Self-Interaction Cross Sections in Strongly Interacting Dark Matter

A Data Management Plan created using DMPTuuli

Creator: Jarno Rantaharju

Affiliation: University of Helsinki

Funder: Academy of Finland

Template: Academy of Finland data management plan guidelines

ORCID iD: 0000-0002-0072-7707

Project abstract:

We will use lattice field theory simulations to calculate interaction cross sections between mesons and baryons in two models, SU(2) and SU(5). The calculation is inspired by applications to strongly interacting dark matter.

ID: 12389

Last modified: 28-04-2021

Self-Interaction Cross Sections in Strongly Interacting Dark Matter

1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Also give a rough estimate of the size of the data produced or collected?

Data produced during the project can be divided roughly in three categories:

1. Software developed to produce configurations of the gauge field and derive measurements from these configurations
2. Values of measurements
3. Gauge configuration

We treat software as a scientific output and endeavour to publish it under an open source license with a unique DOI for each version used in a publication and make it available and easily findable. We do not anticipate any licencing concerns that would prevent this.

Measurements are produced by the software and are fully synthetic. The values are saved in ASCII format and published using the ZENODO service under a unique DOI in the form used to produce in each publication. Any analysis software used to produce values and plots quoted in the paper will be included here. These data files are not large and the overall amount produced is likely to be less that 200GB.

Gauge configurations are intermediate results produced by the simulation. They can be used to produce any measurement in a given model and set of parameters and may be useful for other projects. They require a large amount of disk space, approximately 1GB each. We plan to produce at least 1000 configurations, using at least 1TB of space. The data is stored in raw binary form.

1.2 How will the consistency and quality of data be controlled?

All data our is synthetic, as described above. The consistency and the quality of the data depends on two factors: the correctness of the software and the quality of the simulation that produced the configurations. The correctness of the software will be ensured by through standard software development methods, including automated unit testing, test cases and regression tests.

The quality of the observables will be quaranteed by monitoring several observables related to unphysical effects. The statistical independence of the configurations will be ensured by calculating autocorelation times of physical observables during the simulations.

2. Ethical and legal compliance

2.1 What ethical issues are related to your data management, for example, in handling sensitive data, protecting the identity of participants, or gaining consent for data sharing?

Before beginning any software development or accepting changes from any new developer, their consent to publish under the MIT licence will be sought. Any part of the software not publishable under the MIT licence will be published under the least restrictive licence possible.

We collect no personal data and expect no privacy concerns to arise. All data except software is fully synthetic.

2.2 How will data ownership, copyright and IPR issues be managed? Are there any copyrights, licences or other restrictions that prevent you from using or sharing the data?

In order to facilitate continuing development in the long term, each group member who participates in creating any software or data will retain be given full licence to use and distribute the entire work. Each member will consent to the PI, Jarno Rantaharju, publishing any data and software under the MIT licence, the Creative Commons Attribution licence or under any more restrictive licence. In this way each member can continue to develop, use and publish their version and the PI will be able to maintain the original version. Any future changes to the main version will require the same grant of licence.

3. Documentation and metadata

3.1 How will you document your data to make them findable, accessible, interoperable and reusable for you and others? What kinds of metadata standards, README files or other documentation will you use to help others understand and use your data?

Measurements are published in ASCII form together with a README file describing the contents, including references to the version of the software used to produce it, gauge configurations used and any other parameters necessary to exactly reproduce the data.

Configurations are kept and distributed in the binary form produced by the program, together with the output of the simulation used to produce it and a README file describing the contents of said file. Unfortunately there is no generally accepted data format for gauge configurations. We will include a complete description of the data format used and methods of reading the contents of the file.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will they be backed up?

Each version of software and measurement data will be stored using the ZENODO service, operated by CERN and the OpenAIRE project, and kept locally. The ZENODO copy is expected to remain archived in perpetuity. The local copy will be kept at least for the lifecycle of the project.

The configuration will be kept locally and archived securely for the duration of the project and at least two years after using the IDA service provided by The Finnish Ministry of Education and Culture.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

There are no privacy concerns related to the data and we endeavour to keep the data open and public. The PI is primarily resposible for maintaining open access to the data.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or their metadata, be made available?

All measurement data and software will be made openly available and published under a minimally restrictive licence. Both will be stored and made available using the ZENODO service, operated by CERN and the OpenAIRE project, at or before the publication of any articles based on them.

Configurations will be made available under a minimally restrictive license but cannot be stored on ZENODO due to their size. They will be stored as long as possible, initially using the IDA service. If reproducing this data becomes computationally trivial and is storing is no longer practically useful, only the metadata will be retained.

5.2 Where will data with long-term value be archived, and for how long?

Measurement data and software are scientific outcomes and have long term value. They will be archived by the ZENODO service on publication. The archive will in princliple be maintained in perpetuity. We will retain a local copy of both at least for the duration of the project using services provided by CSC and by the University of Helsinki.

6. Data management responsibilities and resources

6.1. Who will be responsible for specific tasks of data management during the research project life cycle? Estimate also the resources (e.g. financial, time and effort) required for data management.

The PI is responsible for maintaining data and software and versioning and publishing them as used to produce any published articles. This is mainly achieved through a proper implementation of version control and continuous integration of data and software during the project, a practice that reduces rather than increases the time and effort necessary to complete the project. Versioning and publishing data and software requires a few hours of work time. No financial cost is expected.