Inferring nuclear modification of charm and beauty hadrons from simulated electron data

Unfolding non-photonic electron observables

Andrew M. Adare
University of Colorado
for the PHENIX collaboration

PHENIX

University of Colorado
Boulder

XXIV QUARK MATTER DARMSTADT 2014

Nuclear modification of heavy-flavor hadrons

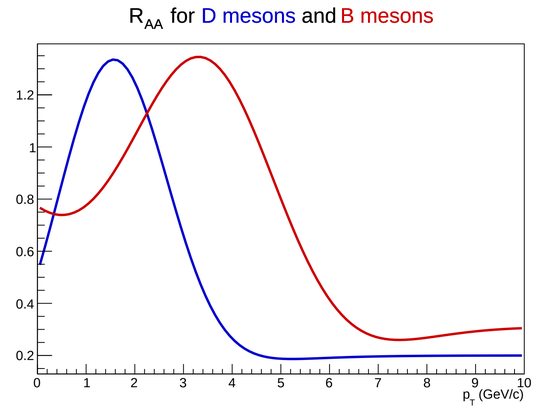
The goal of this poster is to demonstrate that it is possible to infer the modification of heavy-flavor hadrons from electron observables, namely (a) the displaced vertices of tracks identified as electrons and (b) electron invariant yields vs p_T.

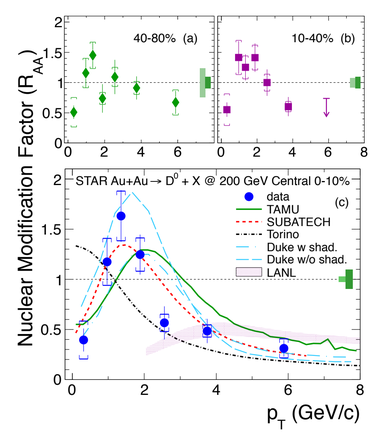
In this exercise, heavy-flavor hadron p_T spectra are generated by pythia, then modified by a blast-wave inspired estimate of R_{AA} [PLB 557 (2003) 26-32] (near right). Their decay electrons are used as the unfolding input "data".

This form for the modification has gained support from a recent result by the STAR collaboration [1404.6185] (far right).

The PYTHIA generator also provides the heavy-flavor decay kinematics, represented as matrices of decay probabilities. This information, along with the (modified) electron data, forms a linear system that can be solved using Bayesian inference.

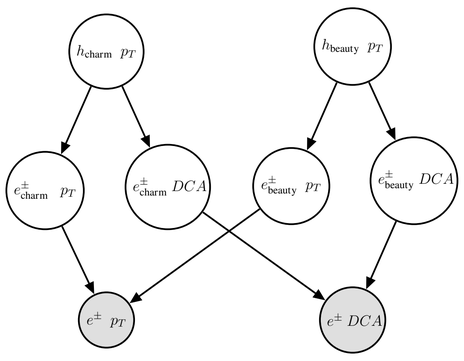
R_{AA} for D mesons and B mesons





A generative model for heavy-flavor decays

Each node in this directed graph represents a probability distribution. The arrows represent conditional dependence.



The objective is to infer the distributions over the latent variables $h_{\text{charm}}\ p_T$ and $h_{\text{beauty}}\ p_T$ from the (shaded) electron observables.

Bayesian unfolding using displaced vertices and spectra

See 1201.4612v4 by G. Choudalakis. His terminology is used here:

T truth vector (length N_t). $\widetilde{\mathbf{T}}$ is the modeled truth (e.g. from MC).
R reconstructed N_r-vector (again, e.g. in MC).
D measured data.

Bayes' theorem says $p(\mathbf{T}|\mathbf{D}) \propto L(\mathbf{D}|\mathbf{T}) \cdot \pi(\mathbf{T})$. In words:
the posterior probability $\propto$ the likelihood $\times$ the prior probability.

The problem amounts to assuming $\pi(\mathbf{T})$ and computing $L(\mathbf{D}|\mathbf{T})$:

$$L(\mathbf{D}|\mathbf{T}) = \prod_{r=1}^{N_r} \frac{R_r^{D_r}}{D_r!} e^{-R_r} \quad (for\ Poisson\ data) \tag{1}$$

The result is **not** a spectrum of points with covariance.
Instead, an N_t-dimensional posterior probability is obtained.
In each t bin, a 1D posterior is marginalized (integrated) from $p(\mathbf{T}|\mathbf{D})$.

Bayesian unfolding: implementation

First, the mapping of $\mathbf{T} \rightarrow \mathbf{R}$ is established, yielding a truth spectrum $\widetilde{\mathbf{T}}$ and a matrix M containing $P(r|t)$ values. $P(r|t)$ is the probability for an object from bin t to be reconstructed in bin r.

Also, select $\pi(\mathbf{T})$. Non-constant prior $\Rightarrow$ bias. Regularization!

Then:

1. A trial $\mathbf{T}$ point is pulled from an N_t-dimensional sampling volume
2. $\mathbf{R} = M\mathbf{T}$
3. $\pi(\mathbf{T})L(\mathbf{D}|\mathbf{T})$ (or, in practice, the log) is computed from $\pi(\mathbf{T})$, $\mathbf{R}$ and $\mathbf{D}$ (eq. 1)
4. $\mathbf{T}$ and $L(\mathbf{D}|\mathbf{T})$ are stored (TTree)
5. Repeat 1-4 until $p(\mathbf{T}|\mathbf{D})$ is well sampled
6. Marginalize: project $p(\mathbf{T}|\mathbf{D})$ to 1D posteriors $p_t(T_t|\mathbf{D})$

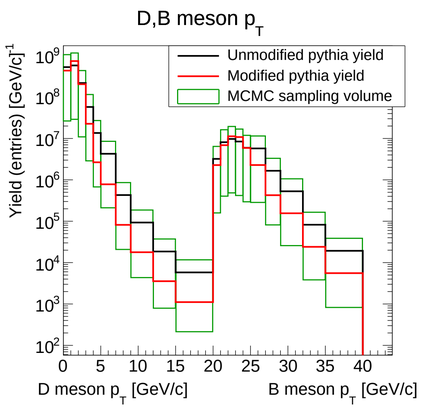
The trickiest part is step 1.

Sampling $L(\mathbf{D}|\mathbf{T})\pi(\mathbf{T})$ in N_t dimensions

The initial sampling volume must be large enough to enclose the "answer" conservatively.

On the other hand, the hyper-volume grows enormously as the boundaries are expanded. Grid sampling and uniform MC can quickly become prohibitive.

The solution is **Markov chain Monte Carlo** (MCMC). It is ergodic (it visits the whole space). Moreover, it samples in direct proportion to $p(\mathbf{T}|\mathbf{D})$!



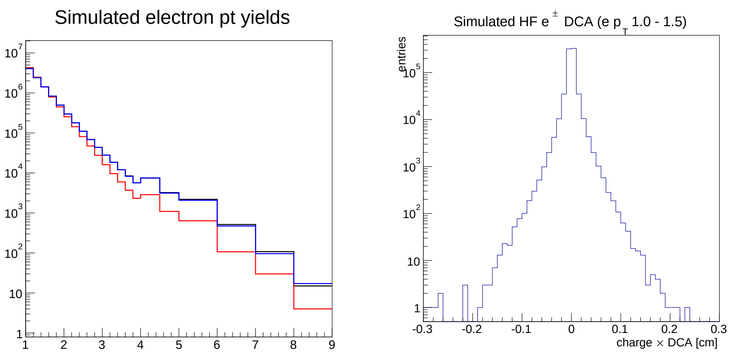
A **sketch** of the Metropolis-Hastings algorithm with a uniform $\pi(\mathbf{T})$:

- Start with $\mathbf{T}_0 = \widetilde{\mathbf{T}}$, and pick a large "hyperbox" around it. Save L_0.
- Propose a new point T_1 near T_0. Compute L_1.
- Is it better ($L_1 > L_0$)? Keep it, update $T_0 \leftarrow T_1$, and resample.
- If not, T_1 gets a second chance. Roll the dice again and accept it with a probability L_1/L_0.
- Repeat…

After equilibration, a Markov chain has randomly toured the whole box, but has climbed to the highest-likelihood regions most often.

Simulated PYTHIA data

The simulation setup includes 10 million electrons sampled from a distribution obtained from the PYTHIA generator. The samples are independent of those used to model the heavy-flavor decays (see below).



Blue/black: unmodified; Red: $\times R_{AA}$

Combining datasets (Electron spectra + DCA)

This problem is a simultaneous unfolding of displaced vertices and electron spectra. This involves computing the joint likelihood for each monte carlo sample and comparing it to data.

$$L(n|\mu) = \prod_{i=1}^{6} w_i \prod_{j=1}^{N_{DCA}} \text{Pois}(n_{i,j}^{DCA}|\mu_{i,j}^{DCA}) \times \prod_{k=1}^{N_{e^{\pm}p_T}} w_k \text{Pois}(n_k^{e^{\pm}p_T}|\mu_k^{e^{\pm}p_T}) \tag{2}$$

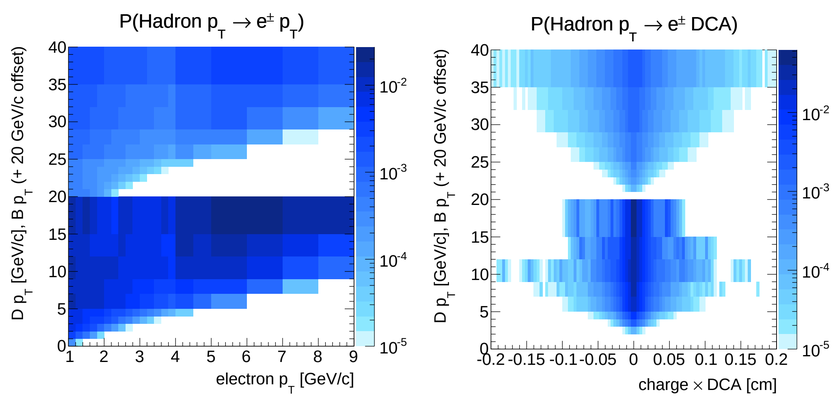
Where $w_i = 1/12$ and $w_k = 1/2$.

If the efficiencies of the DCA samples vs. $e^{\pm}\,p_T$ are unavailable, each trial μ can be scaled to match the integral of the DCA distribution.

This removes the dependence on $||\vec{\mu}||_2$, and only the shape of the DCA "guess" matters.

D,B $p_T \rightarrow e^{\pm}\,p_T$ and DCA matrices

These matrices represent the probability for a heavy-flavor hadron at a given p_T to decay to an electron at a given p_T and/or DCA.



x axis: $e^{\pm}\,p_T$ bin

y axis: Dp_T (0-10) and B p_T (10-20). D,B p_T both 0-10 GeV/c

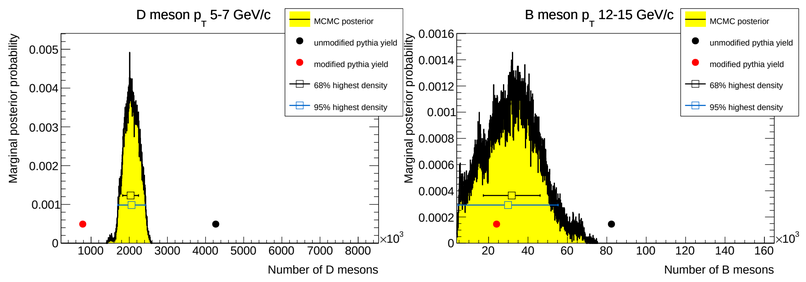
Regularization

Due to statistical fluctuations in the data and ill-conditioned transfer matrices, unbiased unfolding results typically exhibit large variances. The problem grows with model complexity (i.e. number of free parameters).

To deal with this, a prior distribution is included that penalizes results whose ratio to the initial guess has a large total curvature (second derivative). The regularization strength is an adjustable parameter, requiring careful study and transparent disclosure.

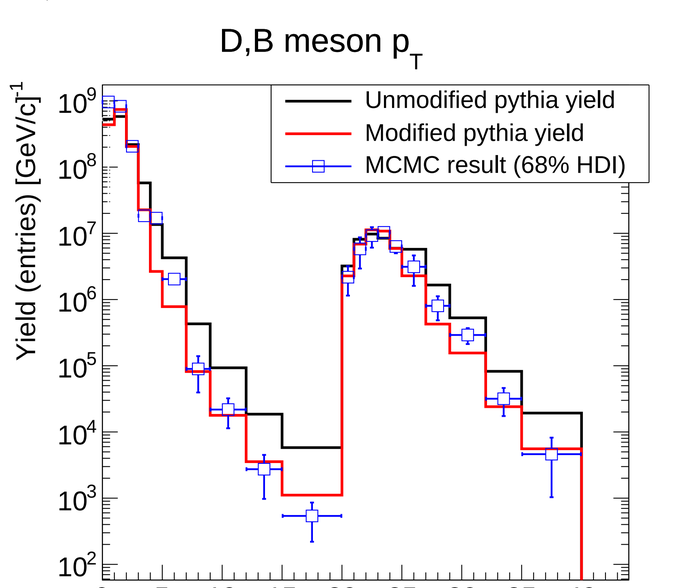
Unfolding results: examples

The output of the sampling algorithm is a 20-dimensional posterior probability distribution. Two marginal distributions from the joint posterior are shown here.



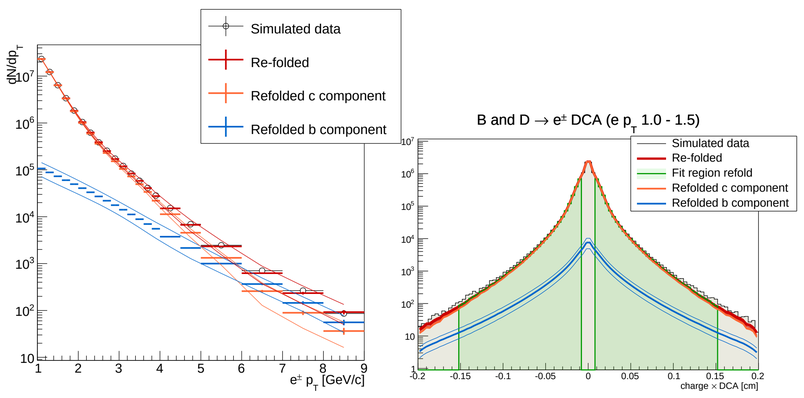
Unfolding results: summary

The shortest interval containing 68% of the samples is shown. This interval is used to summarize the distribution for each hadron p_T dimension, as shown here:



Re-folding the results

A necessary requirement is that the output from the unfolding calculation, when "refolded" (i.e. multiplied by the decay matrices), agree closely with the observed data.

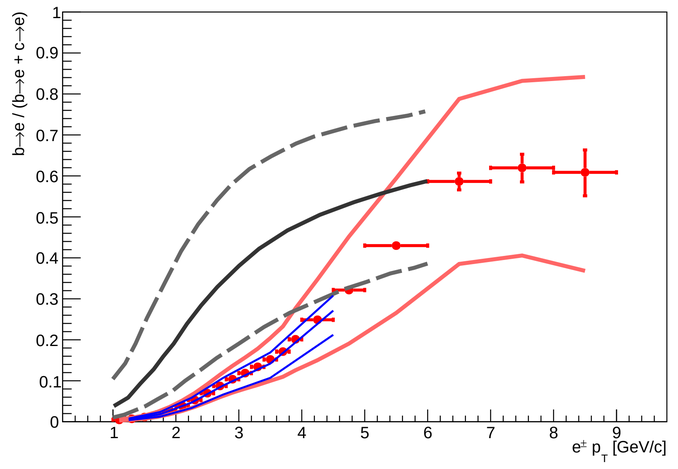


This should be the case for both datasets, and the result should be self-consistent. This self-consistency is shown below.

Beauty fraction vs. electron p_T

The red curve is from the re-folded electron p_T spectrum, and the blue curve is from the re-folded DCA distributions.

The black curve is the result of a FONLL calculation [PRC 84, 044905 (2011)].



Summary

We leave you with the following points:

- Even with no "knowledge" of R_{AA} in the unfolding matrices, the unfolding algorithm approximately recovers the correct modified hadron spectrum when given a noisy, modified electron dataset.
- The result is not perfect. Regularization is required to impose some degree of smoothness on the result. This is a bias introduced by the experimenter based on prior expectations.
- The agreement of a re-folded result with the input dataset is a necessary, but not sufficient requirement for an accurate result.

We're doing this with real data too! Keep an eye out for a PHENIX publication.

Mail: andrew.adare@colorado.edu