The Future Interplay of Monte Carlo and experiment

Stefan Prestel

Future Trends in Nuclear Physics Computing
Jefferson Lab (remotely), March 16, 2016
1. Monte Carlo Event Generators
2. Data / theory comparisons
3. Data / theory comparisons in LHC phenomenology
4. Making theory results available
Different reasons for Monte Carlo

**Experimentalist:** Estimate detector acceptance corrections.

**Phenomenologist:** Assess search strategies and detector development.

**Theorist:** Compare theory to data.

**Common denominator:**
Need detailed and accurate modelling of scattering events.
Event generators as theory repository

New discoveries can only be claimed if we compared data with our “best” theory knowledge before.

General Purpose Monte Carlo Event Generators (MCEGs) aim at being repositories for the most accurate theory knowledge.

Common high-energy MCEGs are HERWIG, PYTHIA and SHERPA. They include knowledge of $O(40$ research years), and need to be

- Developed and improved …or they stop being useful.
- Maintained …or they stop being useful.
- Modular and extendable design …to enable the community can contribute.
Example: Include perturbative QCD frontier for LHC

\( \mu_R = \mu_F = \mu_f^* / 2 \)

\( \sqrt{s} = 7 \text{ TeV} \)

\( W^+ + 5 \text{ jets} + X \)

\( \text{BlackHat} + \text{Sherpa} \)

**F:** \((W + 5 \text{ jets})\) at next-to-leading order. Massive computational task (see Stefan Höche’s talk).
Example: Include perturbative QCD frontier for LHC

**F:** \((W + 0 \text{ jet})@NLO + (W + 1 \text{ jet})@NLO\) combined in PYTHIA 8 vs. ATLAS.

**F:** \((W + 5 \text{ jets})\) at next-to-leading order. Massive computational task (see Stefan Höche’s talk).
Example: Include perturbative QCD frontier for LHC

F: (W + 5 jets) at next-to-leading order. Massive computational task (see Stefan Höche’s talk).

F: Hardest jet in W-production events in CMS. Massive computational task (see Stefan Höche’s talk).
Pieces of the MC puzzle

High-energy colliders

- Parton production
- Jet evolution
- Multiple interactions
- Hadronization
- Beam/target remnant
- Hadron decays

Lepton colliders

Nuclear physics needed!
Pieces of the MC puzzle

High-energy colliders

- Parton production
- Jet evolution
- Multiple interactions
- Saturation?
- Spin structure?
- Hadronization
- Beam/target remnant
- Hadron decays
- Nuclear and medium effects?

Lepton colliders

Nuclear physics needed!
Feedback loop between data and theory

Realistic model? New phenomena?
Improvements necessary?

Data parametrizations may only enter MC if theory cannot provide a model. \(\Rightarrow\) Monte Carlo tuning
Data must eventually lead to better theory.
To compare theory and data, we need to simulate all effects entering the data.
To compare theory and data, we need to simulate all effects entering the data.

This is never fully possible for an outsider. **Gold standard: Unfolded data** ⇒ Detector details irrelevant.
To compare theory and data, we need to simulate all effects entering the data.

This is never fully possible for an outsider. **Gold standard: Unfolded data** ⇒ Detector details irrelevant.

What if simulation does not match data? **Need fast open-source analysis framework for quick progress.**
Lessons from LHC

The progress of perturbative QCD initiated by the LHC relies on

- Unfolded data
- Standardized interfaces (HEPEVT, HepMC)
- Open-source analysis software used/developed by both experiment and Monte Carlo (RIVET, PROFESSOR)

⇒ Quick turnaround possible ⇒ Fast progress

(Note: RIVET offers some “MC smearing” structures if unfolded data is definitely impossible)
The RIVET analysis repository

Extensively tested ROOT-style framework.

Currently ∼340 analyses from SPS, LEP, Tevatron, LHC, RHIC, HERA...

Open source: Experiments provide analyses. Theorists provide analyses.
The RIVET analysis repository

Widely used for prototyping and collaboration, in both experiment and theory.

Well-defined framework allows for systematic and global …theory-data comparisons …tuning (see PROFESSOR tool)

⇒ Quicker progress!
Conversation problems

Experimentalist: Something is wrong (in this plot).

Theorist: What version are you using? My version looks fine.

Experimentalist: Running version 0.VERY.OLD, and cannot run a newer version. What was wrong?

Theorist: A explains it. Thanks for reporting. Please update.

Experimentalist: Thanks.

Outcome:
Information gets lost, same question asked many years later. Start from beginning.

Theory feedback through papers (or emails) is not enough.  
⇒ Make calculations and plots publicly accessible!
Idea: Generate plot ASCII files, store in online data base. Make plots available at (and generate on) web page. Different views of plots to highlight changes over time.

Important as reference, but not time-critical. ⇒ Realization through volunteer computing.

Danger: Human intervention needed when new analysis or new generators are included. Needs to be up-to-date.
mcplots is powered by Rivet analysis repository
mcplots as an example plot generator
mcplots as an example plot generator
mcplots to assess the quality of the theory
mcplots to isolate issues with a few clicks
mcplots to isolate issues with a few clicks
mcplots was the first running project on LHC@home
mcplots was the first running project on LHC@home.

mcplots is powered by LHC@home. Relies on
- BOINC client
- CernVM + Wrapper + Co-pilot.

⇒ Volunteer cloud
Volunteer computing = no deadline computing, but very important for long-lived database
Volunteer computing is excellent to engage public
Volunteer computing is excellent to engage public
Summary

- Monte Carlo is implementation of theory. Needs to be seen and designed as knowledge vessel.
- Theory must be compared against data. Need standardized analysis tools to do that!
- Open-source analysis frameworks needed for quick turnaround.
- Theory needs to provide detailed feedback ...not only through papers ...through code and open access repositories.
  Volunteer clouds possible for such non-urgent HTC tasks.