Annual Report

Funding Programme:	**Helmholtz Joint Research Groups**
Project ID No.:	HRSF-0002
Project Title:	**New avenues in information and data science: advanced imaging applications at the XFEL and cryo-EM frontier**
Principal Investigator:	Prof. Dr. I.A. Vartaniants
Report Period (=Calendar Year):	01/2018-12/2018

1) Group Structure

Please report briefly on the structure and personnel development of your group.

Group members on German side:

Prof. Dr. Ivan A. Vartaniants, PhD, senior scientist, group leader, DESY.
Prof. Dr. Wilfried Wurth, PhD, leading scientist, DESY, Hamburg University
Dr. Adrian Mancuso, group leader, Eu. XFEL
Dr. Sergey Lazarev, staff scientist, DESY
Dr. Max Rose, PhD, staff scientist, DESY
Dr. Luca Gelisio, PhD, staff scientist, DESY
Dr. Young Yong Kim, staff scientist, DESY
Mr. Ruslan Khubbutdinov, PhD student, DESY
Mr. Dmitry Lapkin, PhD student, DESY
Mrs. Dameli Assalauova, PhD student DESY
Dr. Ruslan Kurta, staff scientist, Eu. XFEL
Dr. Giuseppe Mercurio, staff scientist, Eu. XFEL

Group members on Russian side:

Prof. Dr. Vyacheslav Ilyin, PhD, leading scientist, NRC "Kurchatov Institute"
Dr. Alexander Vasiliev, PhD, leading scientist, NRC "Kurchatov Institute"
Anton Teslyuk, leading scientist, NRC "Kurchatov Institute"
Dr. Sergey Bobkov, PhD, staff scientist, NRC "Kurchatov Institute"
Ksenia A. Ikonnikova, PhD student, NRC "Kurchatov Institute"
Timur N. Baymukhametov, PhD student, NRC "Kurchatov Institute"
Evgeniy B. Pichkur, PhD student, NRC "Kurchatov Institute"
Yury M. Chesnokov, PhD student, NRC "Kurchatov Institute"
Sergey I. Zolotarev, PhD student, NRC "Kurchatov Institute"

On German side two PhD students were hired in the frame of this Project, these are: Dameli Assalauova and Dmitry Lapkin.
In the year 2018 Max Rose has defended his PhD thesis on the German side and S. Bobkov defended his thesis on the Russian side.

2) Network/ Meetings

Please describe how the group works together. Have there been any international meetings organized by or attended by the group? What is the contribution of the group to the networking of international partners and the Helmholtz Centre?

The group was actively working together. We have two Workshops organized on a German (DESY) and Russian (Moscow) sides.

First Kick-off Meeting was organized at DESY on February 15-16, 2018.
Second Meeting was organized at Moscow State University on November 27-28 2018.
See for the schedule of the Meetings Attachments.

Besides that we have few short term visits from German side to Russia and from Russia to Moscow for discussions and updates of the project as well as for participation of European XFEL and DESY Photon Science user Meetings and experiments performed at European XFEL.

We had a longer term (three months) visit of S. Zolotarev (NRC "Kurchatov Institute") to DESY, when he was working on the data analysis obtained in the first single particle imaging (SPI) experiments at European XFEL.

Results of the research performed in the frame of this Project were reported on a number of National and International Conferences and Workshops (see for the list Appendixes).

3) Scientific Progress / Milestones

How has your work plan progressed? What important milestones could be achieved during the report period? Is the progress of your work in accordance with original planning or has the work plan been changed?

Description of work performed in the reporting year and scientific results obtained

A joint project supported by the Russian Science Foundation (RSF, grant № 18-41-06001) and the Helmholtz Association (Germany, HRSF-002 grant) has been carried out by the research teams of DESY (Hamburg) and European XFEL (Schenefeld) on German side and NRC "Kurchatov Institute" (Moscow) on the Russian side from the beginning of 2018 on the topic "New avenues in information and data science: advanced imaging applications at the XFEL and cryo-EM frontier".

Two workshops were organized at DESY, Hamburg (Germany) on February 15-16 (2018) and in Moscow on November 27-28 (2018). These Workshops were attended by the majority of the project participants from the Russian and German sides. The list of talks presented at these Workshops may be found in the attached files.

Members of the Russian team (A. Teslyuk and S. Bobkov) together with the members of the German team (I.A. Vartaniants, L. Gelicio, M. Rose, Y. Y. Kim) participated in the p2013

XFEL experiment in December 2017. S. Zolotarev participated in joint work with the DESY team on processing the first data of the "EuXFEL SPI proposal p2013" experiment and participated in the "EuXFEL FXE proposal 2111" experiment on August 23-27.

The Kurchatov Institute is planning to participate in the experiments SPI p2145 and SPI p2146 at the European XFEL in May-June 2019.

In August 2018 M. Rose has defended his thesis on: “Coherent X-ray diffractive imaging of biological samples in 2D and 3D with synchrotron and XFEL radiation” (scientific advisors: Prof. Dr. I. A. Vartaniants (DESY) and Prof. Dr. W. Wurth (DESY)). In October 2018, S. Bobkov defended his PhD thesis on "Classification of diffraction patterns of biomolecules by analyzing their 3D structure using machine learning methods" (scientific advisor V.A. Ilyin (Kurchatov Institute), scientific consultant I.A. Vartanyants (DESY)). The main results of the dissertation work were the basis for the development of the program platform within the project, as well as the basis for the development of the project in the year 2018. The development work in the year 2018 was carried out by A.B. Teslyuk, S.A. Bobkov, E.B., Pichkur, and T.N. Baimukhamedov.

The main goal of the project is to provide data processing and analysis of streaming data obtained in experiments at the European X-ray Free-Electron Laser (European XFEL) and modern cryogenic electron microscopes (cryo-EM) to study the three-dimensional (3D) structure of single nanoscale particles (viruses, large biological molecules and other objects) in the quasi-real-time regime of these data acquired at these experimental facilities.

In 2018, the following results were obtained:

The review and analysis of algorithms and software used for processing experimental data at XFELs and cryo-EM is prepared. The list of references includes 53 publications. The focus was on the platform components of the conveyor data processing at the European XFEL and cryo-EM, which were used in the project development in the year 2018.

On the basis of the computer resources of the NRC Kurchatov Institute a hardware-software stand was created using: containers *Docker Community Edition* version 18.03.1-ce with the support for *Nvidia* accelerators; control systems in a cluster of containers and provide network connectivity to containers based on *Kubernetes* 1.11.0; the virtual network infrastructure on the basis of Flannel 0.10. The results of testing showed excellent prospects for containerization of the platforms under developing with the complete elimination of the manual work. The use of the containerization technology in the *Docker* system solution for the

execution of application modules and their orchestration during the operation of the pipeline platform will allow to build complex composite applications efficiently and provide fast and flexible management of the platforms performance. This technological solution will allow automating the deployment of containerized applications on different computing platforms, providing horizontal scaling, flexible management of the cluster size depending on the load, abstracting the host system level from the application level.

The following machine learning methods were considered for the development in the year 2018:

- maximum likelihood principle for the method of advanced filtering of diffraction patterns
- continuous wavelet transform (*CWT*) based method for estimating the size of the detected particle
- *FSC* (Fourier shell correlation) method for resolution estimate of 3D structure, as well as a half-bit evaluation criterion correlation threshold based on information-entropy approach
- method of linear regression for the alignment of a structure when calculating *FSC*
- support vector machine (*SVM*) for the classification of diffraction patterns
- methods of convolutional neural networks for classification of diffraction images
- K-means method for clustering diffraction images.

These methods were tested on examples of data processing of single particle imaging (*SPI*) experiments performed at the LCLS (Stanford) and cryo-EM microscopes at the NRC "Kurchatov Institute" and CNRS (Grenoble). Testing has shown their high efficiency at different stages of the full cycle of data processing of *SPI* experiments.

An architecture of the pilot software platforms for pipeline processing of streaming data at the European XFEL and cryo-EM was developed. On this basis, the relevant pilot software platforms were created, which were installed at the supercomputer resources of the Kurchatov Institute. With their use, full data processing of the *SPI* experiments at the European XFEL, and experiments at cryo-EM microscopes at the NRC "Kurchatov Institute" and CNRS (Grenoble) for various nanoscale biological objects (viruses) were carried out.

A performance map (calculation and data transfer rates, initial, final and intermediate

data sizes for each application module in these platforms) of these pilot pipeline platforms at the supercomputer resources of the NRC Kurchatov Institute were obtained.

New modules of preprocessing of the first data of the *SPI* experiment (Proposal p2013, December 2017 at the European XFEL) were developed, which allowed to take into account the technical features of the detector used in this experiment. In particular the developed modules of the preprocessing module, the calibration of the detector, the filtering module blank images, the module correction of the background scattering and filtering images by their size. In addition, a module for assessing the resolution of the structure has been developed.

According to the results obtained in 2018, six articles indexed by international databases WoS and Scopus were published. Three publications are included in the first quartile Q1.

DESY contribution to the project implementation

The workflow of the SPI data analysis based on the data obtained at LCLS (Stanford, USA) and European XFEL (Hamburg, Germany) was developed (see Rose *et al.,* IUCrJ (2018)). This workflow consists of the following steps: intensity filtering, classification by the principle component analysis, filtering by the particle size distribution, orientation determination without symmetry constrains, and the final step of the phase retrieval. Special selection procedure based on the Principle Component Analysis (*PCA*) to improve data quality was used. On the next step these data, which were measured at the Linac Coherent Light Source (LCLS) XFEL, were used for orientation determination and three-dimensional (3D) reconstruction of the virus electronic density. The conducted analysis allowed us to identify and quantify nanoscale features of the PR772 virus with a resolution better than 10 nm, which is currently a world record. It was clearly demonstrated that the structure is shrinking in one direction and, as such, does not possess the ideal icosahedral symmetry.

In addition, a working group at DESY worked on the development of various methods of diffraction data analysis and, especially, on the method of angular X-ray cross-correlation analysis (XCCA), which has been applied to the analysis of phase transitions smectic-A−hexatic-B of the first order in the compound of liquid crystals (see Zaluzhnyy *et al.,* Phys. Rev. E (2018)). X-ray investigations revealed a discontinuous first-order transition into the hexatic phase. The temperature region of two-phase coexistence near the phase transition point diminishes with decreasing film thickness. The width of this temperature region as a function of the film thickness was derived on the basis of a Landau mean-field theory in the

vicinity of a tricritical point (*TCP*). Close to *TCP* the surface hexatic-B order penetrates anomalously deep into the film interior.

The work [Rose *et al.,* IUCrJ (2018)] includes among the authors of the project participants from both teams of the project: DESY and RSC "Kurchatov Institute". The members of the German team did the main work of analyzing the experimental data. At the same time contribution of the Russian team consisted mainly in the development and application of classification methods. That emphasizes the high degree of complementarity between the Russian and German teams in the implementation of this project. This work was also highlighted at the DESY Photon Science web site.

The head of the DESY team, I.A. Vartaniants, participated in the international conference XTOP 2018 - the 14th Biennial Conference on High Resolution X-Ray Diffraction and Imaging (http://www.ba.ic.cnr.it/ xtop2018/, Bari, Italy, 3rd to 7th September 2018). At this conference, he made an oral presentation on the basis of the results published in the work [Rose *et al.,* IUCrJ (2018)], where collaboration between German and Russian teams on this project was specially acknowledged.

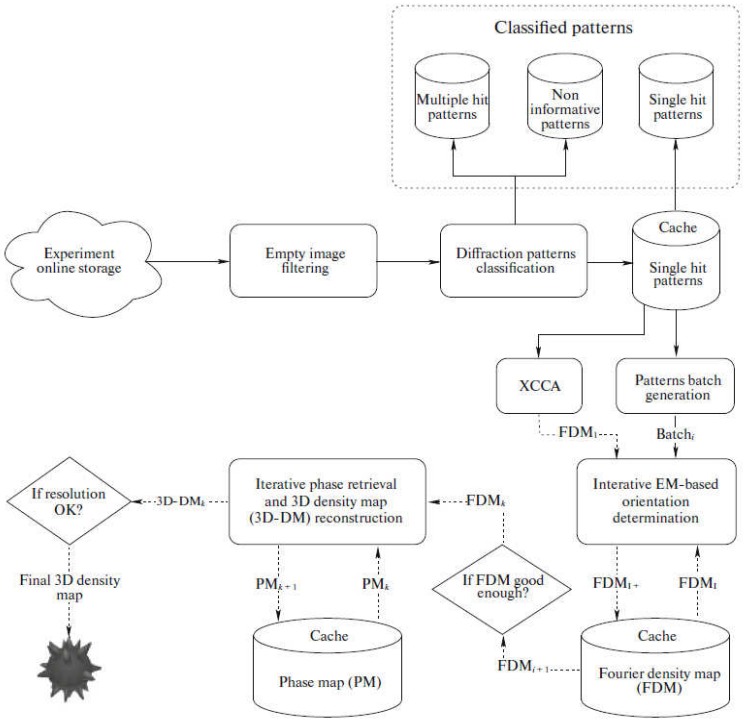

Fig. 1. The workflow of the data processing and analysis for the single particle imaging (SPI) experiments at the X-ray Free-Electron Lasers.

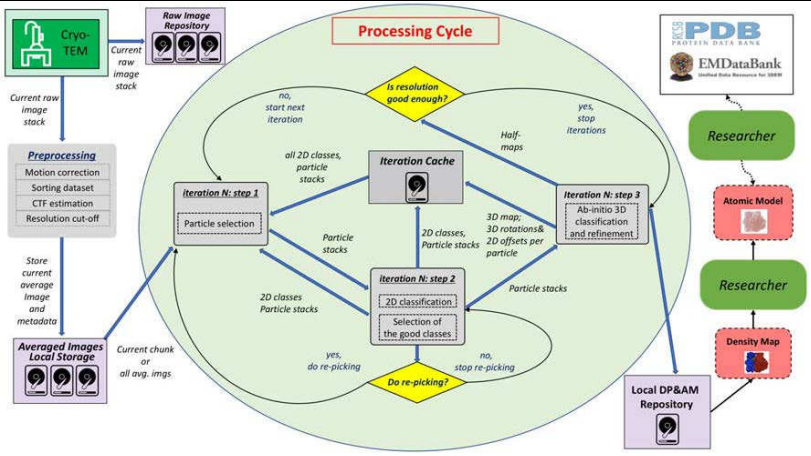

Fig. 2. The workflow of the data processing and analysis for the single particle imaging (SPI) experiments at the cryo-Electron Microscopes.

4) Financial Plan / Time Schedule

Can you comply with the financial plan and time schedule or do you see a need for adjustment?

Financial plan and time schedule looks fine and we do not see any needs for adjustment.

It was some left over of consumables cost money, which was transferred to this year and will be spend this year.

5) Publications of the Group

List of publications in the year 2018 with acknowledgements to grant funding

1. Baymukhametov T.N., Chesnokov Y.M., Pichkur E.B., Boyko K.M., Tikhonova T.V., Myasnikov A.G., Vasiliev A.L., Lipkin A.V., Popov V.O., Kovalchuk M.V. "Three-Dimensional Structure of Cytochrome c Nitrite Reductase as Determined by Cryo-Electron Microscopy" Acta Naturae **10**, 48-56 (2018).
2. Bobkov S.A., Teslyuk A.B., Zolotarev S.A., Rose M., Ikonnikova K.A., Velikhov V.E., Vartanyants I.A., Ilyin V.A. "Software Platform for European XFEL: Towards Online Experimental Data Analysis" Lobachevskii Journal of Mathematics, **39**, 1170–1178 (2018). doi: 10.1134/S1995080218090093
3. Pichkur Evgeny, Baimukhametov Timur, Teslyuk Anton, Orekhov Anton, Kamyshinsky Roman, Chesnokov Yury, Ilyin Viacheslav, Vasiliev Alexander. Velikhov Vasily Towards on-the-fly Cryo-Electron Microscopy Data Processing by High Performance Data Analysis. Journal of Physics: Conf. Series **955** 012005 (2018). doi :10.1088/1742-6596/955/1/012005

4. Rose M., Bobkov S., Ayyer K., Kurta R.P., Dzhigaev D., Kim Y.Y., Morgan A.J., Yoon C.H., Wesphal D., Bielecki J., Sellberg A., Williams G., Maia F.R.N.C., Yefanov O.M., Ilyin V., Mancuso A.P., Chapman H.N., Hogue B.G., Aquil A., Barty A., Vartanyants I.A., "Single-particle imaging without symmetry constraints at an X-ray free-electron laser" IUCrJ **5**, 727-736 (2018). doi: 10.1107/S205225251801120X. Impact factor: 6.575.
5. Boyko Konstantin M., Baymukhametov Timur N., Chesnokov Yuri M., Hons Mishael, Leshchekina Sofya V., Konarev Petr V., Liprin Alexey V., Vasiliev Alexandre L., Masson Patrick, Popov Vladimir O., Kovalchuk Michail V. "3D structure of the natural tetrameric form of human butyrylcholinesterase as revealed by cryoEM, SAXS and MD" Biochimie **156**, 196-205 (2019). doi: 10.1016/ j.biochi.2018.10.017. Impact factor: 3.188.
6. Ivan A. Zaluzhnyy, Ruslan P. Kurta, Nastasia Mukharamova, Young Yong Kim, Ruslan M. Khubbutdinov, Dmitry Dzhigaev, Vladimir V. Lebedev, Elena S. Pikina, Efim I. Kats, Noel A. Clark, Michael Sprung, Boris I. Ostrovskii, and Ivan A. Vartanyants, "Evidence of a first-order smectic-hexatic transition and its proximity to a tricritical point in smectic films" Phys. Rev. E, **98** 052703 (2018). doi: 10.1103/ PhysRevE.98.052703. Impact factor: 2.366.

6) External Funding

We did not requested for any external funding

7) Patent Applications

No. of pending/granted patents

We did not submitted any Patent Applications

8) Awards received by Group Members

The group members did not received any awards

Appendixes: I

Conferences where results in the frame of the grant funding were presented

1. International Conference on Computer Simulation in Physics and beyond (http://csp2018.ac.ru, September 24-27, 2018, Moscow, Russia).
   S. Zolotarev "Analysis of 3D structure resolution limits in single particle imaging with limited data" (oral presentation).

2. The 5-th school of young scientists: “Modern methods of electron and probe microscopy in the study of organic, inorganic nanostructures and nano-biomaterials” (https://www.crys.ras.ru/rcem/shkola, 26-27 August 2018, Chernogolovka, Russia).
   T.N. Baymukhametov "Spatial structure of cytochrome c nitrite reductase from bacteria Thioalkalivibrio nitratireducens studied by by Cryo-Electron Microscopy " (oral presentation).
   Abstracts are published in the school Abstract book, p. 76-77 (2018)
   https://www.crys.ras.ru/document/RCEM-2018/RCEM-2018_abstracts_(school).pdf

3. The 14th Biennial Conference on High Resolution X-Ray Diffraction and Imaging (XTOP 2018) (http://www.ba.ic.cnr.it/xtop2018/, Bari, Italy, 3rd to 7th September 2018). I. A. Vartaniants “Single particle imaging without symmetry constraints at an X-ray free-electron laser” (oral presentation).
   Abstracts are published in XTOP Abstract book, p. 78 (2018).

Appendixes: II

Kick-off Meeting February 15-16, 2018
DESY Bld. 25f, Sem. Rm. 456

February 14, 2018

Arrival to DESY
Visit to European XFEL

February 15, 2018
Open sessions

9:00 – 9:15 E. Weckert (DESY) – Welcome to DESY
9:15 – 9:30 I. Vartaniants (DESY)
The project start: view from DESY
9:30 – 9:45 V. Ilyin (Kurchatov Institute)
The project start: view from Kurchatov Institute
9:45– 10:15 S. Dusterer/W. Wurth (DESY, HH University)
Status of FLASH and challenges for data analysis

10:15 – 10:45 A. Vasiliev (Kurchatov Institute)
Cryo-EM opportunities for high-resolution cryo-microscopy of biological samples and challenges for Big Data evaluation
10:45 – 11:15 Coffee break

11:15 – 11:45 K. Giewekemeyer/ A. Mancuso (European XFEL)
SPB/SFX instrument at European XFEL, first results of operation and challenges of Big Data storage and evaluation
11:45 – 12:15 A. Teslyuk (Kurchatov Institute)
HPC/HTC infrastructure of Kurchatov Institute: opportunities of Big Data storing and evaluating.
12:15 – 14:00 Lunch
14:00 – 14:30 E. Pichkur (Kurchatov Institute)
Towards on-the-fly Cryo-Electron Microscopy Data Processing by High Performance Data Analysis
14:30 – 14:45 R. Kurta (European XFEL)
XCCA analysis of diffraction patterns in SPI experiments
14:45 – 15:00 S. Bobkov (Kurchatov Institute)
Classification of diffraction patterns in SPI experiments
15:00 – 15:15 M. Rose (DESY)
Analysis and reconstruction of virus structure from SPI experiments
15:15- 15:30 G. Mercurio (HH University)
Time-resolved photoemission measurements at FLASH
15:30 – 16:00 Coffee break
16:00 – 16:15 Luca Gelisio (DESY)
Molecular dynamics atomistic and scattering simulations
16:15 – 16:30 S. Zolotarev (Kurchatov Institute)
Reconstruction of orientations with EMC algorithm. Challenges and possible was for improvement.
16:30 – 16:45 K. Ikonnikova (Kurchatov Institute)
Analysis of 3D structure resolution limits in single particle imaging with limited data

16:45 – 17:00 A. Orekhov (Kurchatov Institute)
Cryo-electron microscopy in polymer investigations
17:00 – 17:15 I. Vartaniants (DESY)
Summary of the first day and tasks for the second day
19:00 – Dinner at DESY Bistro

February 16, 2018

Open sessions (continued)

9:00 – 9:15 Yu.Chesnokov (Kurchatov Institute)
Cryo-Electron Tomography and Subtomogram Averaging

9:15 – 9:30 T. Baimukhamedov (Kurchatov Institute)
Cryo-electron tomography and sub-tomogram averaging of eukaryotic polyribosomes

9:30 – 9:45 R. Kamishinsky (Kurchtov Institute)
Single Particle Cryo-EM of Membrane Proteins

Closed sessions for the participants of the Project

9:45 – 10:45 Discussion of the milestones of the Project (mediators I. Vartaniants, V. Il'in, A. Vasiliev)
10:45 – 11:15 Coffee break
11:15 – 12:30 Discussion of the milestones of the Project (mediators I. Vartaniants, V. Ilyin, A. Vasiliev)
12:30 – 13:00 I. Vartaniants (DESY), V. Ilyin&A. Vasiliev (Kurchatov Institute)
Concluding remarks and close of the Meeting
15:00 Visit to CSSB Kay Grünewald

February 17, 2018

Departure from DESY

Appendixes: III

Schedule of the Meeting on Joint RSF-Helmholtz Project

RSF № 18-41-06001, Helmholtz grant № HRSF-002 "New avenues in information and data science: advanced imaging applications at the XFEL and cryo-EM frontier"

November 27-28 2018, Moscow

November 27, 11.00 -18.00

1. V. Ilyin (NRC KI). Open remarks. Status of the Project, view from RF
2. I. Vartaniants (DESY). Status of the Project, view from DESY
3. A. Vasiliev (NRC KI). Status of the Project, KrioEM
4. Sergey Bobkov (NRC KI), "First elements of data flow processing pipeline for XFEL"
5. Sergey Zolotarev (NRC KI), "Methods for processing of low-quality SPI data in context of EuXFEL p2013 experiment"
6. Anton Teslyuk (NRC KI), "A review of containerization and virtualization technologies for HPDA"
7. Max Rose (DESY, Germany), "X-ray single particle imaging without symmetry constraints"
8. Young Yong Kim (DESY, South Korea), "High-resolution 3D reconstruction of biological samples using coherent X-ray diffraction images"
9. Dmitri Lapkin (DESY), "Resolving Crystalline Domains and Domain Boundaries in Nanocrystal Superlattices using angular x-ray cross-correlation approach"
10. Ruslan Khubbutdinov (DESY), "Coherence properties of 4th generation synchrotron sources"
11. Dameli Asalaulova (DESY), "Initial data analysis of single particle imaging experiment at LCLS in 2018"

November 28, 11.00-17.30

12. Jurij Chesnokov (NRC KI), "Contextual data to local astigmatism estimation"
13. Timur Bajmukhometov (NRC KI), "Cryo-EM data processing part in the article "3D structure of the natural tetrameric form of human butyrylcholinesterase as revealed by cryoEM, SAXS and MD".

14. Eugenij Pichkur (NRC KI), "On the fly preprocessing of cryoEM data".
15. Alexander Vasiliev (NRC KI), "Recent advances in KrioEM"

Round table “2019-2020 planning”