Table of Contents

[Optimization scheme involving molecular dynamics 1](#_Toc345433309)

[Simulation server module 2](#_Toc345433310)

[The molmec package 3](#_Toc345433311)

[Force-Field template creator 3](#_Toc345433312)

[Dakota input file generator 4](#_Toc345433313)

[Force-field update script 4](#_Toc345433314)

[The simhost package 5](#_Toc345433315)

[Concentrated LiCl solution: a test study 5](#_Toc345433316)

[Scientific relevance 5](#_Toc345433317)

[Simulated system 5](#_Toc345433318)

[Simulation and analysis tools 6](#_Toc345433319)

[Fitting model 6](#_Toc345433320)

[Fitting Results 6](#_Toc345433321)

[Planned Dakota optimization 7](#_Toc345433322)

# Optimization scheme involving molecular dynamics

Optimization of force-fields employed in molecular dynamics (MD) simulations *via* fit to experimental dynamic neutron scattering structure factors requires optimization of two fundamentally different set of parameters. First, one seeks to optimize those parameters of the force-field that will be subsequently used in production MD runs (partial charges, Lennard-Jones potentials, and bonded force-constants). Second, one must also optimize extra parameters pertaining to the specifics of the neutron scattering experiment (background, beam intensity, and resolution function). Search of the optimal values for the first parameter set require new MD simulations runs every time new values are tried, simulations that usually complete in a time-scale of hours to days. On the other hand, search of the optimal values for the second parameter set require no MD simulation and the optimization completes under a second. Thus, the problem leads itself into an optimization strategy where the two parameter sets are optimized at different stages but whithin a global optimization loop.

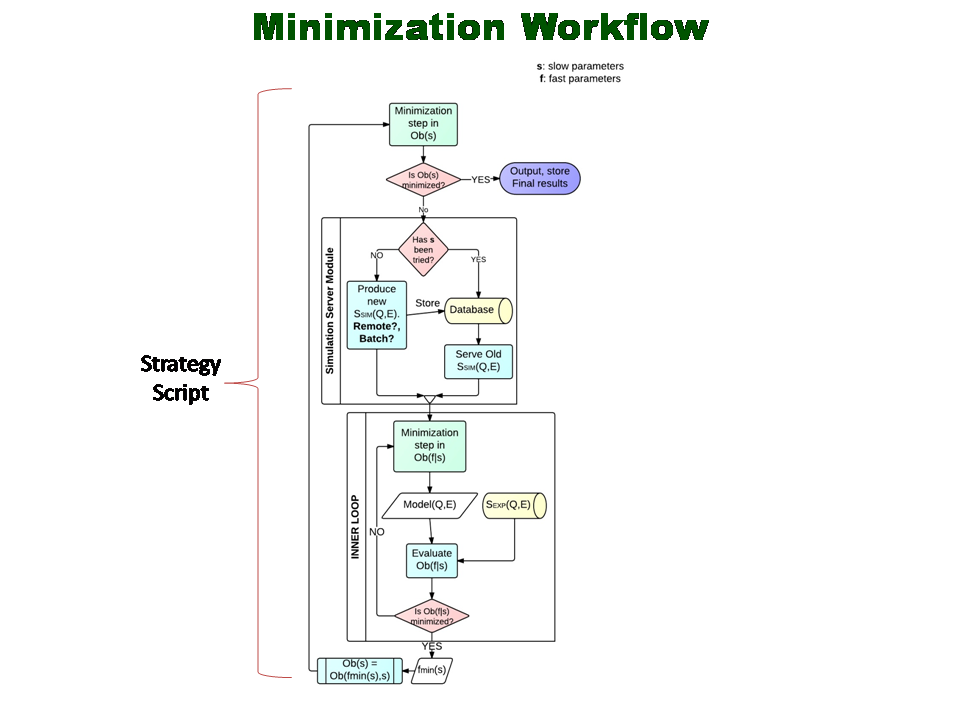


Figure 1: Optimization loop. Optimization of the objective function *Obs(f,s)*, is carried out in two stages. An inner loop optimizes the function under constant force-field parameters *s*, and then the optimal experimental parameters found *f=fmin(s)* are use to yield an objective function that depends only on the force-field parameters.

# Simulation server module

Every time the force-field parameters are changed to new values, a computation of the simulated structure factor must be realized. This computation involves a chain of jobs involving MD simulation(s) and the subsequent computation of the intermediate and final dynamic structure factors. First the MD must be called, which can be a single or a compound run, depending on the system to be simulated and the available computer resources. Next, calculation of the intemediate dynamic structure factor *I(Q,t)* must be called, which in itself is time consuming. Finally, Fourier transform of *I(Q,t)* yields the final dynamic structure factor *S(Q,E)*. A Database server will store the data generated in this chain of jobs. Conversely, the server will provide the data if the force-field parameter values have been already tried.

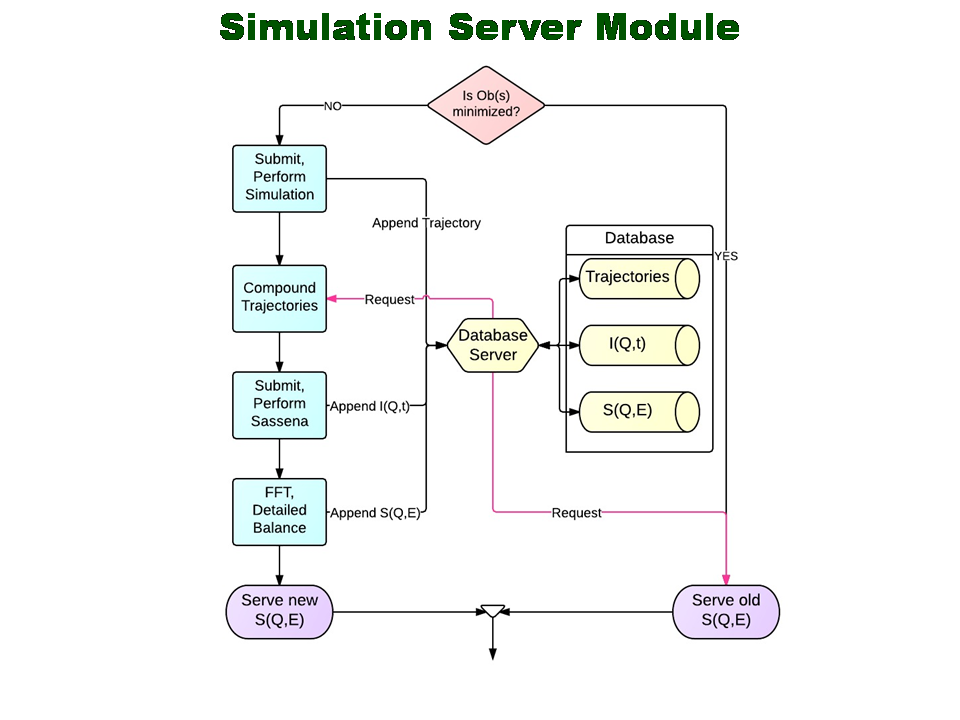


Figure 2. Chain of jobs. A decision is made whether to execute new MD and Sassena simulations to generate a new dynamics structure factor *S(Q,E)* or whether to serve and old *S(Q,E)*.

# The molmec package

We are developing a new open-source python package, molmec, for the purpose of managing all the necessary steps revolving around the molecular mecanics calculations within the overall optimization scheme of force-field parameters. These steps involve, among others: (*i*) preparation and update of force-field files in popular formats (CHARMM, LAMPS, AMBER, GROMACS); (*ii*) preparation of input and output files for the Dakota minimization package; (*iii*) preparation and management of the hierarchy of subdirectories housing the MD and Sassena runs; (*iv*) preparation and update of the input files for each MD run; and (*v*) file format adaptation to the requirements of Sassena.

Configuration of these different steps will be first implemented through simple configuration files, although we plan to extend these with graphical user interfaces as the package gains in stability.

## Force-Field template creator

The current version of the molmec.fftpl module contains the algorithms necessary to select any partial charges in a CHARMM/XPLOR topology file as force-field parameters to be optimized. The module generates a template topology file that can be updated every time the optimization scheme requires trial of new values for the parameter set.

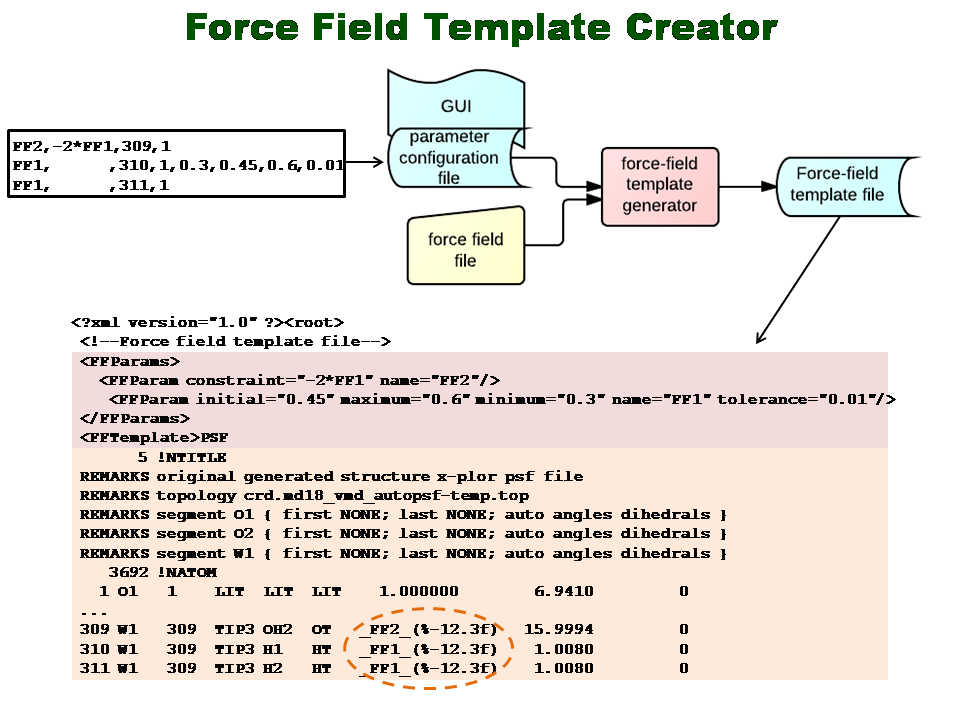


Figure 3. The force-field template creator reads-in a simple configuration file specifying the force-field parameters to be optimized and generates a template CHARMM topology file (in XML format).

## Dakota input file generator

The current version of the molmec.dakota module reads in a templated Dakota input file and inserts the force-field parameters to be optimized with the help of the template topology file. Due to the enormous variety of Dakota options, we plan to offer a limited selection of relevant templated Dakota input files, along with the option allowing experienced users to provide their own. The resulting file will serve as the starting input file for the whole optimization scheme.

## Force-field update script

The current version of the molmec.ffupdate module uses the CHARMM topology template file and the request generated by Dakota in order to produce an updated topology file, with new values for the force-field parameter set to be optimized. This file will be used in the subsequent MD simulations.

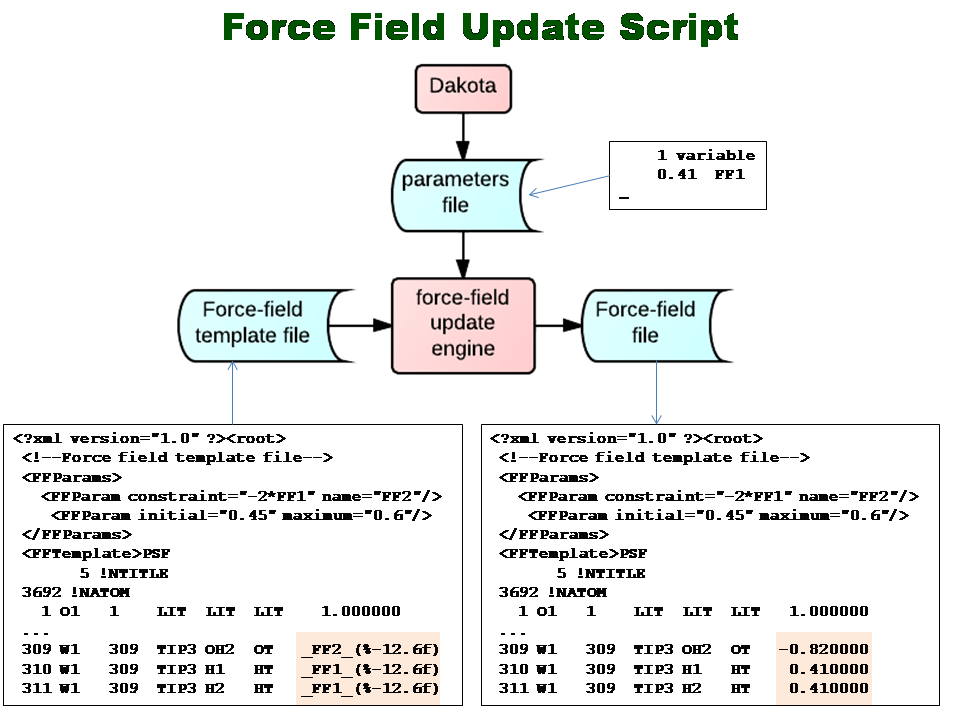


Figure 4. Update of the topology file with a single fit parameter. The templated topology file shows the partial charge of the hydrogen atom in the water molecule as the force-field parameter to be optimized (*FF1*). A second parameter (*FF2*) is required to ensure that the water molecule has a neutral charge. Parameter *FF2* is thus determined by the value of *FF1*. The Force-field update engine takes care of such constraints among force-field parameters. The correct format (%-12.6) is ensured in the template topology file.

# The simhost package

We are developing a new open-source python package, simhost, for the purpose of managing correct scheduling and execution of MD and Sassena simulations in high-performance clusters. The current version of the package contains host, server, and queue modules. Typical tasks to be carried out by the package are: (*i*) query of cluster resources; (*ii*) estimation of required computer time, depending on simulation-system size and cluster resources; (*ii*) creation of on-demand PBS submission scripts for MD and Sassena simulations; and (*iii*) ensure dependencies among the PBS scripts when necessary.

# Concentrated LiCl solution: a test study

We plan to recreate the optimization of the partial atomic charges (PAC) for the three-point model of the water molecule in the presence of a concentrated solution of lithium chloride ions. To this effect, we have carried out MD simulations in a wide range of PAC-values and compared the resulting simulated dynamics structure factors S(Q,E) to experimental results[ref] taken at the BASIS beamline at the SNS facility.

## Scientific relevance

Bulk aqueous solutions of concentrated lithium chloride do not freeze in the 220K -230K temperature range where a dynamic crossover between non-Arrhenius and Arrhenius behavior was experimentally observed for pure water in confined geometries. Thus, the bulk system allows the study of the dynamic crossover temperature with no contamination of results from the confining material which are significant at low Q-values.

From the point of view of our optimization test, polarization of the water molecules may be relevant due to the high ion concentration. Polarization would express itself as a significant deviation of PAC values with respect to those of pure bulk water. We seek to obtain new PAC values for the water molecule consistent with the experimentally observed dynamic structure factor.

## Simulated system

A 30Å x30Å x30Å box containing 1128 water molecules along with 154 Cl- and 154 Li+ ions matching the experiment concentration was prepared with the help of tleap[ref] and VMD[ref] packages. Systems were equilibrated in the NPT ensemble at 1atm and temperatures in the 10K-290K range. Subsequent productions run were executed in the NVE ensemble to prevent atomic decorrelations due to and external heat bath. Each simulation was set for particular values of the water PAC. The charge of the hydrogen atom, *Hq*, was waried in the 0.3e to 0.5e range. As a comparison, the standard values for the TIP3 and SPC/E water models are 0.417e and 0.4238e, respectively.

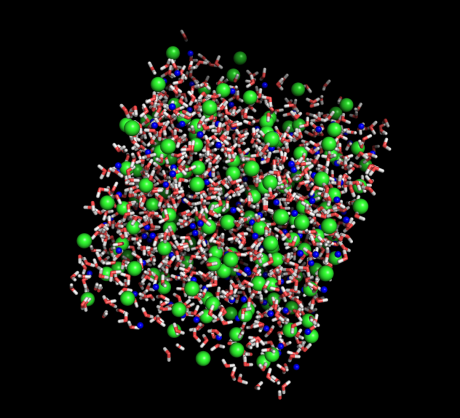


Figure 5. Simulated system: Lithium atoms shown as blue spheres, chloride atoms as green spheres, and water molecules as white and red sticks.

## Simulation and analysis tools

The open-source MC simulation package NAMD was selected for this test study. Functionality for this and other popular MD packages (AMBER, CHARMM, GROMACS, and LAMMPS among others) will be packages will be added. The open-source package Sassena was selected for the calculation of the intermediate scattering quantities. Sassena is developed and maintained at the Center for Molecular Biophysics in ORNL. The open-source framework Mantid, housed in our group, was selected for fast Fourier transform, implementation of the fitting model, and inner optimization of the parameters associated with the experiment.

## Fitting model

A proper comparison between simulated and experimental dynamic structure factors require taking into account the specifics of the experimental set up. We considered: (*i*) an elastic line modeled as the experimental structure factor at low (6K) temperature, *SlowT*; (*ii*) a flat background, *A0*; and (*iii*) a resolution function *Res* derived from the experimental structure factor at the same low temperature. When applied to the simulated dynamic structure factor, *SMD*, we obtain a simulated structure factor *SMD’* that takes into account the specifics of the experiment:



There are three parameters, *h1*, *A0*, and *h2* wich conform the parameter set pertaining to the experiment.

## Fitting Results

In this test study, we ran a set of ten simulations in parallel, each with a different value of the water-hydrogen PAC. Then parameters, *h1*, *A0*, and *h2* were optimized for each simulation to obtain the best fit between *SMD’* and the experimental dynamics structure factor.

|  |  |
| --- | --- |
| ***(a)***  fit_q0.7.jpeg | ***(b)***  fit_q0.7.jpeg |

Figure 6. Comparision between two fits. (a) Fit between experiment (black) and simulated (red) dynamics structure factors for water-hydrogen PAC *Hq*=0.38. The difference between experiment and simulation is shown in green. (b) Same as (a), but now *Hq*=0.42

A cost function was calculated for each simulation, and a cubic interpolation of the points (see Figure 7) suggests a value of the water-hydrogen PAC of *Hq\**=0.43 yields the best comparison between simulations and experiments. This *Hq\** value is higher than those of the standard TIP3 and SCP/E water models, indicating that the presence of LiCl has increased the dipole moment of the water molecule.

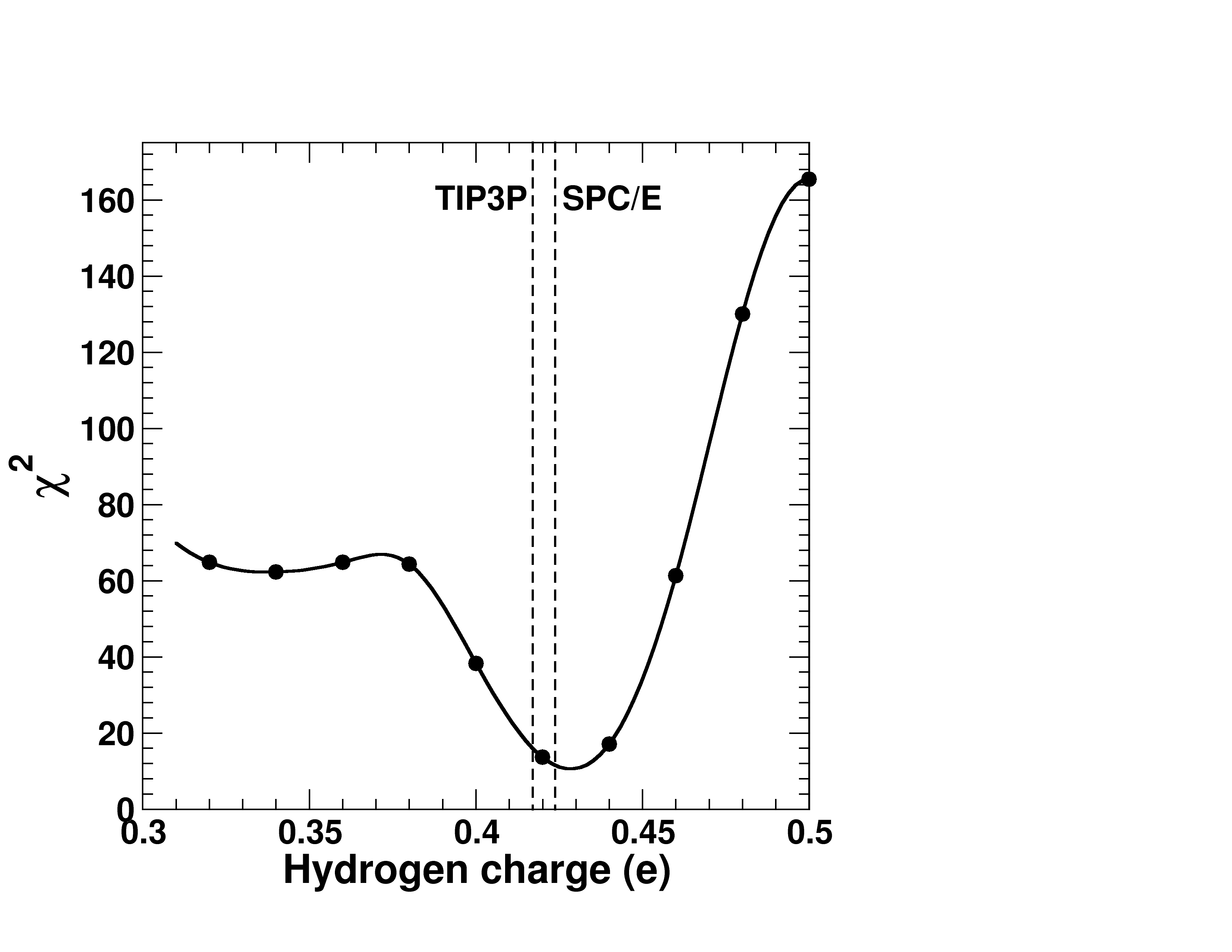


Figure 7. Cost function.

## Planned Dakota optimization

We are now in the process of building the required infrastructure components in order to automatically reproduce the results of Figure 7 in a hands-off mode, with the help of the automation provided by the Kepler workflow capabilities and the Dakota optimization package. Farming of simulations as well as feeding and manipulation of the input/output files to/from the different software components requires considerable time and expertise from the user. These requirements can be avoided with an automatic procedure which will also have the advantage of preventing the occurrence of human error.