

France-Berkeley PHYSTAT Conference on Unfolding

Report of Contributions

Contribution ID: 1

Type: **not specified**

Welcome and Logistics

Monday, June 10, 2024 1:30 PM (15 minutes)

Presenters: BUTTER, Anja (Centre National de la Recherche Scientifique (FR)); NACHMAN, Ben (Lawrence Berkeley National Lab. (US)); BRENNER, Lydia (Nikhef National institute for subatomic physics (NL))

Contribution ID: 2

Type: **not specified**

HEP overview (30'+15')

Monday, June 10, 2024 2:30 PM (45 minutes)

Presenter: GRAS, Philippe (Université Paris-Saclay (FR))

Contribution ID: 3

Type: **not specified**

Statistics overview (30'+15')

Monday, June 10, 2024 1:45 PM (45 minutes)

Presenter: KUUSELA, Mikael (Carnegie Mellon University (US))

Contribution ID: 4

Type: **not specified**

ML overview (30'+15')

Monday, June 10, 2024 3:45 PM (45 minutes)

Presenter: PLEHN, Tilman (Heidelberg University)

Contribution ID: 5

Type: **not specified**

Binned ML methods overview (20'+20')

Tuesday, June 11, 2024 9:00 AM (40 minutes)

Presenter: PAN, Jingjing (Yale University (US))

Contribution ID: 6

Type: **not specified**

Unbinned Discriminative ML methods overview (20'+20')

Monday, June 10, 2024 4:30 PM (40 minutes)

Presenter: PETTEE, Mariel (Lawrence Berkeley National Lab. (US))

Contribution ID: 7

Type: **not specified**

Unbinned Generative ML methods overview (20'+20')

Monday, June 10, 2024 5:10 PM (40 minutes)

Presenter: HUETSCH, Nathan

Contribution ID: 8

Type: **not specified**

Performance / benchmarking with regularisation choice (20'+20')

Tuesday, June 11, 2024 9:40 AM (40 minutes)

Presenter: BRENNER, Lydia (Nikhef National institute for subatomic physics (NL))

Contribution ID: 9

Type: **not specified**

Unfolding in direct BSM search-sensitive regions of phase space (20'+20')

Tuesday, June 11, 2024 2:10 PM (40 minutes)

Presenter: WILLIAMS, Sarah Louise (University of Cambridge (GB))

Contribution ID: **10**

Type: **not specified**

Contributed talks

Contribution ID: **11**

Type: **not specified**

HEP summary

Contribution ID: 12

Type: **not specified**

A 24-dimensional Cross-section Measurement with ATLAS (15+15)

Thursday, June 13, 2024 1:30 PM (30 minutes)

Presenter: PETTEE, Mariel (Lawrence Berkeley National Lab. (US))

Contribution ID: 13

Type: **not specified**

Multidimensional cross-section measurements with H1 (15+15)

Thursday, June 13, 2024 2:00 PM (30 minutes)

Presenter: TORALES ACOSTA, Fernando (Lawrence Berkeley National Lab. (US))

Contribution ID: 14

Type: **not specified**

Closing of the meeting and summary

Thursday, June 13, 2024 3:00 PM (30 minutes)

Presenter: NACHMAN, Ben (Lawrence Berkeley National Lab. (US))

Contribution ID: 15

Type: **not specified**

Welcome reception

Contribution ID: 16

Type: **not specified**

Full Event Particle-Level Unfolding with Variable Length Latent Variational Diffusion (20+20)

Thursday, June 13, 2024 11:20 AM (40 minutes)

Collisions at the Large Hadron Collider (LHC) provide information about the values of parameters in theories of fundamental physics. Extracting measurements of these parameters requires accounting for effects introduced by the particle detector used to observe the collisions. The typical approach is to use a high-fidelity simulation of the detector to generate synthetic datasets that can then be compared directly with experimental data. However, these simulations are often proprietary and computationally expensive. An alternative approach, *unfolding*, statistically adjusts the experimental data for detector effects. Traditional unfolding algorithms require binning data in a small set of pre-selected dimensions. Recent methods using generative machine learning models have shown promise for performing un-binned unfolding in high dimensions, allowing later computation of many observables. However, all current generative approaches are limited to unfolding a fixed set of observables, making them unable to perform *full-event* unfolding in the variable dimensional environment of collider data. A novel modification to the variational latent diffusion model (VLD) approach to generative unfolding is presented, which allows for unfolding of high- and variable-dimensional feature spaces. The performance of this method is evaluated in the context of semi-leptonic $t\bar{t}$ production at the LHC. Additionally, the dependence of the unfolding on the training data prior is assessed by evaluating the model on datasets with alternative priors.

Primary authors: GHOSH, Aishik (University of California Irvine (US)); SHMAKOV, Alexander (University of California Irvine (US)); WHITESON, Daniel (University of California Irvine (US)); GREIF, Kevin Thomas (University of California Irvine (US)); FENTON, Michael James (University of California Irvine (US))

Presenters: SHMAKOV, Alexander (University of California Irvine (US)); GREIF, Kevin Thomas (University of California Irvine (US))

Contribution ID: 17

Type: **not specified**

Response Matrix Estimation in Unfolding Differential Cross Sections (20+20)

Tuesday, June 11, 2024 4:40 PM (40 minutes)

In unfolding problem, the response matrix is the forward operator which models the detector response. In practice, the response matrix is not known analytically. Instead, it needs to be estimated using Monte Carlo simulation, which introduces statistical uncertainty into the unfolding procedure. This raises the question of how to estimate the response matrix in a sensible way. In most analyses at the LHC, this is done by binning the events and counting the corresponding numbers of events from bins to bins. However, this approach can suffer from undersmoothing, especially with a small sample size. To address this issue, we propose a two-step approach to response matrix estimation. First, we estimate the response kernel on the unbinned space. Second, we propagate the estimated response kernel into an integral equation to obtain an estimate for the response matrix.

Primary authors: WASSERMAN, Larry (Carnegie Mellon University); KUUSELA, Mikael (Carnegie Mellon University (US)); ZHU, Richard (Carnegie Mellon University)

Presenter: ZHU, Richard (Carnegie Mellon University)

Contribution ID: **18**

Type: **not specified**

Dealing with Uncertainties (20+20)

Wednesday, June 12, 2024 9:00 AM (40 minutes)

Presenter: CORMIER, Kyle (University of Zurich (CH))

Contribution ID: 19

Type: **not specified**

QUnfold: Quantum Annealing for Distributions Unfolding in High-Energy Physics (20+20)

Thursday, June 13, 2024 9:30 AM (40 minutes)

In High-Energy Physics (HEP) experiments, each measurement apparatus exhibit a unique signature in terms of detection efficiency, resolution, and geometric acceptance. The overall effect is that the distribution of each observable measured in a given physical process could be smeared and biased. Unfolding is the statistical technique employed to correct for this distortion and restore the original distribution. This process is essential to make effective comparisons between the outcomes obtained from different experiments and the theoretical predictions.

The emerging technology of Quantum Computing represents an enticing opportunity to enhance the unfolding performance and potentially yield more accurate results.

This work introduces QUnfold, a simple Python module designed to address the unfolding challenge by harnessing the capabilities of quantum annealing. In particular, the regularized log-likelihood minimization formulation of the unfolding problem is translated to a Quantum Unconstrained Binary Optimization (QUBO) problem, solvable by using quantum annealing systems. The algorithm is validated on a simulated sample of particles collisions data generated combining the Madgraph Monte Carlo event generator and the Delphes simulation software to model the detector response. A variety of fundamental kinematic distributions are unfolded and the results are compared with conventional unfolding algorithms commonly adopted in precision measurements at the Large Hadron Collider (LHC) at CERN.

The implementation of the quantum unfolding model relies on the D-Wave Ocean software and the algorithm is run by heuristic classical solvers as well as the physical D-Wave Advantage quantum annealer boasting 5000+ qubits.

Primary authors: Dr BIANCO, Gianluca (Universita e INFN, Bologna (IT)); GASPERINI, Simone (Universita e INFN, Bologna (IT))

Presenters: Dr BIANCO, Gianluca (Universita e INFN, Bologna (IT)); GASPERINI, Simone (Universita e INFN, Bologna (IT))

Contribution ID: 20

Type: **not specified**

Simplified Template Cross Sections (STXS) (20+20)

Tuesday, June 11, 2024 10:50 AM (40 minutes)

Presenter: BALASUBRAMANIAN, Rahul (Centre National de la Recherche Scientifique (FR))

Contribution ID: 21

Type: **not specified**

Open discussion on challenges and possible solutions in unfolding

Thursday, June 13, 2024 2:30 PM (30 minutes)

Presenter: MALAESCU, Bogdan (LPNHE-Paris CNRS/IN2P3 (FR))

Contribution ID: 22

Type: **not specified**

Unfolding using Denoising Diffusion (20+20)

Thursday, June 13, 2024 10:10 AM (40 minutes)

Unfolding detector distortions in experimental data is critical for enabling precision measurements in high-energy physics (HEP). However, traditional unfolding methods face challenges in scalability, flexibility, and dependence on simulations. We introduce a novel unfolding approach using conditional denoising diffusion probabilistic models (cDDPM). By modeling the conditional probability density between detector-level observations and truth-level particle properties from various physics processes, the cDDPM unfolding performance generalizes across varied simulated processes and kinematic distributions without retraining. We demonstrate proof-of-concept on toy models and evaluate on simulated Large Hadron Collider jets across different physics processes.

Primary author: PAZOS, Camila (Tufts University (US))

Co-authors: KLASSEN, Martin (Tufts University (US)); BEAUCHEMIN, Pierre-Hugues (Tufts University (US)); Prof. AERON, Shuchin (Tufts University); WONGJIRAD, Taritree (Tufts University); CROFT, Vincent Alexander (Nikhef National institute for subatomic physics (NL))

Presenter: PAZOS, Camila (Tufts University (US))

Contribution ID: 23

Type: **not specified**

Unfolding in the context of $g-2$ (20+20)

Wednesday, June 12, 2024 9:40 AM (40 minutes)

Presenter: LELLOUCH, Laurent (CNRS and Aix-Marseille U.)

Contribution ID: 24

Type: **not specified**

Unfolding in the context of a heavy ion analysis (20+20)

Tuesday, June 11, 2024 4:00 PM (40 minutes)

In this study, the process of unfolding is studied in the context of a heavy ion photon-tagged jet analysis. The SVD and D'Agostini unfolding algorithms are compared, and an application of using the MSE to choose the regularization strength is shown. Additionally, the investigation looks into the bias associated with unfolding in relation to prior choice. The performance is evaluated with different theoretical models and the bottom line test.

Primary author: PARK, Molly (Massachusetts Inst. of Technology (US))

Presenter: PARK, Molly (Massachusetts Inst. of Technology (US))

Contribution ID: 25

Type: **not specified**

Unfolding is not unsmearing (20+20)

Tuesday, June 11, 2024 1:30 PM (40 minutes)

In particle physics unfolding methods are employed when the basis used to represent an estimate of the truth is not the basis with statistically independent expansion coefficients. For the discrete unfolding problem the latter is given by the eigenvectors of the Fisher information matrix, which measures the amount of information carried by the data about the truth. In typical cases it is ill-conditioned, with the consequence that the measurements constrain only a small number of the expansion coefficients. This allows for highly efficient data reduction, but only for a biased estimate of the truth. Unfolding methods differ in how they bias the result. A way to quantify this is the posterior response matrix.

Primary author: SCHMELLING, Michael (Max Planck Society (DE))

Presenter: SCHMELLING, Michael (Max Planck Society (DE))

Contribution ID: 26

Type: **not specified**

Moment Unfolding (20+20)

Wednesday, June 12, 2024 11:30 AM (40 minutes)

Presenter: DESAI, Krish

Contribution ID: 27

Type: **not specified**

Profile likelihood unfolding with large number of bins (20+20)

Wednesday, June 12, 2024 10:50 AM (40 minutes)

In this talk we will discuss previous measurements using binned maximum likelihood unfolding focusing on analyses with large numbers of bins and nuisance parameters. We will outline the technical implementation and highlight the challenges and limitations. Furthermore, we showcase the newly-developed approach of “linearized binned likelihood unfolding”, a modified formalism that has a better scaling and allows to perform unfolding on even larger numbers of bins and nuisance parameters.

Primary author: WALTER, David (CERN)

Co-authors: BENDAVID, Josh (Massachusetts Inst. of Technology (US)); DEFRANCHIS, Matteo (CERN)

Presenter: WALTER, David (CERN)

Contribution ID: 28

Type: **not specified**

Using unfolded data in global QCD analyses (20+20)

Tuesday, June 11, 2024 3:20 PM (40 minutes)

Presenter: ZENAIEV, Oleksandr (Hamburg University)

Contribution ID: 29

Type: **not specified**

Likelihood-based unfolding with the CMS Higgs combination tool (20+20)

Tuesday, June 11, 2024 11:30 AM (40 minutes)

The CMS Higgs combination tool is the software package used for statistical analyses by the CMS Collaboration. The package, originally designed to perform searches for a Higgs boson and the combined analysis of those searches, has evolved to become the statistical analysis tool presently used in the majority of measurements and searches performed by the CMS Collaboration. Since Combine has access to the full likelihood function, it can also be used to perform a likelihood-based unfolding. This approach has become the standard unfolding procedure for many analyses within the collaboration.

Primary author: TARABINI, Alessandro (ETH Zurich (CH))

Presenter: TARABINI, Alessandro (ETH Zurich (CH))