

**Helmholtz – RSF Joint Research Groups**

**Joint Project Description Template**

1. **Core data**

#### Deep Learning for Physics with Novel Calorimetry (DeepPhysCAL)

Which priority thematic area is relevant to this project:

x Materials and Emerging Technologies

x Structure and Dynamics of Matter

**Project Partners**

|  |  |  |
| --- | --- | --- |
|  | Name and affiliation of the German Principal Investigator(Spokesperson) | Name and Affiliation of the Russian Principal Investigator |
| Name, title: | Dr. Dirk Krücker | Prof. Dr. Tagir Aushev |
| Host Institution:  | Deutsches Elektronen-Synchrotron (DESY) | Moscow Institute for Physics and Technology (MIPT) |
| AddressContact telephone number and E-Mail address: | Notkestrasse 8522603 Hamburg+49 40 8998 3749 dirk.kruecker@desy.de | Institutsky per. 9141701 Dolgoprudny+7 903 155 7815aushev@phystech.edu |
|  |  |  |
|  |  |  |
|  |  |  |

**Project Summary**

This project is focused on important topics in the field of elementary particle physics. The combination of research with physics data analysis, detector development and computing, bundled with most modern technology as Deep Learning (DL) makes the suggested project exceptional and sustainable. Crucial, but uncovered and therefore original, aspects within the activities at DESY, JINR and the Russian Institutes were identified and connected in common activities to establish strong ties between the group members. This cross-institutional support is a perfect and sustainable network to grow into for young scientists. It will give them expertise in the field and prepare them for the next generation of experiments. With two national laboratories, JINR, DESY, and three Institutes of outstanding ranking a centre of excellence is built.

The overarching goal of this project is the introduction of state-of-the-art DL methods to novel calorimetry and physics data analyses. To fully exploit the emerging opportunities from novel calorimetry with its high-granularity and timing capabilities we will develop Deep Learning methods to reconstruct physics objects precisely and with high efficiency. These objects form the basis of our data analyses, which will also evolve to employ new and innovative Machine Learning methods. These methods will be also followed up in the domain of computing, which will foster connections to the genuine developments in Deep Learning.

The project is organised in 8 work packages. They are partially interwoven to guarantee the cross- fertilization between the different areas. The common aim is the application of DL Methods.

In total four work packages deal with calorimetry:

* WP-Calo-1: Develop novel 5-dimensional (3 space, energy and time) reconstruction of hadronic particles
* WP-Calo-2: Develop reconstruction techniques in the experimental high pile-up environment that aims at techniques to identify interesting signals in overlays of 50-200 collision events
* WP-Calo-3: Develop new calibration and monitoring for highly irradiated Silicon Photomultipliers (SiPMs)
* WP-Calo-4: Develop quality monitoring methods for the production of highly granular SiPM calorimeters with several million channels.

On the physics data analysis side, this project comprises two very fashionable physics topics, which will be addressed in different groups, but with regular common meetings and knowledge exchange:

* WP-Phys-1: Top-quark physics studies with Deep Learning
* WP-Phys-2: Physics with Higgs to tau tau decays

In computing we propose to undertake developments with Deep Learning and other Machine Learning methods in two work packages:

* WP-Comp-1: Intelligent Monitoring of the CMS Tier 1 at JINR
* WP-Comp-2: Artificial intelligence for large scale computing workflow optimisation

The HRJRG will bring together experts for calorimetry, physics analysis and Deep Learning from the very different communities and catalyse the direct transfer of knowledge between them. This ensures that new ideas will find a creative environment for new developments in an early state.

**Key words**

Physics, Elementary Particle Physics, Higgs Physics, Top Quark Physics, Calorimetry, Deep Learning, Machine Learning, Artificial Intelligence, Big Data, Simulation, Reconstruction, HPC, High Performance Computing, GPU

Which describes the aspects of this project:

 x Physics and space sciences

Please select the Helmholtz Research field:

 x Matter

1. **State of the art and preliminary work**

Our applications of Deep Learning methods are motivated by the similarities between research questions in different working areas of our scientific field, which is experimental Elementary Particle Physics, as well in other scientific disciplines. The explosion of new methods from Artificial Intelligence and Deep Learning opens up multiple and innovative opportunities to improve our final scientific results. These improvements will not come from one magic silver bullet but from the careful selection, development and application of the appropriate techniques to the different steps of our data analysis chain.

At the end of 2023, the Large Hadron Collider and its two general-purpose experiments, ATLAS and CMS, will enter a major upgrade, to be ready for the “High-Luminosity” phase [*HL-LHC*], scheduled to start in 2026. At the High-Luminosity LHC the experiments will collect 150 times more data than in Run 1. Already with the brief LHC upgrade in 2019-2020 for the data taking in Run 3, starting in 2021, a dataset of 300 fb-1 is expected. In the “High Luminosity” phase the LHC detectors will have to cope with 150-200 overlapping proton-proton interactions. These largely low energetic pile-up interactions must be separated from the interesting collisions producing particles exhibiting high transverse momenta. To accomplish this task, the CMS detector will be upgraded in several detector components. The most crucial and innovative upgrade is new highly-granular calorimeters (called HGCAL), which measure the energy of produced particles in the endcaps of the experiment [*2017a*]. This novel calorimeter is composed of planes of hexagonal silicon sensors in between lead absorber layers in the front, and scintillator tiles read out individually by SiPMs between steel plates in the back. Particles going through the absorbers will produce showers of secondary particles that will be detected in 3D images with high resolution in energy and now also time.

Machine Learning (ML) is by now regularly used in data analyses in Elementary Particle Physics. The latest CMS results published on the Higgs decay to bottom quarks [*2018a*] and [*2018b*] are state-of-the art examples of the use of Deep Neural Networks. There are three main fields where Deep Learning is applied in this analysis: improved calibration, physics object identification and event selection. Deep Learning is also used as a multivariate regression tool to improve the energy resolution of b-quark jets. Adopting similar approaches for more general aspects is promising for improved calibration or modelling of detector performance overall and over time. Another aspect in the aforementioned analysis is the object identification. CMS is using now highly refined Deep Neural Networks to identify b-quark jets [*2018c*]. Similar multivariate approaches are employed to identify electrons and tau leptons [*2018d*]. In both cases we see excellent opportunities for further development. These kinds of particle identification will strongly benefit from high granularity calorimetry. The third field of application is the selection of the events belonging to the studied process. The example of the above CMS Higgs search shows that traditional Machine Learning tools as Boosted Decision Trees (BDT) are highly efficient, but carefully developed Neural Networks can still improve the experimental sensitivity. Present analyses are typically based on physically motivated variables, for example invariant masses or topological quantities. Deep Neural Networks due have the ability to learn such features from basic input variables [*2014*]. The full potential of Deep Neural Networks is typically only reached by individually trained networks for the individual decay channels. The common DL tools for these studies are Keras [*2015a*] with Tensorflow [*2015b*] as backend together with a broad variety of python-based Machine Learning tools. In the last two years a contributor, PyTorch [*2017b*], has become popular especially in the Machine Learning community since it allows easier experimentation in the development phase of a DL project. These open source tools, originally provided by Google (Tensorflow) and Facebook (pyTorch), are widely used and developed by a large Machine Learning community outside of the flied of Particle Physics. This fact demonstrates the strong potential for cross-disciplinarity and technology transfer between basic research and industry.

The above mentioned ingredients of a data analysis for the final publication of novel physics measurements will be improved by the new abilities of the next generation of hadronic calorimeters. The CMS HGCAL will be the first large scale highly granular calorimeter, which will be a part of a collider detector. It will use silicon sensors and scintillator tiles read out by SiPMs as active material. With about 400.000 SiPM channels it will be a large step, by almost a factor of 20 in size compared to previous projects, and poses complex challenges in all aspects, like for construction, quality assurance, operation, calibration, monitoring, and reconstruction , which are all essential for successful physics analyses and publication of measurements.

CMS has built a prototype for the silicon part of the HGCAL [*2018e*].Data from combined beam tests of the HGCAL silicon prototype with the CALICE SiPM-scintillator calorimeter prototype are available for several particle types. Highly granular calorimeters. like the mentioned CALICE calorimeter, with small cell sizes (5 to 50 mm) were originally proposed already 15 years ago to reach the energy resolution for jets (bundles of particles) as required by future electron-positron linear collider experiments. The CALICE Collaboration develops several different concepts for highly granular calorimeters and constructs and tests prototypes to demonstrate their capabilities and scalability to a large multi-purpose collider detector.. SiPM were developed in a close collaboration between MEPHI and DESY are capable of detecting a single photon. Later these SiPMs could be produced affordable and in larger quantities by the Russian Industry around 2004.This has opened the possibility for scintillator-based calorimeters with unprecedented granularity. Not only the energy and position of the individual hits in a particle shower can be measured, but also the time of the hit, offering new ways to reconstruct these showers in 5 dimensions and reject background with this knowledge. Recently [*2018f*], the CALICE Collaboration, with leading contributions from DESY and Russian groups, has built a SiPM-scintillator calorimeter prototype with nearly 22000 channels and exposed it to beams of several particle types and energies. In order to fully exploit the potential of this technology, new particle reconstruction and identification algorithms based on Machine Learning methods are crucial.

The operation of highly granular calorimeters at hadron colliders is challenging with respect to their radiation tolerance. This is pioneered by the CMS Collaboration with the on-going calorimeter upgrade for Run 3 of the LHC (phase I upgrade, installation 2017-2020) and the recently approved High Granularity CALorimeter HGCAL, the replacement of the calorimeter endcaps for the High Luminosity LHC (phase II upgrade, installation 2024). For the phase I upgrade, parts of the hadronic calorimeter have already been equipped with SiPMs, allowing for an improved readout granularity. Remaining parts will be finished in 2019 and 2020. Calibration and monitoring of the stability of the SiPM signals in the harsh radiation environment at the LHC are essential for the performance of the calorimeter. So far, the single photoelectron resolution (SPE) of the SiPM provide the most robust technique for calibration. With increasing radiation dose, the SiPMs lose the SPE resolution and a new method had to be developed recently. The preliminary results from the already installed SiPMs show the applicability of the new approach, but the expected dose for the remaining phase I upgrade SiPMs is more than a factor of 10 higher than studied so far, and the phase II upgrade will have to operate at up to a factor 1000 higher doses. Deep Learning regression techniques similar to the one applied in the data analysis are a promising approach to build models for radiation damage, the calibration and the monitoring of the calorimeter under such harsh radiation conditions.

The large datasets of Run III and the High Luminosity Phase will need highly efficient computing workflows. The application of Artificial Intelligence methods will allow to automate data processing steps where at present still human intervention is unavoidable. Again, Deep Learning driven decisions techniques may help to speed up the operations, make them efficient with less manpower need.

**1.1 Project-related publications**  (Alphabetically by first author)

[1] M. Berggren, A. Cakir, D. Krücker et al. “*Non-Simplified SUSY: stau-Coannihilation at LHC and ILC,*” Eur. Phys. J. C76 (2016) 183, <https://doi.org/10.1140/epjc/s10052-016-3914-2>

[2] E. E. Boos, V. E. Bunichev, L. V. Dudko, A. A. Markina "*Method of Optimum Observables and Implementation of Neural Networks in Physics Investigations,"* Phys. Atom. Nucl. 71 (2008) 383-393, <https://doi.org/10.1134/S1063778808020191>

[3] K. Borras, D. Krucker and I. Melzer-Pellmann, M. Stein and P. Schleper, “*Energy weighting for the upgrade of the CMS HCAL,*” Phys. Atom. Nucl. 75 (2012)1088-1090, <https://doi.org/10.1134/S1063778812090049>

[4] A. Chernoded, L. Dudko, I. Myagkov and P. Volkov, “*Deep Learning Neural Networks in Data Analysis,*” EPJ Web of Conferences 158 (2017) 06008, <https://doi.org/10.1051/epjconf/201715806008>

[5] G. Eigen et al. [CALICE Collaboration], "*Hadron shower decomposition in the highly granular CALICE analogue hadron calorimeter*," JINST 11 (2016) no.06, P06013, <https://doi.org/10.1088/1748-0221/11/06/P06013>

[6] C. Issever, K. Borras and D. Wegener, “*An Improved weighting algorithm to achieve software compensation in a fine grained LAr calorimeter*.”, Nucl. Instrum. Meth. A545 (2005) 803-812, DESY-04-127., arXiv:physics/0408129, <https://doi.org/10.1016/j.nima.2005.02.010>

[7] D. Krücker, K. Schwank, P. Fuhrmann, B. Lewendel and D.M. South, “*Data preservation for the HERA experiments at DESY using dCache technology,*” J. Phys.: Conf. Ser. 664 (2015) 042029, <https://doi.org/10.1088/1742-6596/664/4/042029>

[8] O.M. Sahin, D. Krücker and I.A. Melzer-Pellmann, “*Performance and optimization of support vector machines in high-energy physics classification problem*s,” Nucl. Instrum. Meth. A838 (2016) 137, arXiv:1601.02809; <https://doi.org/10.1016/j.nima.2016.09.017>, continued in A. Elwood and D. Krücker, “*Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders*”, 2018, arXiv:1806.00322

[9] I. Adachi et. al. “*Precise measurement of the CP violation parameter sin 2*$ϕ$*1 in* $B^{0}\rightarrow (c\overbar{c})K\^0$ *decays,*” Phys. Rev. Lett. 108 (2012) 171802,<https://doi.org/10.1103/PhysRevLett.108.171802>

[10] T. Aushev et. al., “*A scintillator based endcap KL and muon detector for the Belle II experiment,*” Nucl. Instrum. Meth. A 789 (2015) 134-142, <https://doi.org/10.1016/j.nima.2015.03.060>

[11] T. Aushev, A. Kozhinov, Y. Tsybin, “*Least-squares fitting of time-domain signals for Fourier transform mass spectrometry,*” J. Am. Soc. Mass Spectrom. 2014 Jul, 25(7):1263-73, <https://doi.org/10.1007/s13361-014-0888-x>

[12] H. L. Tran, K. Krüger, F. Sefkow, S. Green, J. Marshall, M. Thomson and F. Simon, "*Software compensation in Particle Flow reconstruction,*" Eur. Phys. J. C77 (2017) no.10, 698, <https://doi.org/10.1140/epjc/s10052-017-5298-3>

1. **Objectives and work programme**

**2.1. Anticipated total duration of the project**

36 months

* 1. **Objectives**

The overarching goal of this project is the application of the methods of Deep Learning in physics analyses, recording and processing of experimental data with focus on the calorimetric detector components. We concentrate on the physics analysis hot topics, single top studies and the analyses of the decay of the novel discovered Higgs boson into two tau-leptons. Potentially we will employ the Vector Boson Fusion production of the Higgs which has prominent signatures in the calorimeters in the endcap region of the experiment.

On the hardware and reconstruction side we focus on the novel high-granularity calorimetry and the innovative inclusion of the timing information in the signal reconstruction, the monitoring of the radiation damage as well as calibration aspects to guarantee a high-quality signal. Such a novel high-granular calorimeter will be installed in the future phase II upgrade of the endcaps of the CMS Experiment.

One strong pillar of the project is the exploitation of the opportunities offered by this novel calorimetry with its high granularity and timing capabilities by Machine Learning methods. High granularity calorimeters have been proposed to reach the jet energy precision goals at future electron-positron colliders, but also offer superior possibilities for particle identification and reconstruction, background rejection and pile-up removal, having made them the option of choice for LHC detector upgrades. These aspects can be further improved by including the information of hit times with sub-ns precision, that has become achievable recently. With the huge amount of data with the energy and time information provided by several millions of calorimeter channels, Deep Learning methods are particularly suited to develop new reconstruction algorithms for the physics objects. On the other hand, the large number of channels is a challenge for the detector calibration and the monitoring of radiation damage, which are essential to guarantee a high-quality signal. Also here Deep Learning will be employed to develop and improve the procedures. Finally, already in the construction phase of highly granular calorimeters a stringent quality control of the components is necessary, which can benefit from Deep Learning methods.

One of the advantages of the project is that the methods and instruments developed for the physics analyses could be used in various applications other physics branches as well as beside physics. For example, the Machine Learning algorithms developed for cluster reconstruction in calorimeters could be used for the image recognition in general and more particularly for the fast human-independent cancer spots recognition. Since SiPMs find a wide range of applications nowadays from astroparticle physics to medical applications, e.g. PET scanners, also the calibration and monitoring techniques developed in this project can benefit other areas of science and society.

* 1. **Work programme incl. work packages, milestones and proposed research methods**

##### For a tabular overview see Table 1

|  |  |
| --- | --- |
| Work Packages  | **Table 1 - Overview work packages** (see text for abbreviations) |
| Half Years | 1/1 | 1/2 |   | 2/1 | 2/2 |   | 3/1 | 3/2 | Involved Institutes |
| Meetings | Kick-off(virtual) | Workshop DESY |   | Workshop Moscow |   | Workshop DESY |
| **WP-Calo-1** | pion-cal | val-simu | pion-obs | pion-algo | pion-pub | pfa-pub | DESY MEPhI |
| **WP-Calo-2** |  | clus-algo | pid-algo | pile-up-cor | reco-algo | pile-up-pub | JINR MIPT DESY |
| **WP-Calo-3** |  | gain-method |  | gain-auto |  | gain-opt | MEPhI |
| **WP-Calo-4** |  | qc-proc | qc-run | - | - | - | DESY |
| **WP-Phys-1 (top)** | theor-sim | set-design | arch-opt-train | data-anal | data-anal | conf-report | SINP MSU |
| **WP-Phys-2 (tau)** | pseudo-data | pseudo-data-ana | runIII-data-ana | tauID-pub |  higgs-ana | higgs-pub | DESY |
| **WP-Comp-1** | collect | algo | monitor | fault-dat | fault-mod | auto | JINR |
| **WP-Comp-2** | info-pre | sys-dec | train-ev | 1st-conf | dev-auto | 2nd-conf & implement | DESY |

##### WP-Calo-1: Reconstruction of hadron showers (leading institue DESY)

Novel high-granularity calorimeters with timing information for each individual hit provide unique 5-dimensional (3 space dimensions, energy and time) information on shower development and structure. This information is especially interesting for hadron showers, where a variety of processes is involved in the shower development that happen on time scales between picoseconds and hundreds of nanoseconds. Several reconstruction techniques have already been developed for highly granular calorimeters, such as software compensation for an improved energy reconstruction of hadronic showers, and particle flow algorithms for the reconstruction of jets, which heavily relies on the ability to separate energy depositions of close-by showers.

The beam-test data collected with the CALICE SiPM-scintillator calorimeter prototype and with the combined setup together with the CMS HGCAL silicon prototype provide a unique basis for the development and optimisation of 5-dimensional shower reconstruction algorithms. The amount and complexity of the information make the application of ML and DL techniques very attractive. A new hadron-reconstruction algorithm based on Machine Learning opens up also very promising opportunities for improving the reconstruction of jets with Particle Flow Algorithms. The goal is to first develop a new DL reconstruction algorithm with improved energy resolution, based on the beam -test data. We plan to use this then in order to improve the Pandora PFA [*2009*], which is widely used for the design of future electron-positron collider detectors. The Pandora optimisation analysis will be based on simulated electron-positron collision events.

1. **Year 1**: The work will start with the preparation of the beam-test data. The data need to be calibrated and corrected for SiPM saturation effects, and algorithms to select hadrons from the data with mixed particle content need to be applied. Then the software to simulate the test-beam prototype will be set up, and simulation parameters will be tuned to allow a comparison to data. For the work on the PFA optimisation, the first step is to get a detailed knowledge of the Pandora software framework, identifying the reconstruction routines where DL techniques can be employed for the reconstruction of hadrons.

**Milestone 1.1**: calibration and selection of pion test-beam data (**pion-cal**)
**Milestone 1.2**: validation of the test-beam simulation with data (**val-simu**)

1. **Year 2**: In the next step, observables suitable for use in DL methods will be defined. Several DL architectures will be tested, and an appropriate one will be chosen to develop a new hadron reconstruction algorithm. The DL methods will be tested in parallel on measured and simulated pion data from test beams, and on simulated single hadron events in an electron-positron collider detector.

**Milestone 2.1**: identify observables for DL methods (**pion-obs**)

**Milestone 2.2**: new hadron reconstruction algorithm applied to measured and simulated pion events (**pion-algo**)

1. **Year 3:** The application of the new reconstruction algorithm to the measured pion beam test- data and the comparison to the simulated events form the content of a publication. Based on the DL single hadron reconstruction method, the jet reconstruction in Pandora will be optimised and the results will be documented in a publication.

**Milestone 3.1:** publication of the test-beam results (**pion-pub**)

**Milestone 3.2**: publication of the optimised particle flow reconstruction (**pfa-pub**)

**WP-Calo-2: Calibration and particle reconstruction in High-Pile-Up environment (leading institute: JINR)**

The high granularity of the HGCAL detector will substantially increase the complexity of common tasks, such as calibration, particle reconstruction and event simulation. Convolutional Neural Networks (CNNs) are a natural candidate to accomplish these tasks at best, while minimizing the amount of needed resources. This project aims on developing CNN-based solutions to the main tasks associated to the reconstruction of this detector. There are several classes of tasks where developing of Deep Learning solutions is promising:

* Particle clustering: using unsupervised clustering algorithms, the developed applications should be capable of associating recorded hits to the shower originated by a certain incoming particle. Given the detector geometry, both 2D and 3D clustering can be pursued, seeded by the reconstructed tracks in the inner CMS detector, as well as unseeded algorithms.
* Particle identification and measurement: the showers produced by particles traversing a calorimeter have characteristic shapes, associated to the particle energy and incoming direction as well as on the particle nature. One could then use image-detection techniques, such as deep classifiers and regressions based on Convolutional Neural Networks [*2012*], to accomplish these tasks.
* Pile-up subtraction for event reconstruction with denoising algorithms: due to the presence of as many as 200 simultaneous collisions, reconstructing a particle in HGCAL will imply separating hits associated to the *right* collision from those originating from the others. This could be seen as a noise-reduction problem, similar to those faced in computing-vision applications. Modern denoising algorithms (e.g., denoising autoencoders [*2010*,*2012*]) can be a way to suppress the pile-up contribution with fast data processing.
* Pile-up correction in energy calibration: Currently, the in-situ calibration of the CMS hadron calorimeter includes several steps, starting from radiation damage corrections, equalization of the response of subdetector segments followed by the energy scale calibration with isolated charged hadrons from collision data. The in-situ energy scale calibration in the endcap region is challenging due to pileup contributions and is achieved by using loose isolation conditions accompanied by dedicated techniques of correction for pileup. We plan to develop machine-learning-based techniques to correct the energy of a loosely isolated charged hadrons in high pileup environment. Such a technique will help to increase the calibration precision and the range of in-situ calibrated subdetectors of the CMS hadron calorimeter.

We plan to investigate the application of Deep Learning algorithms in these areas

1. **Year 1**: Studying different clustering algorithms. Evaluating their performance in terms of energy containment, resolution, etc.

**Milestone 1**: Choose and implement a clustering algorithm to be used. (**clus-algo**)

1. **Year 2**: Develop particle identification algorithms for HGCAL. Study possible working points. Implement denoising to fight pile-up. Development of a DL technique to correct for pileup for loosely isolated simulated charged hadrons

**Milestone 2.1**: Implement particle identification and pile-up rejection for HGCAL and study their performance (**pid-algo**)

**Milestone 2.2**: preliminary trained DL algorithm to correct for pile-up using simulated samples with and without pileup (**pile-up-cor**)

1. **Year 3:** The last stage of algorithm development will include the optimisation of the algorithm trained on simulations, and - if possible - the application to test-beam data

**Milestone 3.1**: optimisation of the reconstruction algorithm, comparison to previous results (**reco-algo**)

**Milestone 3.2**: application of the new algorithm, publication (**pile-up-pub**)

**WP-Calo-3: Calibration and monitoring of highly irradiated SiPMs (leading institute: MEPhi)**

Recent generations of SiPMs are radiation tolerant such that they can be used in the environment of hadron calorimeters at LHC or HL-LHC. The calibration and monitoring of the devices in the detector becomes more and more challenging with accumulated irradiation, as they lose their ability to resolve single pixels peaks due to increased noise. The most important quantity for monitoring is the gain since it is determined directly by the overvoltage applied to the SiPM and allows independent monitoring of the operational conditions. Moreover, all important SiPM parameters (light detection efficiency, interpixel crosstalk, afterpulsing) are sensitive to the applied overvoltage. For strongly irradiated SiPMs an additional effect becomes important, a change of the excess noise factor (ENF). It is expected that the SiPMs in the HGCAL will be strongly irradiated and the ENF will change. For the SiPMs installed in the CMS phase I upgrade during 2017, the analysis of the spectra showed that during the year of operation in 2018 the ENF did not change noticeably. The irradiation conditions for which SiPMs start to change their breakdown voltage and ENF are not yet well understood and are a field of active research with SiPM samples irradiated in various facilities. In order to gain further insights, and to fully exploit the possibilities SiPMs offer in radiation environments, the development of new characterization and monitoring methods based on deep learning are envisaged.

1. **Year 1**: First the new method for the gain extraction by using the mean value and the dispersion of the distribution will be thoroughly tested for different background levels, in order to find the limit of applicability in terms of neutron fluence and temperature operation conditions

**Milestone 1**: document new gain extraction method and its limits of applicability (**gain-method**)

1. **Year 2**: With the data obtained from the CMS phase I calorimeter upgrade, a new automatic procedure for the extraction of the gain for thousands of SiPMs with various levels of irradiation will be developed.

**Milestone 2**: automatic gain extraction procedure (**gain-auto**)

1. **Year 3:** The automatic gain extraction procedure will be further optimised to be able to cope with highly irradiated SiPMs as expected after operation in the HGCAL.

**Milestone 3**: optimised automatic gain extraction procedure (**gain-opt**)

**WP-Calo-4: Production techniques for highly granular SiPM-scintillator calorimeters (leading institute: DESY)**

The construction of a highly granular calorimeter for a collider detector with several 100.000 to several million channels poses new challenges. Since a manual inspection and test of the components is impossible, automatic testing procedures are essential to reach the necessary quality and uniformity for good measurements and physics results. DESY will setup infrastructure to demonstrate how these challenges can be met for the HGCAL for wrapping scintillator tiles into reflector foil and assembly of tiles onto electronics boards. With the experience of construction of the CALICE beam-test prototype, DESY is in a unique position for this task. The objective of this work package is a running prototype setup of the assembly and quality tests by summer 2020, before the HGCAL module production starts. The knowledge will then be transferred to the Russian groups who will assemble the HGCAL modules.

1. **Year 1:** The quality control is planned to consist of automated visual inspection by image processing of photographs using DL methods, for unwrapped and wrapped tiles, and for fully assembled boards with arrays of tiles. An adaptive algorithm is needed to efficiently account for about 30 tile formats and about 30 board shapes. In addition, for tests of the electronics, an analysis of the SiPM spectra is planned, similar to what is used for the gain extraction (WP-calo-3).

**Milestone 1**: Development of an assembly and quality control procedure for HGCAL SiPM-scintillator tile boards (**qc-proc**)

1. **Year 2**: knowledge transfer to Russian groups who will do the assembly

**Milestone 2**: first assembly and quality tests performed at production site (**qc-run**)

**WP-Comp-1: Intelligent Monitoring of the CMS Tier 1 (leading institute: JINR)**

To ensure a correct and reliable operation of the CMS Tier 1 computer complex at JINR, a special monitoring system has been designed. The system allows to collect data on different Tier 1 operation aspects, like climate and power-supply systems control and monitoring of local network equipment, telecommunication links and computing nodes. The Tier 1 performance is also monitored using the standard tools of the WLCG and CMS collaboration. These tools allow to monitor grid and other site services, evaluate the accessibility, effectiveness and usability of the services and the entire Tier 1 site. The readiness status of the site is taken into account by the CMS CCU (Central Control Unit) when planning activities within the CMS collaboration. The monitoring system of the sites readiness consists of the following three major subsystems.

* Monitoring of the site availability (Site Available Monitoring – SAM). This includes tests which check the basic services of the site.
* Monitoring of test data processing flows.
* Monitoring of the data transmission in order to compute the data transmission capability between sites.

The readiness of the site is evaluated on the basis of indicators that depend on the results of these tests.

In the project, it is proposed to create an intelligent monitoring of the CMS Tier 1 centre at JINR, mainly in the area of infrastructure, grid and experiment services, and operations. Joining the available information on all levels with the help of Machine Learning techniques will allow to create an automated rapid fault discovery system for hardware, services and networking. Another aim is to foresee and predict possible failures on all levels of the Tier 1 centre. At the moment, there are 24/7 shifts where on-duty engineers control the state of the Tier 1. Having an automated system to detect and predict any potential issue will improve the service quality provided by the computer centre.

1. **Year 1**: Monitoring of local hardware and software infrastructure, specific to the CMS Tier 1 at JINR. Preparing all data sources (hardware, network, operating system, storages, batch system). Creation of the infrastructure to collect, store and process the data. Visualization of monitoring data. Designing fault discovery algorithms based on the joint information from all data sources together with the experience of local experts.

**Milestone 1.1**: System to collect and process monitoring data; local data sources support (**collect**).

**Milestone 1.2**: Fault discovering models and algorithms tuned for the CMS Tier1 at JINR (**algo**).

1. **Year 2**: Monitoring of the Tier1 from the CMS perspectives, adding corresponding sources to the monitoring data flow. Preparing application program interfaces and user interfaces. Visualization of the processes related to the operation of Tier 1. Notification system to propagate discovered issues.

**Milestone 2.1**: Use of WLCG and CMS computing data in the monitoring. Application program interfaces to the monitoring system (**monito**).

**Milestone 2.2**: User program interfaces for the joint monitoring; automatic faults discovering and alarms notifications (**fault-dat**).

1. **Year 3**: Intelligent monitoring of Tier 1 using local, WLCG and CMS data. Fault prognosis support. Decision making support for operations. Control and automated actions on routine issues.

**Milestone 3.1**: Algorithms and models to forecast Tier1 related processes and faults prediction basing on local, WCLG and CMS data (**fault-mod**).

**Milestone 3.2**: Decision making support system for JINR Tier 1 operations and CMS computing shifters; automated actions for routine operational issues (**auto**).

**WP-Comp-2: Artificial intelligence for large scale computing workflow optimisation (leading institute: DESY)**

Simulation and reconstruction for high granularity calorimeters is time consuming and the condition of the future HL-LHC will be demanding in respect of track density and related processing times. Already at Run III of the LHC, computing time requirements will be tight and any improvement of computing workflows will be of large value. Large computing campaigns still needs regular interventions of human operators. In CMS computing the central production system runs jobs on more than a hundred different sites worldwide with about 200k of grid cores. Thousands of workflows with thousands of jobs must be controlled. Failures in such a system are unavoidable in which case the manual intervention of the human operator is the necessary. The appropriate actions how to recover the workflow is left to the discretions of the operator. Analysing the available log files of existing installations, e.g. grid instance worldwide, with deep learning methods should allow to improve workflows and provide advice to human decisions. As a first step to an automated recovery procedure an advice system for the operator can be considered. Previous work with based on Deep Learning had shown promising results [2018g] but the efficiency of the system needs to be improved especially for the goal of an automated system. The previous studies are based on job failure codes and information about the site where the jobs are running. Additional information on job types, i.e. Monte Carlo simulations, type of simulation, or data etc. will improve the performance of the system. A critical point here is the handling of unbalanced data. Some job failure types are rare, and the trainings data will suffer from unbalanced categories. There are different versions of advice systems in medicine or business. While a Deep Learning based approach is promising, other methods could be more robust or may have a reduced demand on trainings data. Different approaches, for example a simple k-nearest-neighbour clustering technique, or methods from recommender systems [*2016*] will be explored.

1. **Year 1**: The system needs data from the CMS grid workflows. The necessary processing steps will be setup to get access to different kinds of log files and operator information. The previous studies need to be recovered and additional information sources will be included. The general information workflow and pre-processing will be defined.

**Milestone 1.1**: Trainings data is available (**info-pre)**.

The performance of different algorithms will be tested on the available trainings data to reach a decision on the best system

**Milestone 1.2**: Decision on appropriate algorithms for the advice system (**sys-dec**).

1. **Year 2**: The beta version of the system needs to be developed, trained and evaluated based on the decisions from year 1.

**Milestone 2.1**: a working beta system (**train-ev**). The experience with the beta system on historic and online data will provide the base for a conference report

**Milestone 2.2**: Presentation on conference (CHEP or an ML orientated conference) (**1st-conf**).

If this system can be implemented as a recommendation system for the operator will depend on the achieved performance. It will also be necessary to adapt to LHC Run III conditions.

1. **Year 3**: The final step is the development and implementation of an automated or semi-automated system. In as far this will be possible depend on the achieved performance of the system (and acceptance within CMS). The experience from the second year will help improve the performance of the.

**Milestone 3.1**: Development of the improved adapted system (**dev-auto**).

The least result of this project part will be the implementation of an advice system for the operator, the best case will be an automated system.

**Milestone 3.2**: (**implement)** and presentation on conference (CHEP or an ML orientated conference) and (**2nd-conf**).

**WP-Phys-1: Physics studies with Deep learning (leading institute: SINP MSU)**

Top quark physics

The top quark is one of the most intriguing objects in the Standard Model: its mass exceeds the masses of the lightest quarks by two or more orders of magnitude and it decays via a single channel prior to formation of bound hadronic states. The effects related to the production and decay of top quarks allow one to directly measure the Vector-Axial vector structure of their interactions and verify the standard model predictions. Analyses of top quark properties are closely related to the field of Higgs physics, since in the Standard Model the interaction of the Higgs boson with the quarks is proportional to the quark mass. The thorough exploration of the top quark sector and its role in the Standard Model is one of the most topical problems in modern physics of elementary particles [2018i,2018j]. Intensive research programs on various aspects of top quark physics are carried out and projected in many experiments at colliders: the LHC, Tevatron, the envisaged future International Linear Collider (ILC), LHeC, and Future Circular Collider (FCC). Further studies at the LHC will be able to provide detailed information on the properties of the Higgs boson and reveal possible manifestations of new physics at energies from ~100 GeV to several TeV.

The core objective of the project is the development of novel data analysis methods based on Deep Neural Network (DNN). This technique is able to dramatically increase the efficiency of data analysis and, therefore, the precision of the achieved measurements [*2014*]. We plan to develop methods of more accurate study of the top quark properties, polarization, differential cross sections, mixing matrix elements, and the properties of the top quark interactions. Special attention will be paid to the emerging opportunity of associated production of the top quark with a W boson and of top quark with a Higgs boson in various extensions of the SM.

**Year 1**: The research plan begins with the theoretical description and simulation of the processes of top quark production with necessary effects, including those beyond the SM. The Tensorflow software package [*2015b*] will be used to analyse the observables. The combination of simulated data with deep learning will allow to accomplish the first objective, namely to formulate general principles to design the set of observables, similar to what was considered before [*2008*].

**Milestone 1.1**: Theoretical description and simulation (**theor-sim**).

**Milestone 1**.**2**: design of the software infrastructure and preliminary set of observables: (**set-design)**.

**Year 2:** The second stage is to investigate the DNN architecture and possible optimisations for a HEP data analysis. The implementation of DNN technique in physics data analyses depends on the specific task and will be considered in several examples, as proposed in [*2017c*]. The first process which requires DNN application is the associative tW production with interference terms to the pair top quark production. The DNN technique is necessary to distinguish the electroweak part of the process from the QCD part. This task is very important for the search of the Beyond of Standard Model (BSM) contribution in the electroweak top-quark interactions. The similar and actual process is the associative single top and Higgs production (tHq) which is rather rare and require DNN technique to increase the sensitivity of the measurements.

**Milestone 2.1.:** Optimisation of DNN parameter for data analysis (**arch-opt-train).**

**Milestone 2.2.:**  Data analysis of electroweak top quark production **(data-anal)**.

**Year 3:** The next step is the attempt to take into account the systematic uncertainties of the final result in the criteria to optimise the DNN parameters: input observables, architecture, methods to train and application at the level of statistical analysis. As a result, we plan to develop methods for optimising the DNN data analysis technology and use them to study the top quark production at CMS. The results are planned to be presented at a conference / publication.

**Milestone 3.1.:** Prepare the general recipes to apply DNN technique in optimised data analysis of the collider experiments. Apply developed methods for the analysis of the electroweak top quark production in CMS experiment (**data-anal).**

**Milestone 3.2.:** Publication in a conference report (**conf-report**) and implementation.

**WP-Phys-2: Deep learning for Higgs to tau tau studies (leading institute: DESY)**

The Higgs boson has been a topic of thorough investigation since its discovery in 2012 onwards, since a precise categorisation of its properties may provide hints for new physics beyond the Standard Model. Electrons play a pivotal role in the identification of tau leptons; the latter are a key observable for Higgs studies [*2018h*]. The CMS experiment is developing a high-granularity calorimeter [*2017a*] for Phase II to cope with the high pileup conditions. In October 2018 a detector prototype was successfully tested using test beams (earlier results have been published in Ref. [*2018e*]). The proponents from the DESY laboratory play a key role and have unique expertise on these test beam measurements. We outlay an enduring strategy for measuring tau leptons at the LHC with a PhD student, which will strengthen the expertise of the existing research groups.

The first objective is to develop DL techniques to identify electrons over a large background; emphasis is placed on using the unique novel timing and high granularity capabilities of the calorimeter. To this purpose, pion test-beam results will be overlaid to mimic high-pileup events, in which electron showers will be embedded. This pseudo-data will allow to test the deep learning methods in a self-consistent manner. Active collaboration and exchanges between Russian and German researchers from WP-Calo-1 and WP-Calo-2 are foreseen.

Optionally, this can be extended with validating and tuning simulation results, and possibly test the identification of hadronically decaying taus. This topic has substantial overlap with the tasks specified in WP-Calo-1.

The second objective is to apply deep learning methods to an analysis of the Higgs boson in tau decays using CMS Run III data (expected in 2021). For this a strong collaboration is foreseen with one of the project proponents who is member of the DESY Higgs to tau group. The combined work on object reconstruction in test-beam data and a physics analysis will strengthen both topics and allow for unique dissemination of expertise.

1. **Year 1:** The priority is to create the pseudo data and develop deep learning methods for electron identification in a high-pileup environment. Optional extensions with studies in simulated data may be included, since the test beam data would allow for a verification of the simulations, and simulated data allow to test different tau-decay modes.

**Milestone 1.1:** create realistic pseudo data from the pion and electron test beam results; potentially develop simulations (**pseudo-data**).

**Milestone 1.2**: Development of the deep learning tau ID with hit timing information for the
 LHC-Phase II. Report in the future study groups of CMS. Publish in a detector paper.

Conference presentation foreseen by postdoc. Potentially analyse simulated events.

(**pseudo-data-anal**).

1. **Year 2:** Priority is to contribute to the tau identification in the new CMS proton-collision data at 14 TeV and to contribute to the official CMS publication on this topic.Optionally, extend with studies of hadronic tau decays in HGCAL simulations.

**Milestone 2.1:** Investigate the application of deep learning techniques to optimise the tau identification in Run III data using simulated data. (**runIII-data-ana**).

**Milestone 2.2:** Contribute to the envisaged CMS paper on tau identification (a reference can be found at [*2018d*]). A conference presentation is foreseen (**tauID-pub**).

1. **Year 3:** The priority is to do an analysis on Higgs boson in tau decays at the increased centre of mass energy of 14 TeV. This is envisaged via the tau-decay mode to electrons, in which the participating student will have developed unique expertise. Apply Machine Learning methods to enhance the sensitivity of the analysis.

**Milestone 3.1:** Apply Deep Learning techniques to optimise the Higgs-to-tau-tau analysis with emphasis on the tau to electron decay mode (**higgs-ana**).

**Milestone 3.2:** Publish on the first Higgs-to-tau-tau results in proton-proton collisions at a centre of mass energy of 14 TeV. Consolidate results in PhD thesis (**higgs-pub**)

* 1. **Data handling**

The data produced and used in this project are owned by the international experimental collaborations, who distribute the data to all collaborators (>5300 members) worldwide. The collaborations follow the guidelines for data preservation and open access policies, which are imposed by the international Funding Agencies.

The main goal of the CALICE collaboration is to develop, test and establish technological solutions for highly granular calorimeters. Results from beam tests are regularly published in international journals. In addition, there is close collaboration with the GEANT4 [*2003*] team to contribute to an improved modelling of hadronic interactions, which is also of interest beyond elementary particle physics applications.

* 1. **Added value of international cooperation**

The support of this Helmholtz-RSF Joint proposal opens the possibility for the Helmholtz and the Russian Institutes to bundle their competences, jointly address the forefront technology in a familiar framework, transfer the developments and gained experience to the new areas of the experiment upgrade and accomplish with this a major role within the international collaboration. This will build new broad competence on both sides which is a specific quality gain for all involved institutes. The proposed project, closely linking physics, novel calorimetry and the development of high-tech computing technologies, opens up great prospects for the long-term partnership between the Helmholtz and the Russian Institutes for next generation of colliders.

Such a long-term partnership will enable high visibility of Germany and Russia in the international collaborations. The scale of the internationality is demonstrated by the facts, that the CMS Collaboration consists of more than 5300 members from more than 220 institutes in more than 50 different countries. The partnering institutes will gain influence in the mamagement of the international collaboration and more generally on the future direction of particle physics.

In this project the novel techniques of Deep Learning are explored within Elementary Particle Physics. These developments are a strategically releveant research topic also for other science areas and the knowledge transfer will significantly contribute to the advancement of basic research.

**2.6 Young Scientists**

The project is deliberately dominated by young scientists and PhD students in order to educate future experts in the novel technologies of AI and DL. Presently, these expertes are rarely found. Training young scientists with experts from DESY and the participating Russian Institutes will guarantee a deep knowledge transfer to the new generation. In order to acquire an excellent curriculum vitae for their future as physicists and versatile scientists, the PhD students and young scientists will work on the frontier computing technology, novel calorimetry, as well as in complex physics data analyses. Our young scientists will participate in cutting edge technology developments and its further evolution.

We will give high priority to the young scientists when deciding on conference speakers or poster presenters. Within the large collaborations they will have high visibility on an international level, which will provide them education in international soft skills and opens up bright career prospects on a global level in science as well as in business and industry. The senior scientist will provide mentoring to the young scientists at the beginning on a regular basis and thn at least once a year.

important of young scientists

Part of the team is a recent chair of the Young Scientist group of the CMS collaboration (YSC). The YSC team together with the CMS Career committee organise networking events for CMS alumni in industry such as Google, Shell, Jaguar-Landrover, the European Patent Office, Intel and many more. The young scientists of our project will benefit from this exchange with these high-tech companies.

**2.7 Sustainability**

The proposed Helmholtz-RSF Joint Research Group will establish strong links between new partner institutes. This configuration is a perfect seed for a broader undertaking, namely a Helmholtz International Research School (HIRS). Such a HIRS will be a novum in the collaboration of Russia and Helmholtz, even Germany, leading to very tight ties between the two partners.

Similar to this proposed project, the HIRS will focus on Deep Learning and its application to forefront detector technologies, continuing in calorimetry but enlarging to aspects of the trigger. The trigger is extremely crucial in the experiment, it selects the rare events of scientific interest in the flood of data. For the concentration on these events the recording has to suppress the flood of data by 6 orders of magnitude. This school will also enlarge the scope of the physics data analysis and will include theoretical and phenomenological topics.

The aspect of technology transfer to industry with the mutual benefit of common developments can be picked up and pursued with this broader scope. These techniques are widely applicable to the ongoing big data revolution in technology companies, banks, transport and other industries.

Such HIRS are especially concentrating on educating young scientists, preparing them very well for the future challenges with conveying crucial competences and international soft skills and knowledge. Having the former CMS wide Young Scientist Chair on board is opening unique chances to create a very fruitful and successful environment for the growth of the scientists in this envisaged HIRS.

It is envisaged to build a strong triangle between DESY - JINR - CERN, the three national laboratories in Russia, Germany and Europe. Therefore, it is guaranteed, that the students will receive a broad scope and excellent infrastructure for their training and their own scientific activity. A perfect preparation for a future bright career. The envisaged PI will be Kerstin Borras, who will have finished her W2/W3 Helmholtz Professorship program by then and is participating part-time in this HRJRG. With her connection to the RWTH Aachen University a strong tie to a German University will be built. In order to avoid double funding of similar projects the envisaged HIRS will cover additional topics and a broader scope and pursue the topics of this Helmholtz-RSF Joint project after the end of its duration.

**2.8 Descriptions of proposed investigations involving experiments on humans, human materials or animals**

This project does not include experiments on humans, human materials or animals.

1. **Bibliography**

[***HL-LHC***] http://home.cern/topics/high-luminosity-lhc

[***2003***] S. Agostinelli et al., “*Geant4 - a simulation toolkit,*” Nuclear Instruments and Methods in Physics Research A 506 (2003) 250-303
[***2008***] E. E. Boos, V. E. Bunichev, L. V. Dudko, A. A. Markina "Method of Optimum Observables and Implementation of Neural Networks in Physics Investigations," Phys. Atom. Nucl.,71 (2008) 383-393, https://doi.org/doi:10.1134/S1063778808020191

[***2009***] Thomson M 2009 Nucl. Instrum. Meth. A611 (2009) 25–40, arXiv:0907.3577

[***2010***] P. Vincent et al., “*Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion*,” JMLR 11(2010) 3371-3408

[***2012***] A. Krizhevsky et al., “*ImageNet Classification with Deep Convolutional Neural Networks,*” 2012 NIPS proceedings

[**2013**] D.P. Kingma and M. Welling, “*Auto-Encoding Variational Bayes*,” https://arxiv.org/abs/1312.6114
[***2014***] P. Baldi, P. Sadowski, D. Whiteson “*Searching for exotic particles in high-energy physics with deep learning,*” Nature Commun. 5 (2014) 4308

[***2015a***] F. Chollet et al., "*Keras,*" 2015 available on <https://github.com/keras-team/keras>

[***2015b***] M. Abadi et al., “*TensorFlow: Large-scale machine learning on heterogeneous systems,*”

2015. Software available from <https://www.tensorflow.org/>

[***2016***] F. Ricci, L. Rokach and B. Shapira “*Recommender Systems Handbook,*” (2 ed.), 2016, Springer

[***2017a***] CMS Collaboration, “*The Phase-2 Upgrade of the CMS Endcap Calorimeter,*” CERN-LHCC-2017-023, https://cds.cern.ch/record/2293646

[***2017b***] A. Paszke et al., “*Automatic differentiation in PyTorch,*” 2017, <https://pytorch.org/>

[***2017c***] Chernoded A., Dudko L., Myagkov I., Volkov P., “*Deep Learning Neural Networks in Data Analysis,*” EPJ Web of Conferences 158, (2017) 06008

[***2018a***] CMS Collaboration, “*Observation of Higgs Boson Decay to Bottom Quarks,*” Phys. Rev. Lett. (2018) 121, 121801

[***2018b***] CMS Collaboration, “*Search for* $t\bar{t }$*H production in the H*$b\bar{b}$*-decay channel with leptonic* $t\bar{t }$*-decays in proton-proton collisions at* $\sqrt{s}=13$*TeV,*” CMS-PAS-HIG-17-026, http://cds.cern.ch/record/2308267

[***2018c***] CMS Collaboration, “*Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,*” 2017, arXiv:1712.07158, JINST 13 (2018) P05011 and

CMS Collaboration, “*Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector,*” CMS-DP-2018-033

[**2018d**] CMS Collaboration, “*Performance of reconstruction and identification of 𝜏 leptons decaying to hadrons and* $υ\_{τ} $*in pp collisions at* $\sqrt{s}=13 $*TeV,*” arxiv:1809.02816, JINST 13 (2018) P10005

[***2018e***] N. Akchurin et. al., “*First beam tests of prototype silicon modules for the CMS High Granularity Endcap Calorimeter,*” JINST 13 (2018) P10023,https://doi.org/10.1088/1748-0221/13/10/P10023

[***2018f***] F. Sefkow, F. Simon, CALICE Collaboration, "*A highly granular SiPM-on-tile calorimeter prototype,*" proceedings of CALOR 2018, Eugene, Oregon, USA, May 21-25, arXiv:1808.09281

[**2018g**] C. Contreras et al., “*CMS Workflow Failures Recovery Panel Towards AI-assisted Operation,” Computing in High-Energy Physics* (CHEP) 2018, Sofia, Bulgaria, July 9-13

[***2018h***] CMS Collaboration, “*Observation of the Higgs boson decay to a pair of 𝜏 leptons with the CMS detector,*” Phys. Lett. B Volume 779 (2018) 283-316

[***2018i***] CMS Collaboration, “*Observation of overline ttH production,*” Phys. Rev. Lett. 120 (2018) 231801

[***2018j***] ATLAS Collaboration, “*Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*,” Phys. Lett. B 784 (2018) 173-191

1. **Requested modules/ funds – Justification**

German project part – funds requested (in EUR)

Description:

The Helmholtz Funds will be used to employ two PhD students, one for the activities in the area of calorimetry, and the other one for the area of Deep Learning. In addition, the salary of the PI will be paid to 30% from this Helmholtz Fund.

For an effective collaboration frequent travels between the partner institutes, as well as to important meetings and activities at CERN, where the CMS experiment and the test beam experiment are taking place, are crucial to keep close communication.

In addition, it is planned to organize annual workshops of the research team during the duration of the joint research group. After a first virtual kick-off workshop, a workshop at DESY will serve as a chance for the team, including the by then hired PhDs, to becoming familiar with each other, the next workshop will take place in Moscow and a last one at DESY will conclude the funding period with taking measures to sustain the grown relationships. It is also necessary that the participants receive the appropriate training, i.e. participation in schools, where necessary.

DESY will contribute on the German side with the engagement of one senior staff scientist (20%), as well as one leading scientist to 10%. This group of scientists, including the PI, will be supported by one PhD student working to 100% for calorimetry, one Postdoc Fellow working to 80% on calorimetry and physics analysis and a second PostDoc Fellow to 50% on Deep Learning methods.

It is of high importance to note, that DESY is the National Laboratory for Particle Physics in Germany. In this role DESY has many facilities and research activities, like Theory, which will be easily accessible for the group members and lead to mutual benefit in the scientific research. The own contribution from the German side will be complemented by the usage and support of the High-Performance Computing Centre, including a GPU farm, at DESY. This contribution will be flexible and is not specified in monetary numbers. In addition, the group has access to the DESY infrastructure for their research activities.

Table:

|  |  |
| --- | --- |
|  | **German project part (in Euro)** |
|  | **Year 1** | **Year 2** | **Year 3** | **Sum** |
| **Funds requested** |  |  |  |  |
|    personnel expenses | 100 000 | 100 000 | 100 000 | 300 000 |
|    travel / workshops | 30 000 | 30 000 | 30 000 | 90 000 |
|    investments  | 0 | 0 | 0 | 0 |
| **Total requested** | 130 000 | 130 000 | 130 000 | 390 000 |
|  |  |  |  |  |
| **Own contribution German part** (if applicable) |  |  |  |  |
|    personnel costs | 155 | 155 | 155 |  |
|    costs of materials and     supplies | 0 | 0 | 0 | 0 |
|    investments       GPUs ? in NAF? (5k/one) | XXX | XXX | XXX | XXX |
| **Total own contribution** (optional) |  |  |  |  |

Please indicate only direct project-related costs/expenses. Overheads are not eligible for funding.

Russian project part – funds requested (in RUR)

MIPT will contribute on the Russian side with the engagement of the Principal Investigator (30%). Also the GPU server with up to 20 khours CPU time per year will be provided by MIPT for the needs of the project. The travel expenses will be covered by the money from the Ministry of science and higher education received by MIPT to cover the travel expenses to CERN.

|  |  |
| --- | --- |
|  | **Russian project part (in RUR)** |
|  | **Year 1** | **Year 2** | **Year 3** | **Sum** |
| **Funds requested** |  |  |  |  |
|  personnel expenses | 5 200 000 | 5 200 000 | 5 200 000 | 15 600 000 |
|  expenses for materials and supplies - travel / workshops | 800 000 | 800 000 | 800 000 | 2 400 000 |
|  investments  | 0 | 0 | 0 | 0 |
| **Total requested** | 6 000 000 | 6 000 000 | 6 000 000 | 18 000 000 |
|  |  |  |  |  |
| **Own contribution Russian part** (if applicable) |  |  |  |  |
|  personnel costs | 2 000 000 | 2 000 000 | 2 000 000 | 6 000 000 |
|  costs of materials and  supplies – travel expenses | 400 000 | 400 000 | 400 000 | 1 200 000 |
|  investments | 0 | 0 | 0 | 0 |
| **Total own contribution** (optional) |  |  |  |  |

It is of high importance to note, that JINR is the National Laboratory for particle physics in Russia. JINR is not only participating, but also taking over the role as host for the full Russian institute cluster. This fact is already a guarantee for excellence in itself and opens the possibility to benefit from the excellent facilities and competences at JINR.

As in the case of Germany the own contribution from the Russian side will be complemented by the usage and support of the High-Performance Computing Centre, including GPU farms, at JINR/DUBNA. This contribution will be flexible and is not specified in monetary numbers. Same is true for the usage of the GPU Farm at MIPT, the available contingent corresponds to the continuous use of about 30 GPUs.

1. **Project requirements**

**5.1. Employment status information**

Krücker, Dirk, Dr., employed by DESY, 5 years contract (PI on the German side, Spokesperson)

Aushev, Tagir, Dr., employed by MIPT, 3 years contract (PI on the Russian side)

**5.2. Composition of the project group**

The project will be carrying out by strongly motivated well-structured international research team under the leadership of brilliant PIs. Having one PI from Germany and one from Russia, and also very experienced and successful group leaders, the aspect of joint research and external representation with close communication and regular meetings is guaranteed.

More than 10 young scientists involved in the project will gain in-depth practical experience on specific topics as part of a Work Package to which they will contribute and will also be able to expand areas of scientific interest for their future careers. Especially it should be noted that in the Russian team, almost 70% of participants are younger than 39 years old.

Participating institutes are: DESY, JINR, MIPT, and MSU …. Full names.

German side: DESY

German PI and Spokesperson

Dirk Krücker, Dr.,

DESY, Research Scientist, 5 year contract,

Core competence Team at DESY for Deep Learning; Leading role for the Machine Learning activities of the CMS group at DESY; Member of the *Helmholtz Inkubator*: “Information and Data Science for section Machine Learning”;

Current research focus is on application of deep learning to the different steps of the data analysis chain at the present LHC, as well as the future high luminosity upgrade.

Experiment patricipation: H1 & HERA-b (HERA), ILC Machine,CMS (LHC), ,

Leading role in DPHEP (data preservation effort) for the HERA experiments, coordinator of the DESY Remote Centre operation for the CMS Data Quality Monitoring;

Fellowships: PhD fellowship from the DFG;

Borras, Kerstin, Prof.,

DESY / RWTH Aachen University, Leading Scientist, permanent,

since 2015 common Professorship. at DESY and RWTH Aachen University (Helmholtz W2/W3); Experiment participation: H1 & ZEUS (HERA), CDF (Tevatron), CMS (LHC);

Leading roles: Deputy Spokesperson of CMS, Head of the CMS Engagement Office, PI of HRJRG-002, coordinator for different calorimeters and physics groups;

International committees: LHC-C, ACOT (TRIUMF), ECSG (FAIR), KET (Germany);

Awards: American Physical Society (APS) Fellow since 2018 “*For outstanding contributions to particle physics including providing exemplary leadership at DESY/RWTH Aachen University, Fermilab, and CERN.*”

Krüger, Katja, Dr.,

DESY, Research Scientist, permanent; ; Experiment participation: H1 (HERA), CALICE (R&D), CMS (LHC), former: OPAL, ATLAS;

Leading roles: Technical Board Chair of CALICE, formerPhysics Coordinator of H1;

International Committees: LHC-C

Awards: CERN research fellow; Scientific assistant University of Heidelberg (Young Investigator Group Leader supported by the Landesstiftung Baden-Württemberg);

Van de Klundert, Merijn, Dr.

DESY, PostDoc fellow, 3 year contract;

MSc in experimental physics (Utrecht University) and Theoretical physics (Amsterdam University); Doctoral degree at Antwerp University on very forward jets in the CASTOR calorimeter at the CMS experiment.

Leading roles: f

Fellowships from a.o. Shell, McKinsey, VFB foundation, CERN, and BNL.

Bocharnikov, Vladimir,

DESY, PhD student, 3 year contract, enrolled also as PhD student at MEPhI

MSc at MEPhI on Higgs boson CP properties research at the future ILC in 2017,

 working on energy reconstruction in a highly granular SiPM-on-tile calorimeter.

Hamed Bakhshiansohi, Dr.,

DESY, PostDoc fellow, 2 year contract (replacement or prolongation after that);

Member of CMS;

Awards: at the Universite Catholique de Louvain, Belgium, awarded with the Move-In-Louvain grant;

Working on Search for anomalous Higgs boson interactions with top quark, CMS to be published.

PhD Student 1 (to be hired from Helmholtz Fund)

PhD Student 2 (to be hired from Helmholtz Fund)

Russian Side: MIPT, MePhI, MSU, JINR

MIPT: full name

Russian PI, Groupleader MIPT

Aushev Tagir, Prof. Dr.

corresponding member of the Russian Academy of Sciences, head of the laboratory of high energy physics (MIPT); member of the Belle Collaboration since 1999, leader of the MIPT group in the Belle and Belle II Collaborations since 2015, KEK, Japan, leader of the MIPT group in CMS Collaboration @CERN since 2015; in 2015-2017 being vice-rector of MIPT was responsible for the research and strategic development of the University, under his lead the University improved its position in the Times Higher Education Physical Sciences ranking from 100+ to the 48th. Awards: Medal of the Russian Academy of Sciences (2005), grant of the President of the Russian Federation (2005). Owner of the grant for Switzerland-Russian scientific cooperation (2012-2013). Since 2007 – co-convener of the major research group at the Belle experiment dedicated to the CP violation measurements. The most recognizable result is the sin2beta measurement (2012), which is still the best measurement of this parameter in the World.

Filatov Oleg,

master student, MIPT, working in the field of B physics @CMS with the machine learning successfully applied in the context of searching for rare Bs meson decays. Good familiarity with Higgs to tau tau analysis in the emu channel as a part of the summer student project @DESY. Currently interested in application of deep adversarial learning in HEP.

,

MePhI …. Full name

Groupleader MePhI

Parygin Pavel

Engineer, MEPhI, engineer in microelectronics and solid-state electronics. For his PhD he works on the Development of radiation tolerant silicon photomultipliers for high energy physics experiments, with data obtained from CMS the detector and in the laboratory.

Bychkova Oksana

PhD student, MEPhI, graduated with honors from MEPhI last year. The subject of her master's thesis was “Calibration and monitoring of SiPM-based detection modules for CMS Hadron Calorimeter”. Now she is a 2nd year PhD student at MEPhI and her work is connected with research and development of the SiPM-based neutron monitor for the CMS experiment.

SINP MSU… full name

Groupleader:

Lukina, Olga, Dr.,

SINP MSU staff Senior Scientist. Experiment participation: ZEUS (HERA), CMS (LHC), former: MIRABELLE (IHEP, Protvino), E632 (FNAL). SINP MSU Group Leader at ZEUS, coordinator for the national forward physics group at CMS supported by the MSHE of the RF, strong experience in detector technology development as well as complex physics analyses. Current interests are in experimental measurements on new physics searches.

,

JINR .. full name - Hosting Institute

Groupleader:

Korenkov, Vladimir, Prof., Dr. of Engineering,

Director of the Laboratory of Information Technologies, JINR. Director of the Laboratory “Cloud technologies and Big Data analytics” of the Plekhanov Russian University of Economics. Has high achievements in the fields of computing and networking, grid technologies, high-performance and distributed computing, parallel computations, visualization and multimedia systems and database applications.

Gorbunov, Ilya, Dr.

PhD, JINR, 5 years contract; higher education at Moscow State University at High Energy Physics chair in 2010. The MSc thesis was devoted to distributed computing in Grid using virtualization to run non-standard environment. After the university my main field of studies were Drell-Yan in muon final state and its angular distributions. PhD thesis was on measurement of Forward-Backward Asymmetry in muon final state at 8 TeV at CMS. I am also taking part in HGCAL reconstruction and performance studies. In particular studies of the shower containment and separation with existing clustering.

**5.3. Cooperation with other researchers**

Because of its use and development of novel technology and techniques as well as the need for extraordinary high energies and intensities, particle physics is bundling all forces and funds on a very international and global level. For example the CMS Collaboration at the LHC at CERN comprises more than 5300 members from more than 220 institutes from more than 50 different countries. The CALICE Collaboration is working on a similar international level.

In the same spirit also the Computing and development of novel techniques for processing Big Data with Artificial Intelligence is an international endeavour.  Beside the internationality the High Performance Computing Centers at DESY and at JINR are serving different communities for various physics research as well as other scientific branches and nonetheless also collaboration with industry for common development.

This project is part of the CMS collaboration. We regulary present machine Learning related results to The Inter-experimental Machine Learning (IML) Working Group provides a forum for the machine learning community at the LHC.

**5.3.1. Researchers with whom you have agreed to cooperate on this project**

SiPM Radiation damage: Dr. Elena Popova, MEPhI; Prof. Erika Garutti, University of Hamburg

Calorimeter reconstruction: Dr. Marina Chadeeva, LPI

Physics: Elisabetta Gallo, Alexei Rapereza both DESY

Computing: Christoph Wissing,DESY

**5.3.2. Researchers with whom you have collaborated scientifically within the past three years**

# DK please add appropriate Colab.

We are presently members of different large collaborations: CALICE, CMS and H1, ZEUS(?).

and former members of the ATLAS, OPAL and DPHEP collaborations. Our  names therefore appears on a large number of papers and with a large number of researchers. Here, we only list these persons, we work within smaller groups excluding students who left academia.

#Please add names

DESY: M. Berggren, J. List, I.A. Melzer-Pellmann, P. Fuhrmann, B. Lewendel, D.M. South,

T. Mkrtchyan, P. Millar, Alexander Glazov, Claudia Seiz

RWTH Aachen/DESY, Kerstin Borras

Istanbul Technical University, Altan Cakir

University of Sussex, B. Safarzadeh Samani

**5.4. Scientific equipment**

For the German project partners, DESY will provide access to the National Analysis Facility for computing, mass storage and Grid access, as well as to the HPC Maxwell cluster. In addition, normal notebooks and Desktop PC will be provided. Due to the Deep Learning related parts of project there is a need for GPU resources. At DESY the CMS group has access to 4 Nvidia P100 GPUs.

It is imortant to note that we greatly benefit from the cooperation in this area since JINR and MIPT can provide much larger GPU resources. The Govorun supercomputer at JINR contains 5 NVIDIA DGX-1 Volta units. Each unit contains 8 Tesla V100 GPUs. NVIDIA Volta GPU microarchitecture was developed especially for ML purposes. These GPUs are used for the computing needs of JINR and its member states. JINR can offer 4 Tesla V100 GPUs for the needs of algorithms development and testing for the DeepPhysCal project. MIPT will provide access to a large GPU farm and continous use of 30 GPUs.

**6. Additional information**

[Text]

**Signature of the Russian Principal Investigator**

**Signature of the German Principal Investigator**