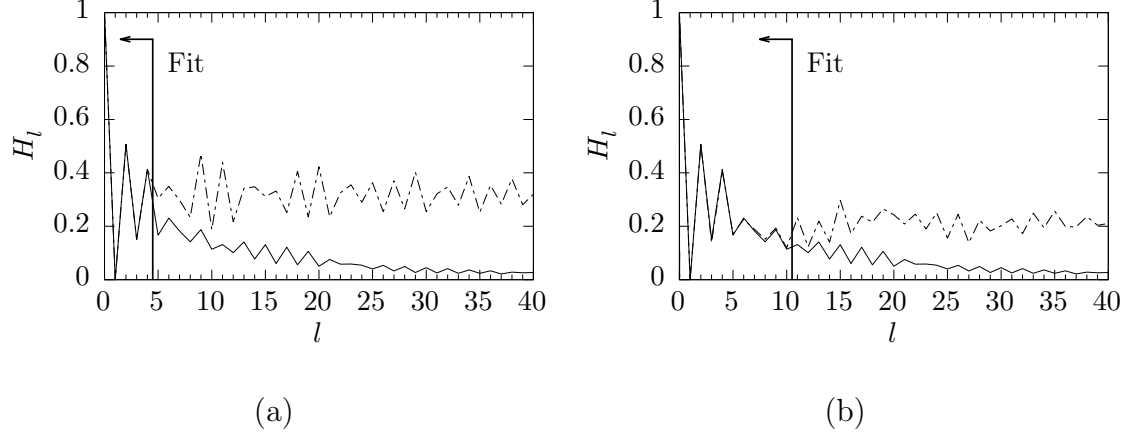


Figure 5.1. A 3-jet-like,  $q\bar{q}g$  event with (solid)  $H_l^{\text{obs}}$  and (dot-dashed)  $H_l^{\text{jet}}$ , fit using  $\delta$ -function jets in (a) a 3-jet model and (b) a 7-jet model. The “Fit” line depicts  $l_{\text{max}}$  [151].

even moderate  $l_{\text{max}}$ , a large number of  $\delta$ -function “jets” are needed (many more than one would expect for a  $q\bar{q}g$  event). Even worse, beyond  $l_{\text{max}}$ , the fit spectrum  $H_l^{\text{jet}}$  stabilizes to a much higher multiplicity plateau than the observed  $H_l^{\text{obs}}$ .

This is the final clue we need to clarify the feature of the power spectrum that remained unexplained at the end of Chapter 4. The *slow* attenuation of  $N$  measurable particles to a lower plateau than their  $n$  originating partons is due to jets’ physical extent — their shape from *soft* QCD radiation. As we saw in Figure 4.23, the power spectrum of extensive objects exhibits a slow attenuation in asymptotic power. The four extra  $\delta$ -function jets in Figure 5.1b were helping to fill out the shape of the three main jets. But this solution is not only costly (each massless jet requires three additional degrees of freedom, which runs the risk of overfitting), it only works at low  $l$ . As the angular resolution increases with  $l$ , the fit diverges from the observation because  $\delta$  functions are not a good approximation for soft QCD. Just as in Section 4.4, we need a continuous jet shape function  $h(\hat{r})$  in our  $\rho(\hat{r})_{\text{jet}}$  model.

In Section 5.1, we will use the tools we developed in Section 4.4 to derive a first approximation for this jet shape function. Fleshing out further details of the

fit in Section 5.2, we will show in Section 5.3 that the power jets fit can accurately reconstruct the event kinematics, even when there is a large amount of pileup. Finally, in Section 5.4 we will discuss the newest features of the power jets fit — which remain under active development (e.g., how will it work at the LHC) — and the promising future of harnessing the correlations in the QCD power spectrum.

### 5.1 Jet shape

To accommodate jet shape, we once again define the event shape as a set of  $n$  localized distributions, one for each jet

$$\begin{aligned}\rho(\hat{r})_{\text{jet}} &= \rho_{(1)}(\hat{r}) + \rho_{(2)}(\hat{r}) + \dots + \rho_{(N)}(\hat{r}), \\ &= f_1 h_{(1)}(\hat{r}) + f_2 h_{(2)}(\hat{r}) + \dots + f_N h_{(N)}(\hat{r}).\end{aligned}\tag{4.68}$$

Let us now propose that each individual shape function  $h_{(i)}(\hat{r})$  is jet-like; its energy is flowing outward from the origin in a definite direction, with its *centroid* (its central axis) defined via its center of momentum  $\vec{p}_i = f_i \int d\Omega \hat{r} h_{(i)}(\hat{r})$ . This shape also gives jets *mass*. It will be extremely important to relate this Lorentz invariant observable to jet shape, which is why we will not simply reuse the pseudo-Gaussian shape function (Eq. 4.63), but will instead derive a jet shape function from first principles.

To keep things simple for the initial deployment of power jets, we will stick to azimuthally symmetric jet shape functions. This is a good approximation of the underlying physics, and will allow us to compute  $H_l^{\text{jet}}$  using the tools already developed in Section 4.4.2. Of course, we do expect anisotropic jet shapes. A major source arises from hard QCD radiation within the jet (e.g.,  $g \rightarrow q\bar{q}$  at somewhat wide angles). Such behavior can be approximated in  $\rho(\hat{r})_{\text{jet}}$  via two sub-jets, with the plane formed by their two centroids defining the anisotropy axis. While this approximation is a bit crude, azimuthal symmetry for each object in  $\rho(\hat{r})_{\text{jet}}$  is extremely useful for

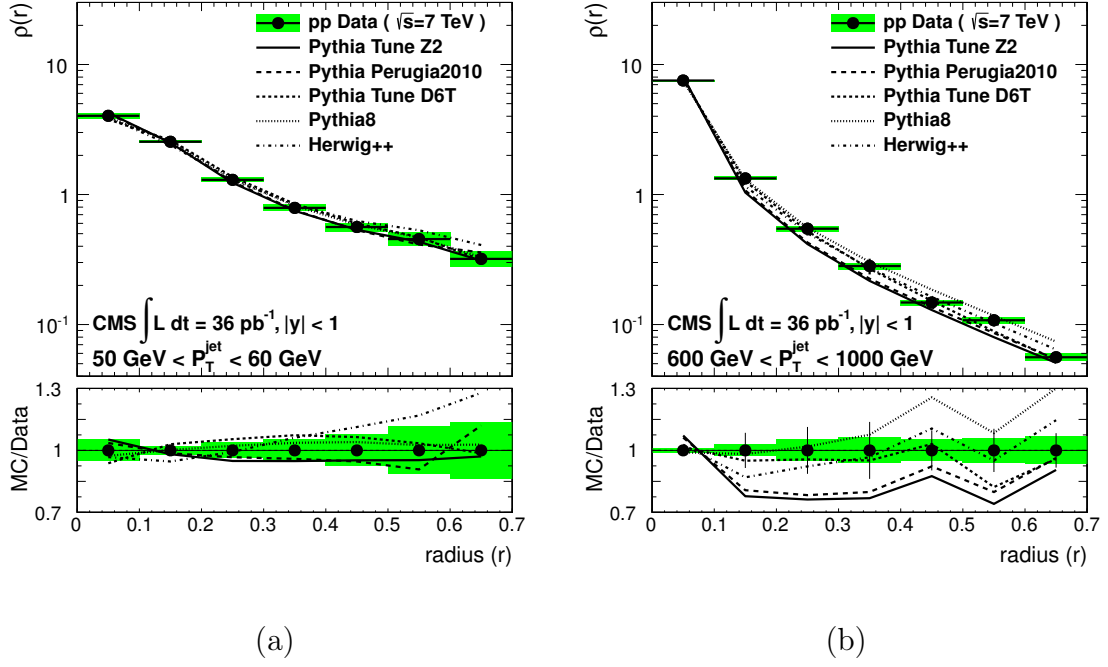


Figure 5.2. For anti- $k_T$  jets of radius  $R = 0.7$  at CMS, the fraction  $\rho$  of jet  $p_T$  residing in an annulus of radius  $r$  from the jet's centroid (with inner and outer radius denoted by the horizontal error bar) [152]. Results are shown for jets with (a) low energy and (b) high energy.

the prompt<sup>19</sup> calculation of  $H_i^{\text{jet}}$ .

This leaves two major effects which affect a jet's shape: (i) soft QCD radiation, which gives jets mass and (ii) jet boost. Since the jet is a massive, extensive object we can imagine boosting into its CM frame and examining its CM shape. This shape is not trivial to calculate from first principles, which is why a showering Monte Carlo is

<sup>19</sup>The nonlinear power jet fit needs to numerically estimate the Jacobian  $J_{lk}$  of the residuals. Since  $J_{lk}$  is a partial derivative with respect to parameter  $b_k$ , each parameter must be varied independently. For  $M$  fit parameters, this requires  $\mathcal{O}(M)$  recalculations of  $H_i^{\text{jet}}$ . Thus, a fast fit requires a rapid calculation of  $H_i^{\text{jet}}$ .

Some will argue that the raw speed of a useful method is not important, only the asymptotic complexity of the calculation, which in this case is  $\mathcal{O}(n^2)$ . This is true for a proven technique, but a novel technique — especially one which relies on inexact numerical tools (such as a non-linear fit) — needs testing and adjustment, and effective prototyping requires fast evaluation.

generally used in most phenomenological studies to produce approximate jet shapes. On the other hand, the effect of the boost is easy to predict — it will collimate the jet into a cone which tightens as the boost increases. This collimation is clearly evident in the data shown in Figure 5.2, where the more energetic jets are more centrally concentrated by their large boost, a property that roughly matches the predictions of showering Monte Carlo calculations (connecting lines).

We can study the boost effect by examining how  $\theta'$ , the polar angle relative to the jet's centroid (with the prime denoting CM variables) maps to  $\theta$  in the lab frame. Since we are restricting our attention to azimuthally symmetric shape functions  $h(\theta')$ , this mapping will allow us to calculate the lab frame jet shape  $h(\theta)$ . Arbitrarily selecting a massless jet constituent in the  $yz$  plane, we can parameterize its direction using only the  $z' = \cos \theta'$  component of its momentum:

$$\mathbf{p}' = E'[1, 0, \sqrt{1 - z'^2}, z'] . \quad (5.5)$$

We then boost the decay particle along the  $z$ -axis and into the lab frame, using the jet's Lorentz boost matrix of

$$\mathbf{\Lambda} = \begin{pmatrix} \gamma & 0 & 0 & \beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta\gamma & 0 & 0 & \gamma \end{pmatrix} , \quad (5.6)$$

$$\mathbf{p} = \mathbf{\Lambda} \cdot \mathbf{p}' = E' [\gamma(1 + \beta z'), 0, \sqrt{1 - z'^2}, \gamma(\beta + z')] . \quad (5.7)$$

Because the decay particles remains massless, we can parameterize the lab frame  $z = \cos \theta = p^3/p^0$  as

$$z = \frac{\beta + z'}{1 + \beta z'} . \quad (5.8)$$

To convert  $h(z')$  to the lab frame,

$$h(z) = \frac{d\Gamma}{\Gamma dz} = h(z'(z)) \frac{dz'}{dz} , \quad (5.9)$$

we will need the inverse change of variable

$$z' = \frac{\beta - z}{-1 + \beta z} \quad (5.10)$$

and the Jacobian

$$\frac{dz'}{dz} = \frac{1 - \beta^2}{(1 - \beta z)^2}. \quad (5.11)$$

For our initial foray into power jets, we should not over-complicate the CM jet shape. In the next section, we will present the simplest possible shape. Of course, we are free to choose more complicated (yet still azimuthally symmetric) jet shapes. In principle, any jet shape function  $h(z')$  will do, provided it satisfies the requirements of a PDF — it must be non-negative and normalized:

$$h(z') \geq 0, \quad \int_{-1}^1 h(z') dz' = 1. \quad (5.12)$$

In Section 5.1.2, I will show how more complicated models can easily be accommodated using a formalism consistent with the power spectrum framework.

**5.1.1 A scalar jet in the CM frame.** The simplest CM jet shape we can write down is a scalar. Lacking vector degrees of freedom to orient its decay in the CM frame, a scalar jet emits daughter particles isotropically;

$$\frac{d\Gamma}{\Gamma d\Omega'} = \frac{d\Gamma}{\Gamma \sin \theta' d\theta' d\phi'} = \frac{1}{4\pi}. \quad (5.13)$$

As a result, the jet's lab frame shape is completely determined by its boost, which provides a direct link between its mass and shape (larger mass, fatter jet). Hence, the shape of each jet in  $\rho(\hat{r})_{\text{jet}}$  will be fully parameterized by the jet's four-momentum  $\mathbf{p}$ .

Orienting the polar/ $z$ -axis along the jet's boost axis, and integrating over the symmetric azimuthal angle, we have

$$\frac{d\Gamma}{\Gamma \sin \theta' d\theta'} = \frac{1}{2}. \quad (5.14)$$



Since  $dz' = -\sin \theta' d\theta'$  (whose sign we flip because  $\theta' \in [0, \pi]$  maps to  $z' \in [1, -1]$ ), we obtain the jet shape in the CM frame

$$h(z') = \frac{d\Gamma}{\Gamma dz'} = \frac{1}{2}. \quad (5.15)$$

Using Equation 5.9, we can convert this into a lab frame shape function

$$h(z) = \frac{1 - \beta^2}{2(1 - \beta z)^2} = \frac{1}{2\gamma^2(1 - \beta z)^2}. \quad (5.16)$$

To use this  $h(z)$  to calculate  $H_l$ , we need its on-axis coefficient (Eq. 4.74)

$$\bar{h}_l = \frac{1}{2\gamma^2} \int_{-1}^1 \frac{P_l(z) dz}{(1 - \beta z)^2}. \quad (5.17)$$

We start by defining the crucial integral

$$A_l \equiv \int_{-1}^1 \frac{P_l(z) dz}{(1 - \beta z)^2}. \quad (5.18)$$

Using the recursive definition of the Legendre polynomial,

$$(n + 1)P_{n+1}(z) = (2n + 1)z P_n(z) - n P_{n-1}(z) \quad \text{or} \quad (5.19)$$

$$n P_n(z) = (2n - 1)z P_{n-1}(z) - (n - 1) P_{n-2}(z), \quad (5.20)$$

we can rewrite the integral as

$$\begin{aligned} A_l &= \frac{1}{l} \left[ (2l - 1) \int_{-1}^1 \frac{z P_{l-1}(z) dz}{(1 - \beta z)^2} - (l - 1) \int_{-1}^1 \frac{P_{l-2}(z) dz}{(1 - \beta z)^2} \right] \\ &= \frac{1}{l} \left[ \frac{2l - 1}{\beta} \int_{-1}^1 \frac{(1 - (1 - \beta z)) P_{l-1}(z) dz}{(1 - \beta z)^2} - (l - 1) A_{l-2} \right] \\ &= \frac{1}{l} \left[ \frac{2l - 1}{\beta} \left[ A_{l-1} - \int_{-1}^1 \frac{P_{l-1}(z) dz}{(1 - \beta z)} \right] - (l - 1) A_{l-2} \right]. \end{aligned} \quad (5.21)$$

To complete  $A_l$  we can now define a new integral

$$B_l \equiv \int_{-1}^1 \frac{P_l(z) dz}{(1 - \beta z)}. \quad (5.22)$$

We can compute  $B_l$  by combining Equation 4.82 with integration by parts (discarding the boundary term because  $P_{l+1}(z) = P_{l-1}(z)$  when  $z = \pm 1$ , as we did in Equation 4.83)

$$\begin{aligned} B_l &= \frac{1}{2l+1} \int_{-1}^1 \frac{dz}{(1-\beta z)} \frac{d}{dz} [P_{l+1}(z) - P_{l-1}(z)] \\ &= \frac{1}{2l+1} \left[ -\beta \int_{-1}^1 \frac{(P_{l+1}(z) - P_{l-1}(z)) dz}{(1-\beta z)^2} \right] \\ &= \frac{\beta}{2l+1} (A_{l-1} - A_{l+1}). \end{aligned} \quad (5.23)$$

Combining the definition of  $A_l$  and  $B_l$ , and after some trivial algebra, we have

$$A_l = \frac{1}{\beta(l-1)} ((2l-1)A_{l-1} - \beta l A_{l-2}). \quad (5.24)$$

We can then convert this into a recurrence relation for  $\bar{h}_l$  (taking  $l \rightarrow l+1$ )

$$\bar{h}_{l+1} = \frac{1}{\beta l} [(2l+1)\bar{h}_l - \beta(l+1)\bar{h}_{l-1}]. \quad (5.25)$$

This recurrence relation requires manually computing the initial conditions

$$\bar{h}_0 = 1, \quad (5.26)$$

$$\bar{h}_1 = \frac{1}{\beta} - \frac{1}{(\beta\gamma)^2} \operatorname{arctanh}(\beta). \quad (5.27)$$

Like the recurrence relation for the Gaussian shape function (Eq. 4.85), this recursion also becomes numerically unstable, but can be corrected using the exact same method discussed in Appendix A.4.

**5.1.2 More complicated jet shape functions.** One naïvely expects the shape of any particular jet to be quite complicated; it is one outcome from an infinite number of possible showers (one parton  $\rightarrow k$  partons), fragmentations ( $k$  partons  $\rightarrow m$  colorless particles) and decay chains ( $m$  particles  $\rightarrow N$  measurable particles). A Standard Model prediction will sum over all possible paths, and thus cannot be expected to accurately reflect any particular jet, but rather the average over all jets. Therefore,

if we wish to measure the shape of a *particular* jet, we should give  $h(z')$  the freedom to occupy one particular shape from the spectrum of possibilities (i.e., allow  $h(z')$  to fit the observed shape, instead of forcing it to use a particular model).

On the other hand, we must be careful; if we give  $h(z')$  too much freedom (e.g., the ability to have any arbitrary shape), then fitting  $h(z')$  to data will fit the sampling noise that arises from our discrete,  $k$ -particle sampling (i.e., spikes corresponding to individual particles). What we would like to fit is a “rough” shape: one which ignores the exact position of each individual particle, but can still resolve the general shape of the jet constituents (e.g., they seem to prefer forward,  $z' > 0$  positions in the CM frame).

The simplest way to do this is a Taylor series,

$$h(z') = a_0 + a_1 z' + a_2 z'^2 + \dots \quad (5.28)$$

However,  $h(z')$  must be non-negative *and* normalized, and constraining the shape coefficients soon becomes non-trivial. Given our domain of validity ( $z' \in [-1, 1]$ ), an obvious solution to the normalization problem is expressing  $h(z')$  as a *Legendre* series

$$h(z') = \sum_{l=0}^{l_{\max}} a_l P_l(z'), \quad (5.29)$$

since the orthogonality condition guarantees that  $P_l(z)$  have null integrals for  $l > 0$ :

$$\int_{-1}^1 P_l(z) dz = \int_{-1}^1 P_l(z) P_0(z) dz = \frac{2}{2l+1} \delta_{l0}. \quad (5.30)$$

Hence, normalization only constrains  $a_0 = 1/2$ , with all other  $a_l$  constrained by  $h(z') \geq 0$ . To keep  $h(z')$  rough,  $l_{\max}$  should be set to some reasonably small value. While we do not attempt to fit anything but a scalar jet shape in this thesis, we expect this Legendre series shape to be useful in future studies.

## 5.2 The missing pieces

In the introduction to this chapter we laid out the power jet concept, to fit jet-like objects to  $H_l$  observations. This required the development of a jet shape function, which we accomplished in the last section — to first approximation, a jet’s physical extent comes from its boost/mass. Before we can start fitting power jets in earnest, we need to tie up a few loose ends.

Given the lessons of Section 4.4, we need a sensible scheme to combine tracks and towers into  $H_l^{\text{obs}}$ ; this scheme will inform our calculation of  $H_l^{\text{jet}}$ . Then we need to choose the parameters  $\vec{b}$  that will control the four-momentum  $\mathbf{p}_i$  for the  $n$  jet-like objects in the model (with their shape defined via  $\gamma_i = E_i/m_i$ ). It is crucial that this choice of  $\vec{b}$  be easy to use autonomously by the non-linear least squares fitting algorithm (since I would prefer not to write a fitting algorithm from scratch).

**5.2.1 The detector filter.** In Section 4.4, we found the importance of smearing the angular position of particles to filter out sampling noise. Smearing at  $\lambda = 1^\circ$  did a reasonable job of filtering the high- $l$  noise from the measurable particles in Figure 4.25 and 4.24, and the same choice should work for tracks (which are well measured in angle). But what about calorimeter towers? Clearly, they capture energy over some patch of solid angle  $\Omega_{\text{twr}}$ , and it would be imprecise to simply treat them like a well-measured track at the tower’s geometric center. They should use a shape function.

But before we determine what a tower’s shape function should be, we can write down the event shape that it creates. In our pseudo-detector, each tower subtends approximately the same solid angle, so to good approximation all will have a common shape. Similarly, if a track’s smearing angle  $\lambda$  does not depend on its physical properties (e.g., its energy, or its reconstructed purity), then all tracks also have a

common shape. Factoring the on-axis coefficients as in Equation 4.87,

$$H_l^{\text{obs}} = (\bar{h}_l^{\text{trk}})^2 H_l^{\text{trk, trk}} + (\bar{h}_l^{\text{twr}})^2 H_l^{\text{twr, twr}} + \bar{h}_l^{\text{trk}} \bar{h}_l^{\text{twr}} H_l^{\text{trk, twr}}. \quad (5.31)$$

Here,  $H_l^{a,b}$  is the raw (unsmeared) power spectrum between separate lists of particles:

$$H_l^{a,b} = \langle f_a | P_l(|\hat{p}_a\rangle \cdot \langle \hat{p}_b|) | f_b \rangle. \quad (5.32)$$

In a more realistic detector, each unique track and tower might have slightly different shape functions, potentially requiring a full expansion of  $H_l^{\text{obs}}$  to  $N^2$  separate terms.

**5.2.1.1 The shape of calorimeter towers.** To use Equation 5.31, we need to define a shape function for calorimeter towers. To do so we must concede that we know how to efficiently calculate  $H_l$  only for azimuthally symmetric shape functions (e.g., a circular cap of solid angle). Unfortunately, the calorimeter towers in our pseudo-detector have a patch of solid angle  $\Omega_{\text{twr}}$  which is rather square. Nonetheless, approximating this square as a circle is not the gravest of sins; the approximation will have no effect on the inter-particle angles, and its entire purpose is to create a low-pass filter, which the edge geometry should affect minimally.

We therefore choose calorimeter towers to be a circular cap of solid angle  $\Omega_{\text{twr}}$  (with angular radius  $\theta_r$ ), uniformly filled with the energy collected by the tower. Aligning the polar/ $z$ -axis with the tower's centroid, the shape function is

$$h(\theta) = \frac{1}{\Omega_{\text{twr}}} \Theta(\theta_r - \theta). \quad (5.33)$$

Here we use the Heaviside  $\Theta$  step function to define the cap's circle, and normalize to the solid angle of the tower:

$$\Omega_{\text{twr}} = \int_0^{\theta_r} \sin \theta \, d\theta \, d\phi = 2\pi(1 - \cos \theta_r). \quad (5.34)$$

It will also become convenient to define the tower's dimensionless share of the total solid angle

$$A_{\text{twr}} \equiv \frac{\Omega_{\text{twr}}}{4\pi}. \quad (5.35)$$

The Heaviside  $\Theta$  function simplifies the calculation of the on-axis coefficient by cutting the integral off at  $\theta_r$  (instead of  $\pi$ ):

$$\bar{h}_l = \frac{1}{\Omega_{\text{tower}}} \int_0^{2\pi} d\phi \int_0^{\theta_r} P_l(z) \sin \theta d\theta = \frac{1}{2A_{\text{tower}}} \int_{\cos(\theta_r)}^1 P_l(z) dz. \quad (5.36)$$

Since Equation 5.34 tells us that  $\cos \theta_r = 1 - 2A_{\text{tower}}$ , it is trivial to calculate

$$\bar{h}_0 = 1. \quad (5.37)$$

For all other  $l$ , we can once again use Equation 4.82 to find

$$\begin{aligned} \bar{h}_l &= \frac{1}{2A_{\text{tower}}(2l+1)} [0 - (P_{l+1}(\cos \theta_r) - P_{l-1}(\cos \theta_r))] \\ &= \frac{1}{2A_{\text{tower}}(2l+1)} [P_{l-1}(1 - 2A_{\text{tower}}) - P_{l+1}(1 - 2A_{\text{tower}})]. \end{aligned} \quad (5.38)$$

Unlike the Gaussian smearing and boosted jet shape, these on-axis coefficients can be computed without recursion, and thus do not require special numerically stabilizing intervention.<sup>20</sup>

**5.2.1.2 Smearing the jet model.** We now know how to calculate the fully smeared power spectrum for a detected event, but do the tracks' smearing  $\lambda$  and calorimeter granularity  $A_{\text{tower}}$  affect  $H_l^{\text{jet}}$ ? Consider the following thought experiment. We have a detector with excellent energy resolution, but terrible angular resolution; it has no tracking, and its calorimeter has only one tower which subtends the entire  $4\pi$  of solid angle. Clearly, we would never expect to observe anything other than  $H_l^{\text{obs}} = 0$  for  $l > 0$  in such a ridiculous detector. This tells us that  $H_l^{\text{jet}}$  from the power jets model must account for the low-pass filter of the detector (the angular smearing of tracks and towers), otherwise it will predict a *much* higher asymptotic power than the detector will observe.

---

<sup>20</sup>The recursion is hidden, as the easiest way to calculate  $P_l$  is via the recurrence relation of Equation 5.19, which is also more numerically stable than a polynomial.

To simulate the angular smearing of tracks and towers in the power jets model, we make the large assumption that the event shape  $\rho(\hat{r})_{\text{jet}}$  splits into charged and neutral particles via some charge fraction  $\tau$ , and that these sub-shapes are identical to the original:

$$\begin{aligned}\rho(\hat{r})_{\text{jet}}^{\text{trk}} &= \tau \rho(\hat{r})_{\text{jet}}; \\ \rho(\hat{r})_{\text{jet}}^{\text{twr}} &= (1 - \tau) \rho(\hat{r})_{\text{jet}}.\end{aligned}\tag{5.39}$$

We can once again factor out the on-axis coefficients (as in Eq. 5.31), so that the power spectrum for the power jet model can be reduced to an  $l$ -dependent prefix times the power spectrum of the shaped jets

$$H_l^{\text{jet}} = (\tau \bar{h}_l^{\text{trk}} + (1 - \tau) \bar{h}_l^{\text{twr}})^2 \times \sum_{i,j} (\bar{h}_i \bar{h}_j) (f_i f_j) P_l(\xi_{ij}),\tag{5.40}$$

Since we are using a pseudo-detector with perfect energy resolution for our initial foray into power jets (as explained in Section 1.3.2) we can use  $\tau = 0.59$  (the predicted charge fraction of Fig. 4.15).

**5.2.2 The fit parameters.** We have defined a jet shape function based on a jet's four-momentum  $\mathbf{p}$  and a formula to calculate  $H_l^{\text{jet}}$ , but we haven't defined the parameters  $\vec{b}$  which the fit algorithm will use to define the  $\mathbf{p}$  of each jet in the model.

The simplest way to parameterize the  $n$  jets in the model is directly through the components  $p_i^\mu$  of their four-momentum. Recall that in Section 4.3.1, the spherical symmetry of  $H_l$  allowed us to arbitrarily choose  $\vec{p}_1 = \hat{z}$  and place  $\vec{p}_2$  in the  $yz$  plane, while  $\mathbf{p}$  conservation constrained  $\mathbf{p}_N$ . The remaining parameters are free;

$$\vec{b} = \{p_2^0, p_2^2, p_2^3, \dots, p_{N-1}^0, p_{N-1}^1, p_{N-1}^2, p_{N-1}^3\}.\tag{5.41}$$

This choice of  $\vec{b}$  works, but it is not a natural space for an autonomous fit. For example, to slightly alter the opening angle between two jets requires the adjustment

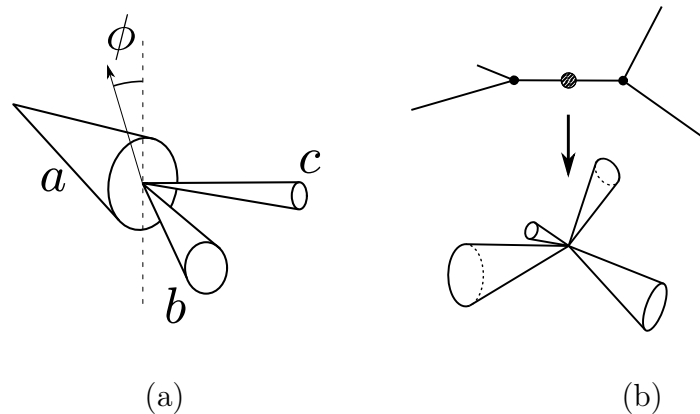


Figure 5.3. Particles splitting into jets. (a) A single  $a \rightarrow b + c$  splitting, depicting the rotation  $\phi$  of the splitting plane. (b) A binary splitting tree with three nodes and four “leaves”, one for each jet.

of *six* degrees of freedom. Furthermore, there is no real hierarchy among the parameters (e.g., we would not expect the second parameter to be more powerful than the seventh), and without a clear hierarchy it is easier for the fitting algorithm to stumble into one of the local minima, rather than the global minimum.

**5.2.2.1 Unconstrained splitting parameters.** We can also define jet momentum  $\mathbf{p}_i$  via a binary tree of particle splittings  $a \rightarrow b + c$ , as depicted in Figure 5.3a:

$$\mathbf{p}_a = \mathbf{p}_b + \mathbf{p}_c . \quad (5.42)$$

Each splitting creates a node in the binary tree, as depicted in Figure 5.3b, with the final “leaves” (un-split legs) defining the jets in the model. In this scheme, the parameters which control the *early* splittings have much more control over the total event shape, and thus more power in minimizing  $\chi^2$ . This creates a clear hierarchy in the fit parameters.

To parameterize this binary splitting tree, we can look to showering Monte Carlo [8], which use such trees to describe QCD radiation in jets. These QCD splitting kernels are commonly parameterized by  $z \equiv E_b/E_a$  [147] (with energy con-



ervation requiring  $E_c = (1 - z)E_a$ .<sup>21</sup> We must obviously restrict  $0 \leq z \leq 1$ , but  $\mathbf{p}$  conservation sets more stringent constraints on  $z$ , which we will now explore.

If we observe the splitting in  $a$ 's CM frame (which, to reduce clutter, we will not denote with a prime, since it is the only frame used by this paragraph), the three four-momenta are quite simple:

$$\mathbf{p}_a = [m_a, \vec{0}]; \quad (5.43)$$

$$\mathbf{p}_b = [E_b, \vec{p}_{bc}]; \quad (5.44)$$

$$\mathbf{p}_c = [E_c, -\vec{p}_{bc}]. \quad (5.45)$$

We know that  $\mathbf{p}_a - \mathbf{p}_b = \mathbf{p}_c$ , and squaring both sides gives

$$m_a^2 + m_b^2 - 2E_b m_a = m_c^2, \quad (5.46)$$

which becomes

$$E_b = \frac{m_a^2 + m_b^2 - m_c^2}{2m_a}. \quad (5.47)$$

Applying the same steps to  $\mathbf{p}_a - \mathbf{p}_c = \mathbf{p}_b$  merely exchanges  $b$  and  $c$ ;

$$E_c = \frac{m_a^2 + m_c^2 - m_b^2}{2m_a}. \quad (5.48)$$

Now squaring both sides of Equation 5.42 and rearranging we find

$$|\vec{p}_{bc}|^2 = \frac{1}{2}(m_a^2 - m_b^2 - m_c^2) - E_b E_c. \quad (5.49)$$

Plugging in Equation 5.47 and 5.48 gives (after some algebra) [8]

$$|\vec{p}_{bc}| = \frac{1}{2}m_a \Delta(m_a, m_b, m_c), \quad \text{where} \quad (5.50)$$

$$\Delta(a, b, c) \equiv \frac{1}{a^2} \sqrt{(a + b + c)(a + b - c)(a - b + c)(a - b - c)}. \quad (5.51)$$

---

<sup>21</sup>I apologize for any  $z$  confusion, given the extensive use of  $z \equiv \cos \theta$  in this thesis, but this splitting energy fraction convention is so common that I hesitate to use anything else.

This passes a simple sanity check;  $|\vec{p}_{bc}|$  becomes imaginary when  $(m_b + m_c) > m_a$ , a configuration which cannot possibly conserve energy.

Returning to the lab frame, particle  $a$  is moving in the  $+\hat{z}$  direction at speed  $\beta_a$ . Given the boost matrix which transforms the CM frame to the lab frame

$$\mathbf{\Lambda} = \begin{pmatrix} \gamma_a & 0 & 0 & \beta_a \gamma_a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta_a \gamma_a & 0 & 0 & \gamma_a \end{pmatrix}, \quad (5.52)$$

the configuration which maximizes(+)/minimizes(-)  $b$ 's lab frame energy  $E_b$  is one in which  $b$  is totally parallel or antiparallel to the boost ( $\vec{p}_{bc} \parallel \pm \hat{z}$ ). Calculating these max/min values of  $E_b$  tells us that  $z \equiv E_b/E_a$  must be constrained to the domain  $z \in [z_-, z_+]$ ;

$$z_{\pm} = \frac{E_{b,\pm}}{\gamma_a m_a} = \frac{1}{2} \left( 1 + \frac{m_b^2 - m_c^2}{m_a^2} \pm \beta_a \Delta(m_a, m_b, m_c) \right). \quad (5.53)$$

Splitting particle  $\mathbf{p}_a$  to two daughters  $a \rightarrow b+c$ , we get  $\mathbf{p}_c$  from  $\mathbf{p}$  conservation. The remaining 4 d.o.f. (i.e.,  $\mathbf{p}_b$ ) can be variously defined, but a natural choice is:

- ( $z$ ): the energy fraction of daughter  $b$ .
- ( $u_{b/c} \equiv m_{b/c}/m_a$ ): the mass fractions of the two daughters ( $u_c \neq 1 - u_b$ , because not all of  $a$ 's energy need manifest as mass).
- ( $\phi$ ): the angle between  $a$ 's splitting plane and the plane of the previous splitting (the splitting which spawned  $a$ ).

Obviously,  $\phi$  has total freedom, but both  $u$  and  $z$  need to be constrained. Since these splitting parameters will be fit with an autonomous algorithm, the algorithm must be informed of these constraints (via inequalities which must be satisfied). This turns out to be rather complicated, especially since the  $z$  constraint depends on the mass fractions, and defining that interplay is non-trivial.

Instead, we can redefine the degrees of freedom such that all constraints are built in to their definition. We denote redefined d.o.f. with a star superscript:

- $z^*$ : the *free* energy fraction of daughter  $b$ . The easiest way to define  $z^*$  is to define how it maps to  $z$ :

$$z = z_- + z^*(z_+ - z_-). \quad (5.54)$$

This mapping always gives a valid  $z$ , provided that  $0 \leq z^* \leq 1$ .

- $(u_{\text{sum}})$ : the fraction of  $m_a$  which continues to manifests as mass. Defining

$$u_{\text{sum}} \equiv \frac{m_b + m_c}{m_a}, \quad (5.55)$$

one must restrict  $0 \leq u_{\text{sum}} \leq 1$ .

- $(u_b^*)$ : the fraction of  $u_{\text{sum}}$  which  $b$  obtains, so that

$$u_b = u_b^* u_{\text{sum}}, \quad (5.56)$$

$$u_c = (1 - u_b^*) u_{\text{sum}}. \quad (5.57)$$

This places  $u_b^*$  in the domain  $0 \leq u_b^* \leq 1$ .

- $(\phi)$ : the angle between  $a$ 's splitting plane and the plane of the previous splitting (the splitting which spawned  $a$ ). Because  $z$  can be both large and small (so that  $b$  and  $c$  can reverse identity), one merely need constrain  $0 \leq \phi \leq \pi$ .

**5.2.2.2 Converting splitting parameters into jets.** We have defined a set of fit parameters  $\vec{b}$  which parameterize a binary splitting tree and are simple to bound. To calculate  $H_i^{\text{jet}}$ , we must convert  $\vec{b}$  into the set of jet four-momenta  $|\mathbf{p}\rangle$  corresponding to the leaves of the tree. Starting at the root node and working outward, we take  $\mathbf{p}_a$  for each node and use its splitting parameters to calculate  $\mathbf{p}_b$  and  $\mathbf{p}_c$ . Solving this conversion for a generic splitting, we can use it at every node in the splitting tree.

As we work out this conversion in the next few paragraphs, I will take great pains to use the most numerically stable expressions (see Appendix A, which explains concepts like floating point “cancellation”). This is because the automated fit algorithm will constantly convert the fit parameters of  $\vec{b}$  into the set of four-momenta  $|\mathbf{p}\rangle$ , and from this set calculate  $H_i^{\text{jet}}$ . Generating this  $|\mathbf{p}\rangle$  via a binary splitting tree will ensure that  $\mathbf{p}$  errors will propagate outward from the root node, so we take extra care to minimize numerical errors; this will ensure that  $H_i^{\text{jet}}$  is as accurate as possible.

Given some  $\mathbf{p}_a$  in the lab frame, it is rather simple to define the two daughter particles in a way that guarantees  $\mathbf{p}$  conservation:

$$\mathbf{p}_a = [E_a, \vec{p}_a], \quad (5.58)$$

$$\mathbf{p}_b = [z E_a, r \vec{p}_a + \vec{k}_T], \quad (5.59)$$

$$\mathbf{p}_c = [(1 - z) E_a, (1 - r) \vec{p}_a - \vec{k}_T]. \quad (5.60)$$

Here we split the mother’s energy via  $z$  and its momentum through  $r$ . These are separate because we have no reason to believe that  $r$  equals  $z$  (although perhaps that’s what we’ll find). We also add some unknown momentum  $\vec{k}_T$ , defined to be transverse to  $\vec{p}_a$ , which will be the primary manifestation of the mother’s mass  $m_a$ .

First we can explicitly convert  $z^*$  to  $z$  (using Eq. 5.53);

$$z = \frac{1}{2} (1 + u_b^2 - u_c^2 + \beta_a (2z^* - 1) \Delta(1, u_b, u_c)) . \quad (5.61)$$

We can now define  $k_T = |\vec{k}_T|$  by squaring Equation 5.59:

$$k_T^2 = (z^2 - r^2 \beta_a^2) E_a^2 - m_b^2 = m_a^2 (\gamma_a^2 (z^2 - r^2 \beta_a^2) - u_b^2) . \quad (5.62)$$

The only undefined d.o.f. is now  $r$ , which we can find by squaring  $\mathbf{p}_a - \mathbf{p}_b = \mathbf{p}_c$ :

$$r = \frac{1}{\beta_a^2} \left( z - \frac{m_a^2 + m_b^2 - m_c^2}{2 E_a^2} \right) = \frac{1}{\beta_a^2} \left( z - \frac{1 + u_b^2 - u_c^2}{2 \gamma_a^2} \right) . \quad (5.63)$$

We can further simplify this expression by plugging in our solution for  $z$  (and using  $\beta^2 = 1 - \gamma^{-2}$ ). We find that  $r$  is closely related to  $z$ , but slightly different;

$$\begin{aligned} r &= \frac{1}{2} \frac{1}{\beta_a^2} \left( \frac{(\gamma_a^2 - 1)}{\gamma_a^2} (1 + u_b^2 - u_c^2) + \beta_a (2z^* - 1) \Delta(1, u_b, u_c) \right) \\ &= \frac{1}{2} \left( 1 + u_b^2 - u_c^2 + \frac{1}{\beta_a} (2z^* - 1) \Delta(1, u_b, u_c) \right). \end{aligned} \quad (5.64)$$

That  $z$  and  $r$  are quite similar when  $\beta_a \approx 1$  creates a problem for  $k_T$  (Eq. 5.62): the term  $z^2 - r^2 \beta_a^2$  will have a floating point cancellation in highly boosted systems (driving  $k_T \rightarrow 0$  in an unstable way). Rewriting the term as  $(z - r\beta_a)(z + r\beta_a)$ , then defining  $u_{bc} \equiv u_b^2 - u_c^2$  and  $z_\Delta \equiv (2z^* - 1)\Delta(1, u_b, u_c)$  for convenience, we find

$$(z \pm r\beta_a) = \frac{1}{2} (1 \pm \beta_a) (1 + u_{bc} \pm z_\Delta). \quad (5.65)$$

Using  $(1 + x + y)(1 + x - y) = 1 + 2x + x^2 - y^2$ , we can now write down the most numerical stable version of  $k_T$  (since it cannot be simplified further, and there are no longer hidden cancellations)

$$\begin{aligned} k_T^2 &= m_a^2 \left( \frac{1}{4} \gamma_a^2 (1 - \beta_a^2) (1 + 2u_{bc} + u_{bc}^2 - z_\Delta^2) - u_b^2 \right) \\ &= \frac{m_a^2}{4} (1 - 2u_b^2 - 2u_c^2 + u_{bc}^2 - z_\Delta^2) \\ &= z^* (1 - z^*) (1 - u_{\text{sum}}^2) (1 - u_{\text{sum}}^2 (2u_b^* - 1)^2) m_a^2. \end{aligned} \quad (5.66)$$

This final expression for  $k_T^2$  passes some necessary sanity checks that were less obvious in Equation 5.62. By construction,  $\vec{k}_T$  is invariant to the boost of  $a$ ; so unlike  $z$  or  $r$ , its magnitude does not depend on  $\gamma_a$  or  $\beta_a$ . Also, none of the multiplicative factors can become negative, so that we always find  $k_T \in \mathbb{R}^3$ . Finally, the conditions which send  $k_T \rightarrow 0$  are consistent with the only kinematic configurations where there is no transverse motion: (i)  $u_{\text{sum}} = 1$ , so that  $a$  fractures into two pieces which preserve its total mass, leaving no leftover energy for  $b$  and  $c$  to move apart (they stay fused

as a composite  $a$ ). (ii)  $z^* = 0$  or  $z^* = 1$ , the configurations which maximize/minimize  $E_b$  because the  $bc$  system is totally parallel/antiparallel to  $a$ 's motion.

### 5.3 Fitting power jets and pileup

We are finally ready to begin fitting power jets to QCD event shapes. The events in this section are simulated at an  $e^+e^-$  collider with  $\sqrt{s} = 400$  GeV. We will see that a robust, repeatable fit is possible, and is also quite resilient to a tremendous amount of pileup. This provides dramatic evidence that power jets can extract useful QCD observables from the vast amount of correlated information embedded in the power spectrum. Continued refinement will permit the extraction of even more information, and the practical deployment of power jets at the LHC.

There are two components to the power jets model: (i) hard QCD radiation is modeled by the binary splitting tree and (ii) soft QCD radiation is modeled by the azimuthally symmetric jet shape (with a jet's "radius" set by its mass). Given this picture, we should concede that the  $n$ -"jet" fit we have been discussing thus far is more accurately an  $n$ -"prong" fit, with each jet composed of one or more extensive prongs. This is because the basic statement of jet-parton duality — a parton with four-momentum  $\mathbf{p}$  creates a cluster of measurable particles which sum to  $\mathbf{p}$  — is only the leading order picture. At next-to-leading order (NLO), there is QCD radiation within the jet, and the hardest radiation creates anisotropic jet substructure.

We roughly depict this NLO picture in Figure 5.4. Modeling this jet substructure is necessary for power jets, because QCD radiation within the jet does not have to be terribly hard to create jet shapes which are poorly fit by azimuthally symmetry. Yet jet substructure is also an increasingly important technique, and being able to handle it naturally (i.e., without reclustering with a smaller radius parameter  $R$ ) is a huge advantage over sequential combination algorithms like anti- $k_T$ .

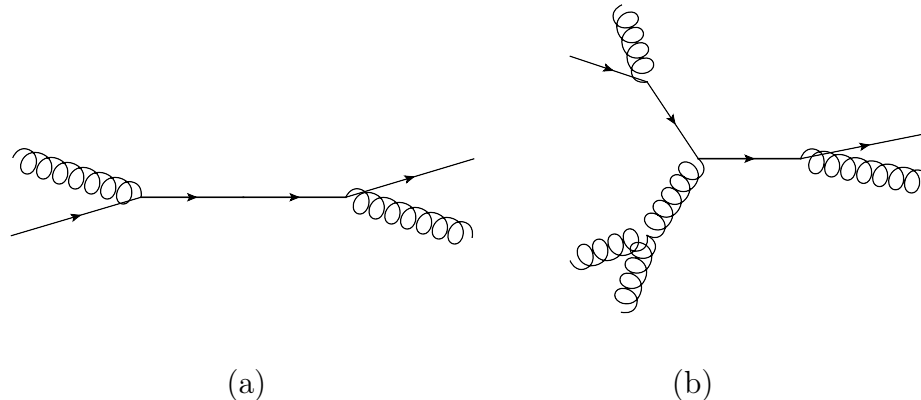


Figure 5.4. A next-to-leading order depiction of (a) 2-jet and (b) 3-jet events. In power jets, hard QCD radiation (partons splitting) is modeled by distinct prongs, which are given extensive shape by soft QCD radiation.

For example, a very boosted Higgs  $h \rightarrow b\bar{b}$  may be the only way to directly measure the Higgs coupling to bottom quarks (due to the huge SM background) [153], but this measurement requires rejecting the still significant  $g \rightarrow b\bar{b}$  background. These signals have very different color structures (singlet versus octet, respectively), which will manifest in their substructure [154] (which is also one of the handles used by boosted top tags [62]). Of course, this  $n$ -prong substructure fit must use  $n \ll N$  to be meaningful, so for the  $q\bar{q}g$  events studied in this section (which are either 2-jet or 3-jet like), we will use a 4-prong or 6-prong final state.

**5.3.1 The fitting procedure.** A good power jets fit needs the NLLS fitting algorithm to repeatedly find the correct, global minimum. This is a bit of an art, and determining the best way to coax the algorithm towards this minimum required a great deal of trial and error. While I eventually settled upon the following procedure for the results presented in this thesis, it is not intended to be the final word.

We start with a 2-prong fit to  $l < 4$ . This provides a crude fit to the event's most general shapes, and establishes which of the two prongs is the more massive. To add the third prong, we split one of the two existing prongs; but which one? Here we

develop a general rule for adding one more splitting and refitting the event:

1. **Split** the most massive prong, since it is most likely to conceal substructure.

This adds up to four new splitting parameters to  $\vec{b}$  ( $u_{\text{sum}}$ ,  $u_b^*$ ,  $z^*$  and  $\phi$ ).

- (a) Initialize the new node's parameters to:

$$u_{\text{sum}} = 1/10, u_b^* = z^* = 1/2, \text{ and } \phi = 0.$$

2. **Constrain** the new parameters using lots of small-angle information:

- (a) Fit out to some “large”  $l_{\text{max}} > 10$ . For this to work,  $l_{\text{max}}$  was chosen as the place where a running average of  $H_l$  drops below 0.05, since beyond this point there is little power left to fit. To avoid fitting too much or too little substructure,  $l_{\text{max}}$  was restricted to  $15 \leq l_{\text{max}} \leq 36$ ; this fits features at least  $24^\circ$  in angular scale, but no smaller than  $10^\circ$  in angular scale.
- (b) Immediately after adding a new node, fix the parameters of the preexisting nodes to the best fit of the previous iteration  $\vec{b}_{k-1}$ , allowing only the *new* node's parameters to vary.
- (c) After fitting the new node, restore full freedom to all parameters and refit the event to the same  $l_{\text{max}}$ .

3. **Refine** the fit using only the most robust information in the event:

- (a) Refit the event using  $l_{\text{max}} = 10$  ( $36^\circ$ ).

Of course, this splitting procedure cannot continue indefinitely; there will be some final refinement, and no further splitting. The final fit must then use the large  $l_{\text{max}}$ , to use as much small-angle information as possible. This requires defining the criteria to terminate the fit.



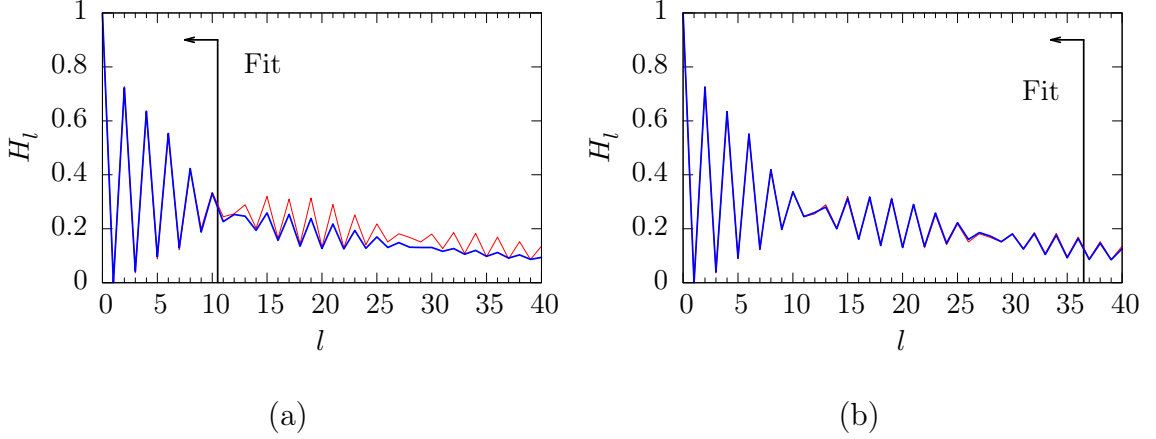


Figure 5.5. The power jet fit for the 2-jet-like  $q\bar{q}g$  event previously seen in Figure 4.8, with (red, thin)  $H_l^{\text{obs}}$  and (blue, thick)  $H_l^{\text{jet}}$ ; (a) the “refined” 3-prong fit ( $l_{\text{max}} = 10$ ) and (b) the final 4-prong fit ( $l_{\text{max}} = 36$ ).

At the LHC, both ATLAS and CMS reconstruct QCD jets with angular radius  $R = \mathcal{O}(20^\circ)$ ; this information is embedded in the power spectrum at  $l \leq 10$  (since the minimum angular scale between jets is  $\sim 2R = 45^\circ$ ). Hence, a useful goodness of fit variable scales the average deviation in the first ten  $H_l$  by the largest  $H_l$  in the event:

$$\zeta = \frac{\sqrt{\chi_{l \leq 10}^2}}{10 \max(H_l)}. \quad (5.67)$$

If the fit has converged to the correct event shape, then  $\zeta$  should remain small (empirically  $\zeta < \mathcal{O}(10^{-4})$ ), even when a large  $l_{\text{max}}$  is used, because the first ten  $H_l$  *largely determine* the subsequent power spectra (since  $H_l$  oscillations are correlated, as we saw in Section 4.2.3.2). In Figure 5.5a, a 3-prong model only supplies a good  $\zeta$  when  $l_{\text{max}} = 10$ . By adding one more splitting in Figure 5.5b, a 4-prong model gives an excellent  $\zeta$ , even when  $l_{\text{max}}$  is large.

If we now turn to the 3-jet-like event of Figure 5.6, we can see that a 3-prong model of Figure 5.6a is totally insufficient for reproducing the event’s shape, because  $\zeta$  is huge. From this, and the large value of  $H_3/H_2 > 1/3$ , we determine that this event is very 3-jet-like, and therefore choose to use a 6-prong model for the final fit (Fig 5.6b). Of course, we’re now faced with an important question; how do we extract

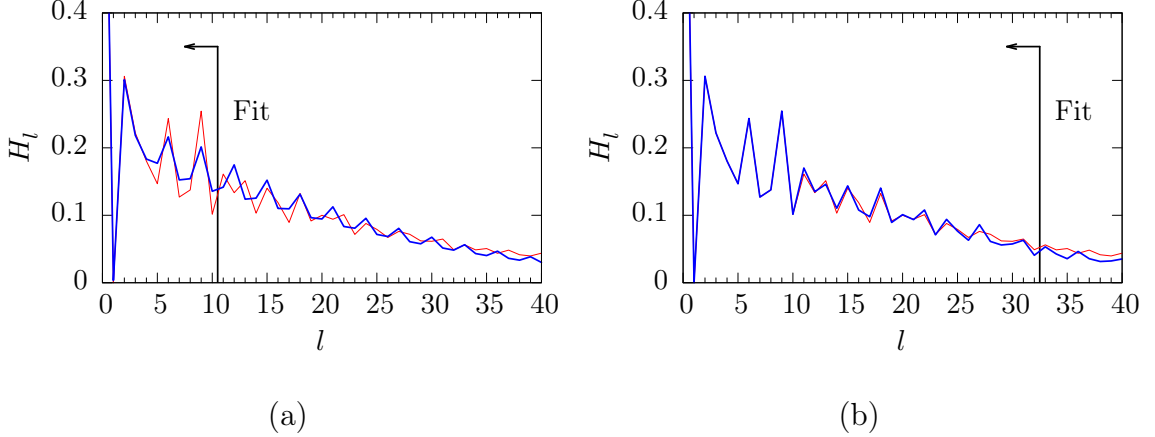


Figure 5.6. The power jet fit for the 3-jet-like  $q\bar{q}g$  event previously seen in Figure 4.9, with (red, thin)  $H_l^{\text{obs}}$  and (blue, thick)  $H_l^{\text{jet}}$ ; (a) the “refined” 3-prong fit ( $l_{\text{max}} = 10$ ) and (b) the final 6-prong fit ( $l_{\text{max}} = 36$ ).

$m$ -jet kinematics from an  $n$ -prong event?

**5.3.2 Power jets have no boundaries.** In the previous section, we showed that  $n$ -prong kinematics are necessary to accurately reproduce the event shape, a natural result of the QCD splitting that gives jets their shape. However, to extract  $m$ -jet kinematics, we will clearly need to rewind this splitting to obtain the original jets. Luckily, we have already encountered a tool designed to do just that.

Recall the  $k_T$ -family of jet definitions from Section 1.4.2. Currently, the most popular, general-purpose jet definition at the LHC is  $R = 0.4$  anti- $k_T$  jets ( $p = -1$ ), since anti- $k_T$  is empirically insensitive to soft QCD radiation. However, anti- $k_T$  is so named because it behaves opposite to  $k_T$  ( $p = 1$ ), which was the original variant of this clustering algorithm.  $k_T$  was designed to rewind QCD showers; it starts by collecting the soft radiation at the edges of jets and sweeping it toward the jet’s center. This definition is exactly how a theorist would like to reconstruct jets, but is problematic in practice due to the irregular boundaries of the jets it creates (as seen in Fig. 5.7). In fact, the exact boundaries of  $k_T$  jets are highly sensitive to the exact details of soft radiation and pileup in the event (the jet boundaries are infrared

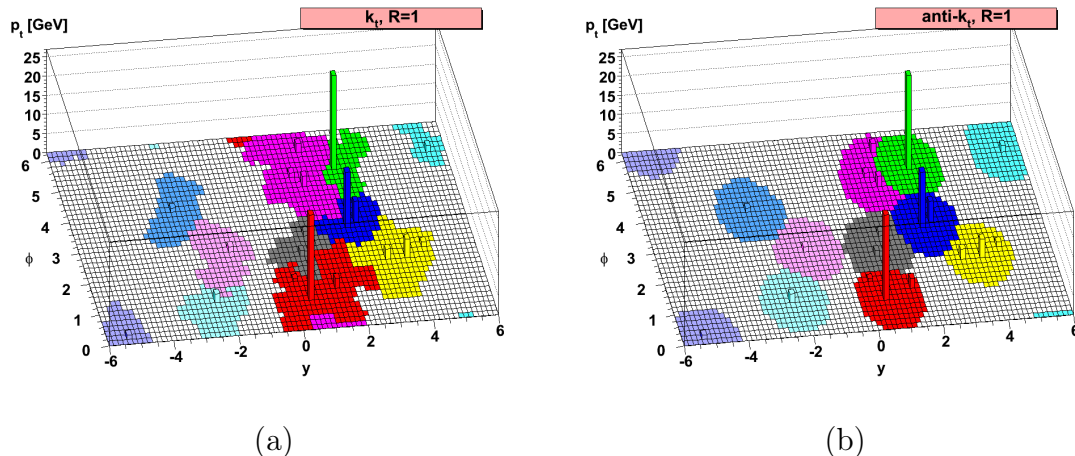


Figure 5.7. The boundaries of (a)  $k_T$  and (b) anti- $k_T$  jets for the same original particles [33]. The irregular boundaries of  $k_T$  jets depend on the fine details of soft QCD radiation, while the “conical” boundaries of anti- $k_T$  jets do not.

unsafe), which is why the nice, conical boundaries of anti- $k_T$  jets makes them the popular choice.

However, this problem of raggedness arises primarily when clustering from raw physics objects (which are individually soft). The prongs of power jets have already absorbed all the soft radiation into their extensive shape. This highlights another extremely important property of power jets — they have no boundaries! They are defined by using *all the correlations* across the entire event. And while these correlations are used to find specific shapes, power jets do not assign specific particles to prongs. Formally, every prong contributes a small amount to every patch of  $d\Omega$ . Of course, this also means that power jets have no arbitrary radius  $R$ ; the physical extent of each prong is driven entirely by the detected event, not some predetermined constant scale. These features address important limitations of existing jet definitions.

Because power jets have no boundaries and no radius, with prongs representing the hard QCD radiation in the event, it is perfectly safe to use the  $k_T$  algorithm to rewind the QCD shower that produces the  $n$ -prong shape. For our initial investigation

of power jets, we need to compare the power jets results to the “truth-level” Monte Carlo information about the original partons. We know that we are reconstructing a 3-jet QCD event, so we keep merging the two prongs which possess the smallest  $k_T$  distance  $d_{ij}$  (see Eq. 1.19) until there are only three prongs left — our three jets. In future studies, merging will terminate based upon some scale choice (e.g., stop merging when the two prongs with the smallest  $k_T$  distance  $d_{ij}$  have a dimensionless invariant mass  $m_{ij}/\sqrt{s}$  exceeding some jet scale  $q_{\text{jet}}$ ).<sup>22</sup>

Given three reconstructed jets, we can then compare their energies and the three pairs of dijet invariant masses to those of the three original partons. Invariant mass is the more important observable, since it is a Lorentz scalar and is sensitive to both energy *and* angles. To make a direct comparison between massive jets and massless partons, we define a corrected invariant mass

$$\tilde{m}_{ij}^2 = \frac{1}{\sqrt{s}} \left( (\mathbf{p}_i + \mathbf{p}_j)^2 - \frac{1}{2}(\mathbf{p}_i^2 + \mathbf{p}_j^2) \right). \quad (5.68)$$

This definition is useful because it preserves the property that  $(\tilde{m}_{12}^2 + \tilde{m}_{13}^2 + \tilde{m}_{23}^2) = 1$ , regardless of  $\mathbf{p}$  being massive or massless.

Of course, we have only *asserted* that power jets are insensitive to soft radiation, since describing soft radiation is the intended purpose of prong shape. To prove this, we should re-analyze the same event dozens of times, but with a different sprinkling of soft radiation each time (we do 20 trials throughout this section). In

---

<sup>22</sup>How does this jet scale  $q_{\text{jet}}$  differ from a jet radius parameter  $R$ ? It turns out that *some* choice of scale is unavoidable in jet reconstruction, because determining whether two objects are isolated jets or subjet siblings requires clearly separating the domain of the hard scatter from the domain of QCD showering. Defining this “jet scale” is a prerequisite for an inclusive measurement (where the number of jets is not fixed, but is determined dynamically from the data). Basing this scale on invariant mass, rather than angular radius  $R$ , is a much more physical choice. Furthermore, this scale choice does not affect soft QCD, which power jets absorbs into prong shape — a power jets reconstruction can contain both fat and thin jets in the same reconstruction, whereas anti- $k_T$  or Cambridge-Aachen cannot.

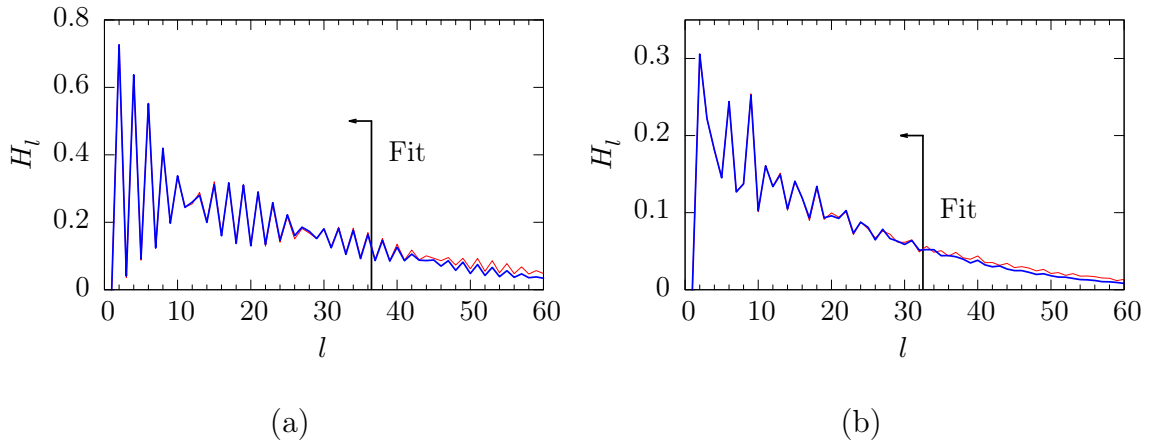


Figure 5.8. The best power jets fit of a  $q\bar{q}g$  jet from twenty trials, each with a different, random isotropic pileup ( $f_{\text{PU}} = 1\%$ ), with (red, thin)  $H_l^{\text{obs}}$  and (blue, thick)  $H_l^{\text{jet}}$ ; (a) 2-jet-like and (b) 3-jet-like event.

in addition to testing that an  $n$ -prong fit is a sensible approach, multiple trials will also test that the autonomous NLLS minimization algorithm can robustly find the same (or kinematically similar) minimum. To create this soft radiation, we will use the pileup model introduced in the next section, contributing  $f_{\text{PU}} = 1\%$  to the event's energy fraction.

Table 5.1. Reconstructed 3-jet kinematics for the 2-jet-like event (Fig. 5.8a).

	$f_1$	$f_2$	$f_3$	$\tilde{m}_{12}^2$	$\tilde{m}_{13}^2$	$\tilde{m}_{23}^2$
parton	0.4753	0.4321	0.0925	0.8150	0.1357	0.0493
power jets	0.4759(0)	0.4354(2)	0.0888(2)	0.8203(3)	0.1267(3)	0.0527(0)
anti- $k_T$	0.4793	0.4269	0.0937	0.8067	0.1419	0.0514

In Table 5.1, we show the result for the 2-jet-like event, whose fit is shown in Figure 5.8a. The table presents power jets' median value, whose error is estimated from the unbiased standard deviation. For comparison, we cluster  $R = 0.4$  anti- $k_T$  jets, then merge them into three jets using the same  $k_T$  procedure used to merge power jets prongs. This is an important sanity test, because anti- $k_T$  has been extensively tested and is known to be reliable in the absence of strong pileup. Both power jets

and anti- $k_T$  do a good job reconstructing the jets' energies, although power jets has a  $-5\%$  discrepancy in  $f_3$ . However, the (dimensionless) dijet invariant mass  $\tilde{m}_{ij}^2$  is a more important observable, since it verifies energy *and* angles; in these columns, it is quite clear that power jets reconstructs the kinematic configuration to within  $+1\%$  for the nearly back-to-back jets, and within  $\pm 5\%$  for the third jet (comparable to anti- $k_T$ 's performance).

Table 5.2. Reconstructed 3-jet kinematics for the 3-jet-like event (Fig. 5.8b).

	$f_1$	$f_2$	$f_3$	$\tilde{m}_{12}^2$	$\tilde{m}_{13}^2$	$\tilde{m}_{23}^2$
parton	0.4074	0.3587	0.2339	0.5322	0.2827	0.1851
power jets	0.4051(1)	0.3658(9)	0.2292(9)	0.5324(16)	0.2846(15)	0.1830(3)
anti- $k_T$	0.4032	0.3691	0.2277	0.5277	0.2820	0.1903

In Table 5.2, we show a similar result for the 3-jet-like event, whose fit is shown in Figure 5.8b. While there is a statistically significant difference between power jets'  $f_2$  and  $f_3$  and the original partons, a difference of a similar magnitude also exists in the anti- $k_T$  reconstruction, indicating that it is likely a systematic effect (e.g., from showering). One immediately obvious source of this misalignment is the fact that extensive jets are massive, while the original partons are massless, so the showering Monte Carlo must move some energy around in order to generate jet mass. This is another reason the dijet invariant mass a more useful observable, and these columns show that power jets reconstruct the entire kinematic configuration to within  $\pm 1\%$ . Furthermore, the errors in these masses indicate that the reconstruction is repeatable and robust; not only is the NLLS able to find the global minimum, this minimum is insensitive to a small amount of soft radiation.

**5.3.3 Power jets naturally accommodate pileup.** We have just seen that power jets can reconstruct three jet events without defining jet boundaries, or assigning specific particles to specific jets, and the procedure is repeatable and robust to a

small amount of soft radiation. However, one of the major challenges of current and future collider physics is a *large* amount of soft radiation, namely pileup. In order for the LHC to see increasingly rarer events, it needs higher luminosity — a higher probability of proton collisions per beam crossing — so that we may more often detect an event with  $Q = \mathcal{O}(500 \text{ GeV})$ . Yet increased luminosity has the side effect of producing dozens of soft, (in)elastic proton scatterings at  $Q = \mathcal{O}(5 \text{ GeV})$ , unrelated to the hard scatter, but overlapping it in the detector.

Tracking provides a great deal of pileup mitigation, because the primary vertex of the hard scatter can be located with millimeter resolution, and tracks emanating from any other primary vertex are pileup (by definition). However, this procedure does not work for neutral particles, whose primary vertex cannot be easily determined, and so other techniques must be used to remove pileup energy from jets (lest the pileup be treated as signal, smearing out the event’s observables). For anti- $k_T$  jets, pileup mitigation can be broadly characterized as *subtraction*; pileup (or more generally, soft energy) is subtracted either before clustering, after clustering, or both. While these techniques can be quite successful, they invariably throw away real information about the hard scatter, and are susceptible to random, local fluctuations in pileup intensity. For example, the area subtraction technique takes a specific event and determines the average pileup energy per  $d\Omega$ . It then calculates the area of a hard jet and subtracts the average amount of pileup. When a particular jet “gets lucky,” and isn’t heavily struck by the pileup, too much energy is subtracted.

Since power jets sum correlations across the entire event, they offer a more global approach to pileup. Each pileup event is a manifestation of the total (in)elastic  $pp \rightarrow X$  process, which has a consistent total cross section  $\sigma$ . The *shape* of this cross section  $\frac{d\sigma}{\sigma dQ d\eta}$  depends on the details of the beam, but these details do not change for long periods of run time. Hence, pileup can be treated as some general shape,

with variable energy fraction that depends on the total number of pileup collisions accompanying any given hard scatter, so that the event shape becomes

$$\rho(\hat{r}) = \rho(\hat{r})_{\text{hard}} + \rho(\hat{r})_{\text{PU}} = (1 - f_{\text{PU}}) \sum_j f_j h_j(\hat{r}) + f_{\text{PU}} h_{\text{PU}}(\hat{r}). \quad (5.69)$$

Thus, when power jets fit  $H_l$ , they can fit  $f_{\text{PU}}$  alongside the  $n$ -prongs of the hard scatter, provided that  $h_{\text{PU}}(\hat{r})$  is known. The beauty of this formulation is that  $h_{\text{PU}}(\hat{r})$  does not need to be calculated from theory; it can be measured directly in the detector, by waiting for events which have lots of pileup and little to no hard scatter. Such “min bias” events constitute a large portion of the data discarded every second at the LHC, because they are simply too mundane to write to disk. Furthermore, since  $h_{\text{PU}}(\hat{r})$  should be the same for every bunch crossing, many min bias events can be summed into a master shape that can be updated throughout the day (to account for variations in the beam’s shape, energy, etc.)

Of course, in any particular event the pileup is discrete; each pileup event is independent, and so their total count follows a Poisson distribution; this determines the pileup energy fraction  $f_{\text{PU}}$ . Furthermore, the continuous shape of pileup  $h(\hat{r})_{\text{PU}}$  is discretely sampled (like the hard scatter), and so is subject to sampling error. In spite of these limitations, there is an interesting balance. When there is just a little pileup, a continuous  $\rho(\hat{r})_{\text{PU}}$  is a bad approximation, but this doesn’t matter because  $f_{\text{PU}}$  is so small that  $\rho(\hat{r})_{\text{PU}}$  has a minimal effect on the observables extracted from the power jets fit. Conversely, when there is a lot of pileup, so that pileup mitigation becomes increasingly necessary,  $\rho(\hat{r})_{\text{PU}}$  becomes a *good* approximation, and the many pileup-pileup correlations from across the event help nail down pileup’s contribution to the total event shape.

**5.3.3.1 Isotropic pileup.** If pileup can be defined via some universal shape  $h_{\text{PU}}(\hat{r})$  (which must be azimuthally symmetric about the beam axis for unpolarized beams) then its contribution to  $H_l$  can be calculated using the tools discussed thus far. This



means that for testing purposes, the exact pileup model chosen is not very important. Since we are still testing power jets at an  $e^+e^-$  collider, where pileup is generally not a large issue, there is no phenomenological pileup model which can be easily turned on in the event generator. Hence, we can choose the simplest pileup model possible; totally isotropic pileup. Since this distribution has a trivial on-axis coefficient  $\bar{h}_l = 0$  for  $l > 0$ , the pileup simply scales  $H_l^{\text{jet}}$  by a factor of  $(1 - f_{\text{PU}})^2$ .

The energy distribution of our pileup model is moderately important. Pileup is generally low energy, but occasionally produces energetic outliers; combined with spatially discrete sampling, this can produce significant Poissonian fluctuations above or below the smooth expectation of  $h(\hat{r})$ , looking less like pileup and more like real QCD radiation. To simulate this high-energy tail, we use an exponential distribution of pileup energy fraction  $f$  (relative to the energy of the hard scatter)

$$g(f) = \exp(-f/\text{Ex}(f))/\text{Ex}(f) . \quad (5.70)$$

We choose  $\text{Ex}(f) = 10^{-3}$ , so that 100 pileup particles will contribute 10% of the energy of the hard scatter. We additionally assume that charged pileup can be effectively removed, so that pileup contributes only to the calorimeter towers.

We now test the robustness of power jets to pileup by repeatedly filling events with random pileup (random isotropic direction and random energy from  $g(f)$ ) and repeating the fit to  $H_l$ . Quite strikingly, even when using relatively high amounts of pileup (so that the signal-to-noise  $S/N$  is quite small, with  $f_{\text{PU}} = 1/(1 + S/N)$ ), power jets are still able to fit the power spectrum, as depicted in Figure 5.9 for the 2-jet-like event and Figure 5.11 for the 3-jet like event. The most striking feature of these fits is their ability to squeeze out all the correlated information embedded in  $H_l$  and accurately and repeatedly reconstruct the basic kinematic shape (energies and angles), even in the face of high pileup. In the inset, the reconstructed kinematics is shown in blue (solid), with the original 3-parton kinematics behind in red (dashed).

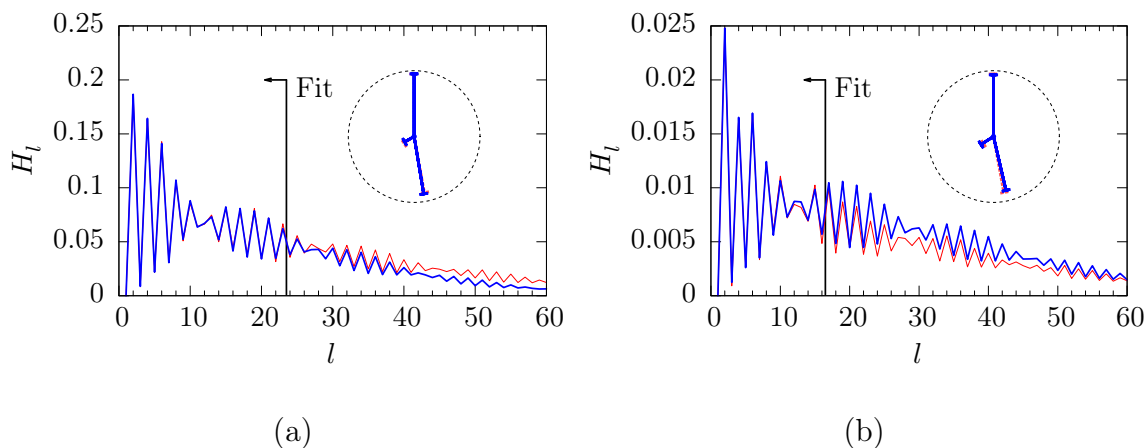


Figure 5.9. The  $H_l$  fit of the 2-jet-like event, with (red, thin)  $H_l^{\text{obs}}$  and (blue, thick)  $H_l^{\text{jet}}$ , and a kinematic depiction of the fit. Random pileup is added until (a)  $S/N = 1$  and (b)  $S/N = 1/5$ .

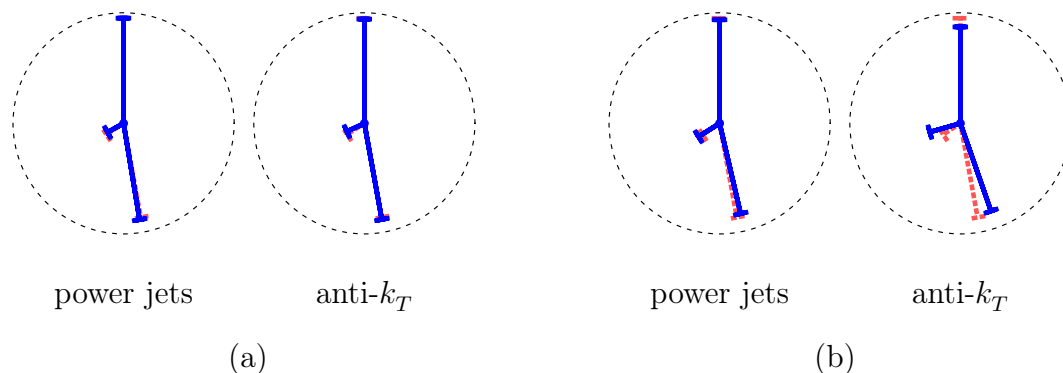


Figure 5.10. The reconstructed kinematics (blue, solid) of the 2-jet-like event, versus (red, dashed) the three original partons, for power jets and anti- $k_T$  with random pileup; (a)  $S/N = 1$  and (b)  $S/N = 1/5$ .

The difference is just barely visible in the high pileup sample.

Curiously, it is pileup itself which aids in its own abatement; the significant pileup-pileup correlations betray their very consistent shape, which can be treated as a single entity. Note also the  $y$ -axis, which shows the massively reduced power of these events, which is nonetheless resolvable due to the relatively high event multiplicity (essentially every calorimeter tower has some energy). When we compare the power jets reconstruction to that of anti- $k_T$ , in Figures 5.10 and 5.12, the susceptibility of

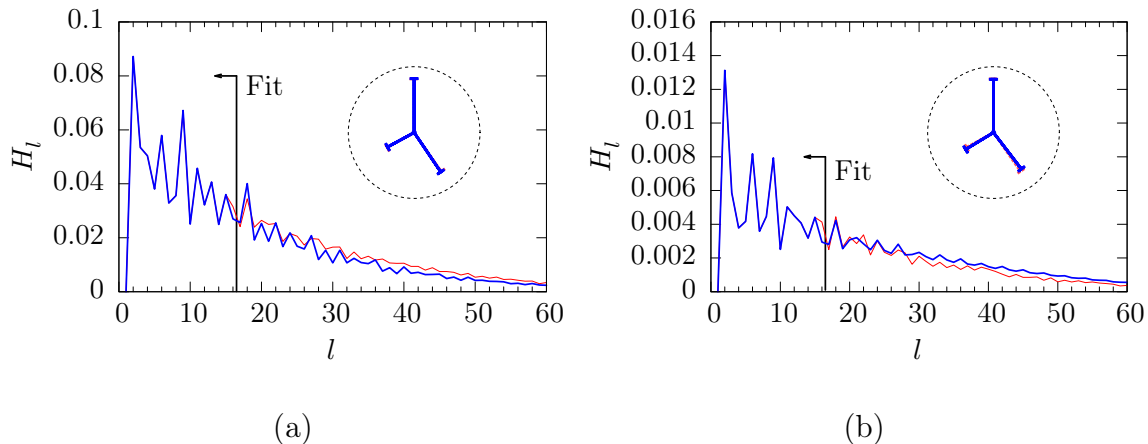


Figure 5.11. The  $H_l$  fit of the 3-jet-like event, with (red, thin)  $H_l^{\text{obs}}$  and (blue, thick)  $H_l^{\text{jet}}$ , and a kinematic depiction of the fit. Random pileup is added until (a)  $S/N = 1$  and (b)  $S/N = 1/5$ .

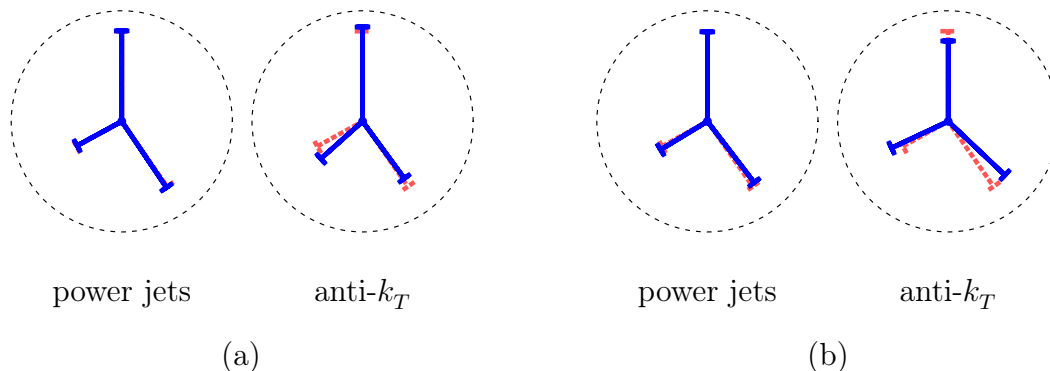


Figure 5.12. The reconstructed kinematics (blue, solid) of the 3-jet-like event, (red, dashed) versus the three original partons, for power jets and anti- $k_T$  with random pileup; (a)  $S/N = 1$  and (b)  $S/N = 1/5$ .

sequential clustering to pileup is clearly visible. Note that pileup subtraction was not attempted on the anti- $k_T$  jets, but neither was it attempted for power jets — the pileup was *fit*. And in all cases, the  $f_{\text{PU}}$  extracted from the fit matched the amount added to the event within a few percent.

This brings us back to the original power jets fit of Figure 5.8, which used  $f_{\text{PU}} = 1\%$  to test the robustness of the basic fit; we now have the nomenclature to discuss one of its features. For its twenty fits, full freedom was given to  $f_{\text{PU}}$ , and

$f_{\text{PU}} \approx 1\%$  was the best-fit solution. This is an important property, because it means that the amount of pileup in the event does not have to be known a priori (although a good starting guess can easily be made based upon observables in the detector). Thus, power jets' pileup model is well suited to handle the Poissonian fluctuation in total pileup intensity that one expects to find in individual events.

#### 5.4 The future of the QCD power spectrum

We started down this path with the aim of extracting new information from the correlated QCD radiation spectrum, and ended with a jet definition that is robust to pileup. It turns out that the QCD power spectrum did not look like the CMB, and is dominated by the shape of jets. We accounted for this shape by fitting prongs to hard QCD radiation, and the shape of prongs to soft QCD radiation. Extracting more subtle shapes (e.g., long-range QCD correlations like the same-side ridge) will be the focus of future work.

Now that we have begun successfully applying the power jets fit, we have an approach that seems well suited to study the jet substructure caused by QCD showering. In addition to using substructure to identify signals like boosted  $h \rightarrow b\bar{b}$  and boosted top, we may also be able to study higher-order, less perturbative effects like color connections and hadronization. Such studies may improve the tuning of Monte Carlo generators, whose accuracy limits the ability of detector experiments to understand their systematic errors. With most of the major engineering out of the way, it is time to start calculating power jet predictions directly from perturbative QCD, and determining useful measurements for the LHC to make.

However, *all* the engineering challenges are not yet solved; first, the power jets fit as currently implemented loses the absolute orientation of the event, which is *probably* useful; second,  $H_l$  is not boost invariant, so it will not work at the LHC in

its current form. Yet these problems are not intractable. In fact, both will likely be solved by extending the NLLS fit to absorb four more degrees of freedom.

**5.4.1 Fitting the orientation.** The current iteration of power jets loses an event's absolute orientation because it uses  $H_l$ , which is rotationally invariant. This rotational invariance is the primary reason that  $H_l$  is infrared and collinear safe (since it loses sensitivity to local fluctuations and small changes in orientation), and also vastly simplifies the calculation of  $H_l$  (versus calculating  $\rho_l^m$  individually, which contain orientation information).

There is a relatively simple way to hack the orientation out of  $H_l$ . Imagine an event. Using the methods of the previous sections, we solve for its shape by writing down an  $n$ -prong model and fitting  $H_l^{\text{jet}}$  to  $H_l^{\text{obs}}$ . We now know the event's shape, and are merely lacking its absolute orientation, which is parameterized by three Euler angles. We can now define a hybrid event shape

$$\rho(\hat{r})_{\text{hybrid}} = \frac{1}{2}(\underbrace{\rho(\hat{r})_{\text{jet}}}_{\text{rotated}} + \rho(\hat{r})_{\text{obs}}). \quad (5.71)$$

If we now rotate the  $n$ -prong jet model by three Euler angles, we can calculate the power spectrum  $H_l^{\text{hybrid}}$  of this hybrid event shape. Only when we have found the *correct* three angles (so that power jets' prongs overlap the tracks and towers which they represent) will  $H_l^{\text{hybrid}}$  match  $H_l^{\text{obs}}$ . Thus, we can *fit*  $H_l^{\text{hybrid}}$  to  $H_l^{\text{obs}}$  to recover the absolute orientation.

I am in the early stages of testing this scheme, and it is sometimes successful; the major problem is that the fit can relatively easily get stuck in a local minimum (imagine a 3-jet event, where the leading jet is correct, but the second and third jet have swapped positions, which corresponds to a local minimum from which the fit cannot climb out). I expect to solve this problem in the coming weeks.

**5.4.2 Power jets at the LHC.** Even though power jets were designed with the problems of the LHC in mind, they cannot currently accommodate one of the main properties of a hadron collider: the longitudinal boost  $y_{\text{cm}}$  of the hard scatter's CM frame. Yet given the proposed solution for recovering the absolute orientation, and the success of the NLLS fit so far, we can also imagine fitting  $y_{\text{cm}}$ .

A good initial guess for  $y_{\text{cm}}$  can be obtained from the CM frame of the leading jets (summing over the anti- $k_T$  jets which account for 90% of the event's  $p_T$ )

$$\mathbf{p}_{90\%} = \sum_i \mathbf{p}_i. \quad (5.72)$$

The  $y_{\text{cm}}$  of  $\mathbf{p}_{90\%}$  will be used to start the power jet fit, after which  $y_{\text{cm}}$  will be given some freedom to move around. In each iteration of the fit, the value of  $y_{\text{cm}}$  can be used to boost the massless tracks and towers into their supposed CM frame, to calculate  $H_i^{\text{obs}}$ . Once the final fit is obtained, the reconstructed power jets can be boosted back into to the lab frame. Of course, the beam holes create an immediate problem;  $y_{\text{cm}}$  has the ability to push energy into them, or pull unobserved energy out of them. Hence, it is likely that a power jets fit will have to treat forward jets differently. This will be an interesting nut to crack.

## CHAPTER 6

### THESIS SUMMARY

This thesis has been an exploration of tools necessary for high energy physics to fully utilize jets — the primary tool for studying QCD, and an invaluable component in the search for new physics. That we need new tools is a symptom of our progress; success requires setting your sights beyond what is currently possible. When the LHC was proposed, the community did not have the tools to tag TeV  $b$  jets or remove the detritus of 40 pileup events from our jets. Not surprisingly, as LHC start-up grew nearer, the tools necessary to see jets at  $\sqrt{S} = 8$  TeV were completed. But staying ahead of the curve requires constant vigilance, and my contributions to high energy physics fall into this category. My talent seems to be noticing loose threads and pulling at them till I drive down to the heart of an issue. This thesis contains three such threads:

- The  $\mu_x$  boosted-bottom jet tag of Chapter 2, which uses the boosted kinematics of  $B$  hadron decay to define an observable  $x$  for  $b$  jet identification. The most important property of the  $\mu_x$  tag is the  $\mathcal{O}(100)$  ratio of the signal efficiency to the light jet fake rate, a ratio which does not depend on jet  $p_T$  or pileup. I show in Chapter 3 that the  $\mu_x$  tag has great potential to look for new physics at very large invariant mass, a region of parameter space which is difficult to probe because existing  $b$  tags contaminate the boosted  $b$  jet signal with far too many light jets.
- Power jets, the first result from my investigation into the QCD power spectrum in Chapters 4 and 5. This research direction builds upon some of the original ways that jets were studied 40 years ago, and while it starts with the power

spectrum  $H_l$  introduced by Fox and Wolfram, it is an entirely novel approach that seems much more practical and portable. A major development is the realization that shape functions are necessary to discard correlations smaller than the intrinsic angular scale of a discrete sample. Reconstructing jets requires a similar mindset: simulate hard QCD radiation with discrete prongs and soft QCD radiation with continuous shape functions. Not only does this framework produce accurate and precise jet kinematics, it naturally accommodates pileup. And by using all the information in the detector, the pileup-pileup correlations can reveal pileup's effect, so that jet reconstruction remains accurate and precise even when pileup is severe ( $S/N = 1/5$ ). These exciting results are the first in what I envision to be a larger program of harnessing the QCD power spectrum for studying particle physics phenomena.

- My robust inversion method (see Appendix A.1), which generates random samples with maximal floating point precision. This recipe grew out of a deep respect for the difference between real and rational arithmetic, since only the latter is possible on a computer.

Each of these threads took significant effort and tenacity to pull, and I think they collectively reflect my ability to find new ways to look at old problems. My hope is that these tools will also allow particle physics to grow more precise. This will be especially important since supersymmetry remains elusive at the LHC, and a 100 TeV proton collider may be required to render a final ruling. The  $\mu_x$  tag and power jets will likely be invaluable tool at such a machine, where nearly all jets will be boosted and bathed in pileup. By using both tools simultaneously, a new layer of unexplained phenomena may be uncovered — a thread in the data which, when pulled, reveals an important truth of Nature that no one expected to find.



APPENDIX A  
NUMERICAL STABILITY

This appendix introduces some of the tools I have developed to ensure the floating point accuracy of my physics research. They are not always essential to the physics, but they form an independent branch of study which, in my opinion, has made me a more complete scientist. Indeed, Section A.1 is an important discovery that will help keep Monte Carlo simulations precise as their complexity continues to grow.

Since Spring Break 2015, my main scientific *hobby* has been floating point arithmetic. I do not recall exactly how I stumbled into this minefield, but I'm certain it spun-off of from my fascination with engineering disasters. There were two notable engineering disasters caused by floating-point failures (or, more accurately, failures of software engineers to account for limitations in floating point arithmetic); a Patriot missile that couldn't find its target in 1991, and an Ariane 5 heavy-lift rocket which flew off course in 1996 (requiring its self-destruction) [155].

Knowing very little about floating point arithmetic, I was surprised that such disasters were possible. After reading Goldberg's excellent monograph [156], I was surprised at how ignorant I was about how computers actually do real-number arithmetic (i.e., they don't). And after three years of tinkering and surveying the knowledge of my peers, I am surprised that errors caused by floating point ignorance don't happen more often. Furthermore, I am appalled that scientists anywhere are allowed to graduate without formal training (however brief) in the math behind their data.<sup>23</sup>

## A.1 A robust inversion method

Monte Carlo simulation is an important tool for modeling highly nonlinear systems (like particle colliders and cellular membranes), and random, floating-point

---

<sup>23</sup>I did not attend IIT as an undergrad, so I never took PHYS 240 (Computational Science), nor any equivalent class at Carthage. All of my floating point wisdom was acquired through independent investigation.

numbers are their fuel. These random samples are frequently generated via the inversion method, which harnesses the mapping of the quantile function  $Q(u)$  (e.g., to generate proposal variates for rejection sampling). Yet the increasingly large sample size of these simulations makes them vulnerable to a flaw in the inversion method;  $Q(u)$  is *ill-conditioned* in a distribution's tails, stripping precision from its sample. This flaw stems from limitations in machine arithmetic which are often overlooked during implementation (e.g., in popular C++ and Python libraries). This section introduces a *robust* inversion method, which reconditions  $Q(u)$  by carefully drawing and using uniform variates. PQRAND, a free C++ and Python package, implements this novel method for a number of popular distributions (exponential, normal, gamma, and more).

**A.1.1 Introduction.** The inversion method samples from a probability distribution  $f$  via its quantile function  $Q \equiv F^{-1}$ , the inverse of  $f$ 's cumulative distribution  $F$  [157, 158].  $Q$  is used to transform a random sample from  $U(0, 1)$ , the uniform distribution over the unit interval, into a random sample  $\{f\}$ ;

$$\{f\} = Q(\{U(0, 1)\}). \quad (\text{A.1})$$

This scheme is powerful because quantile functions are formally exact. But any real-world implementation will be *formally inexact* because: (i) A source of true randomness is generally not practical (or even desirable), while a repeatable pseudo-random number generator (PRNG) is never perfect. (ii) The uniform variates  $u$  and their mapping  $Q(u)$  use finite-precision machine arithmetic. The first defect has received the lion's share of attention, leaving the second largely ignored. As a result, common implementations of inversion sampling lose precision in the tails of  $f$ .

This leak must be subtle if no one has patched it. Nonetheless, the loss of precision commonly exceeds *dozens* of ULP (units in the last place) in a distribution's tails. Contrast this to library math functions (`sin`, `exp`), which are painstakingly

crafted to deliver no more than *one* ULP of systematic error. When the inversion method loses precision, it produces inferior, repetitive samples, to which Monte Carlo simulations *may* become sensitive as they grow more complex, drawing ever more random numbers. Proving that the effect is negligible is incredibly difficult, so the best alternative is to use the most numerically stable sampling scheme possible with floating point numbers — *if* it is not too slow. The robust inversion method proposed here is 80–100% as fast as the original.

To isolate the loss of precision, we examine the three independent steps of inversion sampling:

1. Generate random bits (i.i.d. coin flips) using a PRNG.
2. Convert those random bits into a uniform variate  $u$  from  $U(0, 1)$ .
3. Plug  $u$  into  $Q(u)$  to sample from the distribution  $f$ .

The first two steps do not depend on  $f$ , so they are totally generic (a major virtue of the method). Of them, step 1 has been exhaustively studied [159–161], and is essentially a solved problem — when in doubt, use the Mersenne twister [161, 162]. Step 3 has been validated using real analysis [157, 163], so that known quantile functions need only be translated into computer math functions.

This leaves step 2 which, at first glance, looks like a trivial coding task to port random bits into a real-valued  $Q$ . Yet computers cannot use real numbers, and neglecting this fact is dangerous — using this as its central maxim, this section conducts a careful investigation of the inversion method from step 2 onward. Section A.1.2 begins by using the condition number to probe step 3, finding that a distribution’s quantile function is numerically unstable in its tails. This provides a sound framework for Section A.1.3 to find the subtle flaw in the canonical algorithm for drawing uniform variates (step 2). A *robust* inversion method is introduced to fix

both problems, and is empirically validated in Section 3.2.2 by comparing the near-perfect sample obtained from the PQRAND package to the deficient samples obtained from standard C++ and Python tools.

**A.1.2  $\mathbb{Q}$  are ill-conditioned, but they do not have to be.** Real numbers are not countable, so computers cannot represent them. Machine arithmetic is limited to a countable set like rational numbers  $\mathbb{Q}$ . The most versatile rational approximation of  $\mathbb{R}$  are *floating point* numbers, or “floats” — scientific notation in base-two ( $m \times 2^E$ ). The precision of floats is limited to  $P$ , the number of binary digits in their mantissa  $m$ , which forces relative rounding errors of order  $\epsilon \equiv 2^{-P}$  upon every floating point operation [156]. The propagation of such errors makes floating point arithmetic formally inexact. In the worst case, subtle effects like cancellation can degrade the *effective* (or de facto) precision to just a handful of digits. Using floats with arbitrarily high  $P$  mitigates such problems, but is usually emulated in software — an expensive cure. Prudence usually restricts calculations to the largest precision widely supported in hardware, *binary64* ( $P = 53$ ), commonly called “double” precision.

Limited  $P$  makes the intrinsic stability of a computation an important consideration; a result should not change dramatically when its input suffers from a pinch of rounding error. The numerical stability of a function  $g(x)$  can be quantified via its condition number  $C(g)$  — the relative change in  $g(x)$  per the relative change in  $x$  [155]

$$C(g) \equiv \left| \frac{g(x + \delta x) - g(x)}{g(x)} \right| / \left| \frac{\delta x}{x} \right| = \left| x \frac{g'(x)}{g(x)} \right| + \mathcal{O}(\delta x). \quad (\text{A.2})$$

When an  $\mathcal{O}(\epsilon)$  rounding error causes  $x$  to increment to the next representable value,  $g(x)$  will increment by  $C(g)$  representable values. So when  $C(g)$  is large (i.e.,  $C(g) \rightarrow 2^P$ ),  $g(x)$  is *ill-conditioned* and imprecise; the tiniest shift in  $x$  will cause  $g(x)$  to hop over an *enormous* number of values — values through which the real-valued function passes, and which are representable with floats of precision  $P$ , but which cannot be

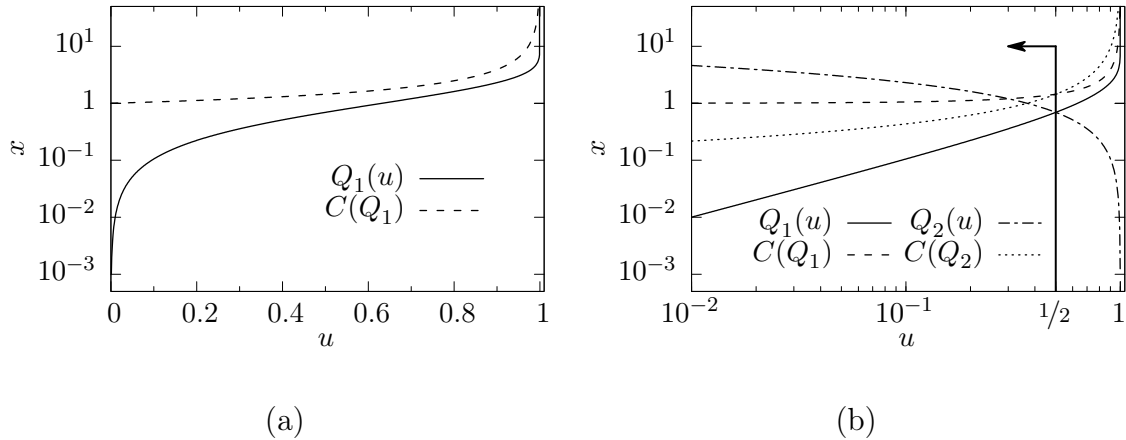


Figure A.1. The  $\lambda = 1$  exponential distribution  $f(x) = e^{-x}$ ; (a) the quantile function  $Q_1$  (solid) and its condition number (dashed) and (b) the “quantile flip-flop” — in the domain  $0 < u \leq 1/2$ , each  $Q$  maps out half of  $f$ ’s sample space while remaining well-conditioned.

attained via the floating point calculation  $g(x)$ . The condition number should be used to avoid such numerical catastrophes.

We now have a tool to uncover possible instability in the inversion method, specifically in its quantile function  $Q$  (step 3). As a case study, we can examine the exponential distribution (the time between events in a Poisson process with rate  $\lambda$ , like radioactive decay);<sup>24</sup>

$$f(x) = \lambda e^{-\lambda x}, \quad (\text{A.3})$$

$$F(x) = 1 - e^{-\lambda x}, \quad (\text{A.4})$$

$$Q_1(u) = -\frac{1}{\lambda} \log(1 - u) = -\frac{1}{\lambda} \text{log1p}(-u), \quad (\text{A.5})$$

$$C(Q_1) = -\frac{u}{(1 - u) \text{log1p}(-u)}. \quad (\text{A.6})$$

A well-conditioned sample from the exponential distribution requires  $C(Q_1) \leq \mathcal{O}(1)$  everywhere, but Figure A.1a clearly reveals that  $C(Q_1)$  (dashed) becomes large as  $u \rightarrow 1$ . Why is  $Q_1$  ill-conditioned there? According to Equation A.2, a function

---

<sup>24</sup> $\text{log1p}(x)$  is an implementation of  $\log(1 + x)$  which sidesteps an unnecessary floating point cancellation [164].

can become ill-conditioned when it is *steep* ( $|g'/g| \gg 1$ ), and  $Q_1$  (solid) is clearly steep at both  $u = 0$  and  $u = 1$ . These are  $f$ 's “tails” — a large range of sample space mapped by a thin, low probability slice of the unit interval. Yet in spite of its steepness,  $Q_1$  remains well-conditioned throughout its small-value tail ( $u \rightarrow 0$ ) because floats are denser near the origin — reusing the same set of mantissae, but with smaller exponents — and a denser set of  $u$  allows a more continuous sampling of a rapidly changing  $Q_1(u)$ . This extra density manifests as the singularity-softening factor of  $x$  in Equation A.2. Unfortunately, the same relief cannot occur as  $u \rightarrow 1$ , where representable  $u$  are not dense enough to accommodate  $Q_1$ 's massive slope.

Because  $Q_1$  is ill-conditioned near  $u = 1$ , the large- $x$  portion of its sample  $\{f\}$  will be imprecise; many large- $x$  floats which should be sampled are skipped-over by  $Q_1$ . This problem is not unique to the exponential distribution; it will occur whenever  $f$  has two tails, because one of those tails will be located near  $u = 1$ . Luckily,  $U(0, 1)$  is perfectly symmetric across the unit interval, so transforming  $u \mapsto 1 - u$  produces an equally valid quantile function;

$$Q_2(u) = -\frac{1}{\lambda} \log(u), \quad (\text{A.7})$$

$$C(Q_2) = -\frac{1}{\log(u)}. \quad (\text{A.8})$$

The virtue of using two valid  $Q$ 's is evident in Figure A.1b; for  $u \leq 1/2$ , each version is well-conditioned, with  $Q_1$  sampling the small-value tail ( $x \leq \text{median}$ ) and  $Q_2$  the large-value tail ( $x \geq \text{median}$ ). Since the pair collectively and stably spans  $f$ 's entire sample space,  $f$  can be sampled via the composition method; for each variate, randomly choose one version of the quantile function (to avoid a high/low pattern), then feed that  $Q$  a random  $u$  from  $U(0, 1/2]$ .

This “quantile flip-flop” — a randomized, two- $Q$  composition split at the median — is a simple, general scheme to recondition a quantile function which becomes unstable as  $u \rightarrow 1$ . It is also immediately portable to antithetic variance reduction,

a useful technique in Monte Carlo integration where, for every  $x = Q(u)$  one also includes the opposite choice  $x' = Q(u')$  [165]. A common convention is  $u' \equiv 1 - u$ , which can create a negative covariance  $\text{cov}(x, x')$  that decreases the overall variance of the integral estimate. Generating antithetic variates with a quantile flip-flop is trivial; instead of randomly choosing  $Q_1$  or  $Q_2$  for each variate, always use both.

**A.1.3 An optimally uniform variate is maximally *uneven*.** The condition number guided the development of the quantile flip-flop, a rather simple way to stabilize step 3 of the inversion method during machine implementation. Our investigation now proceeds to step 2 — sampling uniform variates. While steps 2 and 3 seem independent, we will find that there is an important interplay between them; a quantile function can be destabilized by sub-optimal uniform variates, but it can also wreck itself by mishandling *optimal* uniform variates.

---

**Algorithm 1** Canonically draw a random float (with precision  $P$ ) from  $U[0, 1)$

---

**Require:**  $B \in \mathbb{Z}^+$  ▷  $B$  must be a positive integer  
1:  $A \leftarrow \text{FLOAT}(2^B)$  ▷ Convert  $2^B = (j_{\max} + 1)$  to a float (exactly).  
2: **repeat**  
3:    $j \leftarrow \text{RNG}(B)$  ▷ Convert  $B$  random bits into an integer from  $U[0, 2^B)$ .  
4:    $a \leftarrow \text{FLOAT}(j)$  ▷ Convert  $j$  to a float with precision  $P$ .  
5: **until**  $a < A$  ▷ The algorithm should not return 1 if  $j$  rounds to  $A$ .  
6: **return**  $a/A$

---

The canonical method for generating uniform variates is Algorithm 1 [158–160, 164, 166–169]; an integer is randomly drawn from  $[0, 2^B)$ , then scaled to a float in the half-open unit interval  $[0, 1)$ . Using  $B \leq P$  produces a completely uniform sample space — each possible  $u$  has the same probability, with a rigidly even spacing of  $2^{-B}$  between each. Using  $B = P$  gives the ultimate *even* sample  $\{U_E[0, 1)\}$ , as depicted in Figure A.2E (which uses a ridiculously small  $B = P = 4$  to aide the eye). When  $B > P$ , line 4 will be forced to round many large  $j$ , as the mantissa of  $a$  is not large enough to store every  $j$  with full precision. As  $B \rightarrow \infty$ , this rounding saturates the floats available in  $U[0, 1)$ , creating the *uneven*  $\{U_N[0, 1)\}$  depicted Figure A.2N. This



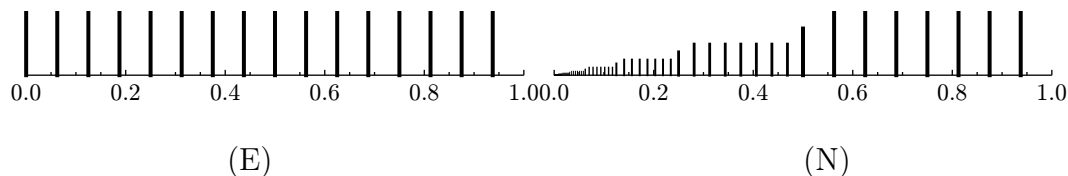


Figure A.2. A visual depiction (using floats with  $P = 4$  for clarity) of each possible  $u$  for (E) even  $\{U_E[0, 1]\}$  and (N) uneven  $\{U_N[0, 1]\}$ . The height of each tic indicates its relative probability, which is proportional to the width of the number-line segment which rounds to it.

uneven sample space is still uniform because large  $u$  are more probable, absorbing more  $j$  from rounding (due to their coarser spacing).

Depending on the choice of  $B$ , Algorithm 1 can generate uniform variates which are either even or uneven, but which is better? There seem to be no definitive answers in the literature — which is likely why different implementations choose different  $B$  — so we will have to find our own answer. We start by choosing the *even* uniform variate  $\{U_E[0, 1]\}$  as the null hypothesis, for two obvious reasons: (i) Figure A.2E certainly *looks* more uniform and (ii) taking  $B \rightarrow \infty$  does not seem practical. However, we will soon find that perfect evenness has a subtle side effect — it forces all quantile functions to become ill-conditioned as  $u \rightarrow 0$ , even if they have an excellent condition number!

The condition number implicitly assumes that  $\delta x$  is vanishingly small. This is true enough for a generic float, whose  $\delta x = \mathcal{O}(\epsilon x)$  is much smaller than  $x$ . But the even uniform variates have an *absolute* spacing of  $\delta u = \epsilon$ . To account for a finite  $\delta x$ , we define a function's *effective* precision

$$P^*(g) \equiv \left| \frac{g(x + \delta x) - g(x)}{g(x)} \right| = \delta x \left| \frac{g'(x)}{g(x)} \right| + \mathcal{O}(\delta x^2). \quad (\text{A.9})$$

Like  $C(g)$ , a large effective precision  $P^*(g)$  indicates an ill-conditioned calculation. For a generic floating point calculation  $\delta x = \mathcal{O}(\epsilon x)$ , so  $P^*$  reverts back to the condition number ( $P^*(g) \approx \epsilon C(g)$ ). But feeding even uniform variates into a quantile function

gives  $\delta u = \epsilon$ , so

$$P_{\text{E}}^*(Q) = \epsilon \left| \frac{Q'(u)}{Q(u)} \right| + \mathcal{O}(\epsilon^2). \quad (\text{A.10})$$

Calculating  $P_{\text{E}}^*(Q)$  for the quantile flip-flop of Figure A.1b indicates that *both*  $Q$  become ill-conditioned as  $u \rightarrow 0$  (where  $Q$  becomes steep), in stark opposition to their excellent condition numbers. That using even uniform variates will break a quantile flip-flop is a problem not unique to the exponential distribution; it occurs whenever  $f$  has a tail (so that  $|Q'/Q| \rightarrow \infty$  as  $u \rightarrow 0$ ).

The reduced effective precision  $P_{\text{E}}^*(Q)$  caused by even uniform variates creates sparsely populated tails; there are many extreme values which  $\{f\}$  will never contain, and those which it does will be sampled too often.  $\{U_{\text{E}}[0, 1]\}$  is simply *too finite*;  $2^P$  even uniform variates can supply no more than  $2^P$  unique values. This implies that the *uneven* sample  $\{U_{\text{N}}[0, 1]\}$  will restore quantile stability, since its denser input space ( $\delta u = \mathcal{O}(\epsilon u)$ ) will stabilize  $P_{\text{N}}^*(Q)$  near the origin. These small  $u$  expand the sample space of  $\{f\}$  many times over, making its tails *far less* repetitive. And since uneven variates correspond to the limit where  $B \rightarrow \infty$  in Algorithm 1, they are equivalent to sampling  $U[0, 1 - \epsilon)$  from  $\mathbb{R}$  and rounding to the nearest float — the next best thing to a real-valued input for  $Q$ .

The virtue of using uneven uniform variates also follows from information theory. The Shannon entropy of a sample space  $X$  counts how many bits of information are conveyed by each variate  $x$ ;

$$H(X) \equiv - \sum_i \Pr(x_i) \log_2 \Pr(x_i). \quad (\text{A.11})$$

The sample space of the even uniform variates ( $B = P$ ) has  $n = \epsilon^{-1}$  equiprobable members, so

$$H_{\text{E}} = - \sum_{i=1}^n \epsilon \log_2(\epsilon) = - \log_2 \epsilon = P. \quad (\text{A.12})$$

This makes sense, since each even uniform variate originates from a  $P$ -bit pseudo-random integer.

The sample space of the uneven  $\{U_N[0, 1)\}$  contains every float in  $[0, 1)$ , which is naturally partitioned into sub-domains  $[2^{-k}, 2^{-k+1})$  with common exponent  $-k$ . Each domain comprises a fraction  $2^{-k}$  of the unit interval, and the minimum exponent  $-K$  depends on the floating point type (although  $K \gg 1$  for *binary32* and *binary64*). The uneven entropy is then the sum over sub-domains, each of which sums over the  $n/2$  equiprobable mantissae<sup>25</sup>

$$\begin{aligned} H_N &= - \sum_{k=1}^K \left( \sum_{i=1}^{n/2} 2^{-k} (2\epsilon) \log_2 (2^{-k} (2\epsilon)) \right) \\ &= \sum_{k=1}^K 2^{-k} (P - 1 + k) \approx P + 1 \quad (\text{for } K \gg 1). \end{aligned} \quad (\text{A.13})$$

*One more bit* of information than even variates is not a windfall. But  $H_E$  and  $H_N$  are the entropies of the *bulk* sample  $\{U[0, 1)\}$ . What is the entropy of the tail-sampling sub-space  $U[0, 2^{-k})$ ?

Rejecting all  $u \geq 2^{-k}$  in the even sample  $\{U_E[0, 1)\}$ , we find that smaller  $u$  have less information

$$H_E(k) = P - k \quad (\text{for } u < 2^{-k}). \quad (\text{A.14})$$

This lack of information in even variates is inevitably mapped to the sample  $\{f\}$ , consistent with the deteriorating effective precision as  $u \rightarrow 0$ . But for *uneven* uniform variates, the sample space is *fractal*; each sub-space looks the same as the whole unit interval, so that  $H_N(k) = P + 1$  as before! Every  $u$  has maximal information, and a high-entropy input should give a high-precision sample.

Both the effective precision  $P^*(Q)$  and Shannon entropy  $H$  predict that us-

---

<sup>25</sup>Ignoring the fact that exact powers of 2 are  $3/4$  as probable, which makes no difference once  $P \gtrsim 10$ .

ing even uniform variates will force a well-conditioned quantile function to become ill-conditioned, precluding a high-precision sample. Switching to uneven uniform variates will *recondition* it. But there is an important caveat; uneven variates are *very* delicate. Subtracting them from one mutates them back into *even* variates (with opposite boundary conditions);

$$1 - \{U_N[0, 1)\} \mapsto \{U_E(0, 1]\}. \quad (\text{A.15})$$

This is floating point *cancellation*. The subtraction erases any extra density in the uneven sample, because it maps the very dense region (near zero) to a region where floats are intrinsically sparse (near one). Conversely, the sparse region of the uneven sample (near one) has no extra information to convey when it is mapped near zero, and remains sparse. This is why  $Q_1$  (Eq. A.6) *must* use `log1p`.

**A.1.4 Precision: lost and found.** In Section A.1.2 we conditioned an intrinsically imprecise quantile function using a two- $Q$  composition. Then in Section A.1.3 we determined that uneven uniform variates are required to *keep*  $Q$  well-conditioned. These two practices comprise our *robust* inversion method, whose technical details we have deliberately left for Appendix B.2 because we have yet to prove that it makes a material difference. If indiscreet sampling decimates the precision of  $\{f\}$ , it should be quite evident in an experiment!

The quality of a real-world sample  $\{f\}$  can be assessed via its Kullback-Leibler divergence [170]

$$D_{\text{KL}}(\hat{P}||\hat{Q}) = \sum_i \hat{P}(x_i) \log_2 \frac{\hat{P}(x_i)}{\hat{Q}(x_i)}. \quad (\text{A.16})$$

$D_{\text{KL}}$  quantifies the *relative* entropy between a posterior distribution  $\hat{P}$  and a prior distribution  $\hat{Q}$  (c.f. Eq. A.11). The empirical  $\hat{P}$  is based on the count  $c_i$  — the number of times  $x_i$  appears in  $\{f\}$

$$\hat{P}(x_i) = c_i/N, \quad (\text{A.17})$$

where  $N$  is the sample size. The *ideal* density  $\widehat{Q}$  is obtained by mapping  $f$  onto floats, using the domain of real numbers  $(x_{i,L}, x_{i,R})$  that round to each  $x_i$ ;

$$\widehat{Q}(x_i) = \int_{x_{i,L}}^{x_{i,R}} f(x) dx = F(x_{i,R}) - F(x_{i,L}). \quad (\text{A.18})$$

$D_{\text{KL}}$  does not sum terms where  $\widehat{P}(x_i) = 0$  (i.e.,  $x_i$  was not drawn), because  $x \log x$  goes smoothly to zero when  $x \rightarrow 0$ .

The  $D_{\text{KL}}$  divergence is not a *metric* because it is not symmetric under exchange of  $\widehat{P}$  and  $\widehat{Q}$  [170]. And while  $D_{\text{KL}}$  is frequently interpreted as the information *gained* when using distribution  $\widehat{P}$  instead of  $\widehat{Q}$ , this is not true here. Consider a PRNG which samples from  $\widehat{Q} = U(0, 1)$ , but samples so poorly that it always outputs  $x = 0.5$  (and thus emits zero information). Its  $D_{\text{KL}} \approx P$  is clearly the precision *lost* by  $\widehat{P}$  (the generator). In less extreme cases, since  $\widehat{Q}$  is the most precise distribution possible given floats of precision  $P$ , any divergence denotes how many bits of precision were *lost*.

Our experiments calculate  $D_{\text{KL}}$  for samples of the  $\lambda = 1$  exponential distribution generated via the inversion method. We use GNU's `std::mt19937` for our PRNG ( $B = 32$ ), fully seeding its state from the computer's environmental noise (using GNU's `std::random_device`). Calculating  $D_{\text{KL}}$  requires recording the count for each unique float, and an accurate  $D_{\text{KL}}$  requires a very large sample size ( $N \gg P$ , so that  $\widehat{P} \rightarrow \widehat{Q}$  in the case of perfect agreement). To keep the experiments both exhaustive and tractable, and with no loss of generality, we use `binary32` ( $P = 24$ , or single precision). Since double precision is governed by the same IEEE 754 standard [171], and both types use library math functions with  $\mathcal{O}(\epsilon)$  errors, the  $D_{\text{KL}}$  results for `binary64` will be identical.<sup>26</sup>

---

<sup>26</sup>A `binary64` experiment is tractable, just not exhaustive. Memory constraints require intricate simulation of tiny sub-spaces of the unit interval, to act as a representative sample of the whole.

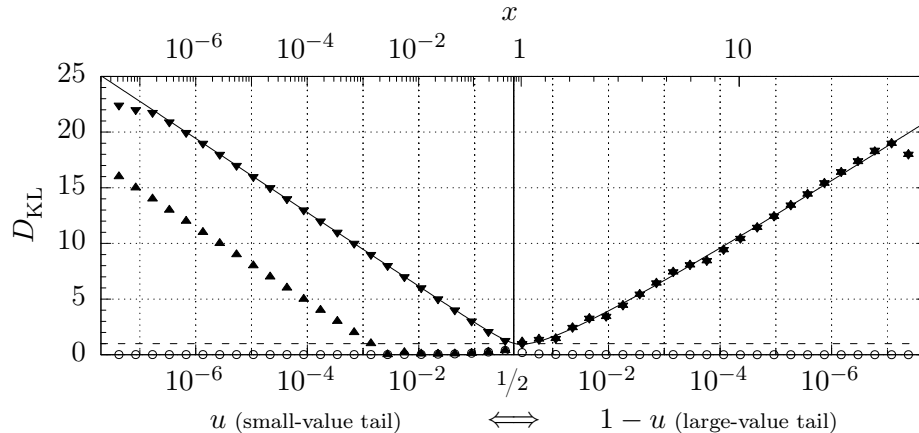


Figure A.3. The bits of precision *lost* ( $D_{\text{KL}}$ ) when sampling the  $\lambda = 1$  exponential distribution via ( $\blacktriangledown$ ) GNU, ( $\blacktriangle$ ) GNU + `log1p`, and ( $\circ$ ) our robust inversion method. The median ( $u = 1/2$ ) bisects the sample-space into two tails, with improbable values near the edges. The sampled variate  $x = Q(u)$  is shown on the top axis.

The first implementation we test is GNU’s `std::exponential_distribution`, a member of the C++11 `<random>` suite, which obtains its uniform variates from `std::generate_canonical` [172, 173]. Given our PRNG, these uniform variates are equivalent to calling Algorithm 1 with  $B = 32$  and  $P = 24$ . This creates a *partially* uneven sample  $\{U_P[0, 1)\}$ , with  $B - P = 8$  bits more entropy than even variates. GNU’s implementation feeds these uniform variates into  $Q_1$  (Eq. A.6), *but without removing its cancellation* by using `log1p`. As predicted by Equation A.15, the cancellation strips any extra entropy from the partially uneven variates ( $B > P$ ), converting them into even ones ( $B = P$ ).

Figure A.3 shows the bits of precision lost by three samples using the same PRNG seed, with the median at the center and increasingly improbable values near the edges — a format which becomes easier to understand by referring to the top axis, which shows the  $x = Q(u)$  sampled by the various  $u$ . Each data point calculates  $D_{\text{KL}}$  for a domain  $u \in [2^{-k}, 2^{-k+1})$ , with a sample size of  $N \approx 10^9$  for each point. The solid line is *not a fit*, but the loss of precision predicted by Equation A.14 (scaled by

$C(Q)$ , because precision is lost at a slower pace when  $Q'/Q < 1$ ). The dotted line is a 1-bit threshold.

GNU's implementation of `std::exponential_distribution` (▼) exhibits a clear and dramatic loss of precision as variates gets farther from the median (and more rare). This imprecision agrees exactly with the prediction of  $H_E(k)$  (Eq. A.14, solid line) — even uniform variates have limited information, and every time  $u$  becomes half as small (so that  $x$  is half as probable), one more bit of precision is lost. This loss of precision in the sample is clearly caused by using uniform variates, which will always happen if  $Q_1$  neglects to use `log1p` internally. Since both Python's `random.expovariate` [166] and Numpy's `numpy.random.exponential` [168] also commit this error, their samples are equally imprecise.

But GNU's `std::exponential_distribution` could have done better; it drew *partially uneven* uniform variates ( $B = 32$ ,  $P = 24$ ), then spoiled them via cancellation. Enabling `log1p` in  $Q_1$  and regenerating GNU's sample (▲) permits Figure A.3 to isolate the two sources of imprecision identified in Sections A.1.2 and A.1.3. (i) Using `log1p`,  $Q_1$  is allowed to be well-conditioned as  $u \rightarrow 0$ , so only the uniform variates themselves can degrade the small-value tail. Moving left from the median, the *partially uneven* variates maintain maximal precision until their 8-bit entropy buffer runs dry. (ii) Conversely,  $Q_1$  is intrinsically ill-conditioned for  $u > 1/2$  in the large-value tail, so the quality of the uniform variates is irrelevant; an ill-conditioned quantile function causes an immediate loss of precision.

PQRAND generates its sample (○) via our robust inversion method, feeding high-entropy, uneven uniform variates  $U_N(0, 1/2]$  into a quantile flip-flop which is always well-conditioned ( $Q_1$  samples  $x \leq \text{median}$  and  $Q_2$  samples  $x \geq \text{median}$ ). Switching to a quantile flip-flop for this final data series means that, to the right of the median, the small values shown on the bottom horizontal axis are now  $u$  instead of

$1 - u$ . The sample’s tails exhibit ideal performance, in stark contrast to the standard inversion method, and precision is lost only near the median, where the composite  $Q$  is a tad unstable ( $C(Q) \gtrsim 1$ , see Figure A.1b). That  $D_{\text{KL}} \approx 0$  everywhere, and never exceeds 1 bit, is clear evidence that our robust inversion method fulfills its existential purpose, delivering the best sample possible with floats of precision  $P$ . Furthermore, this massive boost in quality arrives at  $\sim 80/100\%$  the speed of GNU’s `std::exponential_distribution` for *binary32/64* ( $\sim 30/40$  ns per variate on an Intel i7 @ 2.9 GHz with GCC 6.3, optimization O2).

Similar samples for any rate  $\lambda$ , as well as many other distributions (uniform, normal, log-normal, Weibull, logistic, gamma) are available with PQRAND, a free C++ and Python package hosted on GitHub [174]. PQRAND uses optimized C++ to generate uneven uniform variates (see Appendix B.2), with Cython wrappers for fast scripting. Yet the usefulness of PQRAND is not restricted to the rarefied set of distributions with analytic quantile functions; PQRAND uses rejection sampling for its own normal and gamma distributions. Rejection sampling gives access to *any* distribution  $f(x)$ , provided that one can more easily sample from the proposal distribution  $g(x) \geq f(x)$ . Since the final sample  $\{f\}$  is merely a subset of the proposed sample  $\{g\}$ , a high-precision  $\{f\}$  requires a high-precision  $\{g\}$ , which can be obtained via our robust inversion method.

**A.1.5 Conclusion.** Using the exponential distribution as a case study, we find two general sources of imprecision when sampling a probability distribution  $f$  via the inversion method: (i) When  $f$  has two tails (two places where  $Q'/Q \gg 1$ ), its quantile function  $Q(u)$  becomes ill-conditioned as  $u \rightarrow 1$ . (ii) Drawing uniform random variates using the canonical algorithm (Algorithm 1) gives too finite a sample space, making  $Q(u)$  ill-conditioned as  $u \rightarrow 0$  (even if  $Q(u)$  has a good condition number there). Both problems can lose dozens of ULP of precision in a sample’s tails,



and they are especially problematic for simulations using single precision — in the worst case,  $\sim 0.5\%$  of variates will lose at least a third of their precision. This vulnerability is found in popular implementations of the inversion method (e.g., GNU’s implementation of C++11’s `<random>` suite [172], and the `python.random` [166] and `numpy.random` [168] modules for Python, and more).

This section introduces a *robust* inversion method which reconditions  $Q$  by combining (i) *uneven* uniform variates (Algorithm 2, see Appendix B.2) with (ii) a quantile flip-flop (a two- $Q$  composition split at the median). Our method produces the best sample from  $f$  possible with floats of precision  $P$ , and is significantly faster than schemes which “exactly” sample distributions to arbitrary precision [175–177]. The precision of a random sample is especially important for large, non-linear Monte Carlo simulations, which can draw *so many* numbers that they may be sensitive to this vulnerability. Since it is difficult to exhaustively validate large simulations — in this case, to prove that a loss of precision in the tails creates only negligible side effects — the best strategy is to use the most numerically stable components at every step in the simulation chain, provided they are not prohibitively slow. To this end, we have released PQRAND [174], a free C++ and Python implementation of our robust inversion method, which is 80–100% as fast as standard inversion sampling.

## A.2 Relativistic kinematics

When working with relativistic systems on a computer (i.e., boosting to and fro between the CM and lab frames) it is quite easy to burn through half the floating point precision in a single calculation (to go from 16 digits of precision to only 8). This is because relativistic quantities like  $\beta$  and  $\gamma$  compress certain kinematic regimes into very small ranges;  $\gamma \approx 1$  in the Newtonian limit and  $\beta \approx 1$  in the relativistic limit. This means that vastly different systems are described by nearby numbers, and the more nearby they become, the more we need to worry about digitization.

In this section we will look at expressions that repeatedly show up in relativistic calculations, and explore the best way to calculate them on a computer. Our primary tool will be a single floating point substitution — the only one that every scientist should commit to memory. But don't take my word for it; read Goldberg's excellent monograph [156].

**A.2.1 Cancellation.** Cancellation is the bane of floating point calculations; when you have only a limited number of digits to work with, you have to round your answers. One will therefore find that for some small floating point number  $\delta$ ,

$$1 - (1 - \delta) \neq \delta. \quad (\text{A.19})$$

Rounding  $(1 - \delta)$  to some number near 1 requires wasting digits storing leading 9s (e.g.,  $1 - \pi \times 10^{-8} = 0.9999999685840735$ ), and storing these 9s requires discarding information about  $\delta$ . This information cannot be recovered when we undo the subtraction in the next step (*canceling* out all those leading 9s we unnecessarily stored). A calculation as pointless as  $1 - (1 - \delta)$  is easy for a human to spot, but not necessarily for a computer, so we often have to hold their hands.

An extremely common expression with cancellation is  $1 - \sqrt{1 - x^2}$  for small  $x$ . Fortunately, some quick numerical surgery removes the cancellation entirely

$$1 - \sqrt{1 - x^2} = (1 - \sqrt{1 - x^2}) \frac{1 + \sqrt{1 - x^2}}{1 + \sqrt{1 - x^2}} = \frac{x^2}{1 + \sqrt{1 - x^2}}. \quad (\text{A.20})$$

This is an exact replacement which now utilizes the full floating precision over the entire domain  $x \in [0, 1]$ , whereas the original rounds to zero as soon  $x < \sqrt{\epsilon}$ . An immediate use case of this expression is the relativistic quantity

$$1 - \beta = 1 - \sqrt{1 - \gamma^{-2}} = \frac{1}{\gamma^2 + \gamma\sqrt{\gamma^2 - 1}}. \quad (\text{A.21})$$

This is just one example; the rest of this section finds the most stable way to calculate common relativistic quantities when only limited information is available about the

system.

**A.2.2 The most stable relativistic expressions.** Given limited information about a system, we often need to calculate other relativistic properties, and we should always choose the most stable way.

*Note: This section exists mainly for personal reference, so I stop re-deriving these.*

**A.2.2.1 From  $\gamma$ .** Knowledge of  $\gamma$  provides excellent description of ultra-relativistic systems, since there are nearly unlimited floats available to describe  $\gamma \gg 1$ . Conversely, there is poor information about classical systems, since there are limited floats near  $\gamma \approx 1$ . As such,  $\beta$ ,  $\beta\gamma$  and  $\gamma - 1$  lose precision when  $\gamma \rightarrow 1$ .

$$\beta = \sqrt{(1 - \gamma^{-1})(1 + \gamma^{-1})} \quad (\text{A.22})$$

$$\beta\gamma = \sqrt{(\gamma - 1)(\gamma + 1)} \quad (\text{A.23})$$

$$1 - \beta = 1 - \frac{\sqrt{(\gamma - 1)(\gamma + 1)}}{\gamma} = \frac{\gamma - \sqrt{(\gamma - 1)(\gamma + 1)}}{\gamma} \quad (\text{A.24})$$

$$= \frac{1}{\gamma(\gamma + \sqrt{(\gamma - 1)(\gamma + 1)})} \quad (\text{A.25})$$

**A.2.2.2 From  $\beta$ .** The speed  $\beta$  is complementary to  $\gamma$ ; it provides excellent description of classical systems (nearly unlimited floats to describe  $\beta \approx 0$ ), yet poor descriptions of ultra-relativistic systems (limited floats when  $\beta \rightarrow 1$ ). When  $\beta \rightarrow 1$ , the following expressions (and  $1 - \beta$ ) cannot be determined precisely.

$$\gamma = \sqrt{\frac{1}{1 - \beta^2}} \quad (\text{A.26})$$

$$\beta\gamma = \sqrt{\frac{\beta^2}{1 - \beta^2}} \quad (\text{A.27})$$

$$\gamma - 1 = \sqrt{\frac{1}{1 - \beta^2}} - 1 = \frac{1 - \sqrt{1 - \beta^2}}{\sqrt{1 - \beta^2}} = \frac{\beta^2}{\sqrt{1 - \beta^2}(1 + \sqrt{1 - \beta^2})} \quad (\text{A.28})$$

**A.2.2.3 From  $E$  and  $\vec{p}$ .** If energy and momentum are known, then it is straightforward to calculate  $\beta = \sqrt{p^2}/E$  and reuse the equations from Section A.2.2.2 (multiply-

ing top and bottom by  $E$  to reduce unnecessary FLOPS). As such, these expressions also lose precision when  $\beta \rightarrow 1$ .

$$\gamma = \frac{E}{m} = \frac{E}{\sqrt{E^2 - p^2}} \quad (\text{A.29})$$

$$\beta\gamma = \sqrt{\frac{p^2}{E^2 - p^2}} \quad (\text{A.30})$$

$$\gamma - 1 = \frac{p^2}{\sqrt{E^2 - p^2}(E + \sqrt{E^2 - p^2})} \quad (\text{A.31})$$

**A.2.2.4 From  $m$  and  $\vec{p}$ .** Mass and momentum provide the best description of a system, because it is possible to precisely define both small  $\beta$  and large  $\gamma$ . We can reuse the results of Section A.2.2.3 (replacing  $E \rightarrow \sqrt{p^2 + m^2}$  and  $\sqrt{E^2 - p^2} \rightarrow m$ ), and all expressions have full precision.

$$\beta = \sqrt{\frac{p^2}{p^2 + m^2}} \quad (\text{A.32})$$

$$\gamma = \frac{\sqrt{p^2 + m^2}}{m} \quad (\text{A.33})$$

$$\beta\gamma = \frac{\sqrt{p^2}}{m} \quad (\text{A.34})$$

$$\gamma - 1 = \frac{p^2}{m(\sqrt{p^2 + m^2} + m)} \quad (\text{A.35})$$

$$1 - \beta = \frac{\sqrt{p^2 + m^2} - \sqrt{p^2}}{\sqrt{p^2 + m^2}} = \frac{m^2}{\sqrt{p^2 + m^2}(\sqrt{p^2 + m^2} + \sqrt{p^2})} \quad (\text{A.36})$$

### A.3 A precise vector class

Given my investigation into unsafe operations in floating point arithmetic, I decided to write my own C++ vector classes, so that I could be assured that they internally respect floating point limitations (e.g., the options available with ROOT use `acos` to find interior angle). My `Vector2`, `Vector3`, and `Vector4` classes store 2, 3, and 4-vectors (respectively), and are highly modified extensions of classes originally written by Z. Sullivan. In addition to defining basic functionality (add, scale, dot,

cross, etc.), I needed tools to rotate 3-vectors and boost 4-vectors; this section presents the mathematical formalism I chose to use to accomplish those latter tasks.

*Note: The Rodrigues formula (both for rotations and boost) is shown for reference, whereas Section A.3.1.2 presents my own work.*

**A.3.1 Rotating vectors.** The most natural method to rotate 3-vectors is via  $3 \times 3$  square matrices. However, it can be cumbersome to implement these matrices for an arbitrary rotation, since they require hard-coding large trigonometric expressions. The size of these expressions lends itself to the possibility of floating point cancellation, and ensuring their numerical stability is a headache. However, one does not have to use matrices to define rotations; the Rodrigues formula rotates using only the dot and cross product.

**A.3.1.1 The Rodrigues rotation formula.** To implement an active, right-handed (RH) rotation of vector  $\vec{w}$  by some angle  $\psi$  about some axis  $\hat{x}$ , the victim is projected into two pieces: its longitudinal length and transverse vector

$$w_L = \vec{w} \cdot \hat{x}, \quad (\text{A.37})$$

$$\vec{w}_T = \vec{w} - w_L \hat{x}. \quad (\text{A.38})$$

We can independently calculate the transverse  $\vec{w}_T$  rotated by  $90^\circ$

$$\vec{w}_{T'} = \hat{x} \times \vec{w}. \quad (\text{A.39})$$

The two transverse  $\vec{w}$  form a basis for the rotation, and  $\hat{x}$  is unaltered by it; this allows us to calculate the rotated vector quite simply

$$\vec{w}' = R(\vec{w}) = w_L \hat{x} + \cos \psi \vec{w}_T + \sin \psi \vec{w}_{T'}. \quad (\text{A.40})$$

We can validate this scheme by showing it preserves the defining properties of

rotations. First, the angle to the axis is unaltered;

$$\begin{aligned} \vec{w}' \cdot \hat{x} &= (\vec{w} \cdot \hat{x}) \hat{x} \cdot \hat{x} + \cos \psi (\vec{w} \cdot \hat{x} - (\vec{w} \cdot \hat{x}) \hat{x} \cdot \hat{x}) + \sin \psi ((\hat{x} \times \vec{w}) \cdot \hat{x}) \\ &= \vec{w} \cdot \hat{x}. \end{aligned} \quad (\text{A.41})$$

(since  $\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b})$ ). Second, the vector's length is unaltered;

$$\begin{aligned} |\vec{w}'|^2 &= (\vec{w} \cdot \hat{x}) + (\sin^2 \psi + \cos^2 \psi)(|\vec{w}|^2 - (\vec{w} \cdot \hat{x})^2) \\ &= |\vec{w}|^2. \end{aligned} \quad (\text{A.42})$$

This result uses the identity  $|\vec{u} \times \vec{v}|^2 = |\vec{u}|^2 |\vec{v}|^2 - (\vec{u} \cdot \vec{v})^2$  and implicitly draws upon the mutual orthogonality of  $\vec{w}_L$ ,  $\vec{w}_T$  and  $\vec{w}_{T'}$ .

**A.3.1.2 The hidden degree for freedom when rotating  $\vec{u}$  to  $\vec{v}$ .** Instead of an axis  $\hat{x}$  and angle  $\psi$  as our input degrees of freedom, we may want the rotation that takes vector  $\vec{u} \rightarrow \vec{v}$ . We can reuse Equation A.40, and define:

$$\cos \psi = \hat{u} \cdot \hat{v}, \quad (\text{A.43})$$

$$\sin \psi = |\vec{u} \times \vec{v}|, \quad (\text{A.44})$$

$$\hat{x}_1 = \frac{\hat{u} \times \hat{v}}{|\hat{u} \times \hat{v}|} = \frac{\hat{u} \times \hat{v}}{\sin \psi}. \quad (\text{A.45})$$

However,  $\hat{x}_1$  is only one *possible* rotation axis which takes  $\vec{u} \rightarrow \vec{v}$ . Consider a rotation of  $\psi = \pi$  about the axis bisecting the two normalized vectors

$$\hat{x}_2 = \frac{\hat{u} + \hat{v}}{|\hat{u} + \hat{v}|} = \frac{\hat{u} + \hat{v}}{\sqrt{2(1 + \hat{u} \cdot \hat{v})}}. \quad (\text{A.46})$$

This rotation also clearly takes  $\vec{u} \rightarrow \vec{v}$ . In fact, any axis which satisfies

$$\hat{u} \cdot \hat{x} = \hat{v} \cdot \hat{x} \quad (\text{A.47})$$

defines a valid rotation (since  $\vec{u}$  and  $\vec{v}$  will have the same latitude relative to  $\hat{x}$ , and thus trace out the same circle during the rotation). The “rotation which takes  $\vec{u} \rightarrow \vec{v}$ ” is ambiguous!

This ambiguity stems from a hidden degree of freedom. We are free to choose a coordinate system where

$$\vec{u} = [0, 0, 1], \quad (\text{A.48})$$

$$\vec{v} = [\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta]. \quad (\text{A.49})$$

We can get from  $\vec{u}$  to  $\vec{v}$  in two steps: (i) a RH rotation about  $\hat{z}$  by angle  $\phi$  and (ii) a RH rotation about  $\hat{y}'$  (the new  $y$ -axis, after the first rotation) by angle  $\theta$ . This procedure uses only two of the three Euler angles required to cover  $\text{SO}(3)$  (the group of 3-dimensional rotations). To complete the coverage, we need a final RH rotation about  $\hat{z}''$  (the final  $z$ -axis, which in our case is the newly minted  $\vec{v}$ ).

By construction, this post-rotation about  $\vec{v}$  by angle  $\omega$  cannot alter  $\vec{v}$ , so it does not spoil the original purpose of this rotation (take  $\vec{u} \rightarrow \vec{v}$ ). Instead,  $\omega$  determines what happens to *every other* vector, and does so by selecting *one* axis  $\hat{x}$  from the set which map  $\vec{u} \rightarrow \vec{v}$ . If we can find this  $\hat{x}$ , we can describe the complete operation as a single rotation (instead of two sequential rotations). We have already found two valid  $\hat{x}$  (Eqs. A.45 and A.46), and they are fortuitously orthogonal, so we can use them to construct a basis that parameterizes all possible axes of rotation

$$\hat{x} = a \hat{x}_1 + b \hat{x}_2. \quad (\text{A.50})$$

Our task is now clear; given  $\vec{u}$ ,  $\vec{v}$  and  $\omega$ , determine  $a$  and  $b$  to find  $\hat{x}$ .

$R_1$  is the rotation about  $\hat{x}_1$  by  $\theta = \arccos(\hat{u} \cdot \hat{v})$  and  $R_2$  is the rotation about  $\hat{x}_2$  by  $\omega$ . Applying them consecutively produces a composite rotation which cannot alter the axis of rotation;

$$\hat{x} = R_2(R_1(\hat{x})). \quad (\text{A.51})$$

Since this defining property applies to both of  $\hat{x}$ 's component individually (and linearly), we can solve for  $a$  and  $b$  by using unitarity as one equation (i.e.,  $a^2 + b^2 = 1$ ),

then obtain the other equation by calculating

$$\begin{aligned} a &= R_2(R_1(a \hat{x}_1 + b \hat{x}_2)) \cdot \hat{x}_1 \\ &= a R_2(R_1(\hat{x}_1)) \cdot \hat{x}_1 + b R_2(R_1(\hat{x}_2)) \cdot \hat{x}_1. \end{aligned} \quad (\text{A.52})$$

Beginning with  $\hat{x}_1$ , we are lucky that  $R_1$  does not alter its own axis, while  $R_2$  creates only one term parallel to  $\hat{x}_1$ ;

$$R_1(\hat{x}_1) = \hat{x}_1 ; \quad (\text{A.53})$$

$$R_2(R_1(\hat{x}_1)) \cdot \hat{x}_1 = ((\hat{x}_1 \cdot \hat{v})\hat{v} + \cos \omega(\hat{x}_1 - 0) + \sin \omega \underbrace{(\hat{v} \times \hat{x}_1)}_{\perp \text{ to } \hat{x}_1}) \cdot \hat{x}_1 = \cos \omega. \quad (\text{A.54})$$

The effect on  $\hat{x}_2$  is slightly more complicated;

$$\begin{aligned} R_1(\hat{x}_2) &= ((\hat{x}_2 \cdot \hat{x}_1)\hat{x}_1 + \cos \theta(\hat{x}_2 - 0) + \sin \theta(\hat{x}_1 \times \hat{x}_2)) \\ &= \frac{1}{\sqrt{2(1 + \cos(\theta))}} \left( \cos \theta (\hat{u} + \hat{v}) + \sin \theta \frac{(\hat{u} \times \hat{v})}{\sin \theta} \times (\hat{u} + \hat{v}) \right) \\ &= \frac{1}{\sqrt{2(1 + \cos(\theta))}} (\cos \theta (\hat{u} + \hat{v}) + \hat{v} - \hat{u} \cos \theta + \hat{v} \cos \theta - \hat{u}) \\ &= \frac{1}{\sqrt{2(1 + \cos(\theta))}} ((2 \cos \theta + 1)\hat{v} - \hat{u}), \end{aligned} \quad (\text{A.55})$$

using  $(\vec{a} \times \vec{b}) \times \vec{c} = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{a}(\vec{b} \cdot \vec{c})$ . Before we put all of Equation A.55 through  $R_2$ , we should recall that the final equation uses only terms parallel to  $\hat{x}_1$ .  $R_2$  returns only terms which are (i) parallel to the incoming vector (which in this case is in the  $uv$ -plane, and thus orthogonal to  $\hat{x}_1$ ), (ii) parallel to  $\hat{v}$  (also in the  $uv$  plane) and (iii) perpendicular to  $\hat{v}$  (via  $\hat{v} \times \vec{w}$ ). Hence, only the 3<sup>rd</sup> piece is meaningful, and since  $\hat{v} \times \hat{v} = 0$ , we only need to give it Equation A.55's  $\hat{u}$  term;

$$\begin{aligned} R_2(R_1(\hat{x}_2)) \cdot \hat{x}_1 &= \frac{1}{\sqrt{2(1 + \cos(\theta))}} R_2(-\hat{u}) \cdot \hat{x}_1 = \frac{1}{\sqrt{2(1 + \cos(\theta))}} (\hat{v} \times (-\hat{u})) \cdot \hat{x}_1 \\ &= \sin \omega \frac{\sin \theta}{\sqrt{2(1 + \cos(\theta))}}, \end{aligned} \quad (\text{A.56})$$



using  $\hat{u} \times \hat{v} = \sin \theta \hat{x}_1$ . Combining Equation A.52, A.53 and A.56 we get our system of equations

$$a = a \cos \omega + b \sin \omega \frac{\sin \theta}{\sqrt{2(1 + \cos(\theta))}}, \quad (\text{A.57})$$

$$1 = a^2 + b^2. \quad (\text{A.58})$$

Having done the hard work, we can plug our system of equations into Mathematica to obtain

$$a = 2 \cos(\omega/2) \frac{\sin(\theta/2)}{c(\theta)}, \quad (\text{A.59})$$

$$b = 2 \sin(\omega/2) \frac{1}{c(\theta)}, \quad (\text{A.60})$$

$$c(\theta) = \sqrt{3 - \cos(\omega) - \cos(\theta)(1 + \cos(\omega))}. \quad (\text{A.61})$$

I then used Mathematica to validate this solution by checking that  $\hat{x}$  matches the eigenvector of the composite rotation  $R_2(R_1(\vec{v}))$  (up to the parity operation  $\hat{x} \rightarrow -\hat{x}$ , an irreducibly ambiguity of the eigenvector).

However,  $c(\theta)$  is numerically unstable if used naïvely, due to the cosine cancellations. These terms should be rewritten in a form which *doubles* the precision of the floating point result

$$1 - \cos(t) \mapsto 2 \sin^2(t/2). \quad (\text{A.62})$$

This gives us

$$c(\theta) \mapsto \sqrt{2 \sin^2(\theta/2) + 2 \sin^2(\omega/2) + (1 - \cos(\theta) \cos(\omega))}. \quad (\text{A.63})$$

The final cancellation can be corrected using

$$\cos(\theta) \cos(\omega) = \frac{1}{2}(\cos(\theta + \omega) + \cos(\theta - \omega)); \quad (\text{A.64})$$

$$(1 - \cos(\theta) \cos(\omega)) \mapsto \frac{1}{2}(1 - \cos(\theta + \omega)) + \frac{1}{2}(1 - \cos(\theta - \omega)). \quad (\text{A.65})$$

Again using Equation A.62, we obtain the final expression

$$c(\theta) \mapsto \sqrt{2 \sin^2(\theta/2) + 2 \sin^2(\omega/2) + \sin^2(\theta + \omega) + \sin^2(\theta - \omega)}. \quad (\text{A.66})$$

Given  $\vec{u}$ ,  $\vec{v}$  and  $\omega$ , we now have the tools to find the axis  $\hat{x}$  about which the composite rotations occur.<sup>27</sup> But what is the angle  $\psi$  of rotation? We can determine  $\psi$  empirically by projecting  $\hat{u}$  and  $\hat{v}$  into the plane of rotation (e.g.  $\hat{u}_\perp = \hat{u} - (\hat{u} \cdot \hat{x})\hat{x}$ ). The rotation angle is then defined via

$$\psi = \text{atan2}(\sin \psi, \cos \psi) = \text{atan2}(\text{sign}(a) |\hat{u}_\perp \times \hat{v}_\perp|, \hat{u}_\perp \cdot \hat{v}_\perp). \quad (\text{A.67})$$

It is best to use  $\text{atan2}$  because it is more precise for angles near  $0$ ,  $\pi/2$  and  $\pi$ . Note that we have to inject the *sign* of  $a$  into  $\sin(\psi)$ , because when  $a < 0$ , the RH rotation becomes larger than  $\pi$ , so we must instead use a *negative* RH rotation.

I have tested an implementation of this algorithm and it works quite well (it is both length and angle preserving). I have additionally validated that rotating once about  $\hat{x}$  gives the same result as the two-step, composite rotation.

**A.3.2 Boosting between frames.** Since  $\text{SO}(3)$  is a sub-group of the Lorentz group, it is not surprising that there is a boost analog of the Rodrigues formula. A four momentum  $\mathbf{p}$  (with energy  $p^0$  and 3-momentum  $\vec{p}$ ) can be boosted by some speed  $\vec{\beta}$  by splitting  $\vec{p}$  into its longitudinal  $p_L$  and transverse momentum  $\vec{p}_T$  relative to  $\hat{\beta}$  (see Eqs. A.37 and A.38). Unlike the Rodrigues rotation formula, we do not need to find a second perpendicular axis, because the “rotation” is always between the longitudinal momentum and energy/time. This allows us to use the standard definition of the Lorentz boost matrix, but with a coordinate system defined by  $\hat{\beta}$ ;

---

<sup>27</sup>There is one class of system where  $\hat{x}$  remains ambiguous; when  $\vec{u}$  and  $\vec{v}$  are antiparallel,  $\hat{x}_1$  and  $\hat{x}_2$  are both null. If  $\vec{x}$  is supplied externally,  $\omega$  rotates it around the shared  $uv$  axis, but  $\vec{x}$  cannot be determined from  $\omega$  alone.

writing  $p^0$  and  $p_L$  as scalars, but keeping the transverse momentum  $\vec{p}_T$  as a vector

$$\mathbf{p} = \begin{pmatrix} \gamma & 0 & \beta\gamma \\ 0 & \mathbb{1} & 0 \\ \beta\gamma & 0 & \gamma \end{pmatrix} \begin{pmatrix} p^0 \\ \vec{p}_T \\ p_L \end{pmatrix} = \begin{pmatrix} \gamma p^0 + \beta\gamma p_L \\ \vec{p}_T \\ \beta\gamma p^0 + \gamma p_L \end{pmatrix} \quad (\text{A.68})$$

$$= \begin{pmatrix} p^0 \\ \vec{p}_T \\ p_L \end{pmatrix} + \begin{pmatrix} (\gamma - 1)p^0 + \beta\gamma p_L \\ \vec{0} \\ \beta\gamma p^0 + (\gamma - 1)p_L \end{pmatrix}. \quad (\text{A.69})$$

Hence, adding the correct four-momenta implements the desired boost;

$$\mathbf{p} = \mathbf{p} + \mathbf{b}, \quad (\text{A.70})$$

$$b^0 = (\gamma - 1)p^0 + \beta\gamma p_L, \quad (\text{A.71})$$

$$\vec{b} = (\beta\gamma p^0 + (\gamma - 1)p_L) \hat{\beta}. \quad (\text{A.72})$$

There are intrinsic floating point limitations to this boost scheme. When an ultra-relativistic particle  $\mathbf{p}$  is boosted back to its CM frame, machine  $\epsilon$  relative rounding errors when adding  $\mathbf{b}$  to  $\mathbf{p}$  can cause relative errors of  $\mathcal{O}(\gamma\epsilon)$  in  $p_{\text{cm}}^0$  (and absolute errors of the same order in  $\vec{p}_{\text{cm}}$ , which should always be a null vector in the CM frame).

#### A.4 Stable recursion

The on-axis coefficients  $\bar{h}_l$  for the Gaussian shape function (Eq. 4.63) were calculated in Section 4.4.2.1 via the following recursion;

$$\bar{h}_0 = 1, \quad (\text{A.73})$$

$$\bar{h}_1 = \frac{1}{\tanh(\lambda^{-2})} - \lambda^2, \quad (\text{A.74})$$

$$\bar{h}_{l+1} = -(2l + 1)\lambda^2 \bar{h}_l + \bar{h}_{l-1}. \quad (\text{A.75})$$

In Figure A.4 we plot  $\bar{h}_l$  for several choice of  $\lambda$ , which reveals an immediate problem with this recurrence relation;  $\bar{h}_l$  experience a near-exponential decay, but as it becomes

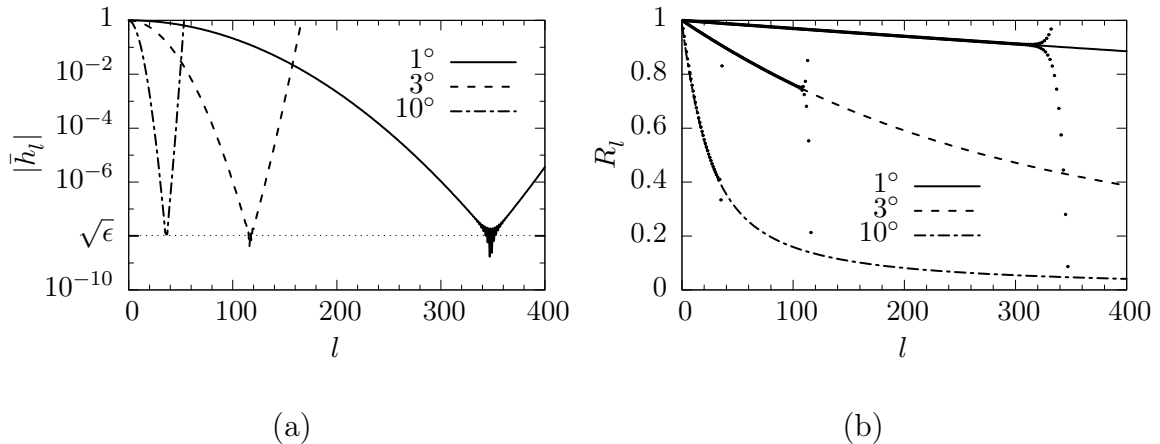


Figure A.4. Recursive  $\bar{h}_l$  for the Gaussian shape functions (Eq. 4.63). (a)  $\bar{h}_l$  for three smearing angles, connected by lines. (b)  $R_l$  (as discrete points) for the same angles, with the prediction of Equation A.81 plotted with lines.

*very small* (approaching the square root of machine epsilon  $\epsilon$ ), it suddenly becomes negative, then begins to *grow* exponentially (in the figure we plot  $|\bar{h}_l|$  so that all negative values can appear on a log-scale).

The fact that it becomes negative when  $\bar{h}_l \sim \sqrt{\epsilon}$ , and the large oscillation visible in the  $\lambda = 1^\circ$  transition, is highly suggestive of numerical instability. Looking at the recurrence relation (Eq. A.75), we can see that the only way that  $\bar{h}_{l+1}$  can go to zero is via floating point cancellation between the  $l$  and  $l - 1$  terms. Therefore, the recurrence loses precision when  $\bar{h}_l \rightarrow \sqrt{\epsilon}$ . Inevitably,  $\bar{h}_l$  will become accidentally negative, which causes an immediate positive feedback loop by giving  $\bar{h}_l$  and  $\bar{h}_l$  opposite signs, so that the terms add constructively. This oscillating sign pattern persists, causing  $|\bar{h}_{l+1}| > |\bar{h}_l|$ .

Now, one could correct this problem by simply truncating  $\bar{h}_l$  to zero when it begins to oscillate; it *is* admittedly small when the problem occurs. However, this will create side effects, and it is hard to categorically prove that they are negligible. Instead, we will find a solution which is only slightly more complicated to implement than naïve truncation.

We can study the instability via the ratio of adjacent  $\bar{h}_l$ ;

$$R_{l+1} \equiv \bar{h}_{l+1}/\bar{h}_l. \quad (\text{A.76})$$

Figure A.4b plots  $R_l$  (as dots) for the recursive  $\bar{h}_l$  data in Figure A.4a, and the onset of instability is quite evident. Note that  $R_l$  is not *nearly* as small as  $\bar{h}_l$ , so perhaps the recurrence relation for  $R_l$  will be more stable. We can define it by dividing Equation A.75 by  $\bar{h}_l$ ;

$$R_1 = \bar{h}_1, \quad (\text{A.77})$$

$$R_{l+1} = -(2l+1)\lambda^2 + \frac{1}{R_l}. \quad (\text{A.78})$$

Unfortunately, the recursive  $R_l$  is *less* stable than  $\bar{h}_l$ , so it is not immediately helpful.

But note that  $R_l$  does not change rapidly with  $l$ ; in fact,  $R_{l+1} \approx R_l$  is a good approximation for adjacent  $l$ . Replacing  $R_{l+1}$  with  $R_l$  in Equation A.78 gives

$$R_l = -(2l+1)\lambda^2 + \frac{1}{R_l}, \quad (\text{A.79})$$

which we can solve using the quadratic equation's positive solution

$$R_l = \frac{1}{2} \left( -(2l+1)\lambda^2 + \sqrt{(2l+1)^2\lambda^4 + 4} \right). \quad (\text{A.80})$$

This equation has a cancellation, and can be reformulated using Equation A.20

$$R(l) = \frac{2}{(2l+1)\lambda^2 + \sqrt{(2l+1)^2\lambda^4 + 4}}. \quad (\text{A.81})$$

Plotting this approximation through  $R_l$  in Figure A.4b, it works unusually well. It also manifestly defines a decaying  $\bar{h}_l$ , since  $0 \leq R(l) \leq 1$ . Hence, a stable  $\bar{h}_l$  is

$$\bar{h}_{l+1} = R(l+1)\bar{h}_l. \quad (\text{A.82})$$

However, since  $R(l)$  assumes that  $R_l$  is approximately flat, it is indeed an *approximation*; note the slight divergence from the  $\lambda = 10^\circ$  data points for  $l < 10$ . Hence, Equation A.81 should only be used once the  $\bar{h}_l$  recursion begins to fail, permitting  $\bar{h}_l$  to decay smoothly to zero.

APPENDIX B  
SMALL DETAILS

This appendix contains important details whose technical nature distracts from the narrative of the main text thus far.

**B.1 A finite, isotropic sample** In Section 4.3.3.1, we studied the power spectrum for random, isotropic particles. This section discusses how to create such a sample, and how to predict  $\text{Ex}(\tilde{f}^2)$  for the sample's scale-free energy fraction  $\tilde{f}$ .

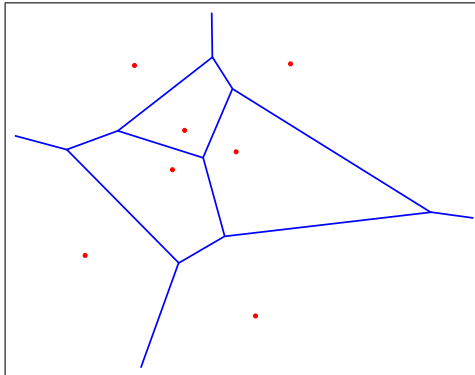


Figure B.1. A Voronoi diagram for a flat square filled with seven points [178].

Sampling  $N$  particles discretely from an isotropic, homogeneous  $\rho(\hat{r})$  breaks both isotropy and homogeneity. However, it is still possible to ensure that the total momentum is zero, so that isotropy is *mostly* preserved. A sample of  $N$  isotropic particles  $\vec{p}$  can be generated via the following algorithm:

1. Isotropically draw  $N$  unit vectors  $\hat{v}_i$ , which will define particle  $\vec{p}_i$ .
2. The energy of  $\vec{p}_i$  (its length) is proportional to the solid angle sampled by its random vector  $\vec{v}_i$ ;

$$f_i = \Omega_i / (4\pi). \quad (\text{B.1})$$

3. To determine  $\Omega_i$ , partition the unit sphere into Voronoi cells. The Voronoi cell for vector  $\hat{v}_i$  is the locus of points on the unit surface for which no other  $\vec{v}$  is closer. Alternatively, if one more vector is drawn and lands in the  $k$ th cell, then

vector  $\hat{v}_k$  will be the nearest neighbor of the new vector. A Voronoi diagram for a flat surface is depicted in Figure B.1; each point is surrounded by a cell where no other points are closer. Note that Voronoi cells are usually not symmetric.

4. Place the  $i$ th particle's direction of travel  $\hat{p}_i$  at the geometric center of the Voronoi cell for  $\hat{v}_i$ .

The simplest way to partition the unit surface into Voronoi cells is via Monte Carlo integration<sup>28</sup>:

1. Isotropically draw  $M \gg N$  unit vectors  $\hat{w}$ .
2. For each  $\vec{w}$ , find the closest  $\hat{v}_i$  (i.e., which Voronoi cell did it strike).
3. Each vector  $\hat{v}_i$  keeps two running sums:
  - (a) A count  $m_i$  of how many  $\hat{w}$  have fallen in its Voronoi cell. After all  $\vec{w}$  are drawn,  $\Omega_i/(4\pi) = m_i/M$ .
  - (b) The sum  $\vec{g}_i = \sum \hat{w}$  of all  $\hat{w}$  that struck the cell, which empirically finds the Voronoi cell's geometric center.

After all  $\vec{w}$  are drawn,

$$\vec{p}_i = \frac{m_i}{M} \frac{\vec{g}}{|\vec{g}|}. \quad (\text{B.2})$$

However, since Monte Carlo integration has errors proportional to  $1/\sqrt{M}$ , we will find that momentum is not totally conserved:

$$\vec{p}_{\text{tot}} = \sum_i \vec{p}_i \neq \vec{0}. \quad (\text{B.3})$$

---

<sup>28</sup>Free codes for Voronoi partitions exist, but they generally work on a flat, disconnected surface. A sphere is both curved and connected, creating topological difficulties that are easily solved by brute force Monte Carlo integration.



To place the system in its CM frame, the momentum imbalance is subtracted from each vector proportional to its share of the final state energy;

$$\vec{p}_i \mapsto \vec{p}_i - f_i \vec{p}_{\text{tot}}. \quad (\text{B.4})$$

Finally,  $f_i$  is recalculated for this new set of balanced particles.

**B.1.1 Predicting the energy fractions of isotropic particles.** The multiplicity plateau was defined by Equation 4.47, which predicted that its height is inversely proportional to particle multiplicity. This is a strong statement, and we should test it in fine detail. This requires calculating  $\text{Ex}(\tilde{f}^2)$  from the probability distribution of the scale-free energy fractions  $h(\tilde{f})$ . For isotropic particles sampled via the procedure of Section B.1, energy fraction is proportional to the solid angle of Voronoi cells, so we need to calculate the distribution of Voronoi areas.

Distributing  $N$  points uniformly on the surface of the unit sphere, what is the probability that a point lands within some solid angle  $\Omega$ ? For simplicity, we will define the dimensionless area  $A$  to be the normalized solid angle

$$A = \frac{\Omega}{4\pi}. \quad (\text{B.5})$$

The probability that  $A$  receives a point is then the complement of the probability that no points land in

$$\text{Pr} = 1 - (1 - A)^N = AN + \mathcal{O}(A^2). \quad (\text{B.6})$$

To define a Voronoi cell from  $A$ , it must contain only one point, and the extent of its cell's area is defined by the "ring" of points which surround it. To first approximation, a Voronoi cell can only occur when (i) *exactly one* point lands in  $A$ , (ii) at least one point lands in the infinitesimal annulus of area  $dA$  (which defines  $A$ ), and (iii) all the other points land beyond the annulus

$$d\text{Pr} = \binom{N}{1} A(1 - (1 - dA)^{N-1})(1 - A)^{\mathcal{O}(N)} = N(N - 1)A(1 - A)^{\mathcal{O}(N)}dA. \quad (\text{B.7})$$

Here, we use  $\mathcal{O}(N)$  in the exponent because we do not know *how many* points lie in the annulus (even though we only keep  $dA$  to first order), so we don't know many are left to fall beyond annulus. As  $dA \rightarrow 0$ , this exponent will approach  $N - 2$ .

Now we must acknowledge that  $A$  is not the area of the central Voronoi cell, because the points that define the annulus have their own Voronoi cells which steal from  $A$ . The simplest thing to do is to assume that, on average, the area of the central Voronoi cell is  $A/b$ , where  $b \geq 1$  is some universal constant that sums over all the irregular shapes that Voronoi cells take on the unit sphere. The energy fraction  $f$  of the isotropic particles is proportional to this Voronoi area,

$$f = \frac{A}{b}. \quad (\text{B.8})$$

We can now convert Equation B.7 from  $A$  to  $f$

$$d\text{Pr} = N(N - 1)b^2 f (1 - bf)^{\mathcal{O}(N)} df. \quad (\text{B.9})$$

But the probability distribution for  $f$  is not very useful (since  $\text{Ex}(f) = 1/N$  by construction) so we switch to the scale-free energy fraction  $\tilde{f} \equiv Nf$  (as defined by Eq. 4.45). Substitution  $\tilde{f}$  into our existing differential probability gives

$$d\text{Pr} = \frac{N(N - 1)}{N^2} b^2 \tilde{f} \left(1 - \frac{b\tilde{f}}{N}\right)^{\mathcal{O}(N)} d\tilde{f}. \quad (\text{B.10})$$

Taking the limit where  $N \rightarrow \infty$ , the difference between  $N - 1$  and  $N$  evaporates, and the product of  $N$  infinitesimal “rotations” becomes the exponential function;

$$d\text{Pr} = b^2 \tilde{f} \exp(-b\tilde{f}) d\tilde{f}. \quad (\text{B.11})$$

To determine the differential energy fraction for isotropic particles, we prescribed the exact conditions that define their Voronoi area. One of these conditions is that the particle landed in the area  $A$ . But this is the probability to strike *a particular*  $A$ ! We put our finger on  $A$  and said, “What is the chance a particle will fall

here?” We do not care *where* each particle falls, and it is possible to have more than one particle with the same Voronoi area. Thus, we must scale  $dPr$  by some weight  $w$  which counts how many particles with a given  $A$  we expect to find. The naïve guess notes that  $M = 1/A$  areas are required to cover the surface, so after sampling  $N$  particles (and using the gross assumption that no two particles land in the same area), we expect that  $w = NA$  of the  $M$  areas will be filled.<sup>29</sup>

This  $w$  is rather crude; a more versatile weight factor is the power law

$$w = (NA)^k = (b\tilde{f})^k. \quad (\text{B.12})$$

To understand why we would want to use such a weight factor, recall that we have yet to define  $b$  (the fraction of  $A$  which belongs to the central Voronoi cell). It is not easy to define  $b$  from first principles; it accounts for the asymmetric shape of Voronoi cells across the entire unit sphere, and thus should be kept an empirical constant — the “fudge factor” in our overly prescriptive model of Voronoi cell formation. But  $\text{Ex}(\tilde{f}) = 1$  by construction, and if  $b$  is the only free parameter in  $h(\tilde{f})$ , then  $b$  is constrained by this condition (and  $\text{Ex}(\tilde{f}^2)$  by proxy). Freeing our fudge factor requires building an additional degree of freedom into  $h(\tilde{f})$ . A power law is an appropriate choice for this emancipation, because it does not add an arbitrary scale to  $h(\tilde{f})$ .<sup>30</sup>

Appending the weight to Equation B.13, we have a prototype distribution  $h(\tilde{f})$

$$\frac{dPr}{d\tilde{f}} = b(b\tilde{f})^{k+1} \exp(-b\tilde{f}). \quad (\text{B.13})$$

This lacks a proper normalization, from the power law’s arbitrary  $k$ . A standard

---

<sup>29</sup>Note that  $NA$  is also the probability that at least one point lands in a *particular*  $A$  (Eq. B.6), but this is simply a numerical coincidence.

<sup>30</sup>Alternatively, I ran the simulation first and plotted the histogram, and found a rising power law at small  $\tilde{f}$  times an exponential decay. This guided the development of the theoretical model for Voronoi area — and since the power law had a non-integer exponent, I knew this was an appropriate solution.

integral identity

$$\int_0^\infty b (b x)^m \exp(-b x) dx = \Gamma(m + 1), \quad (\text{B.14})$$

allows Equation B.13 to be normalized to

$$\frac{d\text{Pr}}{d\tilde{f}} = \frac{b}{\Gamma(k + 2)} (b \tilde{f})^{k+1} \exp(-b \tilde{f}). \quad (\text{B.15})$$

Now we must constrain  $k$  from  $\text{Ex}(\tilde{f}) = 1$ . Reusing Equation B.14, we can write

$$1 = \int_0^\infty \frac{b}{\Gamma(k + 2)} (b \tilde{f})^{k+1} \exp(-b \tilde{f}) \tilde{f} d\tilde{f} \quad (\text{B.16})$$

$$= \frac{1}{b} \int_0^\infty \frac{b}{\Gamma(k + 2)} (b \tilde{f})^{k+2} \exp(-b \tilde{f}) d\tilde{f} = \frac{\Gamma(k + 3)}{b\Gamma(k + 2)} = \frac{(k + 2)}{b}, \quad (\text{B.17})$$

so that  $k = b - 2$ . This gives the final distribution for the relative energy fraction  $\tilde{f}$

$$h(\tilde{f}) = \frac{b^b}{\Gamma(b)} \tilde{f}^{b-1} \exp(-b \tilde{f}). \quad (\text{B.18})$$

Computing the second moment

$$\text{Ex}(\tilde{f}^2) = 1 + \frac{1}{b} \quad (\text{B.19})$$

and plugging it into Equation 4.47, we can predict that random, isotropic particles will have a multiplicity plateau at

$$\text{Ex}(\langle f | f \rangle) = \frac{1}{N} \left( 1 + \frac{1}{b} \right). \quad (\text{B.20})$$

To determine the value of  $b$ , we sampled  $N = 16,384$  isotropic particles via the algorithm described in Section B.1 and binned their  $\tilde{f}$  into a histogram. The resulting  $h(\tilde{f})$  is depicted in Figure B.2 using Poissonian error bars (i.e., the relative error in each bin is estimated to be  $1/\sqrt{\text{count}}$ ). This observed  $h(\tilde{f})$  was fit with Equation B.18 to determine the empirical Voronoi constant  $b = 3.592(43)$  (with the error in  $b$  determined via the asymptotic standard error of the fit). The corresponding prediction for the multiplicity plateau is  $H_l \sim 1.28/N$ .

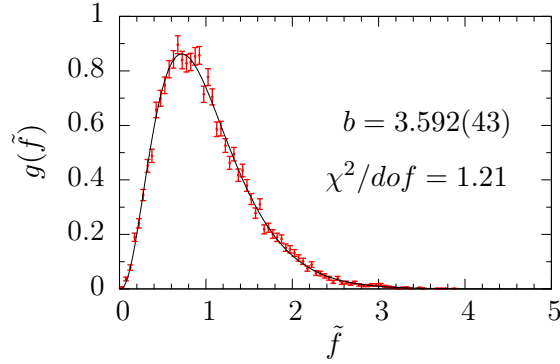


Figure B.2. The  $h(\tilde{f})$  distribution for an  $N = 16,384$  sample of isotropic Voronoi vectors. The shape is fit via Equation B.18 to determine the empirical Voronoi constant.

## B.2 Drawing uneven uniform variates

In Section A.1.3 we saw that the best uniform variates are uneven, obtained by taking  $B \rightarrow \infty$  in Algorithm 1. Since this will take an infinite amount of time, this section develops are more practical algorithm.

A clue to an alternate scheme lies in the bitwise representation of the *even* uniform variate from Algorithm 1, for which every  $u < 2^{-k}$  has a reduced entropy  $H_E = P - k$  (Eq. A.14). When  $B = P$ , Algorithm 1 draws an integer  $M$  from  $[0, 2^P)$ , then converts it to floating point. Inside the resulting float, the mantissa is stored as the integer  $M^*$ , which is just the original integer  $M$  with its bits shifted left until  $M^* \geq 2^{P-1}$ . This bit-shift ensures that any  $u < 2^{-k}$  always has at least  $k$  trailing zeroes in  $M^*$ ; zeroes which contain no information. Filling this always-zero hole with new random bits will restore maximal entropy.

Given the domain required by a quantile flip-flop, Algorithm 2 samples uneven  $\{U_N(0, 1/2]\}$  from the half-open, half-unit interval. It works by taking  $B \rightarrow \infty$ , yet knowing that floating point arithmetic will truncate the infinite bit-stream to  $P$  bits of precision. As soon as the RNG returns the first 1 (however many bits that takes), only the next  $P + 1$  bits are needed to convert to floating point;  $P$  bits to fill the

---

**Algorithm 2** Draw an *uneven* random float (with precision  $P$ ) from  $U(0, 1/2]$

---

**Require:**  $B \geq P$

```

1:  $n \leftarrow 1$                                 ▷ We return  $j/2^n$ . Starting at  $n = 1$  returns  $u \in (0, 1/2]$ .
2: repeat
3:    $j \leftarrow \text{RNG}(B)$                     ▷ Convert  $B$  random bits into an integer from  $U[0, 2^B)$ .
4:    $n \leftarrow n + B$ 
5: until  $j > 0$                                 ▷ Draw random bits until at least one is non-zero.
6: if  $j < 2^{P+1}$  then                          ▷ Require  $S \geq P + 2$  significant bits.
7:    $k \leftarrow 0$ 
8:   repeat
9:      $j \leftarrow 2j$ 
10:     $k \leftarrow k + 1$ 
11:   until  $j \geq 2^{P+1}$                         ▷ Shift  $j$ 's bits left until  $S = P + 2$ .
12:    $j \leftarrow j + \text{RNG}(k)$                 ▷ Fill the  $k$ -bit hole with fresh entropy.
13:    $n \leftarrow n + k$                         ▷  $u$ 's coarse location doesn't change.
14: end if
15: if  $j$  is even then  $j \leftarrow j + 1$       ▷ Make  $j$  odd to force proper rounding.
16: return  $\text{FLOAT}(j)/\text{FLOAT}(2^n)$             ▷ Round  $j$  to a float using R2N-T2E.

```

---

mantissa, and two extra bits for proper rounding. To fix  $u$ 's coarse location, the first loop (line 5) finds the first significant bit. The following conditional (line 6) requires  $S \geq P + 2$  significant bits. If  $S$  is too small,  $j$ 's bits are shifted left until the most significant (leftmost) bit slides into the  $P + 2$  position (line 11). Then the vacated space on the right is filled with new random bits, and the leftward shift is factored into  $n$ , so that only  $u$ 's fine location changes (enhancing precision while preserving uniformity). Finally, the integer is rounded into  $(0, 1/2]$ .<sup>31</sup>

Algorithm 2 needs two extra bits to maintain uniformity when  $j$  is converted to a float. With few exceptions, exact conversion of integers larger than  $2^P$  is not possible because the mantissa lacks the necessary precision. Truncation  $j$  won't work

---

<sup>31</sup>We exclude zero from the output domain of Algorithm 2 because, while theoretically possible, it will *never happen* (given a reliable RNG). Returning zero in *binary32* (single precision) would require drawing more than 150 all-zero bits in the first loop. Given a billion cores drawing  $B = 32$  every nanosecond, that would take  $\mathcal{O}(10^{55})$  years (although the first variate with sub-maximal entropy would only take  $\mathcal{O}(10^{41})$  years). For *binary64*, the numbers get ridiculous.

because  $j < 2^{n-1}$ , so Algorithm 2 would never return  $u = 1/2$ , a value needed by a quantile flip-flop to sample the exact median. Since Algorithm 2 must be able to round  $j$  up, it uses round-to-nearest, ties-to-even (R2N-T2E). Being the most numerically stable IEEE 754 rounding mode, R2N-T2E is the default choice for most operating systems.

Yet R2N-T2E is slightly problematic because Algorithm 2 is truncating a theoretically infinite bit stream to finite significance  $S$ . There are going to be rounding ties, and when T2E kicks in, it will pick *even* mantissae over odd ones, breaking uniformity. To defeat this bias,  $j$  is made odd. This creates a systematic tie-breaker, because an odd  $j$  is always closer to only one of the truncated options, without giving preference to the even option. This system only fails when  $S = P + 1$ , and only the final bit needs removal. In this case,  $j$  is equidistant from the two options, and T2E kicks in. Adding a random buffer bit (requiring  $S \geq P + 2$ ) precludes this failure.

An important property of Algorithm 2 is that  $u = 1/2$  is *half as probable* as its neighbor,  $u = \frac{1}{2}(1 - \epsilon)$ . Imagine dividing the domain  $[1/4, 1/2]$  into  $2^{P-1}$  bins, with the bin edges depicting the representable  $u$  in that domain. Uniformly filling the domain with  $\mathbb{R}$ , each  $u$  absorbs a full bin of real numbers via rounding (a half bin to its left, a half bin to its right). The only exception is  $u = 1/2$ , which can only absorb a half bin from the left, making it half as probable. But recall that  $\{U_N(0, 1/2)\}$  is intended for use in a quantile flip-flop — a regular quantile function folded in half at the median ( $u = 1/2$ ). Since *both*  $Q$  map to the median when they are fed  $u = 1/2$ , the median will be double-counted *unless*  $u = 1/2$  is half as probable.

Not only does Algorithm 2 produce significantly better uniform variates than `std::generate_canonical` (see Fig. A.3), it does so at equivalent computational speed ( $\sim 5$  ns per variate using MT19937 on an Intel i7 @ 2.9 GHz). This is possible because line 6 is rarely true ( $\sim 0.1\%$  when  $N = 64$  and  $P = 53$ ), so the code to top-up

entropy is rarely needed, and the main conditional branch is quite predictable. For most variates, the only extra overhead is verifying that  $S \geq P + 2$ , then making  $j$  odd, which take no time compared to the RNG and R2N-T2E operations.



## BIBLIOGRAPHY

- [1] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.
- [2] Morad Aaboud et al. Search for resonances in diphoton events at  $\sqrt{s}=13$  TeV with the ATLAS detector. *JHEP*, 09:001, 2016.
- [3] Vardan Khachatryan et al. Search for Resonant Production of High-Mass Photon Pairs in Proton-Proton Collisions at  $\sqrt{s} = 8$  and 13 TeV. *Phys. Rev. Lett.*, 117(5):051802, 2016.
- [4] Keith Pedersen and Zack Sullivan.  $\mu_x$  boosted-bottom-jet tagging and  $Z'$  boson searches. *Phys. Rev.*, D93(1):014014, 2016.
- [5] Keith Pedersen and Zack Sullivan. Probing the two Higgs doublet wedge region with charged Higgs boson decays to boosted jets. *Phys. Rev.*, D95(3):035037, 2017.
- [6] The ATLAS collaboration. Measurement of two-particle correlations in  $\sqrt{s} = 13$  TeV proton-proton collisions at the LHC with the ATLAS detector. ATLAS-CONF-2015-027, 2015.
- [7] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [8] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [9] P. Duinker. Review of Electron - Positron Physics at PETRA. *Rev. Mod. Phys.*, 54:325, 1982.
- [10] The ATLAS Collaboration. ATLAS event at 13 TeV - First stable beam, 3 June 2015 - run: 266904, evt: 25884805. General Photo, Jun 2015.
- [11] The CMS collaboration. Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill. <https://cds.cern.ch/record/2231915>.
- [12] Marco Grassi. *Measurement of the Standard Model Higgs Boson Couplings by Means of an Exclusive Analysis of its Diphoton Decay Channel*. PhD thesis, Rome U., 2012.
- [13] Abraham Seiden. Characteristics of the ATLAS and CMS detectors. *Phil. Trans. R. Soc. A*, 370:892–906, 2012.
- [14] Brian Dorney. Anatomy of a jet in CMS. <https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>.
- [15] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.
- [16] S. Agostinelli et al. Geant4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [17] John Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

- [18] J. Allison et al. Recent developments in Geant4. *Nucl. Instrum. Meth.*, A835:186–225, 2016.
- [19] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.
- [20] Keith Pedersen. Modified delphes 3. <https://github.com/keith-pedersen/delphes/tree/MuXboostedBTagging>.
- [21] Reinhard Beck and L Tiator. Subproject h2: Mesonic structure of the nucleon. 2018.
- [22] Benoit Beckers and Pierre Beckers. A general rule for disk and hemisphere partition into equal-area cells. *Comput. Geom. Theory Appl.*, 45(7):275–283, August 2012.
- [23] Edward Farhi. Quantum chromodynamics test for jets. *Phys. Rev. Lett.*, 39:1587–1588, Dec 1977.
- [24] R. Brandelik et al. Properties of Hadron Final States in  $e^+ e^-$  Annihilation at 13-GeV and 17-GeV Center-Of-Mass Energies. *Phys. Lett.*, B83:261–266, 1979.
- [25] Howard Georgi. A Simple Alternative to Jet-Clustering Algorithms. arXiv:1408.1161, 2014.
- [26] Yang Bai, Zhenyu Han, and Ran Lu.  $J_{E_T}^H$ : A Two-prong Jet Finding Algorithm. arXiv:1509.07522, 2015.
- [27] Nikolaos Kidonakis, Gianluca Oderda, and George F. Sterman. Threshold re-summation for dijet cross-sections. *Nucl. Phys.*, B525:299–332, 1998.
- [28] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007.
- [29] W. Bartel et al. Experimental Studies on Multi-Jet Production in  $e^+ e^-$  Annihilation at PETRA Energies. *Z. Phys.*, C33:23, 1986. [,53(1986)].
- [30] S. Catani, Yuri L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber. New clustering algorithm for multi-jet cross-sections in  $e^+ e^-$  annihilation. *Phys. Lett.*, B269:432–438, 1991.
- [31] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.
- [32] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.
- [33] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti- $k(t)$  jet clustering algorithm. *JHEP*, 04:063, 2008.
- [34] Matteo Cacciari and Gavin P. Salam. Dispelling the  $N^3$  myth for the  $k_t$  jet-finder. *Phys. Lett.*, B641:57–61, 2006.
- [35] R. D. Peccei and Helen R. Quinn. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.*, 38:1440–1443, 1977.

- [36] Guido Altarelli, B. Mele, and M. Ruiz-Altaba. Searching for New Heavy Vector Bosons in  $p\bar{p}$  Colliders. *Z. Phys.*, C45:109, 1989. [Erratum: *Z. Phys.*, C47:676, 1990].
- [37] Rabindra N. Mohapatra and Jogesh C. Pati. Left-Right Gauge Symmetry and an Isoconjugate Model of CP Violation. *Phys. Rev.*, D11:566–571, 1975.
- [38] R. N. Mohapatra and Jogesh C. Pati. A Natural Left-Right Symmetry. *Phys. Rev.*, D11:2558, 1975.
- [39] G. Senjanovic and Rabindra N. Mohapatra. Exact Left-Right Symmetry and Spontaneous Violation of Parity. *Phys. Rev.*, D12:1502, 1975.
- [40] Mirjam Cvetič and Jogesh C. Pati.  $N = 1$  Supergravity Within the Minimal Left-right Symmetric Model. *Phys. Lett.*, 135B:57–62, 1984.
- [41] JoAnne L. Hewett and Thomas G. Rizzo. Low-Energy Phenomenology of Superstring Inspired  $E(6)$  Models. *Phys. Rept.*, 183:193, 1989.
- [42] A. Leike. The Phenomenology of extra neutral gauge bosons. *Phys. Rept.*, 317:143–250, 1999.
- [43] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.
- [44] Lisa Randall and Raman Sundrum. An Alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, 1999.
- [45] R. S. Chivukula, Elizabeth H. Simmons, and J. Terning. Limits on noncommuting extended technicolor. *Phys. Rev.*, D53:5258–5267, 1996.
- [46] Bogdan A. Dobrescu and Felix Yu. Coupling-mass mapping of dijet peak searches. *Phys. Rev.*, D88(3):035021, 2013. [Erratum: *Phys. Rev.*, D90(7):079901, 2014].
- [47] Bogdan A. Dobrescu. Leptophobic Boson Signals with Leptons, Jets and Missing Energy. arXiv:1506.04435, 2015.
- [48] Georges Aad et al. Search for high-mass dilepton resonances in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Phys. Rev.*, D90(5):052005, 2014.
- [49] Vardan Khachatryan et al. Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *JHEP*, 04:025, 2015.
- [50] Serguei Chatrchyan et al. Searches for new physics using the  $t\bar{t}$  invariant mass distribution in pp collisions at  $\sqrt{s}=8$  TeV. *Phys. Rev. Lett.*, 111(21):211804, 2013. [Erratum: *Phys. Rev. Lett.*, 112(11):119903, 2014].
- [51] Georges Aad et al. A search for  $t\bar{t}$  resonances using lepton-plus-jets events in proton-proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *JHEP*, 08:148, 2015.
- [52] Vardan Khachatryan et al. Search for resonant  $t\bar{t}$  production in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Phys. Rev.*, D93(1):012001, 2016.

- [53] Georges Aad et al. Search for  $W' \rightarrow t\bar{b}$  in the lepton plus jets final state in proton-proton collisions at a centre-of-mass energy of  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Phys. Lett.*, B743:235–255, 2015.
- [54] Serguei Chatrchyan et al. Search for a  $W'$  boson decaying to a bottom quark and a top quark in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. *Phys. Lett.*, B718:1229–1251, 2013.
- [55] Zack Sullivan. Fully differential  $W'$  production and decay at next-to-leading order in QCD. *Phys. Rev.*, D66:075011, 2002.
- [56] Daniel Duffy and Zack Sullivan. Searching for  $W'$  bosons through decays to boosted-top and boosted-bottom jets. *Phys. Rev.*, D90(1):015031, 2014.
- [57] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie. Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks. *Phys. Rev. Lett.*, 101:142001, 2008.
- [58] The CMS collaboration. A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging. CMS-PAS-JME-09-001, 2009.
- [59] Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top Jets at the LHC. *Phys. Rev.*, D79:074012, 2009.
- [60] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011.
- [61] Christoph Anders, Catherine Bernaciak, Gregor Kasieczka, Tilman Plehn, and Torben Schell. Benchmarking an even better top tagger algorithm. *Phys. Rev.*, D89(7):074047, 2014.
- [62] The CMS Collaboration. Boosted Top Jet Tagging at CMS. CMS-PAS-JME-13-007, 2014.
- [63] Leandro G. Almeida, Mihailo Backović, Mathieu Cliche, Seung J. Lee, and Maxim Perelstein. Playing Tag with ANN: Boosted Top Identification with Pattern Recognition. *JHEP*, 07:086, 2015.
- [64] Gregor Kasieczka, Tilman Plehn, Torben Schell, Thomas Strebler, and Gavin P. Salam. Resonance Searches with an Updated Top Tagger. *JHEP*, 06:203, 2015.
- [65] Emanuele Usai. Boosted top: experimental tools overview. In *Proceedings, 7th International Workshop on Top Quark Physics (TOP2014): Cannes, France, September 28-October 3, 2014*, 2015.
- [66] The CMS Collaboration. Performance of b tagging at  $\sqrt{s} = 8$  TeV in multijet,  $t\bar{t}$  and boosted topology events. CMS-PAS-BTV-13-001, 2013.
- [67] The ATLAS collaboration. b-tagging in dense environments. ATL-PHYS-PUB-2014-014, Aug 2014.
- [68] The CMS collaboration. Algorithms for b jet identification in CMS. CMS-PAS-BTV-09-001, 2009.
- [69] Vardan Khachatryan et al. Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Phys. Rev.*, D91(5):052009, 2015.

- [70] Georges Aad et al. Search for  $W' \rightarrow tb \rightarrow qqbb$  decays in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Eur. Phys. J.*, C75(4):165, 2015.
- [71] The ATLAS collaboration. Calibrating the b-Tag Efficiency and Mistag Rate in  $35 \text{ pb}^{-1}$  of Data with the ATLAS Detector, 2011.
- [72] The ATLAS collaboration. Calibration of  $b$ -tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment. ATLAS-CONF-2014-004, 2014.
- [73] The ATLAS collaboration. Calibration of the performance of  $b$ -tagging for  $c$  and light-flavour jets in the 2012 ATLAS data. ATLAS-CONF-2014-046, 2014.
- [74] Matteo Cacciari. Fragmentation functions of light and heavy quarks. <https://www.physics.smu.edu/scalise/cteq/schools/summer98/cacciari.ps>, 1998.
- [75] The ATLAS collaboration. Commissioning of the ATLAS high-performance  $b$ -tagging algorithms in the 7 TeV collision data. ATLAS-CONF-2011-102, 2011.
- [76] Serguei Chatrchyan et al. Identification of  $b$ -quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
- [77] Georges Aad et al. A study of the material in the ATLAS inner detector using secondary hadronic interactions. *JINST*, 7:P01013, 2012.
- [78] Georges Aad et al. A neural network clustering algorithm for the ATLAS silicon pixel detector. *JINST*, 9:P09009, 2014.
- [79] H. Bachacou, J. F. Laporte, A. Ouraou, S. Hassani, and R. Nicolaidou. Soft muon tagging. ATL-PHYS-PUB-2009-021, ATL-COM-PHYS-2009-210, 2009.
- [80] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [81] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev.*, D93(3):033006, 2016.
- [82] The ATLAS collaboration. ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11. ATL-PHYS-PUB-2011-009, ATL-COM-PHYS-2011-744, 2011.
- [83] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.
- [84] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [85] G. Aad et al. (ATLAS Collaboration). Approved muon plots. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsMuon>.
- [86] Georges Aad et al. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data. *Eur. Phys. J.*, C74(11):3130, 2014.

- [87] Georges Aad et al. Performance of the ATLAS muon trigger in pp collisions at  $\sqrt{s} = 8$  TeV. *Eur. Phys. J.*, C75:120, 2015.
- [88] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Trecani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP*, 01:013, 2007.
- [89] Johan Alwall, Simon de Visscher, and Fabio Maltoni. QCD radiation in the production of heavy colored particles at the LHC. *JHEP*, 02:017, 2009.
- [90] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.
- [91] T. D. Lee. A Theory of Spontaneous T Violation. *Phys. Rev.*, D8:1226–1239, 1973.
- [92] Pierre Fayet. Supergauge Invariant Extension of the Higgs Mechanism and a Model for the electron and Its Neutrino. *Nucl. Phys.*, B90:104–124, 1975.
- [93] Pierre Fayet. Supersymmetry and Weak, Electromagnetic and Strong Interactions. *Phys. Lett.*, 64B:159, 1976.
- [94] Pierre Fayet. Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions. *Phys. Lett.*, 69B:489, 1977.
- [95] Savas Dimopoulos and Howard Georgi. Softly Broken Supersymmetry and SU(5). *Nucl. Phys.*, B193:150–162, 1981.
- [96] Howard E. Haber and Gordon L. Kane. The Search for Supersymmetry: Probing Physics Beyond the Standard Model. *Phys. Rept.*, 117:75–263, 1985.
- [97] Jihn E. Kim. Light Pseudoscalars, Particle Physics and Cosmology. *Phys. Rept.*, 150:1–177, 1987.
- [98] Neil Turok and John Zadrozny. Electroweak baryogenesis in the two doublet model. *Nucl. Phys.*, B358:471–493, 1991.
- [99] Michael Joyce, Tomislav Prokopec, and Neil Turok. Nonlocal electroweak baryogenesis. Part 2: The Classical regime. *Phys. Rev.*, D53:2958–2980, 1996.
- [100] Koichi Funakubo, Akira Kakuto, and Kazunori Takenaga. The Effective potential of electroweak theory with two massless Higgs doublets at finite temperature. *Prog. Theor. Phys.*, 91:341–352, 1994.
- [101] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012.
- [102] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at  $\sqrt{s} = 7$  and 8 TeV. *JHEP*, 08:045, 2016.
- [103] P. M. Ferreira, John F. Gunion, Howard E. Haber, and Rui Santos. Probing wrong-sign Yukawa couplings at the LHC and a future linear collider. *Phys. Rev.*, D89(11):115003, 2014.

- [104] Nathaniel Craig, Francesco D’Eramo, Patrick Draper, Scott Thomas, and Hao Zhang. The Hunt for the Rest of the Higgs Bosons. *JHEP*, 06:137, 2015.
- [105] John F. Gunion and Howard E. Haber. The CP conserving two Higgs doublet model: The Approach to the decoupling limit. *Phys. Rev.*, D67:075019, 2003.
- [106] Ilya F. Ginzburg and Maria Krawczyk. Symmetries of two Higgs doublet model and CP violation. *Phys. Rev.*, D72:115013, 2005.
- [107] P. S. Bhupal Dev and Apostolos Pilaftsis. Maximally Symmetric Two Higgs Doublet Model with Natural Standard Model Alignment. *JHEP*, 12:024, 2014. [Erratum: *JHEP*, 11:147, 2015].
- [108] Georges Aad et al. Search for charged Higgs bosons in the  $H^\pm \rightarrow tb$  decay channel in  $pp$  collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector. *JHEP*, 03:127, 2016.
- [109] D. Dicus, A. Stange, and S. Willenbrock. Higgs decay to top quarks at hadron colliders. *Phys. Lett.*, B333:126–131, 1994.
- [110] Eri Asakawa, Jun-ichi Kamoshita, Akio Sugamoto, and Isamu Watanabe. Production of scalar Higgs and pseudoscalar Higgs in multi-Higgs doublet models at gamma gamma colliders. *Eur. Phys. J.*, C14:335–345, 2000.
- [111] Vardan Khachatryan et al. Search for a charged Higgs boson in  $pp$  collisions at  $\sqrt{s} = 8$  TeV. *JHEP*, 11:018, 2015.
- [112] Jan Hajer, Ying-Ying Li, Tao Liu, and John F. H. Shiu. Heavy Higgs Bosons at 14 TeV and 100 TeV. *JHEP*, 11:124, 2015.
- [113] CEPC-SPPC Study Group. CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector. IHEP-CEPC-DR-2015-01, IHEP-TH-2015-01, IHEP-EP-2015-01, 2015.
- [114] Georges Aad et al. Search for  $t\bar{t}$  resonances in the lepton plus jets final state with ATLAS using  $4.7 \text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 7$  TeV. *Phys. Rev.*, D88(1):012004, 2013.
- [115] Neil D. Christensen and Claude Duhr. FeynRules — Feynman rules made easy. *Comput. Phys. Commun.*, 180:1614–1641, 2009.
- [116] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. FeynRules 2.0 — A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250–2300, 2014.
- [117] Celine Degrande. Automatic evaluation of UV and R2 terms for beyond the Standard Model Lagrangians: a proof-of-principle. *Comput. Phys. Commun.*, 197:239–262, 2015.
- [118] M. Herquet C. Duhr and C. Degrande. 2hdm.fr. <https://feynrules.irmp.ucl.ac.be/wiki/2HDM>.
- [119] The ATLAS collaboration. Expected performance of the ATLAS  $b$ -tagging algorithms in Run-2. ATL-PHYS-PUB-2015-022, Jul 2015.
- [120] Georges Aad et al. Identification of high transverse momentum top quarks in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *JHEP*, 06:093, 2016.

- [121] Marcela Carena, David Garcia, Ulrich Nierste, and Carlos E. M. Wagner. Effective Lagrangian for the  $\bar{t}bH^+$  interaction in the MSSM and charged Higgs phenomenology. *Nucl. Phys.*, B577:88–120, 2000.
- [122] Sally Dawson, Chung Kao, and Yili Wang. SUSY QCD Corrections to Higgs Pair Production from Bottom Quark Fusion. *Phys. Rev.*, D77:113005, 2008.
- [123] Wolfgang Altmannshofer, Joshua Eby, Stefania Gori, Matteo Lotito, Mario Martone, and Douglas Tuckler. Collider Signatures of Flavorful Higgs Bosons. *Phys. Rev.*, D94(11):115032, 2016.
- [124] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.
- [125] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008.
- [126] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008.
- [127] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010.
- [128] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev.*, D81:094023, 2010.
- [129] Ahmed Ali and Gustav Kramer. Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD. *Eur. Phys. J.*, H36:245–326, 2011.
- [130] G. Hanson et al. Evidence for Jet Structure in Hadron Production by  $e^+e^-$  Annihilation. *Phys. Rev. Lett.*, 35:1609–1612, 1975.
- [131] Martin White [on behalf of the Planck collaboration]. First cosmology results from Planck. [http://mwhite.berkeley.edu/Talks/White\\_Planck\\_LBNL13.pdf](http://mwhite.berkeley.edu/Talks/White_Planck_LBNL13.pdf), 2013.
- [132] Franz Zotter. Spherical harmonics up to degree 3, as used in third-order ambisonics. [https://commons.wikimedia.org/wiki/File:Spherical\\_Harmonics\\_deg3.png](https://commons.wikimedia.org/wiki/File:Spherical_Harmonics_deg3.png), 2013.
- [133] Geoffrey C. Fox and Stephen Wolfram. Observables for the analysis of event shapes in  $e^+e^-$  annihilation and other processes. *Phys. Rev. Lett.*, 41:1581–1585, Dec 1978.
- [134] G.C. Fox and S. Wolfram. Event shapes in  $e^+e^-$  annihilation. *Nuclear Physics B*, 149:413, October 1979.
- [135] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.
- [136] Daniel M. Kaplan. The Discovery of the Upsilon Family. *NATO Sci. Ser. B*, 352:359–380, 1996.



- [137] Christoph Berger et al. Observation of a Narrow Resonance Formed in  $e^+e^-$  Annihilation at 9.46-GeV. *Phys. Lett.*, 76B:243–245, 1978.
- [138] C. W. Darden et al. Observation of a Narrow Resonance at 9.46-GeV in electron-Positron Annihilations. *Phys. Lett.*, 76B:246, 1978.
- [139] Walter R. Innes et al. Observation of structure in the  $\Upsilon$  region. *Phys. Rev. Lett.*, 39:1240–1242, 1977. [Erratum: *Phys. Rev. Lett.*, 39:1640, 1977].
- [140] R. D. Field and R. P. Feynman. A Parametrization of the Properties of Quark Jets. *Nucl. Phys.*, B136:1, 1978.
- [141] B. Gittelman and P. Skubic. First Physics Results From the CLEO Detector at CESR. In *Proceedings, 15th Rencontres de Moriond*, pages 13–30, 1980.
- [142] A. Chen et al. Observation of the Decay  $\bar{B}^0 \rightarrow D^{*+}\rho^-$ . *Phys. Rev.*, D31:2386, 1985.
- [143] S. H. Lee et al. Evidence for  $B^0 \rightarrow \pi^0\pi^0$ . *Phys. Rev. Lett.*, 91:261801, 2003.
- [144] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J.*, C74:3026, 2014.
- [145] T. Horiguchi et al. Evidence for Isospin Violation and Measurement of  $CP$  Asymmetries in  $B \rightarrow K^*(892)\gamma$ . *Phys. Rev. Lett.*, 119(19):191802, 2017.
- [146] M. S. Alam et al. Shape studies of quark jets versus gluon jets at  $s^{*}(1.2) = 10$ -GeV. *Phys. Rev.*, D46:4822–4827, 1992.
- [147] R. Keith Ellis, W. James Stirling, and B. R. Webber. QCD and collider physics. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 8:1–435, 1996.
- [148] Georges Aad et al. Measurement of the charged-particle multiplicity inside jets from  $\sqrt{s} = 8$  TeV  $pp$  collisions with the ATLAS detector. *Eur. Phys. J.*, C76(6):322, 2016.
- [149] Georges Aad et al. Measurement of jet charge in dijet events from  $\sqrt{s}=8$  TeV  $pp$  collisions with the ATLAS detector. *Phys. Rev.*, D93(5):052003, 2016.
- [150] Krishnavedala. Plot of a dirac comb with period T. [https://commons.wikimedia.org/wiki/File:Dirac\\_comb.svg](https://commons.wikimedia.org/wiki/File:Dirac_comb.svg), 2012.
- [151] Keith Pedersen and Zack Sullivan. Spherical harmonics for multiparticle final states. In *Meeting of the APS Division of Particles and Fields (DPF 2017) Batavia, Illinois, USA, July 31-August 4, 2017*, 2017.
- [152] Serguei Chatrchyan et al. Shape, Transverse Size, and Charged Hadron Multiplicity of Jets in  $pp$  Collisions at 7 TeV. *JHEP*, 06:160, 2012.
- [153] Albert M Sirunyan et al. Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair. *Phys. Rev. Lett.*, 120(7):071802, 2018.
- [154] The ATLAS collaboration. Boosted Higgs ( $\rightarrow b\bar{b}$ ) Boson Identification with the ATLAS Detector at  $\sqrt{s} = 13$  TeV, 2016.

- [155] Bo Einarsson. *Accuracy and Reliability in Scientific Computing*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.
- [156] David Goldberg. What every computer scientist should know about floating-point arithmetic. *ACM Comput. Surv.*, 23(1):5–48, 1991.
- [157] Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986.
- [158] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C (2nd Ed.): The Art of Scientific Computing*. Cambridge University Press, NY, USA, 1992.
- [159] Donald E. Knuth. *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.
- [160] Pierre L'Ecuyer. Uniform random number generators: A review. In *Proceedings of the 29th Conference on Winter Simulation, WSC '97*, pages 127–134, Washington, DC, USA, 1997. IEEE Computer Society.
- [161] Pierre L'Ecuyer and Richard Simard. TestU01: A C library for empirical testing of random number generators. *ACM Trans. Math. Softw.*, 33(4):22:1–22:40, August 2007.
- [162] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, 1998.
- [163] György Steinbrecher and William T. Shaw. Quantile mechanics. *European Journal of Applied Mathematics*, 19(2):87–112, 2008.
- [164] ISO. *ISO/IEC 14882:2011 Information technology — Programming languages — C++*. International Organization for Standardization, Geneva, Switzerland, 2012.
- [165] D.P. Kroese, T. Taimre, and Z.I. Botev. *Handbook of Monte Carlo Methods*. Wiley, 2013.
- [166] Python Software Foundation. `random.py`. <https://github.com/python/cpython/blob/master/Lib/random.py>, 2017. [Online; accessed 6-JAN-2018].
- [167] Python Software Foundation. `_randommodule.c`. [https://github.com/python/cpython/blob/master/Modules/\\_randommodule.c](https://github.com/python/cpython/blob/master/Modules/_randommodule.c), 2017. [Online; accessed 14-FEB-2018].
- [168] NumPy Developers. `distributions.c`. <https://github.com/numpy/numpy/blob/master/numpy/random/mtrand/distributions.c>, 2017. [Online; accessed 6-JAN-2018].
- [169] NumPy Developers. `randomkit.c`. <https://github.com/numpy/numpy/blob/master/numpy/random/mtrand/randomkit.c>, 2017. [Online; accessed 14-FEB-2018].
- [170] Assaf Ben-David, Hao Liu, and Andrew D. Jackson. The Kullback-Leibler Divergence as an Estimator of the Statistical Properties of CMB Maps. *JCAP*, 1506(06):051, 2015.

- [171] IEEE Task P754. *IEEE 754-2008, Standard for Floating-Point Arithmetic*. IEEE, 2008.
- [172] Free Software Foundation. random.h. [https://gcc.gnu.org/onlinedocs/gcc-6.3.0/libstdc++/api/a01509\\_source.html](https://gcc.gnu.org/onlinedocs/gcc-6.3.0/libstdc++/api/a01509_source.html), 2017. [Online; accessed 11-DEC-2017].
- [173] Free Software Foundation. random.tcc. <https://gcc.gnu.org/onlinedocs/gcc-6.3.0/libstdc++/api/a01510.html>, 2017. [Online; accessed 11-DEC-2017].
- [174] Keith Pedersen. PQRAND. <https://github.com/keith-pedersen/pqRand>, 2018.
- [175] Donald E. Knuth and Andrew C. Yao. The Complexity of Nonuniform Random Number Generation. In J. F. Traub, editor, *Algorithms and Complexity: New Directions and Recent Results*. Academic Press, New York, 1976.
- [176] Philippe Flajolet, Maryse Pelletier, and Michèle Soria. On buffon machines and numbers. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '11*, pages 172–183, Philadelphia, PA, USA, 2011. Society for Industrial and Applied Mathematics.
- [177] Charles F. F. Karney. Sampling exactly from the normal distribution. *ACM Trans. Math. Softw.*, 42(1):3:1–3:14, January 2016.
- [178] Markus Matern. Voronoi-diagramm. [https://commons.wikimedia.org/wiki/File:Voronoi\\_diagram.svg](https://commons.wikimedia.org/wiki/File:Voronoi_diagram.svg), 2009.