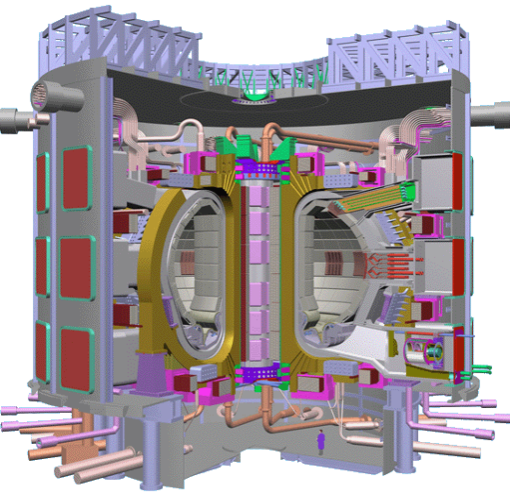


**Использование распределенной среды
REST-сервисов для оптимизационной
идентификации углеродистых структур
по рентгеновским дифракционным
характеристикам наноматериалов**

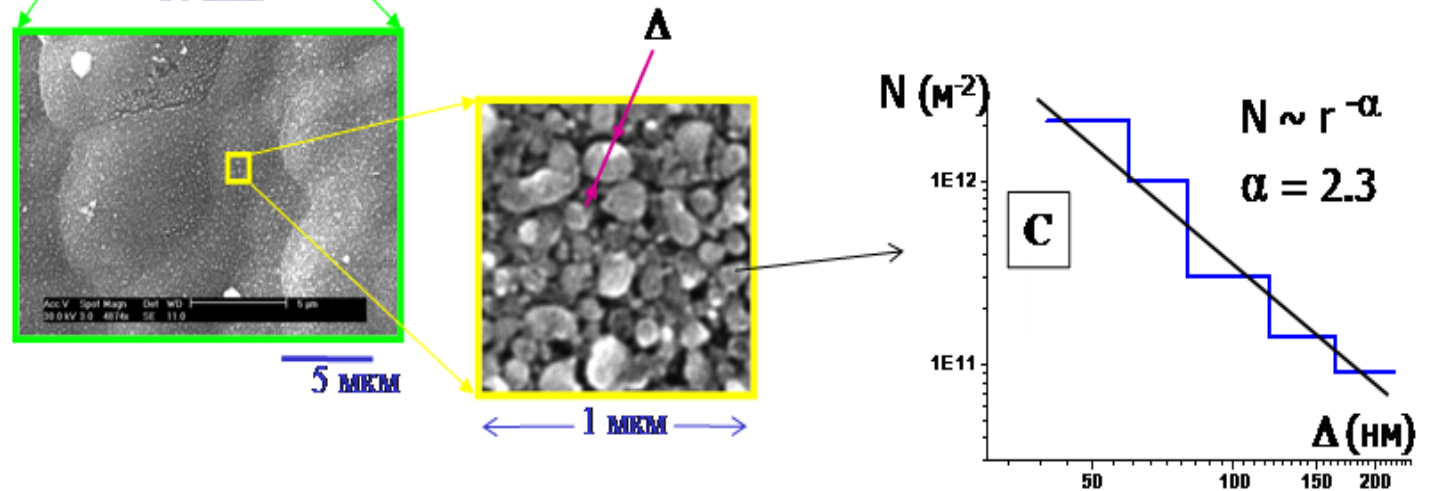
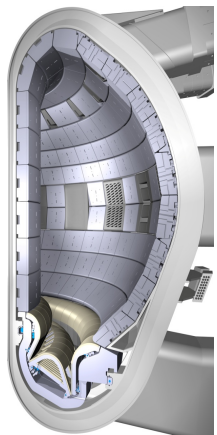
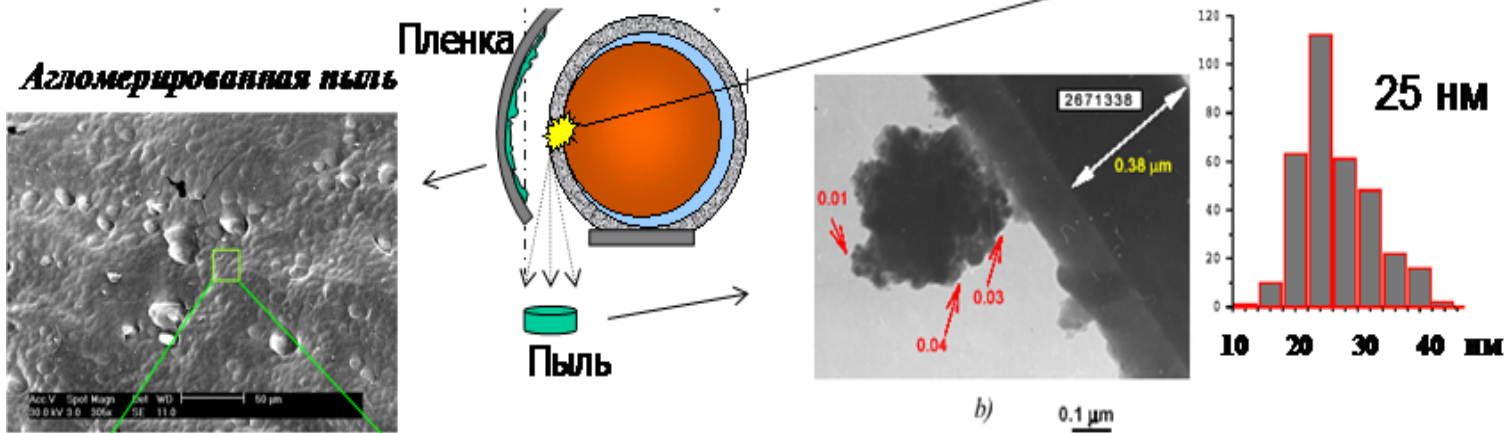
В.В. Волошинов Институт системного анализа РАН

В.С. Неверов РНЦ «Курчатовский институт»

Структурный анализ углеродных плёнок в Токамак Т-10



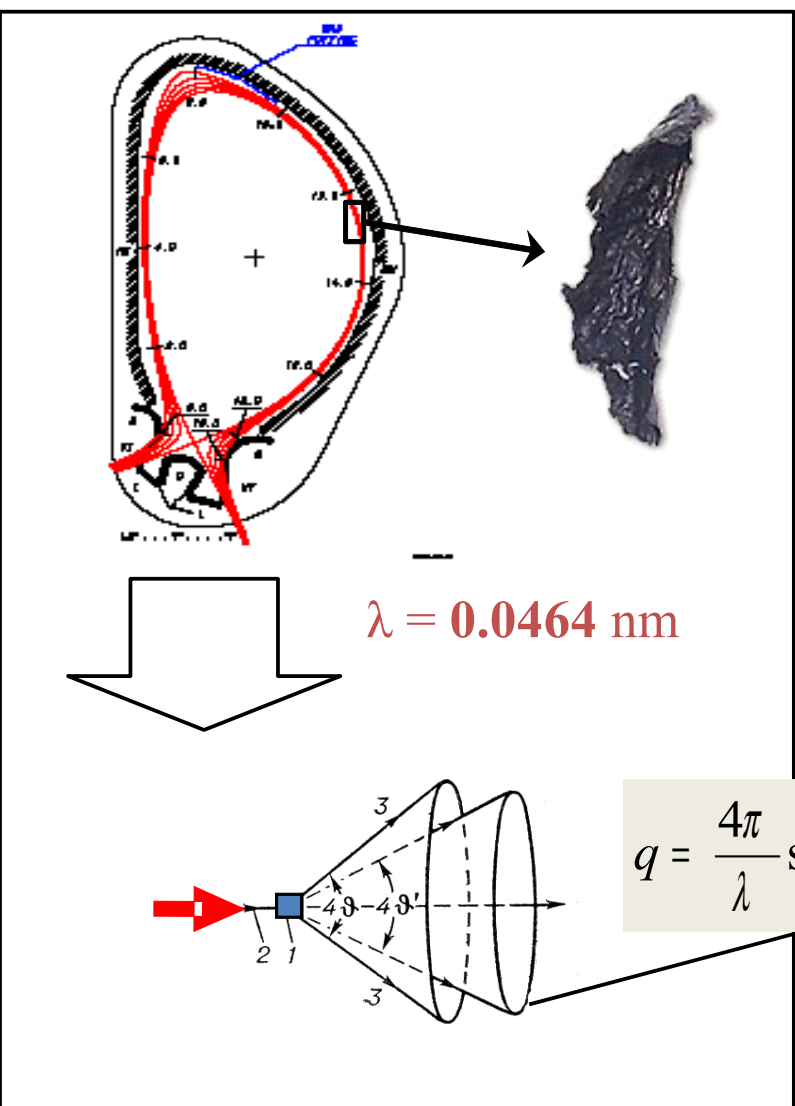
600 кВт × 0.5сек ⇒ (50 МВт/м²) ⇒ 2000 °С - ИТЭР – подобные условия.



В токамаке происходит агломерация наночастиц

Введение

Исследование **углеродных** пленок, образующихся внутри вакуумной камеры **токамака Т-10**, проведённое **методом рентгеновского рассеяния** на **синхротронном источнике** в Курчатовском институте (**А.А. Велигжанин, Я.В. Зубавичус, Н.Ю. Свечников, В.Г. Станкевич**), выявило три основные особенности при значениях модуля вектора рассеяния $q \sim 10, 30$ и 50 нм^{-1} .



Обнаруженный пик из-за своей ширины не может быть объяснён наличием нанокристаллов графита или других типичных примесей.

Сложности анализа результатов рентгеновского рассеяния на наноматериалах:

- порядок в структуре сохраняется лишь на **нескольких нанометрах**,
- отсутствуют **резкие** брегговские пики,
- имеются **широкие** пики, которые уже нельзя трактовать как брегговские,
- дифракционный профиль дает мало информации для восстановления сложного наноматериала.*

Это означает:

- восстановление **электронной плотности, методами Фурье-анализа** не дает исчерпывающей информации об объекте,
- **стандартные методы** кристаллографии **не применимы**,
- **нет** возможности получить **единственное решение**.*

*см. например «The nanostructure problem» <http://physics.aps.org/articles/v3/25>

Вывод: Решение **обратной задачи** — восстановление структуры нанобъектов требует сужения класса возможных решений и применение других диагностических методов, **прямого моделирования рассеяния** на нанобъектах и последующего сопоставления с экспериментом.

Структурный анализ и предлагаемый метод решения.

Задача:

Определить наиболее вероятный топологический состав **углеродных** наноструктур в образце.

Предлагаемый метод решения:

- Предположение о конечном наборе ожидаемых типов структур , возможно ответственных за широкий пик на $q \sim 10$ нм⁻¹, и **сравнение модельных кривых** (до 500 типов) рентгеновского рассеяния **с экспериментом** (больше 1000 точек)
- Определение наиболее вероятного состава **углеродных** наноструктур в образце из заданного набора структур **методом оптимизационной идентификации неизвестных параметров**.

Методика предполагает:

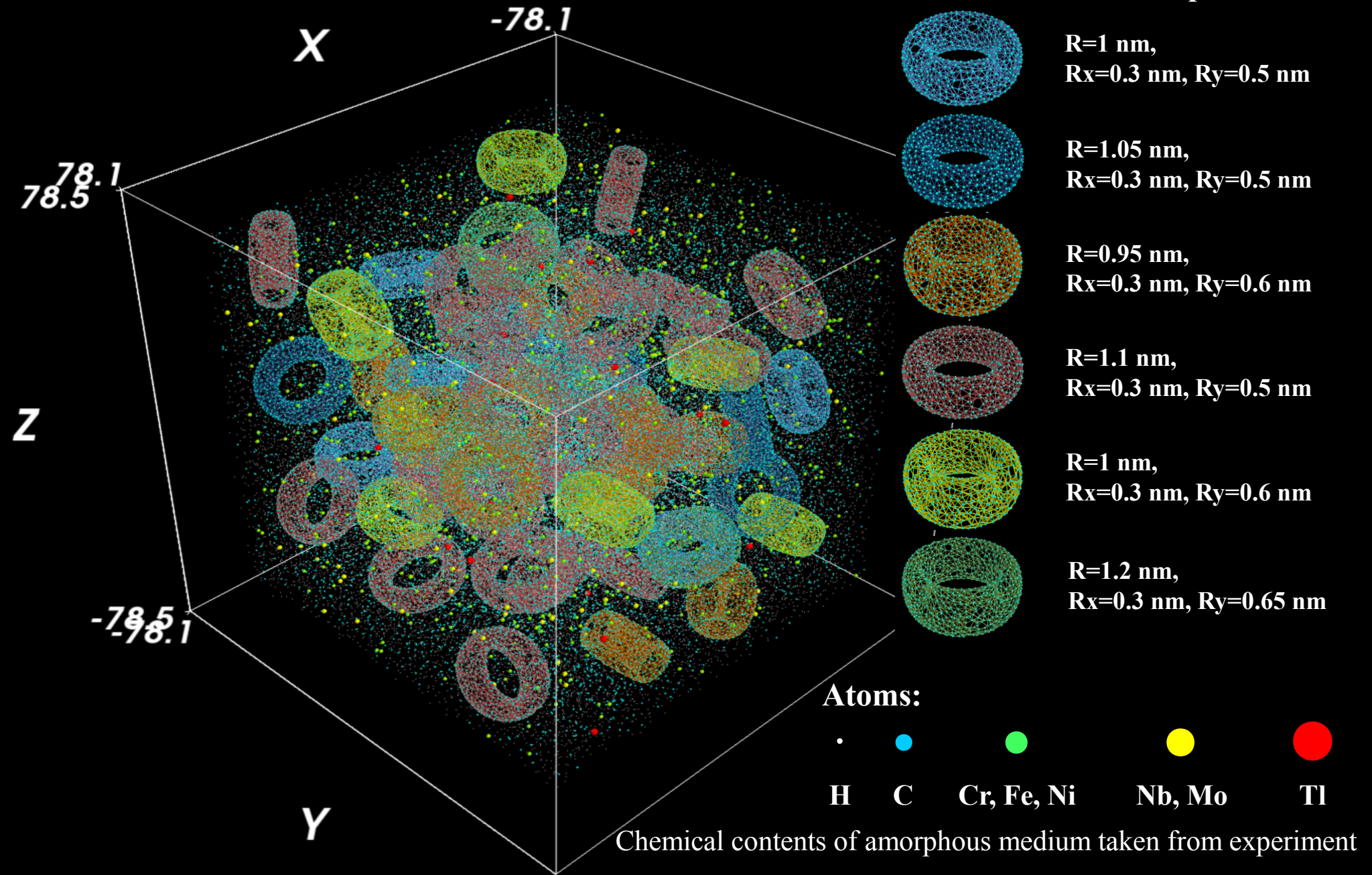
- "Массированные" вычисления модельных кривых рентгеновского рассеяния (C/C++, MPI, OpenMP)
- Применение производительных пакетов оптимизации.

Вывод:

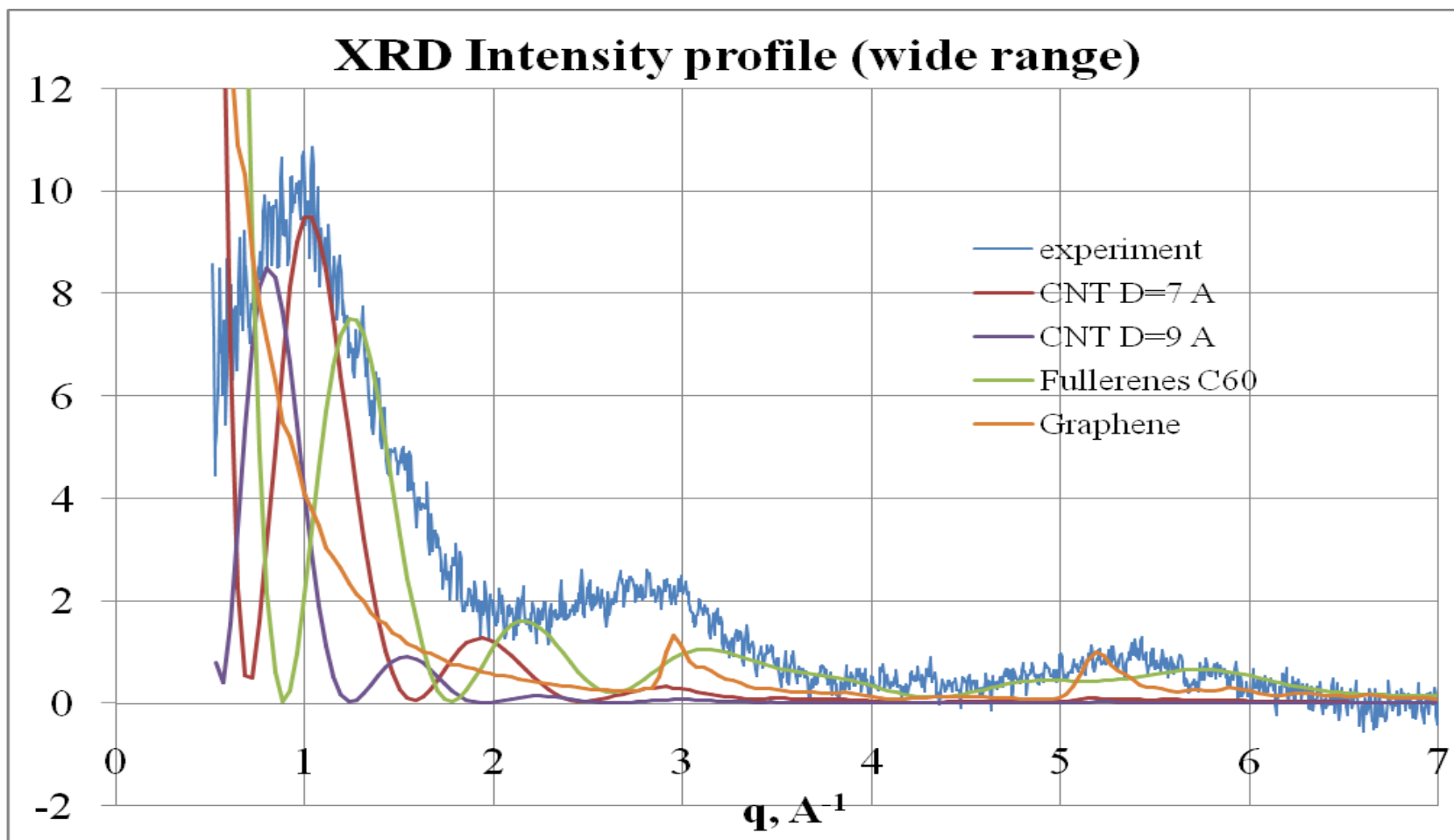
- **Целесообразен перенос вычислений в распределённую вычислительную среду.**

Структурный анализ углеродных плёнок. Смесь структур

L2 optimization



Модель рассеивания с последующей оптимизационной идентификацией по экспериментальным данным



Возможности созданного численного кода по расчету дифракционных характеристик наноструктур.

На Входе:

- Неограниченное число структур в ансамбле.
- Каждая структура представляется элементарным блоком (например, элементарная ячейка кристалла).
- Элементарный блок может иметь неограниченное количество копий, организованных в регулярные (кристаллы) и нерегулярные структуры.
- Нерегулярные структуры могут включать в себя регулярные и наоборот.

На Выходе:

- Одномерный профиль интенсивности рассеяния рентгеновского излучения, рассчитанный по формуле Дебая, с возможностью вывода отдельных интерференционных вкладов.
- Двумерная дифракционная картина с возможностью усреднения по пространственным углам волнового вектора падающего излучения.

Интерфейс:

- Простой XML интерфейс.
- Для генерации начальных данных, обработки и визуализации результатов используется набор скриптов, написанных на Python.

Вычисления:

- Код написан на C/C++ и распараллелен при помощи как MPI, так и OpenMP.

Минимизируется ошибка:

$$Z_j(\mathbf{x}, a, b, A) = S_{\text{exp}}(q_j) - \sum_{i=1}^n S_i(q_j) x_i - a (f_c(q_j))^2 - A \sum_{k=1}^{N_{\text{impur}}} \alpha_k (f_k(q_j))^2 - b, (j=1:m)$$

Дополнительные условия:

$$S_{\text{exp}}(q_j) - a (f_c(q_j))^2 - A \sum_{k=1}^{N_{\text{impur}}} \alpha_k (f_k(q_j))^2 - b \geq -0.5, (j=1:m), \quad \sum_{i=1}^n x_i + a = A, \quad x_i, a \geq 0$$

3 способа минимизации:

\mathbf{L}_1 : $\sum_{j=1}^m |Z_j(\mathbf{x}, a, b, A)|_{\mathbf{x}, a, b, A} \rightarrow \min$ - минимизация суммы абсолютных ошибок

\mathbf{L}_2 : $\sum_{j=1}^m (Z_j(\mathbf{x}, a, b, A))^2_{\mathbf{x}, a, b, A} \rightarrow \min$ - минимизация суммы квадратов ошибок

\mathbf{L}_{inf} : $\max_{j=1:m} |Z_j(\mathbf{x}, a, b, A)|_{\mathbf{x}, a, b, A} \rightarrow \min$ - минимизация максимальной абсолютной ошибки

Для случаев \mathbf{L}_1 и \mathbf{L}_{inf} был использован свободно доступный ЛП-солвер **LP_SOLVE**, (<http://lpsolve.sourceforge.net>)

Для случая \mathbf{L}_2 был использован свободно доступный пакет нелинейной оптимизации **Ipopt** (<http://projects.coin-or.org/Ipopt>)

Оптимизационные алгоритмы

Минимизируется ошибка:

Постоянный фон неизвестного происхождения

$$Z_j(\mathbf{x}, a, b, A) = S_{\text{exp}}(q_j) - \sum_{i=1}^n S_i(q_j) x_i - a (f_c(q_j))^2 - A \sum_{k=1}^{N_{\text{impur}}} \alpha_k (f_k(q_j))^2 - b, (j=1:m)$$

Эксперимент

Модельные кривые

Фон от углеводов и микропримесей

Дополнительные условия

$$S_{\text{exp}}(q_j) - a (f_c(q_j))^2 - A \sum_{k=1}^{N_{\text{impur}}} \alpha_k (f_k(q_j))^2 - b \geq -0.5, (j=1:m), \quad \sum_{i=1}^n x_i + a = A, \quad x_i, a \geq 0$$

3 способа минимизации:

Искомые величины: наиболее вероятное число атомов углерода, принадлежащих i-той структуре и углеводородному фону.

$$L_1: \sum_{j=1}^m |Z_j(\mathbf{x}, a, b, A)| \xrightarrow{x, a, b, A} \min - \text{минимизация суммы абсолютных ошибок}$$

Созданный численный код и предложенные методы решения могут быть применимы и для других задач идентификации структурного состава углеродных наноматериалов, а при некоторой модификации оптимизационной задачи, не только углеродных.

Для случаев L1 и Linf был использован свободно доступный ЛП-солвер LP_SOLVE, (<http://lpsolve.sourceforge.net/>)

Для случаев L1 и Linf имеет смысл предоставление широкого доступа к использованию кода и оптимизационных алгоритмов.

Программный инструментарий для:

- **быстро** преобразования **существующих приложений** (в том числе, Грид-приложений) в удалённо доступные Веб-сервисы
- интеграции сервисов в сложно-составные вычислительные сценарии - новые композитные сервисы, средствами "визуального программирования".
- управления выполнением созданных сценариев

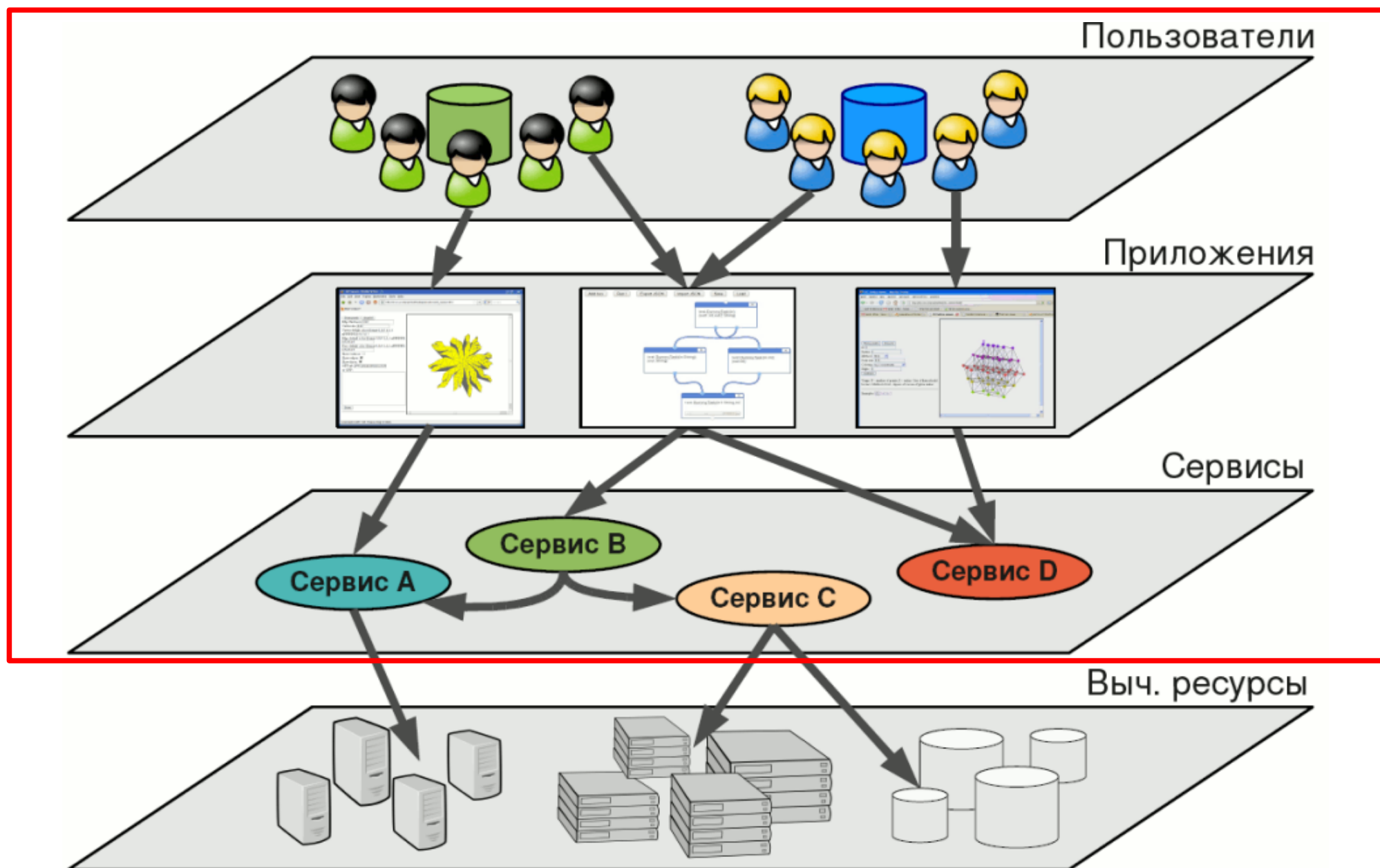
Свободно доступна в исходных кодах:

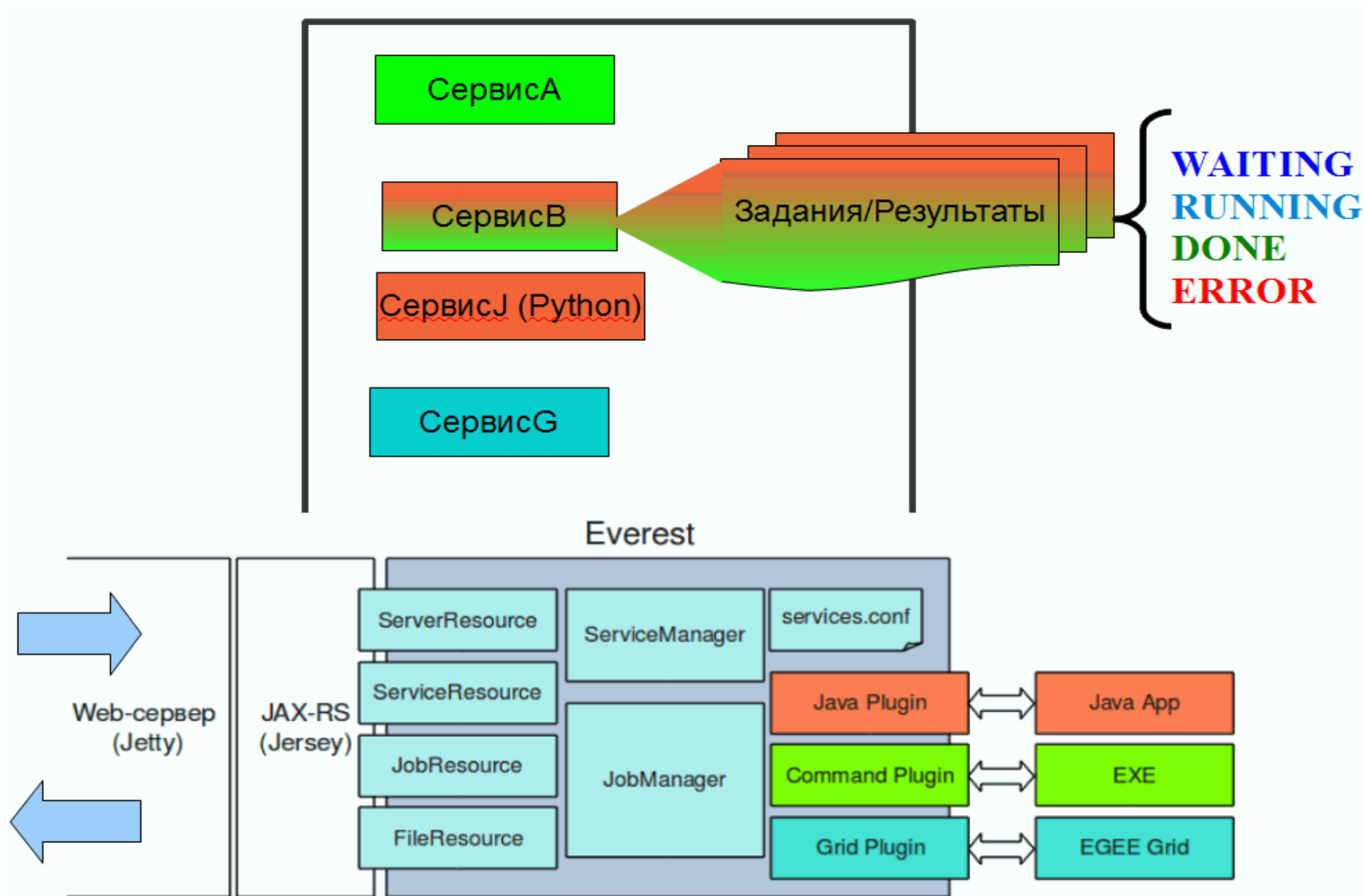
<https://www.assembla.com/spaces/mathcloud/documents> (Apache License, Version 2.0)

Публикации:

<https://www.assembla.com/wiki/show/mathcloud/Publications>

Инструментарий MathCloud. Архитектура.





Инструментарий MathCloud. Система управления сценариями.



- Каждый сервис MathCloud оформляется в виде Веб-сервиса, реализующего **унифицированный интерфейс**
- Этот интерфейс основан на:
 - спецификациях протокола HTTP (GET, POST, DELETE) для обмена сообщениями
 - архитектурном стиле REST (REpresentational State Transfer, принятом в Web 2.0)
 - распространённых форматах представления данных (JSON, HTML)

MathCloud - один из подходов к организации научной кооперации на принципах Web 2.0

Реализация в MathCloud. Подключение сервиса "command"

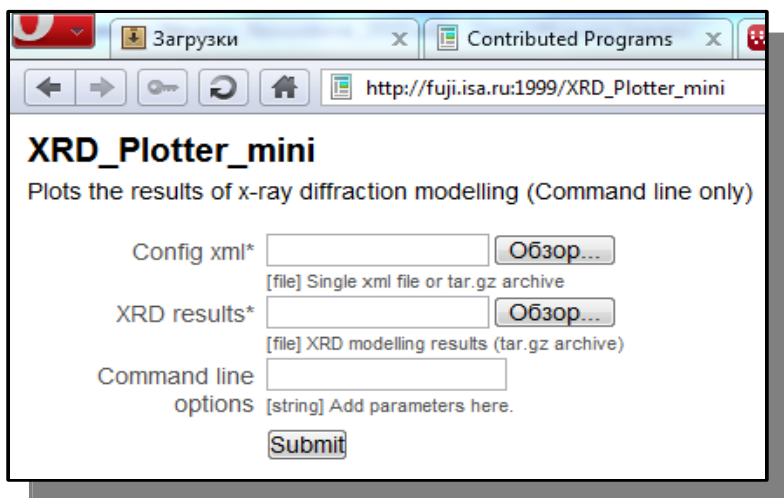
```
name : XRD_Plotter_mini,
description : "Plots the results of x-ray diffraction modelling",
inputs : {
  start : {type:file, title:"Config xml",
    description: "Single xml file or tar.gz archive"},
  results : {type:file, title:"XRD results",
    description: "XRD modelling results (tar.gz archive)"},
  CL : {type:string, title:"Command line options",
    description: "Add parameters here.", optional: true}
},
outputs : {
  pictures : {type:file, title:"Pictures"},
  out : {type:file, title:"Output"},
  err : {type:file, title:"Error"}
},
implementation : {
  type : command,
  command : "python /home/vneverov/XRDnano/plotXRD_web.py $CL",
  input2file:{
    start:start,
    results:results.tar.gz
  },
  file2output:{
    pictures.tar.gz:pictures,
    stdout:out,
    stderr:err
  }
}
```


Реализация в MathCloud. Подключение сервиса "grid"

```
name : XRD_Calculation,
description : "X-ray diffraction modeling (calc. in GRID)",
inputs : {
  N : {type: integer, title:"Number of calculations"},
  start : {type:file, title:"Config xml",
    description: "Single xml file or tar.gz archive"},
  structures : {type:file, title:"Structures archive",
    description: "tar.gz archive"}
},
outputs : {
  results : {type:file, title:"Results"},
  out : {type:file,title:"Output"},
  err : {type:file,title:"Error"}
},
implementation : {
  type : grid,
  vo : gilda,
  job-file : "/home/vneverov/XRDnano/startMass.jdl",
  pre : "python /home/vneverov/XRDnano/preXRD_calc.py",
  post : "/bin/bash /home/vneverov/XRDnano/postXRD_calc.sh",
  input2file:{
    start:start,
    structures:structures.tar.gz
  },
  file2output:{
    results.tar.gz:results,
    stdout.tar.gz:out,
    stderr.tar.gz:err
  }
}
```

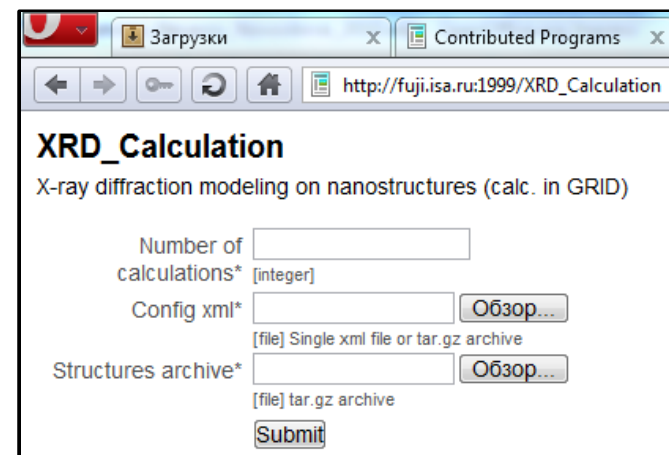
Сервисы (Веб-форма):

- **XRD_Calculation** — массовый расчет дифракционных характеристик наноструктур в гриде.



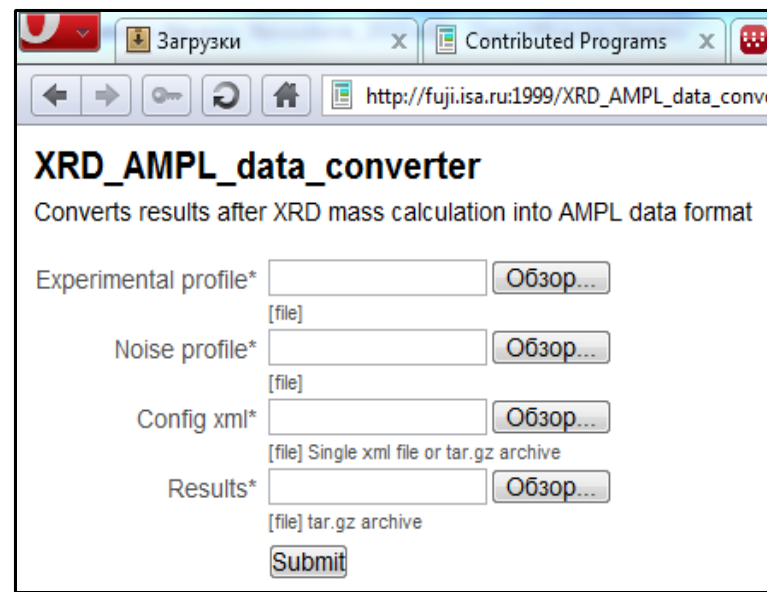
The screenshot shows the 'XRD_Plotter_mini' web form. The title is 'XRD_Plotter_mini' and the subtitle is 'Plots the results of x-ray diffraction modelling (Command line only)'. The form contains three main input sections: 'Config xml*' with a file selection button 'Обзор...', 'XRD results*' with a file selection button 'Обзор...', and 'Command line options' with a text input field and a 'Submit' button. The text input field has a placeholder '[string] Add parameters here.'.

- **XRD_AMPL_data_converter** — преобразование результатов массовых расчетов и экспериментальных данных в единый AMPL файл данных.



The screenshot shows the 'XRD_Calculation' web form. The title is 'XRD_Calculation' and the subtitle is 'X-ray diffraction modeling on nanostructures (calc. in GRID)'. The form contains three main input sections: 'Number of calculations*' with a text input field and a label '[integer]', 'Config xml*' with a file selection button 'Обзор...' and a label '[file] Single xml file or tar.gz archive', and 'Structures archive*' with a file selection button 'Обзор...' and a label '[file] tar.gz archive'. There is a 'Submit' button at the bottom.

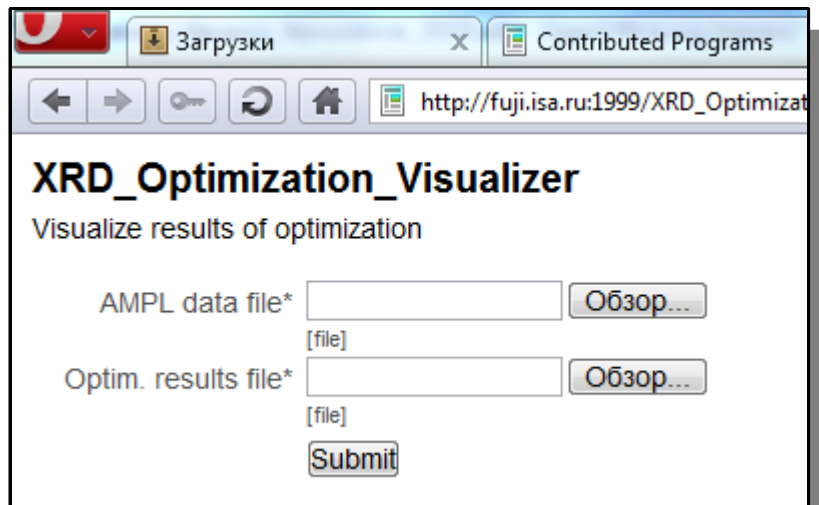
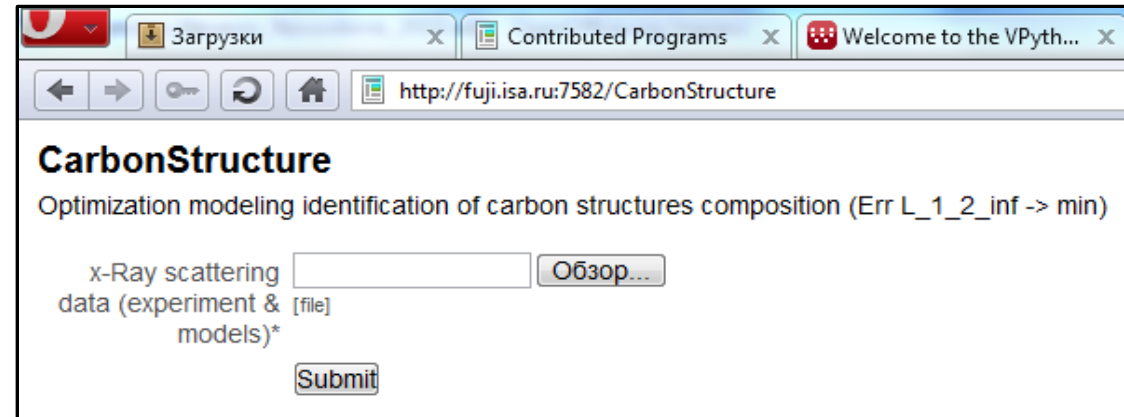
- **XRD_Plotter** — построение графиков по результатам расчетов.



The screenshot shows the 'XRD_AMPL_data_converter' web form. The title is 'XRD_AMPL_data_converter' and the subtitle is 'Converts results after XRD mass calculation into AMPL data format'. The form contains four main input sections: 'Experimental profile*' with a file selection button 'Обзор...' and a label '[file]', 'Noise profile*' with a file selection button 'Обзор...' and a label '[file]', 'Config xml*' with a file selection button 'Обзор...' and a label '[file] Single xml file or tar.gz archive', and 'Results*' with a file selection button 'Обзор...' and a label '[file] tar.gz archive'. There is a 'Submit' button at the bottom.

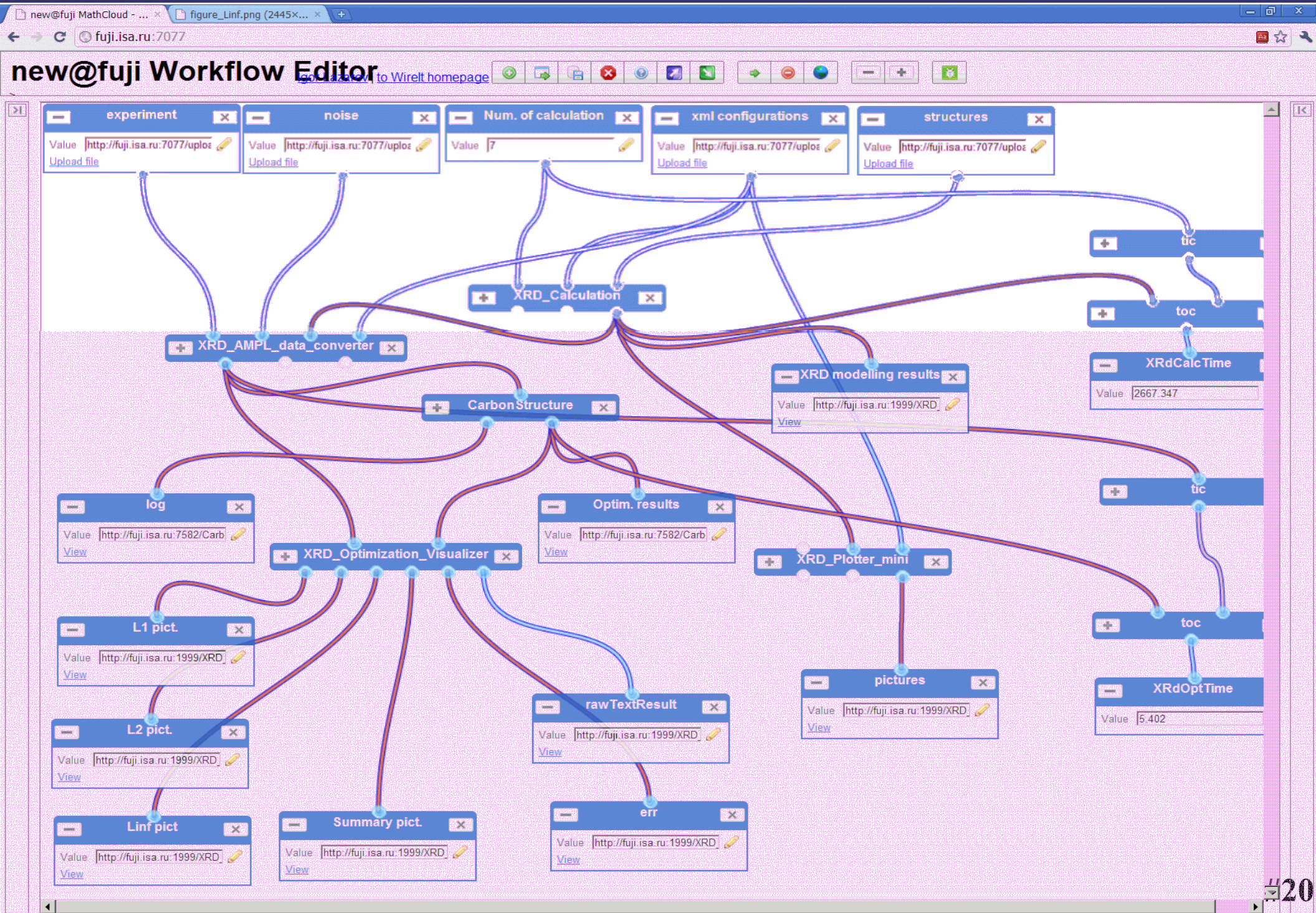
Сервисы (Веб-форма):

- **CarbonStructure** — оптимизационная идентификация по трем критериям, входной параметр — AMPL файл данных.

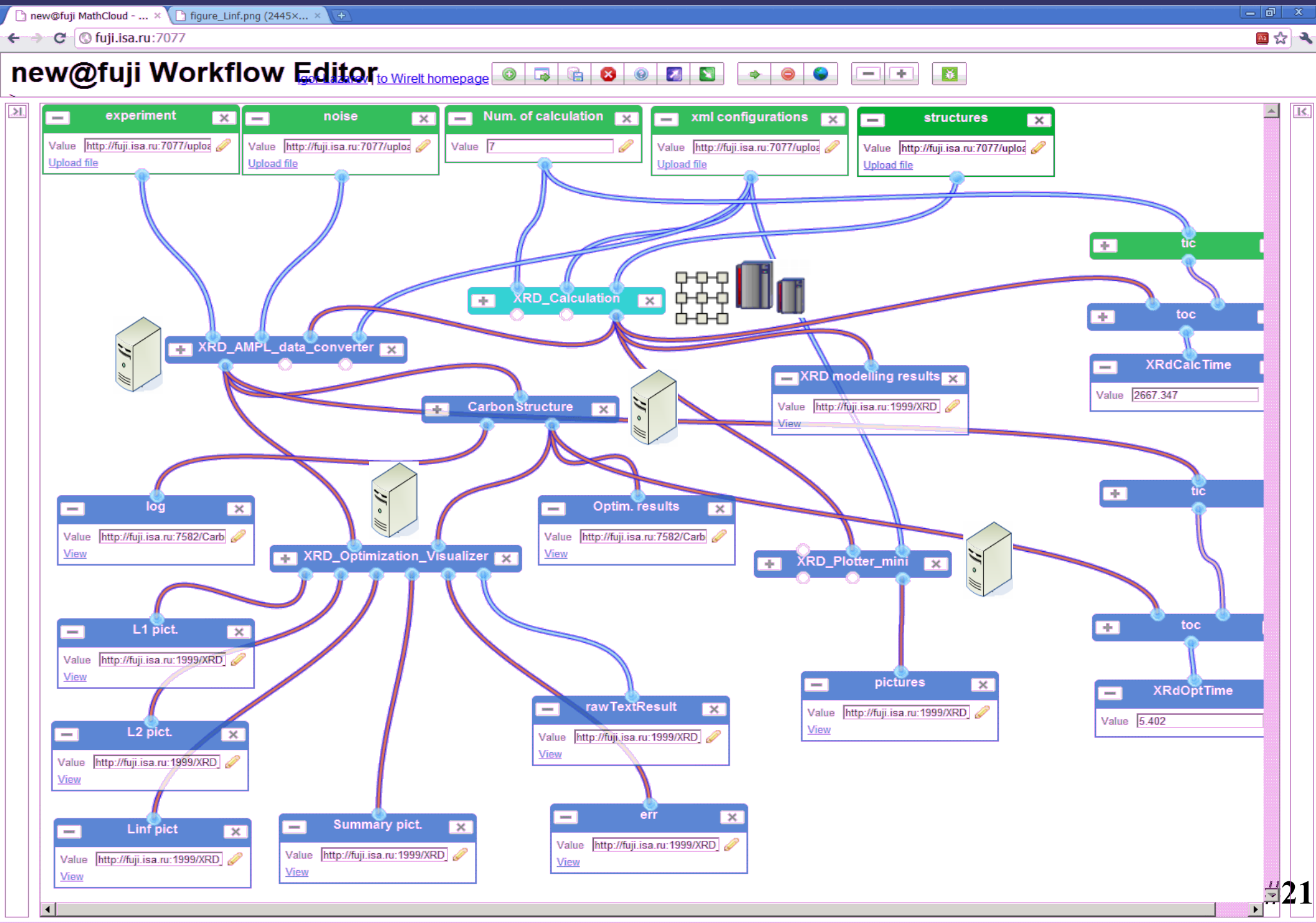


- **XRD_Optimization_Visualizer** — обработка и визуализация результатов оптимизации.

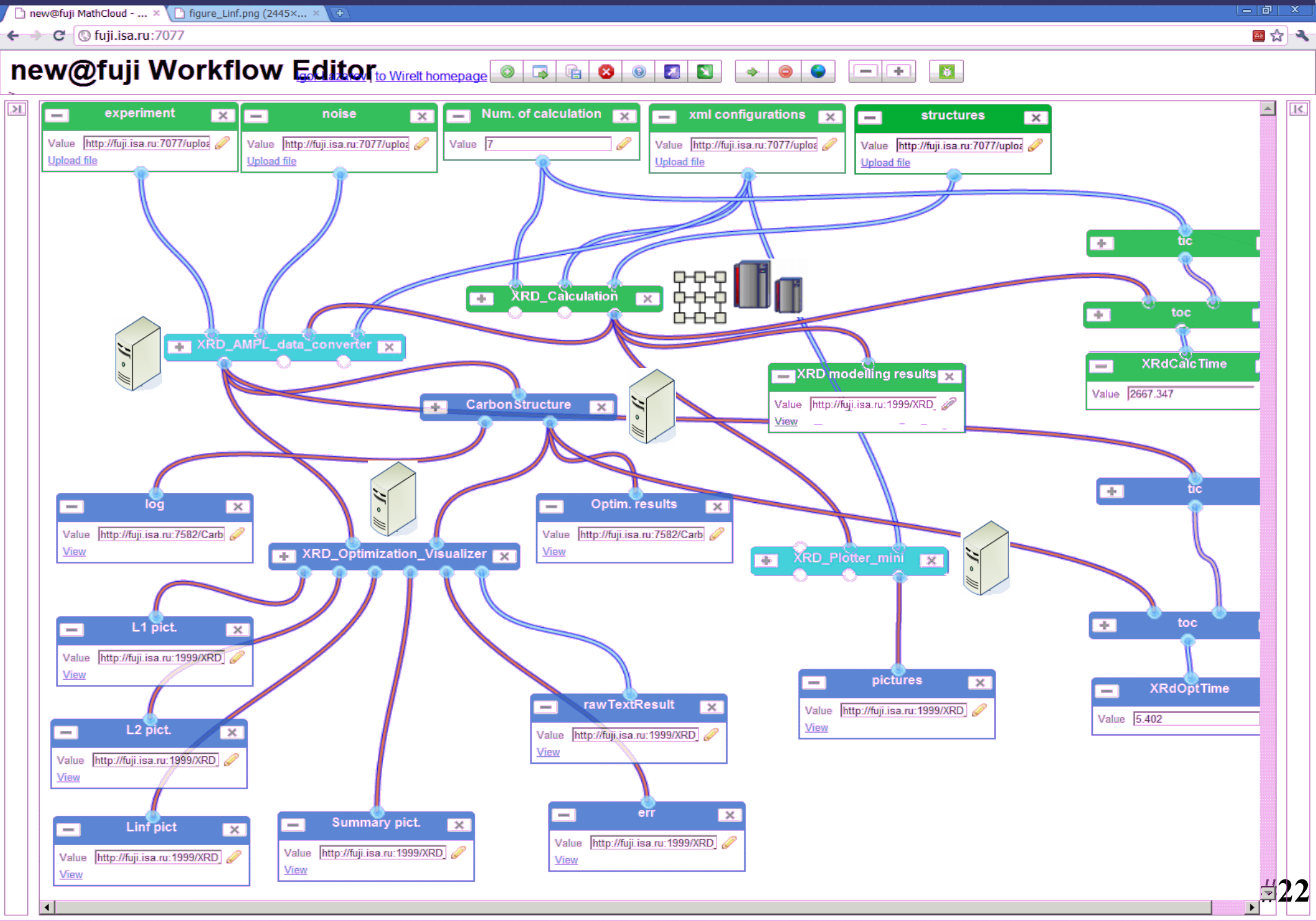
Система структурного анализа в MathCloud



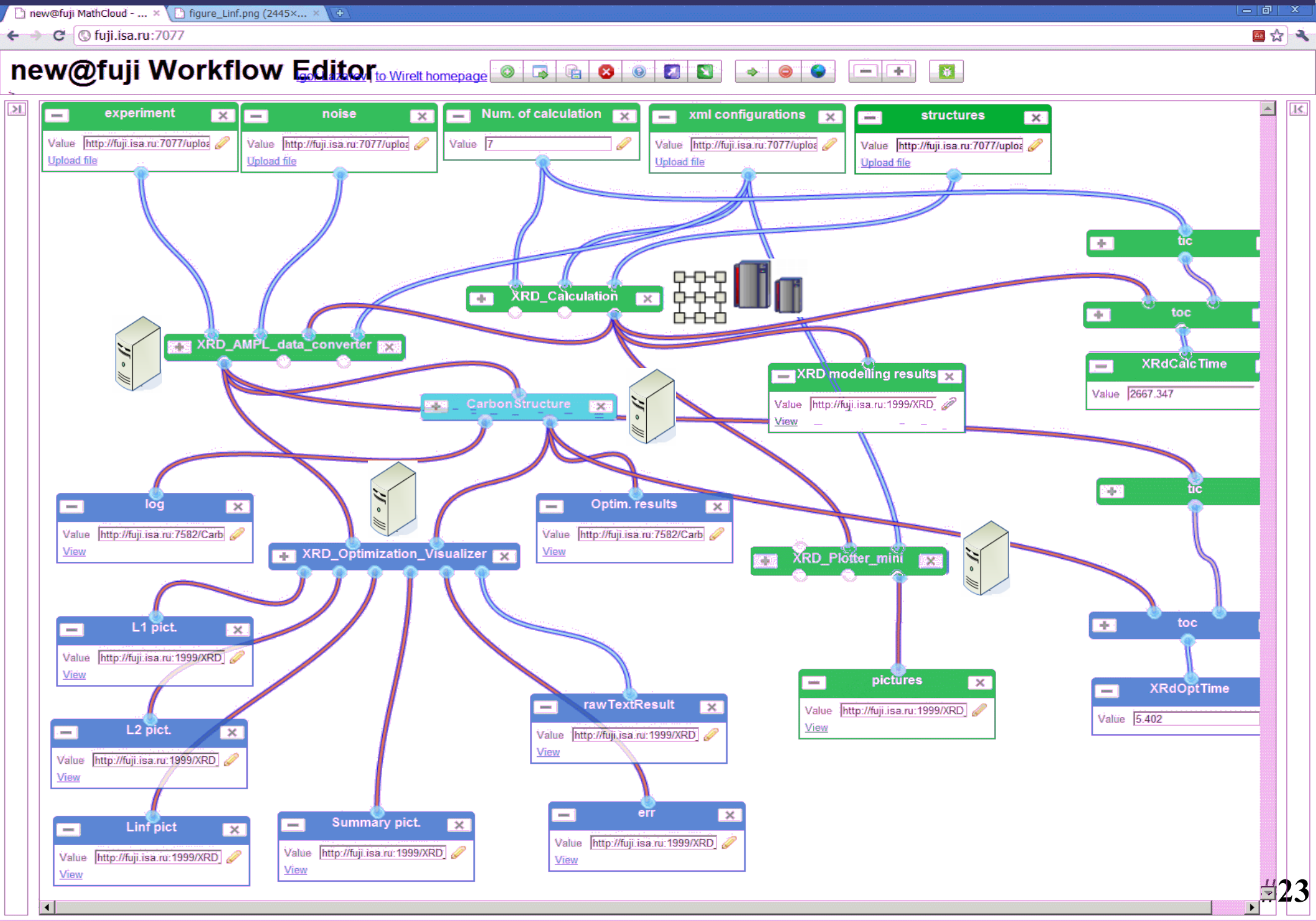
Система структурного анализа в MathCloud



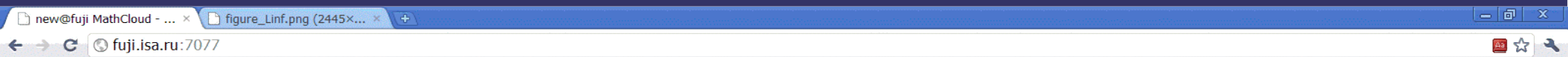
Система структурного анализа в MathCloud



Система структурного анализа в MathCloud

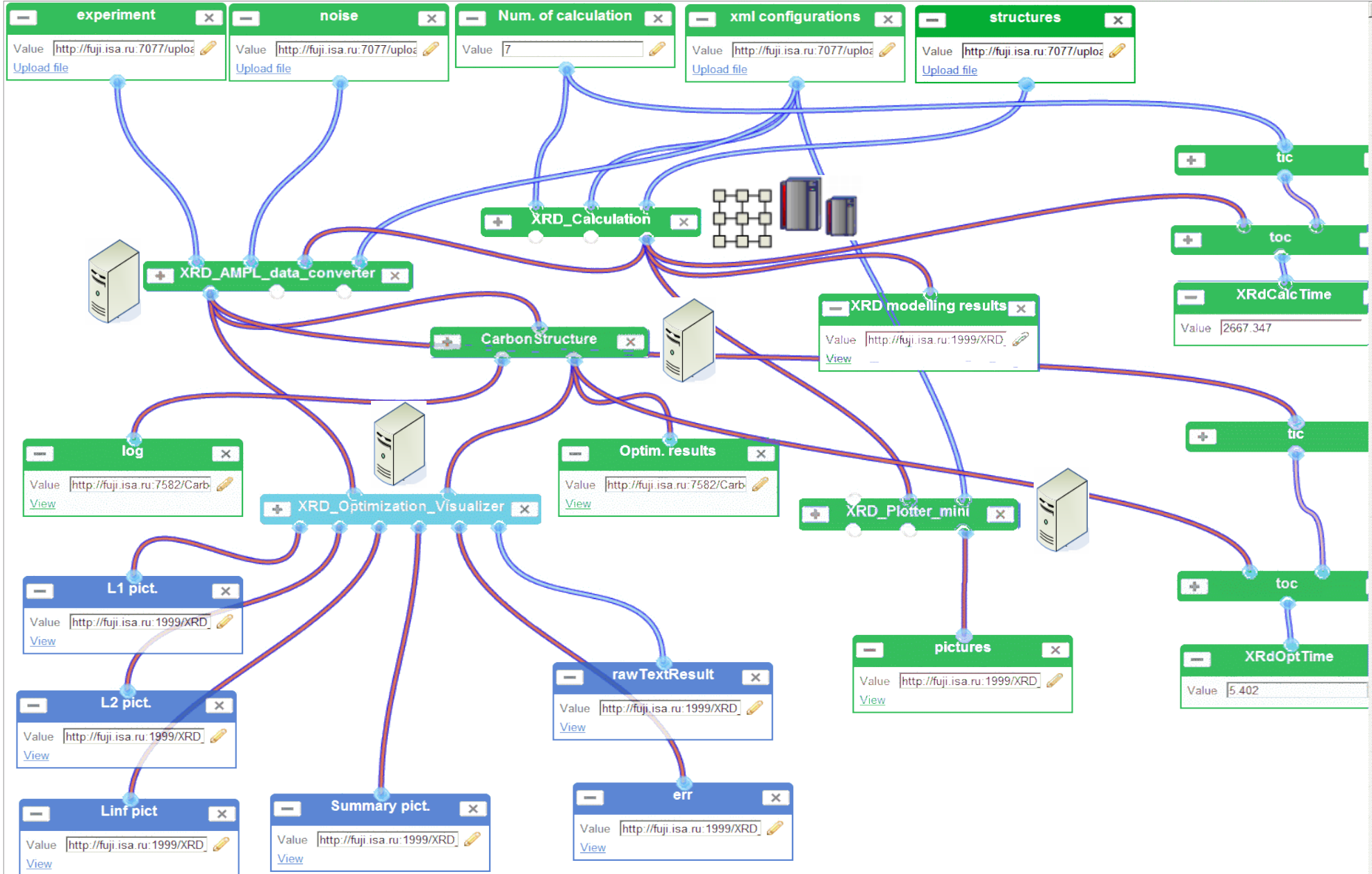


Система структурного анализа в MathCloud

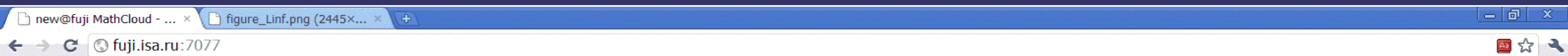


new@fuji Workflow Editor

[Go to Wiret homepage](#)

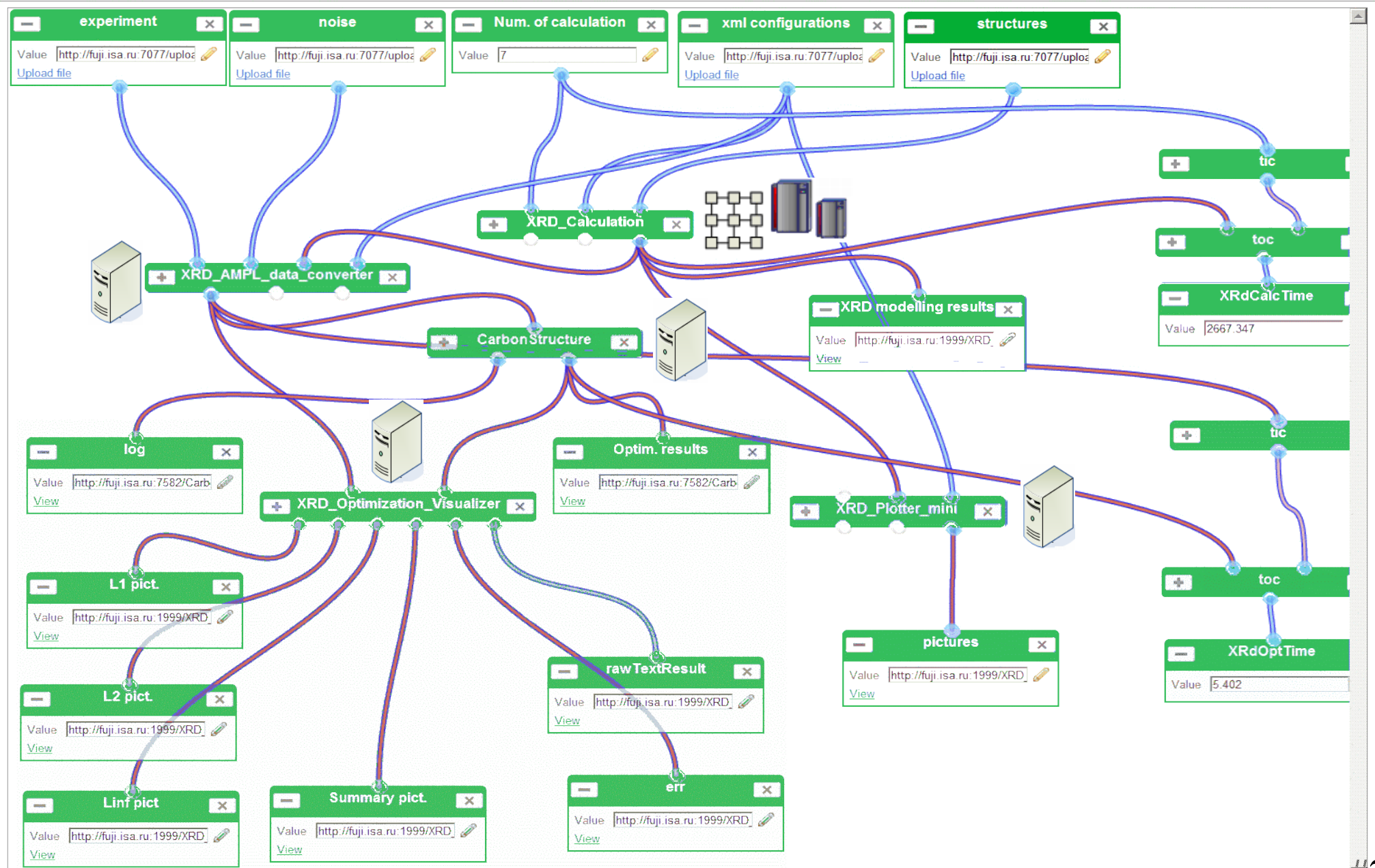


Система структурного анализа в MathCloud

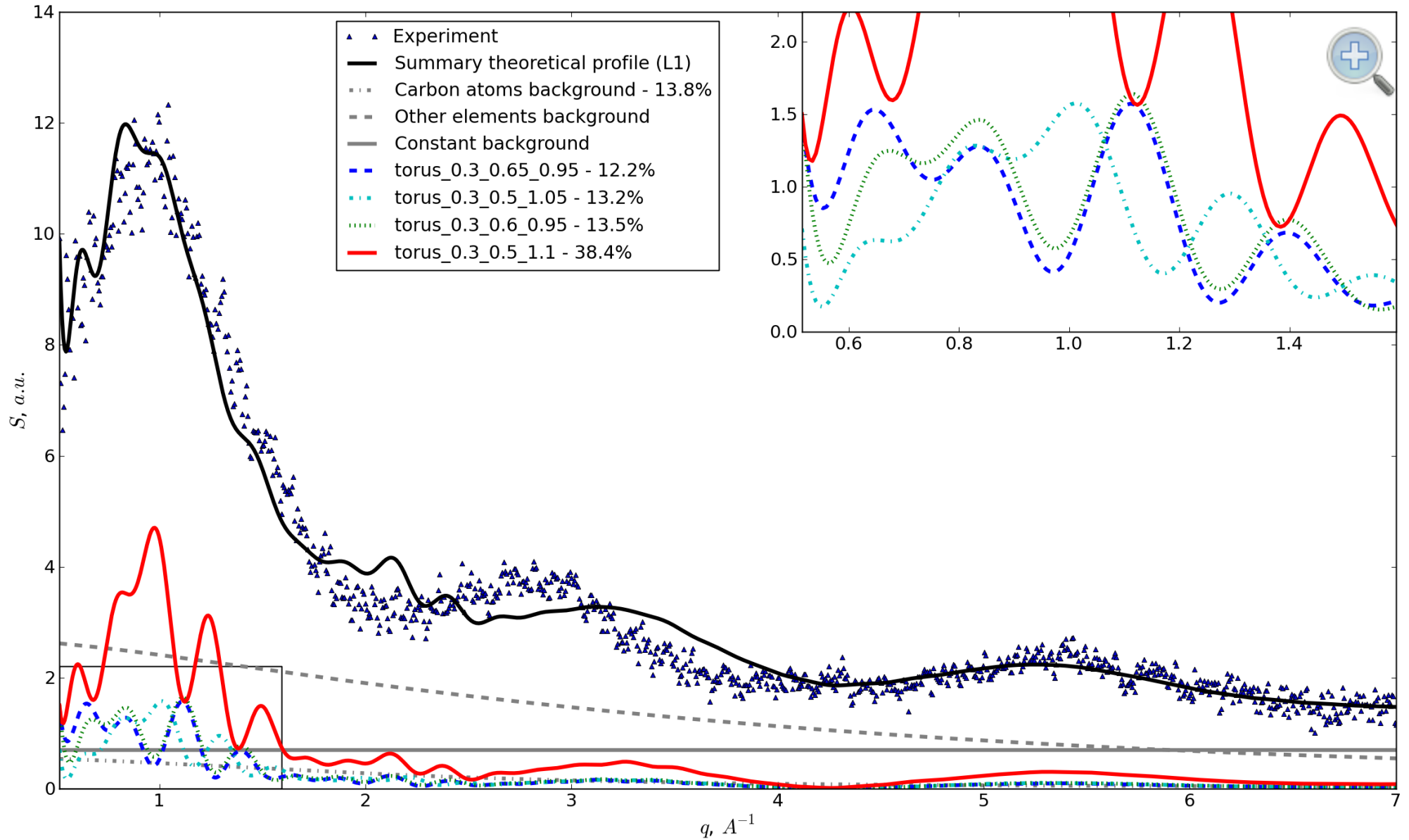


new@fuji Workflow Editor

[Go to MathCloud to WireIt homepage](#)

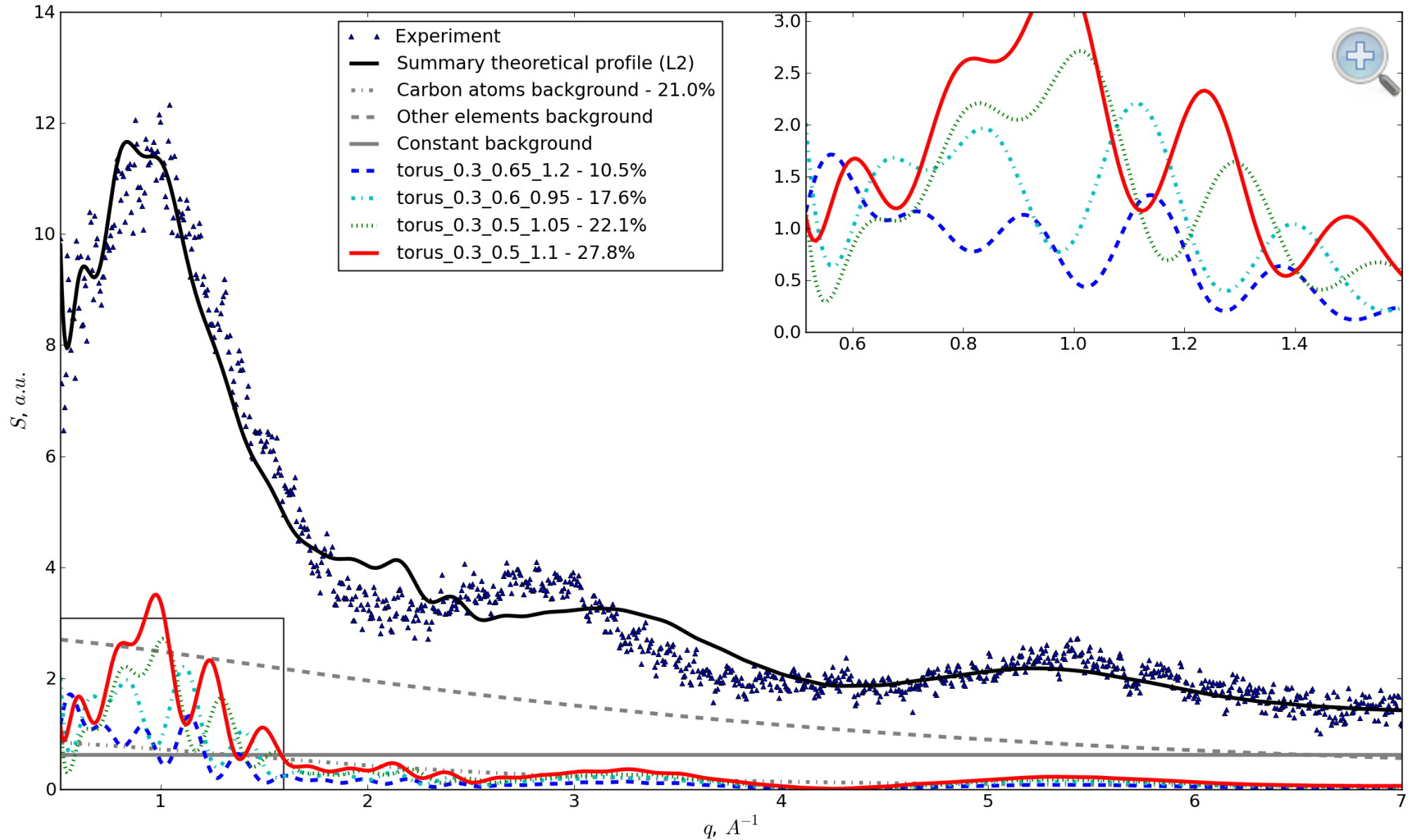


Визуализация результатов



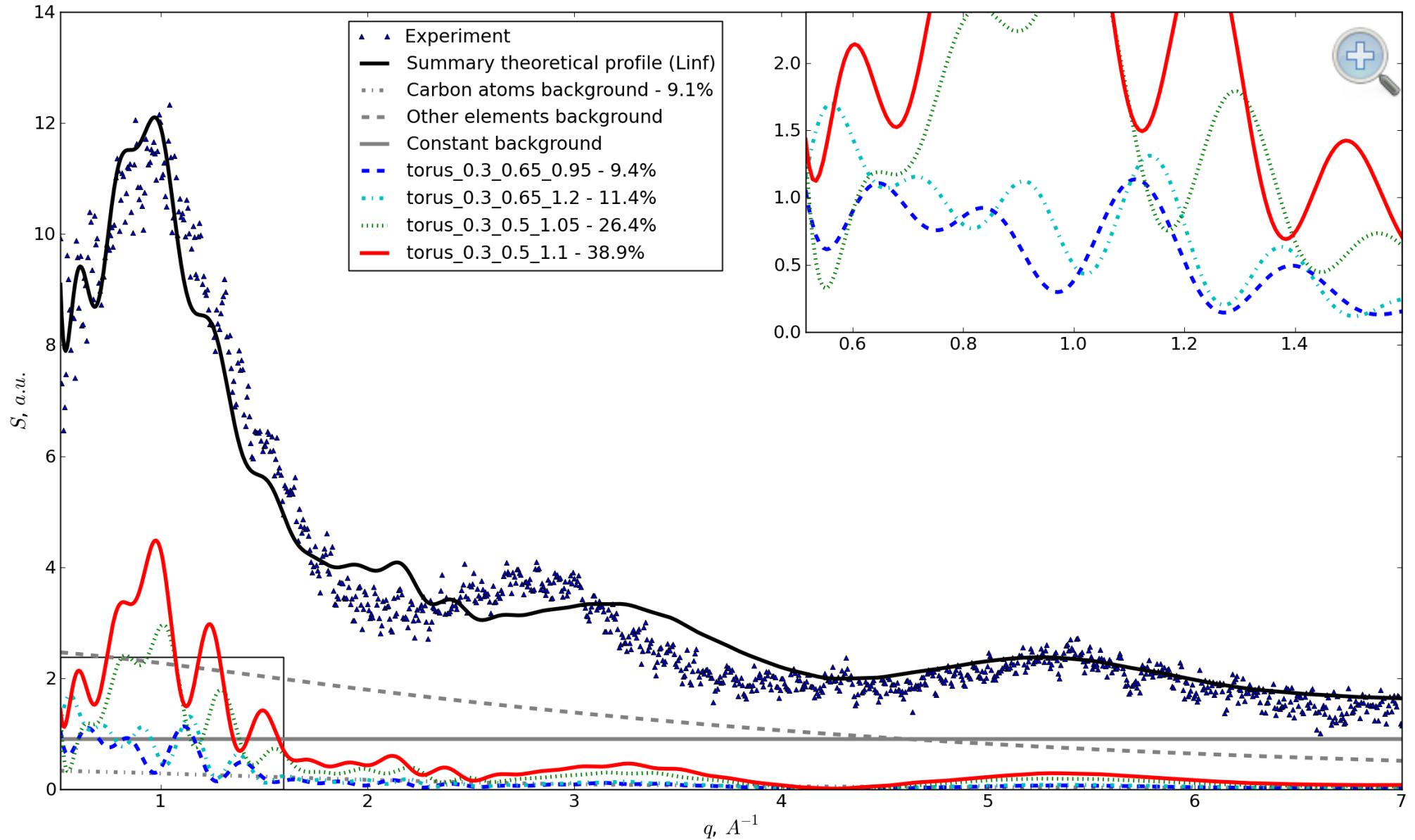
$\mathbf{L}_1: \sum_{j=1}^m |Z_j(\mathbf{x}, a, b, A)| \xrightarrow{x, a, b, A} \min$ - минимизация суммы абсолютных ошибок

Визуализация результатов



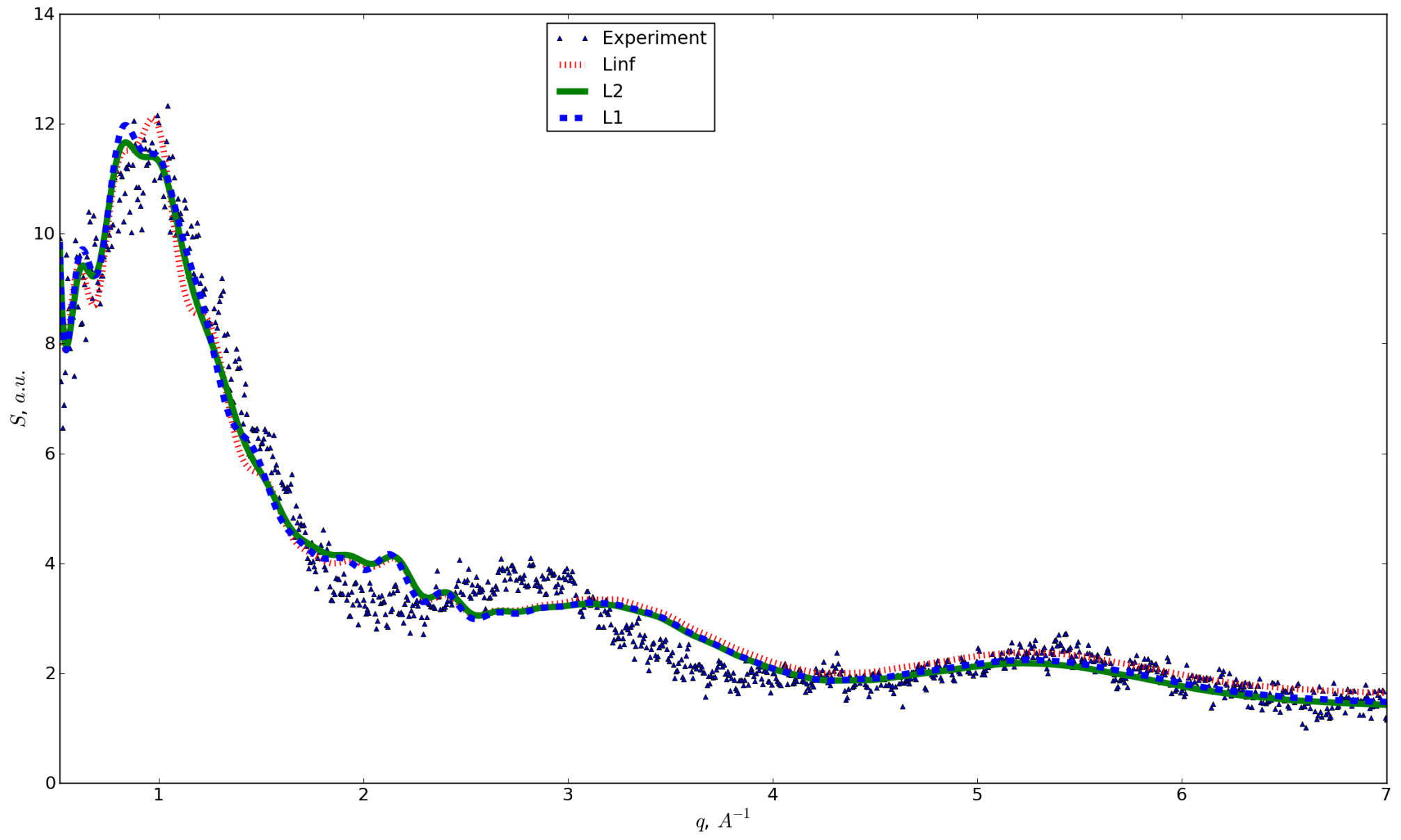
$L_2: \sum_{j=1}^m (Z_j(x, a, b, A))^2 \xrightarrow{x, a, b, A} \min$ - минимизация суммы квадратов ошибок

Визуализация результатов



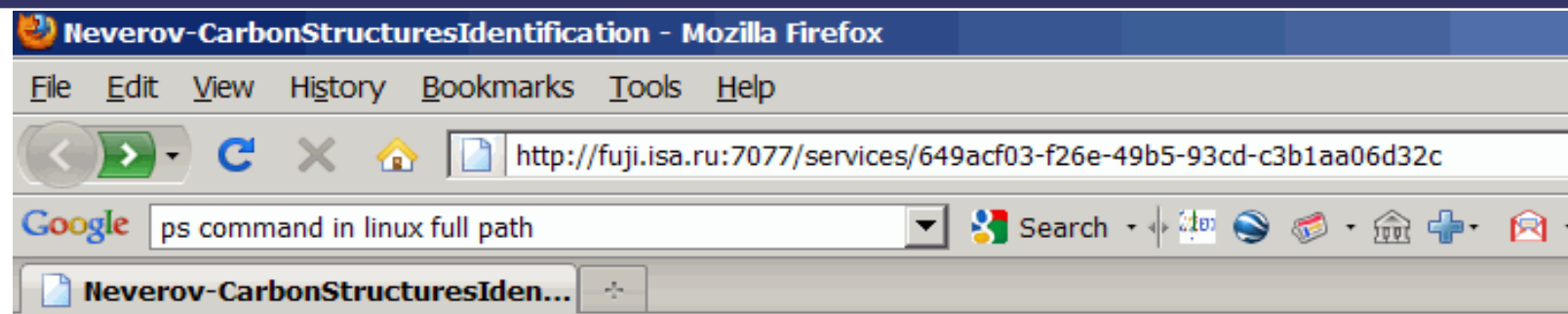
$L_{\text{inf}}: \max_{j=1:m} |Z_j(\mathbf{x}, a, b, A)|_{\mathbf{x}, a, b, A} \rightarrow \min$ - минимизация максимальной абсолютной ошибки

Визуализация результатов



Суммарный график

Веб-форма системы структурного анализа в MathCloud



Neverov-CarbonStructuresIdentification

Config xml*

[file] Single xml file or tar.gz archive

Number of calculations*
[integer]

Structures archive*

[file] tar.gz archive

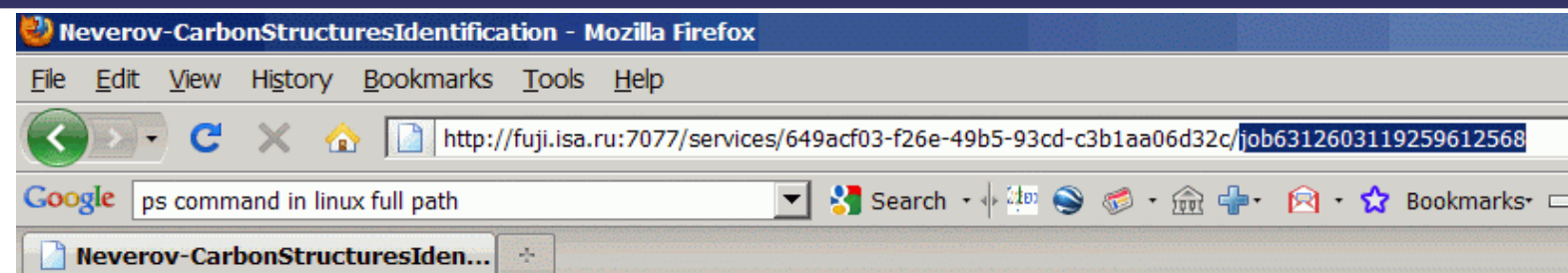
Experimental profile*

[file]

Noise profile*

[file]

Контроль выполнения. Веб-форма.



State: RUNNING

Info: Workflow state: RUNNING

Result:

workflowId: 649acf03-f26e-49b5-93cd-c3b1aa06d32c

start-time: 13 Nov 2010 18:09:27 416

tasks: [{"name":"task0 (XRD_Plotter_mini)",

"state":"WAITING"},

{"name":"task2 (XRD_Calculation)",

"state":"RUNNING",

"info":"Service request state: RUNNING. Info: Grid job status is RUNNING",

"start-time":"13 Nov 2010 18:09:27 716"},

{"name":"task6 (XRD_AMPL_data_converter)",

"state":"WAITING"},

{"name":"task9 (CarbonStructure)",

"state":"WAITING"},

{"name":"task10 (XRD_Optimization_Visualizer)",

"state":"WAITING"},

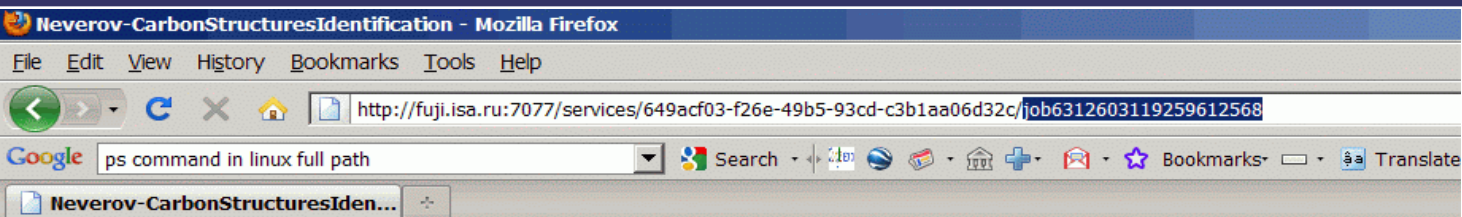
{"name":"task20 (tic)",

"state":"DONE",

"start-time":"13 Nov 2010 18:09:27 717",

"finish-time":"13 Nov 2010 18:09:27 720"},

Результаты выполнения задания. Веб-форма.



State: DONE

Info: Workflow state: DONE

Result:

Pictures	file	
L1 figure	file	
L2 figure	file	
Linf figure	file	
Summary figure	file	
Optimization output file (L_1_2_inf)	file	Results
Results	file	
Output	file	
Runtime log	file	This is the raw output of script
Error	file	
XRdCalcTime	2151.68	
XRdOptTime	5.401	

Заключение

- Реализация вычислительного сценария идентификации углеродистых структур по рентгеновским дифракционным характеристикам наноматериалов в MathCloud, как мы надеемся, позволит в будущем **сторонним исследователям использовать наш подход** при обработке соответствующих экспериментальных данных.
- **Имеющиеся средства инструментария MathCloud** для модификации вычислительных сценариев позволяют потенциальным пользователям легко **совершенствовать и дополнять** созданное распределённое приложение.

Благодарности

Авторы выражают благодарность за помощь в создании работы:

А.Б. Кукушкину, В.А. Вознесенскому, О.В. Сухорослову, А.А. Велигжанину, Я.В. Зубавичусу, Н.Ю. Свечникову и В.Г. Станкевичу.

Работа поддержана РФФИ (проекты 09-07-00469 и 08-07-00430-а) и Европейским проектом EGEE-III (Enabling Grids for E-sciencE, "Развёртывание Гридов для развития вычислительноёмкой науки")