GenEvA: A New Framework For Event Generation

Frank Tackmann

University of California, Berkeley
Lawrence Berkeley National Laboratory

ALCPG Workshop 2007
Fermilab, October 22–26, 2007

Work with Christian Bauer and Jesse Thaler

Outline

Motivation

Motivation

- Growing need and availability of precise theoretical calculations
  - more and more NLO or even NNLO results
  - Subleading log resummation, power corrections (e.g. in SCET)
- Make these available to experiments (more easily and more quickly)
  - Need a generic way to use *inclusive parton-level* calculations to get *exclusive hadron-level* events

Partonic Calculations vs. Algorithmic Tools

- Separate partonic calculations from algorithms
  - Calculations should not have to know specifics of implementations

GenEvA: Generate Events Analytically

Goals and Benefits of GenEvA

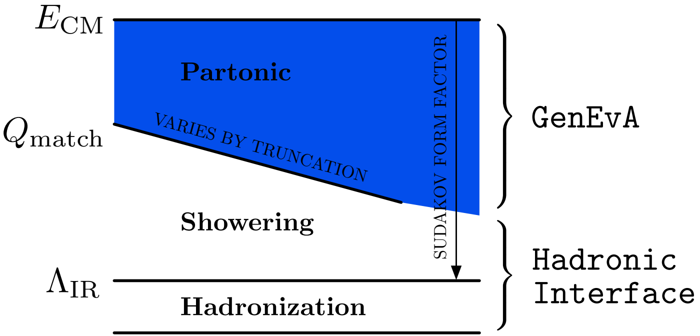
- Provide *generic* and *straightforward* framework to map inclusive parton-level calculations into exclusive hadron-level events
  - Let user think in terms of inclusive partonic cross sections
  - Avoid having to think about algorithms on case by case basis
- Use most accurate available prediction for each part of phase space

Idea behind

1. Use parton shower as phase-space generator to distribute points (events) in multiplicity, flavor and phase space
2. Determine *exact* probability $\mathrm{d}P$ for each point (event) to be generated
3. Reweight to *any desired distribution* $\mathrm{d}\sigma$ on event-by-event basis

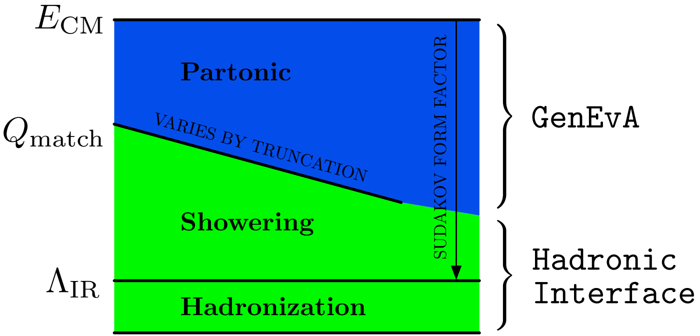
$$w = \mathrm{d}\sigma / \mathrm{d}P$$

Three Regimes of Event Generation



- Partonic Regime: Includes full field theory information
  - *Any* inclusive calculation with the notion of a matching scale

Three Regimes of Event Generation



- Partonic Regime: Includes full field theory information
  - *Any* inclusive calculation with the notion of a matching scale
- Showering Regime: Defined in terms of splitting probabilities
  - *Any* parton shower that can start showering at a given scale
- Hadronization: **Pythia**, **Herwig**

Separation between Partonic and Showering via matching scale Q_{match}

Q_{match} Dependence

Showering has logarithmic Q_{match} dependence

- Leading Q_{match} dependence is property of QCD not a specific shower
- Canceling Q_{match} dependence is theoretical not algorithmic issue
- Sufficient to include neccessary Q_{match} dependence in calculations

In **GenEvA**

- Q_{match} dependence in Partonic can be included
  - Analytically: using resummed calculation (NLO, subleading logs)
  - Numerically: on-the-fly via parton shower phase-space generation
- Value of Q_{match} is fluid and determined on event-by-event basis
  - Increase Q_{match}: Fewer final states in Partonic, more Showering
  - Decrease Q_{match}: More final states in Partonic, less Showering
- Smooth transition between Partonic and Showering

Toy Version of GenEvA

Toy cross sections

$$\frac{\mathrm{d}\sigma_B}{\mathrm{d}x} = B\delta(x)\,, \qquad \frac{\mathrm{d}\sigma_V}{\mathrm{d}x} = a\left(\frac{B}{\epsilon} + V\right)\delta(x)\,, \qquad \frac{\mathrm{d}\sigma_R}{\mathrm{d}x} = aB\frac{R(x)}{x^{1+\epsilon}}$$

$$\sigma_{\mathrm{total}} = B + aV + aB\int_0^1 \mathrm{d}x\,\frac{R(x)-1}{x}$$

Toy shower

Splitting function: $Q(x) \xrightarrow{x\to 0} R(x) \xrightarrow{x\to 0} 1$

Sudakov factor: $\Delta_Q(x_1, x_2) = \exp\left(-a\int_{x_2}^{x_1} \mathrm{d}x\,\frac{Q(x)}{x}\right)$

Splitting probability: $\mathrm{d}P(x; x_{\mathrm{start}}) = a\,\frac{Q(x)}{x}\,\Delta_Q(x_{\mathrm{start}}, x)\,\mathrm{d}x$

Master Formula

$$\sigma = \sigma_p^{(0)}(x_{\rm match}^{(0)})\,\mathrm{MC}(x_{\rm match}^{(0)}) + \int_{x_{\rm match}^{(0)}}^{1} \mathrm{d}x_1\,\mathrm{d}\sigma_p^{(1)}(x_1; x_{\rm match}^{(1)})\,\mathrm{MC}(x_{\rm match}^{(1)})$$

- $\sigma_p^{(0,1)}(x_{\rm match}^{(n)})$: Analytic partonic cross sections for 0 and 1 emissions
  - $x_{\rm match}$ dependence is primarily a *scale dependence*
  - Singularities cancel analytically
- $\mathrm{MC}(x_{\rm match})$: Showering starting at $x_{\rm match}$
- $x_{\rm match}^{(0,1)}$: Matching scale for 0- and 1-parton emission
  - $x_{\rm match}^{(1)} = x_1$ to avoid dead zone
  - $x_{\rm match}^{(0)}$ some low scale (could be as low as $x_{\rm cut}$)

Partonic Cross Sections at Tree Level

Tree-level generator

$$\sigma_p^{(0)}(x_{\text{match}}) = B$$

$$\sigma_p^{(1)}(x) = aB\,\frac{R(x)}{x}$$

Total and one-emission cross sections

$$\sigma = B + aB \int_{x_{\text{match}}}^{1} \mathrm{d}x\,\frac{R(x)}{x}$$

$$\sigma^{(1)}(x) = aB\,\frac{1}{x}\begin{cases} R(x) & (x > x_{\text{match}}) \\ Q(x)\,\Delta_Q(x_{\text{match}}, x) & (x < x_{\text{match}}) \end{cases}$$

- σ correct to $\mathcal{O}(1)$, and $\sigma^{(1)}(x)$ to $\mathcal{O}(a)$
- Logarithmic x_{match} dependence in σ and $\sigma^{(1)}(x)$

Partonic Cross Sections at Tree Level

Tree-level generator Sudakov improved

$$\sigma_p^{(0)}(x_{\mathrm{match}}) = B\,\Delta_Q(1, x_{\mathrm{match}})$$

$$\sigma_p^{(1)}(x) = aB\,\frac{R(x)}{x}\,\Delta_Q(1, x)$$

Total and one-emission cross sections

$$\sigma = B + aB \int_{x_{\mathrm{match}}}^{1} \mathrm{d}x\,\frac{R(x) - Q(x)}{x}\,\Delta_Q(1, x)$$

$$\sigma^{(1)}(x) = aB\,\frac{1}{x}\begin{cases} R(x)\,\Delta_Q(1, x) & (x > x_{\mathrm{match}}) \\ Q(x)\,\Delta_Q(1, x) & (x < x_{\mathrm{match}}) \end{cases}$$

- σ correct to $\mathcal{O}(1)$, and $\sigma^{(1)}(x)$ to $\mathcal{O}(a)$
- Logarithmic x_{match} dependence in σ and $\sigma^{(1)}(x)$ cancels
  - Example of ME/PS merging, but no special algorithm needed

Partonic Cross Sections at NLO

NLO Slicing

$$\sigma_p^{(0)}(x_{\mathrm{match}}) = \tilde{B}$$

$$\tilde{B} = \sigma_{\mathrm{total}} - aB \int_{x_{\mathrm{match}}}^{1} \mathrm{d}x\, \frac{R(x)}{x}$$

$$\sigma_p^{(1)}(x) = aB\, \frac{R(x)}{x}$$

Total and one-emission cross sections

$$\sigma = \sigma_{\mathrm{total}}$$

$$\sigma^{(1)}(x) = a\,\frac{1}{x} \begin{cases} BR(x) & (x > x_{\mathrm{match}}) \\ \tilde{B}Q(x)\,\Delta_Q(x_{\mathrm{match}}, x) & (x < x_{\mathrm{match}}) \end{cases}$$

- σ and $\sigma^{(1)}(x)$ correct to $\mathcal{O}(a)$, no x_{match} dependence in σ
- Two sources of Logarithmic x_{match} dependence in $\sigma^{(1)}(x)$

Partonic Cross Sections at NLO

NLO Slicing Sudakov improved = NLO Subtraction [MC@NLO]

$$\sigma_p^{(0)}(x_{\mathrm{match}}) = \tilde{B}\,\Delta_Q(1, x_{\mathrm{match}})$$

$$\tilde{B} = \sigma_{\mathrm{total}} - aB \int_{x_{\mathrm{match}}}^{1} \mathrm{d}x\, \frac{R(x) - Q(x)}{x}$$

$$\sigma_p^{(1)}(x) = aB\,\frac{R(x) - Q(x)}{x} + a\tilde{B}\,\frac{Q(x)}{x}\,\Delta_Q(1, x)$$

Total and one-emission cross sections

$$\sigma = \sigma_{\mathrm{total}}$$

$$\sigma^{(1)}(x) = a\,\frac{1}{x}\begin{cases} BR(x) + (\tilde{B} - B)Q(x)\,\Delta_Q(1, x) & (x > x_{\mathrm{match}}) \\ \tilde{B}Q(x)\,\Delta_Q(1, x) & (x < x_{\mathrm{match}}) \end{cases}$$

- σ and $\sigma^{(1)}(x)$ correct to $\mathcal{O}(a)$, no x_{match} dependence in σ
- Two sources of Logarithmic x_{match} dependence in $\sigma^{(1)}(x)$ cancel

Partonic Cross Sections at NLO

Simple NLO [POWHEG inspired]

$$\sigma_p^{(0)}(x_{\mathrm{match}}) = \sigma_{\mathrm{total}}\,\Delta_T(1, x_{\mathrm{match}})$$

$$\sigma_p^{(1)}(x) = a\,\sigma_{\mathrm{total}}\,\frac{T(x)}{x}\,\Delta_T(1, x)$$

$$T(x) = \frac{B}{\sigma_{\mathrm{total}}}\,R(x) \qquad \Delta_T = \exp\biggl(-a\int \mathrm{d}x\,\frac{T(x)}{x}\biggr)$$

Total and one-emission cross sections

$$\sigma = \sigma_{\mathrm{total}}$$

$$\sigma^{(1)}(x) = a\,\frac{1}{x}\begin{cases} B R(x)\,\Delta_T(1, x) & (x > x_{\mathrm{match}}) \\ \sigma_{\mathrm{total}}\,Q(x)\,\Delta_T(1, x_{\mathrm{match}})\Delta_Q(x_{\mathrm{match}}, x) & (x < x_{\mathrm{match}}) \end{cases}$$

- σ and $\sigma^{(1)}(x)$ correct to $\mathcal{O}(a)$
- Still no logarithmic x_{match} dependence in $\sigma^{(1)}(x)$

More Partonic Cross Sections

Everything boils down to calculating partonic cross sections.

Straightforward to generalize

- More emissions at tree level
- Combining NLO and additional tree level emissions, e.g.
  - 2+3 jet: analytic NLO matrix element
  - 4 jet: **MadGraph**
  - 5 jet: **AMEGIC++**
  - 6 jet: **ALPGEN**
  - 7 jet: **O'Mega**
  - 8+ jets: Parton Shower with quantum interference [Nagy, Soper]

In principle also straightforward, modulo technical details

- NNLO (+ additional tree level emissions)
- Resummed expressions (NLL, NNLL)

The Parton Shower as Phase Space Generator

Advantages

- Automatically covers all of multiplicity, flavor, phase space
- Automatically has the right singularity structure
- As a by-product, resums leading logarithms
- It's fast!

Challenges

- Need to know precise probability distribution $\mathrm{d}P$

  Use analytic parton shower algorithm [Christian Bauer, FT, arXiv:0705.1719]

  - Momentum conservation at each vertex
  - Full analytic control over $\mathrm{d}P$

- Must take into account that phase space is covered multiple times by parton shower

The Event Weight

$$w = \frac{\mathrm{d}\sigma}{\mathrm{d}P}$$

The Event Weight

$$w = \frac{\sigma(\Phi)\mathrm{d}\Phi}{\mathrm{d}P}$$

- $\mathrm{d}\sigma = \sigma(\Phi)\,\mathrm{d}\Phi$ is function of Lorentz invariant phase space Φ

The Event Weight

$$w = \frac{\sigma(\Phi)\mathrm{d}\Phi}{\mathcal{P}(\Sigma)\mathrm{d}\Sigma}$$

- $\mathrm{d}\sigma = \sigma(\Phi)\,\mathrm{d}\Phi$ is function of Lorentz invariant phase space Φ
- $\mathrm{d}P = \mathcal{P}(\Sigma)\,\mathrm{d}\Sigma$ is function of full parton shower history $\Sigma = \{t_i, z_i\}$

The Event Weight

$$w \equiv w(\Phi) = \frac{\sigma(\Phi)}{\mathcal{P}[\Sigma(\Phi)]J[\Sigma(\Phi)]}$$

- $\mathrm{d}\sigma = \sigma(\Phi)\,\mathrm{d}\Phi$ is function of Lorentz invariant phase space Φ
- $\mathrm{d}P = \mathcal{P}(\Sigma)\,\mathrm{d}\Sigma$ is function of full parton shower history $\Sigma = \{t_i, z_i\}$
- Mapping $\Sigma \to \Phi \equiv \Phi(\Sigma)$, Jacobian $J(\Sigma) = \mathrm{d}\Sigma/\mathrm{d}\Phi$

The Event Weight

$$w \equiv w(\Phi) = \frac{\sigma(\Phi)}{\sum_i \mathcal{P}[\Sigma_i(\Phi)] J[\Sigma_i(\Phi)]}$$

- $d\sigma = \sigma(\Phi)\, d\Phi$ is function of Lorentz invariant phase space Φ
- $dP = \mathcal{P}(\Sigma)\, d\Sigma$ is function of full parton shower history $\Sigma = \{t_i, z_i\}$
- Mapping $\Sigma \to \Phi \equiv \Phi(\Sigma)$, Jacobian $J(\Sigma) = d\Sigma/d\Phi$
- Each Φ can have multiple $\Sigma_i(\Phi)$ that map to it
  - Have to sum over all parton shower histories $\Sigma_i(\Phi)$ that map to the same point Φ in phase space.

Overcounting

$$w \equiv w(\Phi) = \frac{\sigma(\Phi)}{\sum_i \mathcal{P}[\Sigma_i(\Phi)] J[\Sigma_i(\Phi)]}$$

Summing over $\mathcal{P}[\Sigma_i(\Phi)]$ is hard

- Requires to explicitly construct all $\Sigma_i(\Phi)$ for given Φ
- Same problem in subtraction methods

Overcounting

$$w \equiv w(\Sigma) = \frac{\sigma(\Sigma)}{\mathcal{P}(\Sigma)}$$

Instead make weight function of Σ

Overcounting

$$w \equiv w(\Sigma) = \frac{\sigma[\Phi(\Sigma)]}{\mathcal{P}(\Sigma)J(\Sigma)}\,\hat{\alpha}(\Sigma)$$

Instead make weight function of Σ

- Define $\sigma(\Sigma) = \dfrac{\sigma[\Phi(\Sigma)]}{J(\Sigma)}\,\hat{\alpha}(\Sigma)$ with $\hat{\alpha}(\Sigma) = \dfrac{\alpha(\Sigma)}{\sum_i \alpha(\Sigma_i)}$
- $\hat{\alpha}(\Sigma)$ distributes $\sigma(\Phi)$ among $\Sigma_i(\Phi)$, can be chosen freely
  - Trivial but inefficient: $\alpha(\Sigma) = 1$
  - Ideal but hard: $\alpha(\Sigma) = \mathcal{P}(\Sigma)J(\Sigma)$
  - General: Pick $\alpha(\Sigma_i) \approx \mathcal{P}(\Sigma)J(\Sigma)$ and compute $\sum_i \alpha(\Sigma_i)$ using **ALPHA** algorithm [ALPGEN]

Proof of Concept Version

Already works

- $e^+e^- \to$ jets (jet $= u,\ d,\ s,\ c$)
- Use **MadGraph** matrix elements (currently up to 6 final states)
  - Pure matrix elements
  - Sudakov improved resummed matrix elements
- Showering regime covered by underlying analytic parton shower

On paper, but not in C++ yet

- NLO matrix elements
  - resummed NLO (one-loop 2 jet + tree-level 3 jet)
  - + additional resummed tree-level matrix elements

Comparison with MadGraph/MadEvent

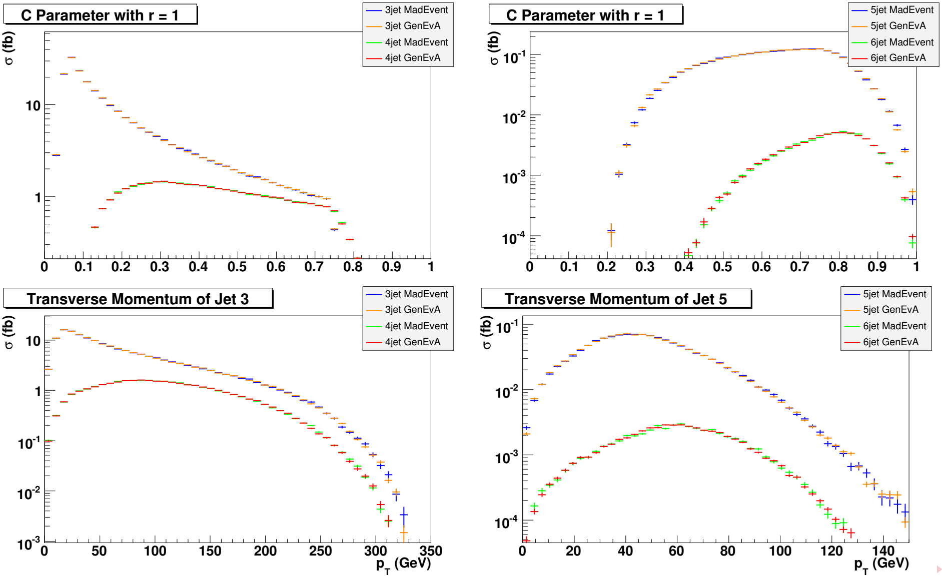
Reweight to pure **MadGraph** ME and compare with **MadEvent**

Cross section in ab for

- $E_{\mathrm{CM}} = 1000\,\mathrm{GeV}$
- $Q_{\mathrm{cut}} = 100\,\mathrm{GeV}$

process	MadEvent	GenEvA
$4j$	36483 ± 49	36439 ± 69
$u\bar{u}gg$	14055 ± 32	14003 ± 44
$d\bar{d}gg$	3490 ± 9	3498 ± 22
$u\bar{u}c\bar{c}$	283.4 ± 1.3	273 ± 7
$u\bar{u}d\bar{d}$	175.9 ± 0.9	184 ± 6
$u\bar{u}u\bar{u}$	131.9 ± 0.9	135 ± 4
$5j$	2540.5 ± 3.3	2550 ± 6
$u\bar{u}ggg$	909.8 ± 2.1	916 ± 3
$d\bar{d}ggg$	227.4 ± 1.0	229 ± 2
$u\bar{u}c\bar{c}g$	54.44 ± 0.31	54 ± 1
$u\bar{u}d\bar{d}g$	33.96 ± 0.31	35 ± 1
$u\bar{u}u\bar{u}g$	25.41 ± 0.16	25 ± 1

Comparison with MadGraph/MadEvent

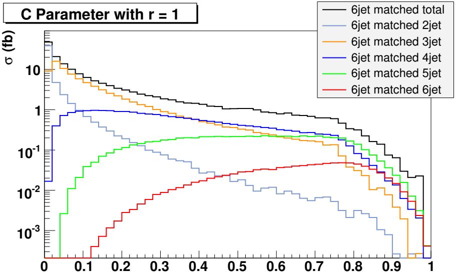


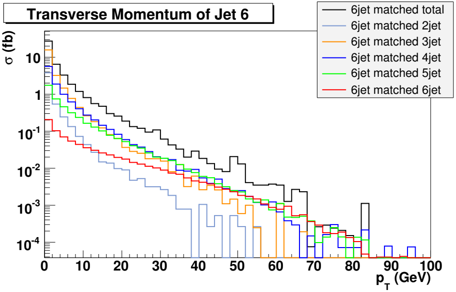
Fully Inclusive Sample

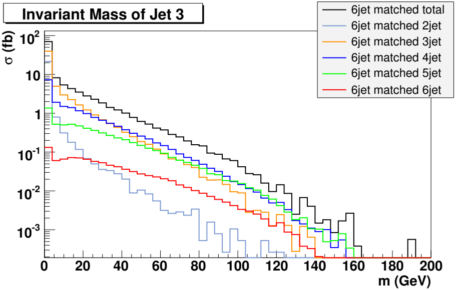
Match to MadGraph ME

Contributions of **2, 3, 4, 5, 6** jet matrix elements for

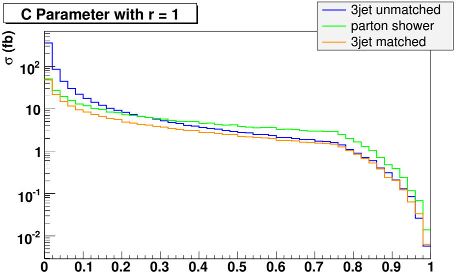
- $E_{\mathrm{CM}} = 1000\,\mathrm{GeV}$
- $Q_{\mathrm{match}} = 50\,\mathrm{GeV}$
- $Q_{\mathrm{cut}} = 10\,\mathrm{GeV}$

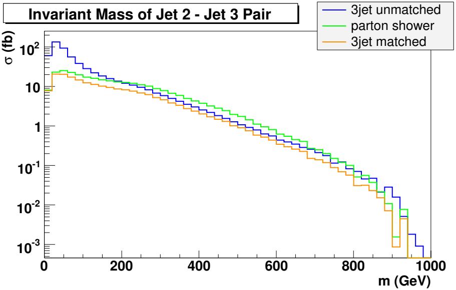




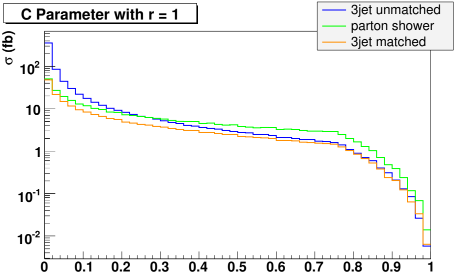


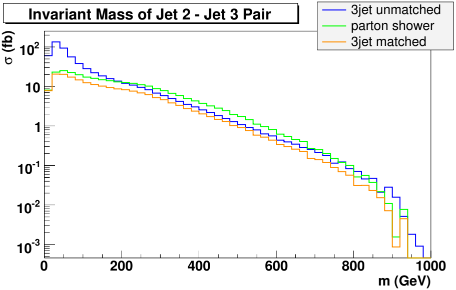
Matching Matrix Elements and Parton Shower

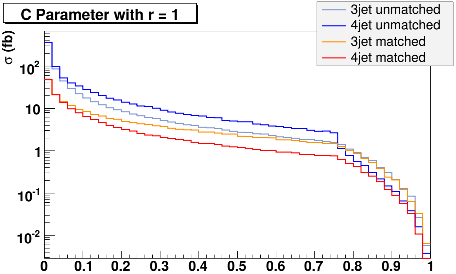


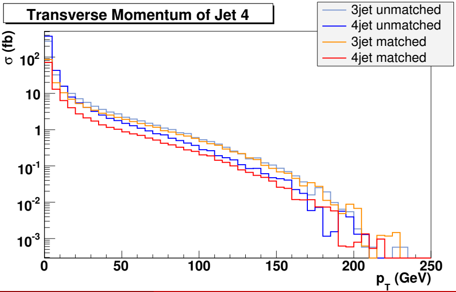


Matching Matrix Elements and Parton Shower







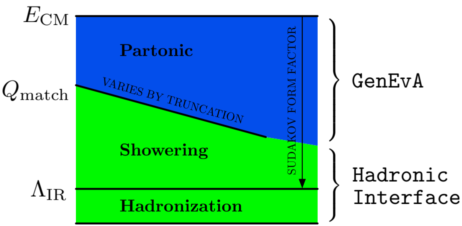


Conclusions

GenEvA is a *framework* to turn theory calculations into exclusive events

- Any matrix element with the notion of a matching scale
- Any parton shower that can start showering at a given scale
- Simple matching, independent of jet algorithm and parton shower
- Analytic parton shower can be used as efficient phase space generator

Need help from experts to extend philosophy to the full range of interesting processes (pp, $W/Z/t$ production, ...)



Outline

Backup Slides

Weighted vs. Unweighted samples

$$N_{\text{eff}} = \frac{\left[\sum_n w_n\right]^2}{\sum_n w_n^2} \qquad \text{vs.} \qquad N_{\text{unw}} = \frac{\sum_n w_n}{w_{\text{max}}}$$

- N_{unw}: expected number of events after unweighting
  - Poor efficiency from rare events with large weights
  - Systematic error due to unknown true w_{max} or truncation
- N_{eff}: measures statistical power of weighted sample
  - Number of events of a *statistically equivalent* unweighted sample
  - If N_{eff}/N too small, partially unweight until $N_{\text{eff}} \lesssim N$
- Always $N_{\text{eff}} > N_{\text{unw}}$, usually $N_{\text{eff}} \gg N_{\text{unw}}$

$e^+e^- \to 5j$	N_{eff}/N	N_{unw}/N	
MadEvent	0.89	0.11	(0.5 truncated)
GenEvA	0.40	0.03	
	0.89	0.12	(partially unweighted)

Master Formula Generalized

$$\sigma = \sigma_p^{(0)}(x_{\mathrm{match}}^{(0)})\,\mathrm{MC}(x_{\mathrm{match}}^{(0)}) + \int_{x_{\mathrm{match}}^{(0)}}^{1} \mathrm{d}x_1\, \mathrm{d}\sigma_p^{(1)}(x_1; x_{\mathrm{match}}^{(1)})\,\mathrm{MC}(x_{\mathrm{match}}^{(1)})$$
$$+ \int_{x_{\mathrm{match}}^{(0)}}^{1} \mathrm{d}x_1 \int_{x_{\mathrm{match}}^{(1)}}^{x_1} \mathrm{d}x_2\, \sigma_p^{(2)}(x_1, x_2; x_{\mathrm{match}}^{(2)})\,\mathrm{MC}(x_{\mathrm{match}}^{(2)}) + \cdots$$

- $\sigma_p^{(n)}(x_1, \ldots, x_n; x_{\mathrm{match}}^{(n)})$: Analytic partonic cross sections for n emissions
  - $x_{\mathrm{match}}^{(n)}$ dependence is primarily a *scale dependence*
  - Singularities cancel analytically
- $\mathrm{MC}(x_{\mathrm{match}})$: Showering starting at x_{match}
- $x_{\mathrm{match}}^{(n)} \equiv x_{\mathrm{match}}^{(n)}(x_n)$: Matching scale for n-parton emission
  - $n = n_{\max}$: $x_{\mathrm{match}}^{(n)} = x_n$ to avoid dead zone
  - $n < n_{\max}$: $x_{\mathrm{match}}^{(n)} < x_n$ arbitrary, could be as low as x_{cut}

Matching Dependence