Scalable cyberinfrastructure applications

Team: J. Brehmer[1 2], K. Cranmer[1 2], **Irina Espejo**[1], S. Macaluso[1 2] and H. Müller[1]

Institutions: [1] Center for Data Science, New York University

[2] Department of Physics, New York University

The SCAILFIN Project

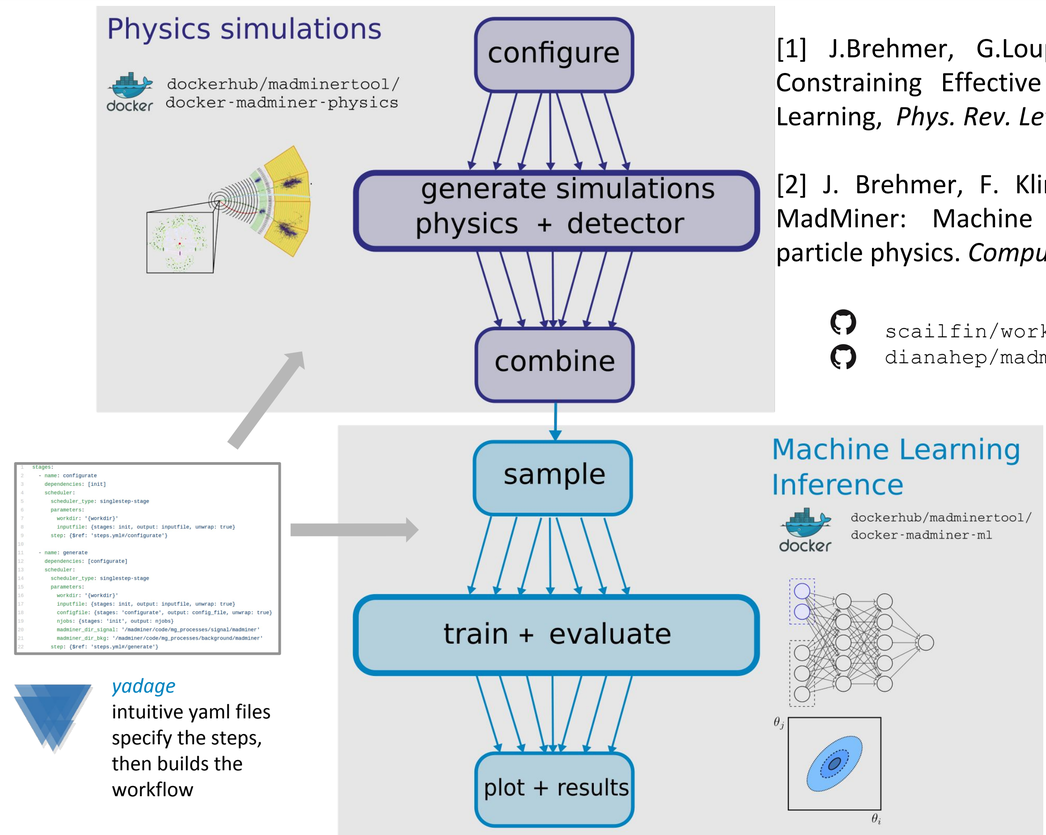
MadMiner on REANA

Simulators provide a *high-fidelity description* of phenomena. If they are too complex the likelihood is intractable leading to *poor statistical analysis*

MadMiner (Machine learning-based inference for particle physics [1], [2]) .
Its goal is to estimate parameters using advances in *Neural Networks*.

To deploy MadMiner in a scalable way we need:

- ✓ containerization
- ✓ parallelization
- ✓ reusability



[1] J.Brehmer, G.Louppe, J.Pavez and K.Cranmer. Constraining Effective Field Theories with Machine Learning, *Phys. Rev. Lett. 2018* 121, 111801

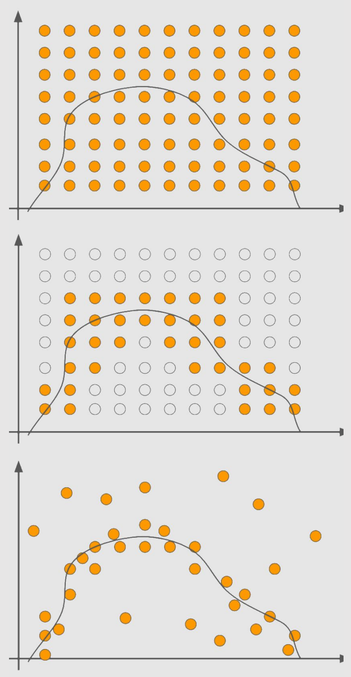
[2] J. Brehmer, F. Kling, I. Espejo and K. Cranmer. MadMiner: Machine learning-based inference for particle physics. *Comput.Softw.Big Sci*. 4 3 (2020)

scailfin/workflow-madminer
dianahep/madminer

Excursion

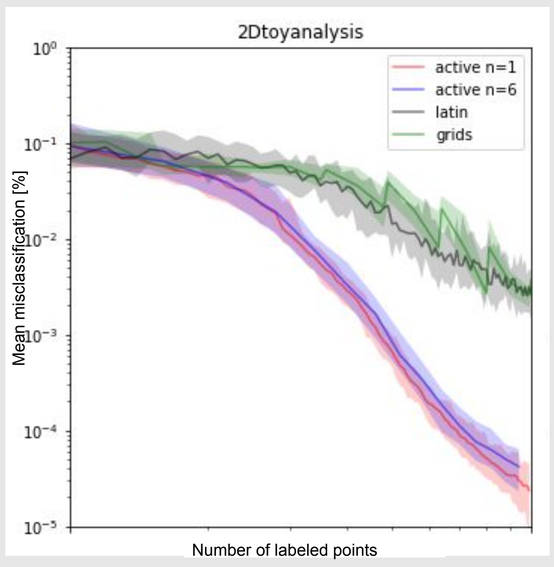
in collaboration with G. Louppe[2] and L. Heinrich[3]
[2]University of Liège, [3]CERN

- Goal is to find *level sets of black-box functions* that are expensive to evaluate. Examples: test statistics from complex simulations.
- Evaluate the black box function at *interesting points only* instead of evaluating at whole regular grid. We use a *Gaussian process* to: interpolate between samples and model uncertainty in the knowledge of the black box function.
  The *acquisition function* regulates the exploration vs exploitation tradeoff. Select one that *minimizes global uncertainty* of the location of the excursion set.
- Future: efforts will focus on *scalability* wrt the dimensionality of the function domain space. Example, likelihood ratio as function of mass, charge, spin,...



diana-hep/excursion
irinaespejo/excursion

… Is it all worth it? Yes! [3]

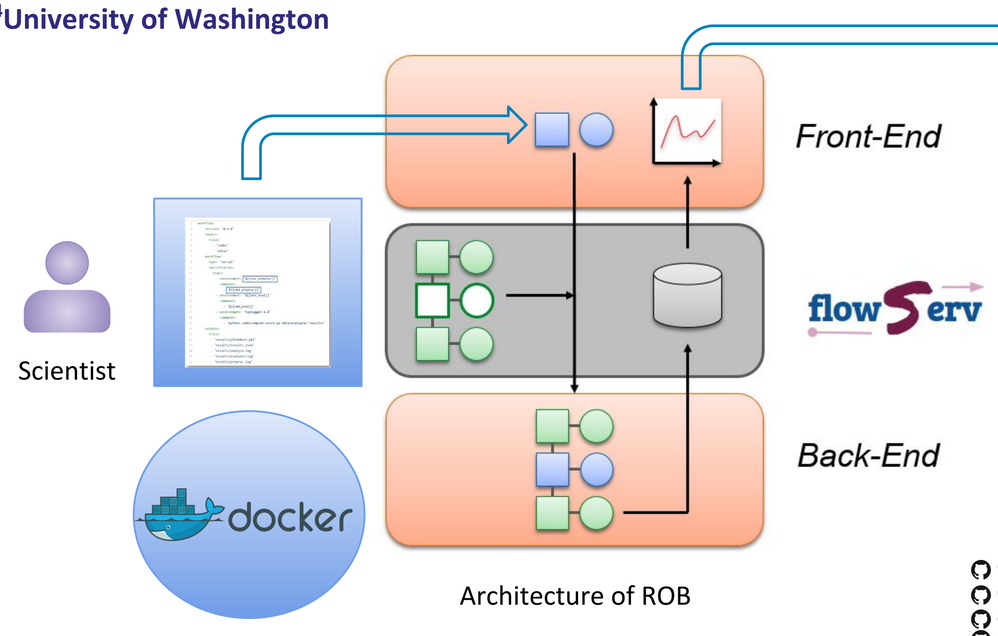


[3] L. Heinrich, G. Louppe, K. Cranmer, *Excursion Set Estimation using Sequential Entropy Reduction for Efficient Searches for New Physics at the LHC*, ACAT 2019

ROB
Reproducible Open Benchmark Platform

in collaboration with S. Hsu[4], A. Maritz[4], A. Rawat[4] and C. Suaysom[4]
[4]University of Washington

- ROB is an experimental prototype for enabling *community benchmarks* of data analysis algorithms. The goal of ROB is to allow user communities to evaluate the performance of their different data analysis algorithms in a *reproducible competition-style* format.
- The *workflow template* and input data are defined by a coordinator. The template contains placeholders for workflow steps that are implemented by the participants (e.g., with Docker containers). The backend processes the submission workflows. The user interface allows participants to *submit new runs* and to view the results.



scailfin/flowserv-core
scailfin/rob-webapi-flask
scailfin/rob-ui
SebastianMacaluso/TopTagComparison

National Science Foundation (OAC-1841471)
National Science Foundation (ACI-1450310)

This project is supported by National Science Foundation under Cooperative Agreement OAC-1836650. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.