ErUM-ARIA
Simulations: Active, Refined and Interpretable Approaches

Methoden zur effizienten Resourcennutzung für physikalische Simulationen

Projektpartner

- Prof. Dr. Johannes Erdmann - III. Physikalisches Institut A, RWTH Aachen
  johannes.erdmann@physik.rwth-aachen.de
- Dr. Christian Haack[1], Erlangen Centre for Astroparticle Physics
  christian.haack@fau.de
- Prof. Dr. Claudio Kopper, Erlangen Centre for Astroparticle Physics
  claudio.kopper@fau.de
- Prof. Dr. Julia Kowalski - Lehrstuhl für Methoden der Modellbasierten Entwicklung in den Computergestützten Ingenieurwissenschaften (MBD), RWTH Aachen
  kowalski@mbd.rwth-aachen.de
- Prof. Dr. Christopher Wiebusch, III. Physikalisches Institut B, RWTH Aachen
  wiebusch@physik.rwth-aachen.de

Assozierter Partner

- Dr. Sebastian Schoenen, ControlExpert GmbH, Langenfeld
  s.schoenen@controlexpert.com

[1] Project coordination

1 Objectives

1.1 Overall goal of the project

The goal of this project is to substantially improve the simulations of physical systems concerning accuracy while reducing their computational cost. We aim to develop extremely fast, parametrized simulations and systematically investigate discrepancies in model descriptions using various state-of-the-art AI methods.

Predictive simulations of physical processes are at the core of state-of-the-art achievements in the natural sciences and engineering. In particle and astro-particle physics, the computational first-principles modeling of physical processes is essential for designing experiments and establishing robust data analysis pipelines that accommodate complex detector responses. In classical engineering, such as model-based product design, computational first-principles modeling of heat and mass transfer allows one to tailor a solution to given requirements. While (Astro-)particle physics and engineering may seem unrelated at first sight, they have in common that both fields critically depend on accurate simulations to translate theoretical models into practical insights.

Today, predictive simulations achieve remarkable accuracy and are essential for scientific discoveries and practical applications. However, their precision and reliability almost always come at the cost of long computational time or additional resources, such as the need for high-performance computing clusters or specialized hardware. This yields a trade-off between simulation precision and computational resources whenever high-statistics simulations are required. This project aims to address this trade-off with multiple approaches.

The first approach is based on **interpretable surrogate models** [1]. Its goal is to develop **ultra-fast simulations** that combine state-of-the-art methods in generative deep learning and interpretability to automatize the parametrization of the detector response. Generative surrogate models employed currently in fast simulations are *a priori* "black boxes" that are very difficult to interpret. This hinders identifying and correcting systematic biases. We propose to address this issue with **Kolmogorov-Arnold-Networks (WP2)** and **Symbolic Regression (WP3)** (SR), which we will use to derive compact and interpretable analytic expressions from the outputs of the surrogate models. In computational engineering, precision is of utmost importance, requiring generative surrogate models to be physically consistent. Rather than verifying consistency a posteriori, we will implement **Bayesian model discovery with constitutive neural networks (WP1)**, which allows the incorporation of invariants and symmetries into the surrogate model architecture during calibration.

Our second approach aims to leverage the full precision of predictive simulations while increasing their computational efficiency. Simulations in particle physics consist of multiple simulation steps (particle interaction, detector simulation, data reduction), where the most expensive step is typically the detector simulation. Examples (simulated events) generated from simulators often vary dramatically in their relevance for downstream analysis tasks, so, in many cases, a large amount of computation time is spent on creating irrelevant examples. We propose to address this issue with a hybrid simulation approach based on **Active Control (WP4)**, which aims to predict the relevance of examples early in the simulation pipeline in real-time. Examples of high predicted relevance can then be assigned a higher priority for downstream simulation, and thus, the overall simulation efficiency can be substantially increased.

However, these approaches trade reduced computational costs with increased prediction errors. Furthermore, the models may not explicitly account for the complexity of real-world conditions or measurements, such as nonlinear interactions, discrepancies in sensor readout, stochastic variations, and environmental influences.

Thus, our third approach focuses on **learning and describing discrepancies** across simulations by addressing **Domain Shifts (WP5)** to enhance the simulation framework. We aim to develop a neural network approach to learn and apply discrepancies between datasets, may they stem from different simulations or measurement data and simulations. This method will improve

simulation accuracy and efficiency and enable the re-purposing of older datasets by aligning them with current knowledge, reducing the need for new simulations.

Advanced techniques, including normalizing flow models or generative adversarial networks, will facilitate feature transfer across detector geometries, enhancing calibration, signal-background separation, and detector response understanding for a more robust simulation pipeline.

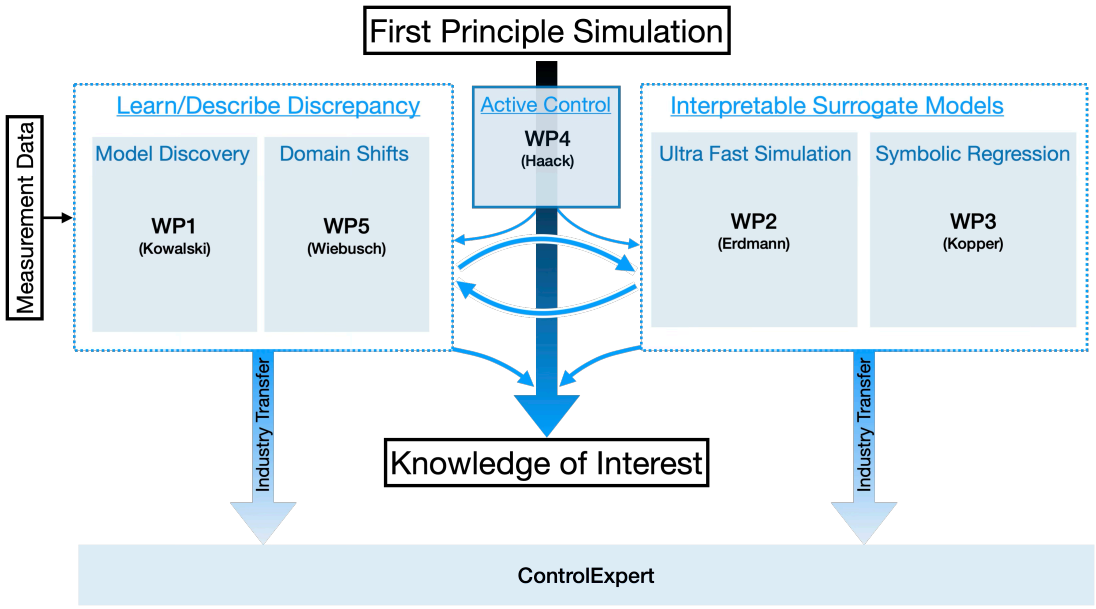

Figure 1: Overview of the work packages and their relations.

The approaches that will be developed are general and apply to a broad range of ErUM experiments and beyond. To show their real-world applicability, we will deploy them to a particle physics experiment (CMS) and an astroparticle physics experiment (IceCube) to problems in engineering and transfer this knowledge to industry.

An overview of the overall project is given in Figure 1. The key objectives of the project are:

Bayesian Model discovery: The goal is to develop a flexible tool that allows Bayesian inference on parameters of an underlying partial differential equation while complying with known physical constraints via constitutive neural networks (WP1).

Ultra-fast detector simulation with KANs: Goal is to develop a method to automize the extraction of parametrized detector responses from detailed detector simulations using Kolmogorov-Arnold Networks (KANs) (WP2).

Ultra-fast detector simulation with Symbolic Regression: The Goal is to utilize symbolic regression to derive interpretable analytic parametrizations of the outputs of complex simulators (WP3).

Active Control for Detector Simulations: The Goal is to implement an active learning paradigm for simulation pipelines. We aim to optimize simulation pipelines such that more computation time is spent on simulating examples of high relevance to downstream analysis tasks. (WP4)

Domain Shifts: The objective is to create methods that can identify and learn the systematic differences between two datasets. These methods allow us to pinpoint the causes of these differences and to adjust the data on an event-by-event basis between the two domains. This process improves the quality of less accurate or outdated datasets, reducing the need for costly computational simulations. (WP5)

All of these objectives have scientific applications in all ErUM research fields that rely on the simulation of physical processes. In addition, there are clear avenues for the transfer of the tools to be developed to industry. We will be working with *ControlExpert*, a company involved in digitizing and speeding up manual processes in automobile insurance claim management. This will enable us to apply, test, and verify our tools well outside our research fields. Our work will propose to enable ControlExpert to substantially save computational resources, and thus contribute to the sustainable transformation in industry.

1.2 Relation of the project to the funding policy goals

The project targets the development of new innovative tools and algorithms in experimental simulations using cutting-edge AI methods. Our goal of improving both the speed and accuracy of detector simulations simultaneously will be immediately beneficial for improving the performances of major research infrastructures and can be verified by the exploitation of our tools within these. However, the need for highly efficient and accurate first-principle simulation goes beyond particle and astroparticle physics but is relevant to many fields beyond, such as model-based engineering. The configuration of the applying consortium maps two major research infrastructures, CMS and IceCube, from two fields within ErUM with engineering and industry. Structurally, this facilitates the development of general-purpose tools beyond the immediate benefit for the studied use cases of the different work packages. Hence, we directly address the goals of interdisciplinary networking and transfer while at the same time fostering digital competence within ErUM in a general context. Moreover, the enhanced computational performance will also reduce the computing load, given more sustainable solutions in data science.

1.3 Scientific and/or technical aim of the project

The project focuses on a general challenge of many experiments in the field of ErUM and beyond. Most analyses of experimental data, as well as the training of machine learning tools, require precise simulations. More precise simulations and more detailed modeling usually require substantially increased computational resources. This means that they can often only be achieved as a vast collaborative effort. Still, physics analyses are often limited by the size of the simulated samples. This project aims to systematically approach this challenge using state-of-the-art AI methods for improving both, reducing the computational demands while improving the accuracy of complex simulations. We aim to develop generalized tools that enable the community to more efficiently use available computing resources with the long-term perspective of substantially reducing the power consumption related to simulation production.

1.4 Explanation of the interdisciplinary collaboration in the pillar "Software and Algorithms"

The project is proposed by an interdisciplinary group of applicants. We cover the two ErUM topics, particles and universe. The partners actively contribute to the major research infrastructures CMS and IceCube, explicitly mentioned in the call. Additionally, we include a group from engineering science with expertise in model-based developments in computational engineering. As an important key for extending the scope and applicability of our tools beyond the scope of ErUM, we collaborate with the industrial partner *ControlExpert*. This company strongly focuses on data science and develops AI based solutions in automobile insurance claim management.

2 Status of science and technology in the relevant field, previous work

2.1 Status of science and technology in the relevant field

Experiments in the ErUM fields rely heavily on Monte Carlo simulations to accurately describe the measurement and data reduction processes. These include, in particular, experiments in particle physics, astroparticle physics, hadron physics, and neutron physics. Mismatches in the simulated and experimental data, as well as limited simulation statistics, cause systematic uncertainties, which can bias the statistical analysis and thus have to be minimized as much as possible.

However, generating a large number of simulated events is computationally very costly. At the ATLAS experiment at the (HL-)LHC, for example, Monte Carlo simulations are expected to account for about 50% of the total CPU resources in 2030 [2], where the main driver of the necessary computing time is the Monte Carlo simulation of the detector response. At the LHC experiments, two main approaches have been developed to address this challenge, which differ in their accuracy: On the one hand, surrogate models have been developed based on deep generative models ("fast simulation") [3, 4, 5]. These models are meant to be used in data analysis and reach an impressive accuracy compared to the full Monte Carlo model. Still, they are costly to train and are often only available to members of the experimental collaborations. In situations where less accurate simulations are required, e.g., for detector design or phenomenological studies, analytic parametrizations can provide an appropriate tradeoff between accuracy and computational cost. In the past, analytic expressions for parametrising complex simulations have typically been derived manually, guided by physical intuition or simplified scenarios [6].

In IceCube, ML-based surrogate models have not yet been incorporated into the simulation pipeline. Here, the main computational cost is associated with simulating the Cherenkov photon transport in the ice. Traditional approximative methods, such as splines or analytic models [7], have been used as an alternative to the first-principles simulations. However, because of the high dimensionality of the problem, these approximative methods have to impose symmetries that do not hold in practice, limiting the precision of these methods.

Simulations for neutrino telescopes typically employ several techniques to improve the efficiency of the simulation. Techniques, such as weighted sampling, can lead to better phase-space coverage in the downstream event samples as more events of high relevance are simulated. Event-based biasing algorithms exist for dedicated simulation codes (e.g. [8] for air-shower simulations with CORSIKA). However, these algorithms are generally hand-tailored for a specific simulation and require deep intrusion into the simulation code. Nevertheless, surrogate models can predict many properties of simulated events, such as their approximative position in the final-level phase space after the detector response simulation, without performing the detailed detector simulation. As proposed in this project, a controlling algorithm will allow for fine control over the sample density of simulated events at the final analysis level, efficiently using available computing resources.

Symbolic Regression (SR) is a subfield of Machine Learning that seeks to discover mathematical expressions representing the relationships in data. SR is commonly employed in regression tasks to model the relationship between a dependent variable (target) and independent variables (predictors). Unlike classical parametric regression, which assumes a predefined functional form (e.g., linear regression), or non-parametric regression methods (e.g., kernel regression or neural networks), which are highly flexible but often opaque, SR does not require the user to specify a functional form beforehand. Instead, it searches the space of possible symbolic expressions to identify models that are both accurate and interpretable, bridging the gap between flexibility and interpretability. SR has been shown to successfully reproduce physical laws, such as learning Newton's Law of Gravity from observed orbital trajectories [9], the discovery of formulas from physics textbooks [10] and the construction of optimal observables for particle-physics experi-

ments [11]. Although using SR for density estimation has previously been explored (see [12]), its use in creating generative models is, to our knowledge, underexplored compared to regression tasks.

Recently, Kolmogorov-Arnold networks (KANs) have been proposed as an interpretable alternative to multilayer perceptrons (MLP) [13]. These networks learn the activation functions on the nodes instead of the weights in MLPs. The learned activation can then be directly interpreted or symbolically regressed to obtain an analytic description. Despite their advantages, SR and KANs have not been widely studied in ErUM research fields. The application of KANs was studied for event classification [14], regression [15] and higher-order calculations in particle physics theory [16]. The application of SR and KANs to detector simulations, as proposed in this project, is a novelty.

Surrogate modeling strategies are by now also an integral part of model-based engineering workflows, which often center around parameter-to-observable mappings aiming for sensitivity studies, uncertainty management, and parameter calibration, all of which require a large number of model evaluations. Approaches to surrogate modeling range from intrusive model-order reduction to regression-inspired statistical approaches to the application of modern machine learning.

Surrogate models that are non-intrusive concerning the underlying simulation method are particularly flexible and are also referred to as the objective bias method [17], for instance, the family of Gaussian process emulators [18, 19, 20], in which a Gaussian process is conditioned to simulation data. Gaussian processes can be evaluated ultra-fast and allow for an error estimation that can be exploited to design active learning strategies [21]. GP training, however, is computationally expensive and feasible only when the underlying parameter space is of moderate dimension. Whenever the surrogate has to account for a high-dimensional parameter space, e.g. describing the structure or geometry of an engineering system, neural networks constitute a better approach. A recent trend is to enrich the neural network with information on the underlying physical and mechanistic processes, either built into the network architecture as CNNs and GNNs, or even GANs, or physics-informed loss functions [17, 22] to account for inherent knowledge of the to-be-surrogated mechanistic processes or data dependencies. Automated model discovery has been proposed recently that enforces thermodynamic consistency through accounting properties, such as material objectivity, material symmetry, and incompressibility [23].

Simulations in a particle physics experiment form the foundation for the physics analysis. Still, they are first-order approximations of real-world measurements. These approximations introduce systematic discrepancies due to simplifications in simulation models and the inability to replicate complex detector responses, which can affect subsequent physics analyses. Additionally, fast-simulation surrogate models, while efficient, may exacerbate the issue by further obscuring event-level differences from the actual ground truth. Domain adaptation methods, such as GANs, diffusion networks, and normalizing flows, translate between domains without directly pairing examples from each domain. This method has already been effectively used to transfer images between domains in computer vision.

Techniques like CycleGAN or DRIT are widely used to map images from one style to another, such as converting photographs into artistic renditions, translating images between different seasons (e.g., summer to winter), or adapting satellite imagery for various environmental conditions [24, 25, 26, 27]. Recently, diffusion networks have gained considerable attention due to their substantial advancements in image generation tasks [28, 28, 29]. Domain adaptations have also been investigated in particle physics [30], while other approaches offer a reweighting of existing simulation data [31] to account for discrepancies to measurement data. While methods for domain adaptation have been studied in particle physics, they are still underutilized and not extensively researched, making this a promising area for further exploration.

2.2 Previous work of the applicants

Profile of the contributing workgroups

WG Erdmann Johannes Erdmann is a Heisenberg professor for big data analytics in physics research. He works on the application of deep learning methods for the data analysis in high-energy physics experiments and gravitational-wave physics. He is a member of the CMS and Einstein Telescope Collaborations. He has a strong background in LHC data analysis in top-quark and Higgs-boson physics, where he made leading contributions to the observation of new processes [32, 33, 34]. At the CMS experiment, he focuses on improving the precision of Higgs-boson measurements with deep learning [35]. He has developed applications for deep learning in the area of collider physics [36, 37, 38, 39] and has recently proposed the first application of Kolmogorov-Arnold networks in particle physics [14]. At the Einstein Telescope, he works on the application of differential programming for the reduction of Newtonian noise [40].

WG Haack As an early career researcher, C. Haack is formally part of C. Kopper's research chair, however, he leads an independent research program. As a member of the IceCube, KM3NeT, and P-ONE collaborations, he has worked on data analysis, reconstruction, and simulation for large-volume neutrino telescopes. In IceCube, he has focused on the development of analysis techniques for the measurement of the Galactic neutrino flux [41, 42, 43], simulations for future detector upgrades [44] and novel reconstruction techniques which have lead to the observation of a neutrino at the Glashow Resonance [45]. He has co-led the *Reconstruction Working Group*, which oversees implementing and applying novel machine-learning techniques. As the author of one of the standard analysis frameworks used in IceCube, he has broad experience in the design and management of software frameworks. He currently focuses on developing techniques for machine-learning-aided detector optimization for future neutrino telescopes, such as P-ONE [46], differentiable programming for optimizing data-analysis pipelines, surrogate models for detector simulation and the integration of machine learning models into data selection pipelines.

WG Kopper C. Kopper's focus is on detector simulation and modern analysis methods for high-energy neutrino detectors. He has been active in the IceCube/IceCube-Gen2, KM3NeT, and P-ONE collaborations. His current focus is on characterizing diffuse astrophysical flux measurements, event reconstruction tools, and developing new data analysis methods in high-energy neutrino detection. He is the main author of the *clsim* GPGPU-based photon propagation tool [47, 48] used extensively in high-energy neutrino telescopes to simulate their detector response. He has also been heavily involved in several of the key discoveries in IceCube, such as the initial discovery of the astrophysical neutrino flux [49, 50] and the first evidence for neutrino emission from a specific source (TXS 0506+056) [51]. He previously acted as Analysis Coordinator of the IceCube collaboration, a key leadership role in the collaboration's organization structure with extensive responsibilities in guiding the broad scientific output of the project. Currently, he co-leads the *Diffuse and Atmospheric Neutrino Fluxes Working Group* within the collaboration. He specializes in data analysis methods and, relevant to this proposal, in detector simulation techniques and ML/AI tools.

WG Kowalski Julia Kowalski's research focuses on the field of computational science & engineering. She and her team develop methods and software for data-integrated simulation models of complex system's including surface transport, heat transfer, phase-change and mixing processes [52, 53, 54, 55, 56]. The team also works on error-controlled surrogate models, e.g., based on Gaussian Processes, to achieve ultrafast evaluation of models, such as needed for sensitivity analyses, uncertainty management, Bayesian parameter calibration, and model selection [57, 58, 59]. Further areas of interest are sustainability and

reproducibility of the developed computational workflows, including software and (benchmarking) data sets [60], and their embedding into digital twins and virtual testbed infrastructure [61, 62]. In the context of DLR's Explorer Initiatives, she develops simulation models for trajectory and performance prediction of ice exploration technology [63] and, in that context, collaborated with Christopher Wiebusch, who is also a PI to this project. Julia Kowalski is a member of the board of directors of the Center for Modeling and Simulation Science of the Jülich Aachen Research Research Alliance (JARA), in which she fosters activities in sustainable computing, e.g., via organizing the Karman Conference on Sustainable Computational Science and Engineering[2]. She is also Steering Committee member of the Profile Area Production Engineering at RWTH and its Digital Twin Initiative.

WG Wiebusch The scientific focus of the research group (RWTH Aachen) is data analysis and detector development in astroparticle physics with emphasis on the measurement of high-energy cosmic neutrinos using the IceCube Neutrino Observatory [64], as well as neutrino oscillation physics with Double Chooz [65] and JUNO [66]. Furthermore, we are involved in technology development for future missions in space physics [67, 68, 69]. Special interest is placed on robust and precise estimation of astrophysical flux parameters [70], which is limited by the simulation-precision of challenging-to-control environmental and detector parameters, such as the depth-dependent transparency of Antarctic ice. We have extensive experience with machine learning in all areas of data selection, classification, and reconstruction, as well as in complex statistical procedures for data analysis. We sign responsible for the *Northern Tracks* selection, a standard data stream in IceCube, which is currently enhanced from a BDT-based to a new DNN-based selection. The group is familiar with various techniques in deep neural networks ranging from graph convolutional networks, see, e.g., [71], to more recent large language models and flow-based generative models. Within the ErUM-funded AISafety project, we have gained expertise in adversarial attacks and adversarial training in fundamental research, e.g., the recently developed algorithm MiniFool [72] that integrates experimental uncertainties into adversarial attacks of deep neural networks for data selection in IceCube [73]. Within the DFG-funded NeuroDOM project, we develop a new transformer-based event reconstruction for the IceCube Upgrade.

3 Detailed description of the work plan

3.1 Necessary project resources

The project is structured in five work packages as detailed in section 3.2. For each work package we request funding for a doctoral researcher or a part-time post-doc. Only those dedicated researchers will be funded through this proposal. The work-groups will be supplemented by master and bachelor students with research topics within ErUM-ARIA and further doctoral students and postdocs from the participating research groups with related research subjects. In total, we apply for five E-13 positions at 75 % salary for the duration of the project — one for each participating group. As the project work is carried out in close thematic proximity to computer science with higher salaries for doctoral researchers, the 75 % salary (above the often used 50 % to 67 %) enables the recruitment of technically competent employees also from the field of computer science or post-docs with experience in data science. Those scientists, if recruited, would work 75% directly on the project and 25% on the analysis of physics data, which would be financed from other funds (own contribution).

Our financial plan is shown in table 1. The personnel costs are based on flat N.N. personnel starting with E13 level-2, where the costs slightly differ between Erlangen and Aachen. This results in a total of 917 k€ for personnel.

[2] https://sustainable-cse.org/

Position	Duration	Principal Investigator	Costs
PostDoc/PhD	TVL-13, 75%, 3 years	Prof. Dr. Kowalski	184 700 €
PostDoc/PhD	TVL-13, 75%, 3 years	Prof. Dr. Erdmann	184 700 €
PostDoc/PhD	TVL-13, 75%, 3 years	Prof. Dr. Kopper	181 315 €
PostDoc/PhD	TVL-13, 75%, 3 years	Dr. Haack	181 315 €
PostDoc/PhD	TVL-13, 75%, 3 years	Prof. Dr. Wiebusch	184 700 €
Destination	**Number**	**Participants**	**Costs**
Int. Conferences & coll. Meetings	1 per year	1 participant per group	30 000 €
Europ. Conferences & coll. Meetings	1 per year	1 participant per group	19 500 €
National community meetings (DPG/ErUM)	1 per year	2 participants per group	18 000 €
Computing/Science school	1	1 participant per group	5500 €
Internal ErUM-ARIA meetings	1-2 per year	2-3 per group	9840 €
Internships at industry partner	1	1 participant per group	11 200 €
Project related consumables			7500 €
Total project costs (excl. Overhead)			1 018 700 €

Table 1: Overview of requested funding.

In addition to the personnel cost, we apply for travel funds restricted to the scientists directly involved in the project. The goal is to enable the participation of each scientist in one international conference and one collaboration meeting per year, except for the last year with only one of the two because of the planned internships (see below). Additionally, we apply for a one-time participation in a national science or computing school or lecture program for each scientist in the first year of the project. We apply for travel funds for the scientists and one master's student per group to participate in the DPG spring meeting or equivalent.

Internally, within the project, we plan for one in-person meeting per year with the participation of all people involved in the project, including master and bachelor students. A central element of the work plan is internships at the industry partner toward the end of the project. Here, we apply for financial support for the required daily allowances for a two-month internship for each scientist. These internships will be co-financed by the industrial partner. The requested travel funds total about 94 k€ which is roughly 10 % of the requested personnel funds. Finally, we estimate a small amount of 1.5k€ per group for general project-related expenses, such as publication costs and consumables not covered by the base funding. The total requested project funding is **1,018,700€** plus additonal 20 % overhead.

3.2 Work plan including milestones

We define the following milestones, which specify the key deliverables for our project.

- **Milestone 1 – Ultra Fast Surrogates developed** – We have utilized SR and KANs to create ultra-fast surrogate models for detector simulations. Performance and efficiency gain compared to first-principles Monte-Carlo simulation have been quantified
- **Milestone 2– Active Simulation Control integrated** – A framework for active simulation control is integrated into simulation pipelines to save computational resources
- **Milestone 3 –Domain Adaptations integrated** – A framework for domain adaption to mitigate domain shifts has been developed and applied to the ultra-fast surrogate models.
- **Milestone 4 – Efficiency Gain Quantified** – We have quantified the efficiency gain & resources savings of our methods
- **Milestone 5 – Methods applied in experiments.** – Surrogate models have been applied to experiments (IceCube, CMS, rheological engineering models).

Our 5 work packages are designed to achieve these milestones. The work breakdown structure of the individual work packages and milestones is shown in fig. 3. Our cooperation structure highlighting the collaboration on individual tasks is summarized in tab. 2 and explained in detail in Chapter 5. The leadership of WP1, WP2, and WP5 is assigned to RWTH Aachen, coordinated

by PIs Kowalski, Erdmann, and Wiebusch, respectively. WP3 and WP4 will be led by PIs Kopper and Haack at FAU Erlangen. C. Haack will lead the overall project.

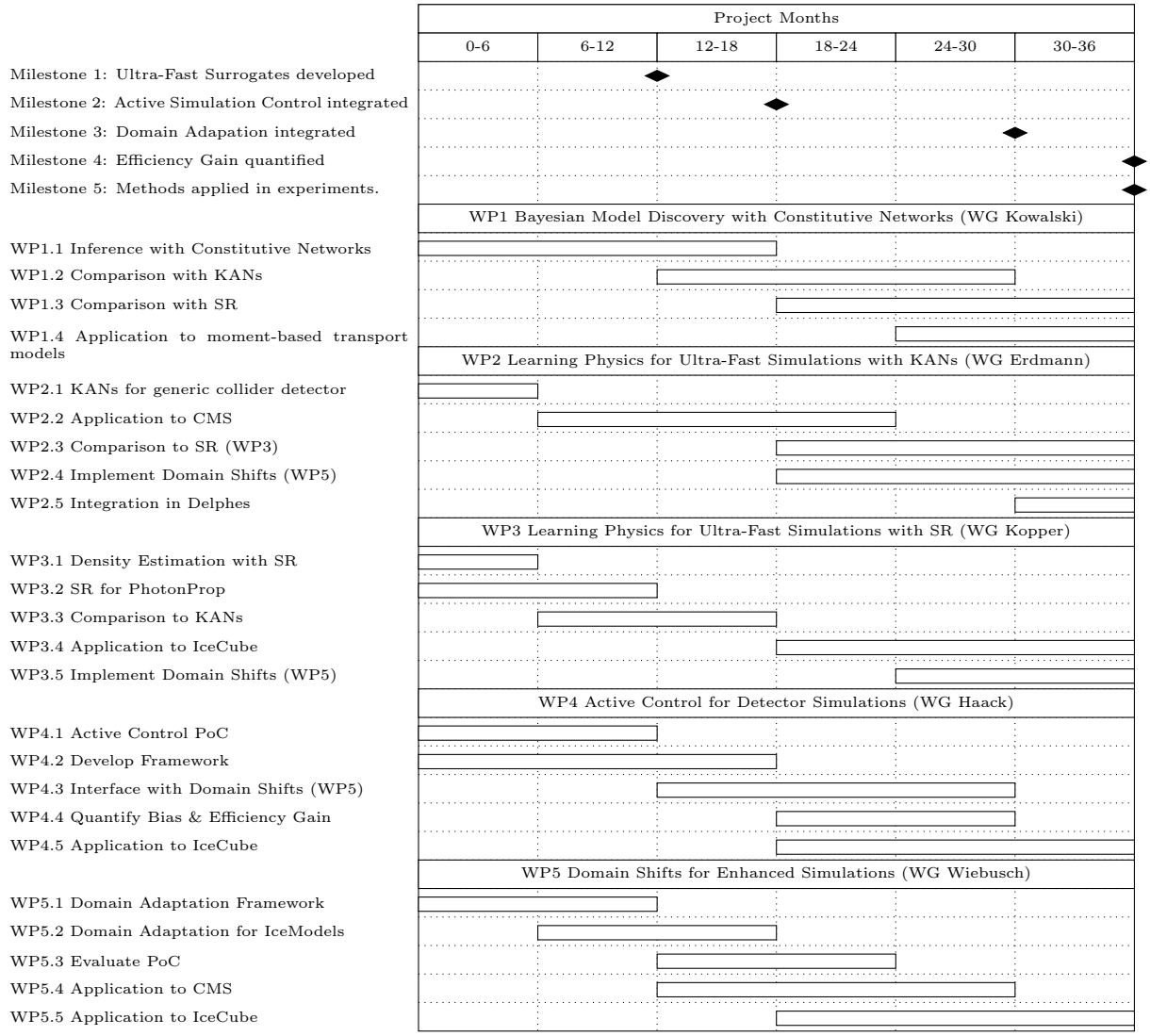

Figure 2: Gantt Chart for the Work Breakdown Structure

3.2.1 WP1: Bayesian Model Discovery with Constitutive Networks

The goal of this project is to utilize recently proposed thermodynamically consistent neural networks as surrogates for built-in rheologies. This will result in novel Bayesian model discovery. Rather than determining step by step the plausibility of candidate rheologies, we will infer its dominant contributions by means of Bayesian updates in a single step. This work package is a key consumer of the surrogate model techniques developed in WP2 and WP3 and demonstrates their applicability in the engineering context. The active simulation control framework developed in WP4 will allow us to increase the efficiency of our simulation pipeline.

Deliverables We will provide a framework for Bayesian model discovery of constitutive models in rheology and demonstrate the efficiency of surrogate models (**Milestone 4**) in describing experimental data (**Milestone 5**).

Work Package	Sub-Tasks	Principal Investigator	Cooperation Partner
Bayesian Model Discovery with Constitutive Networks			
WP1-1:	Inference with Constitutive Networks	Kowalski	Haack
WP1-2:	Comparison with KANs	Kowalski	Erdmann, Kopper
WP1-3:	Comparison with SR	Kowalski	Kopper, Erdmann
WP1-4:	Application to moment-based transport models	Kowalski	Haack
Learning Physics for Ultra-Fast Simulations with KANs			
WP2-1:	Proof of concept based on a generic collider detector	Erdmann	Haack, Kopper, Kowalski
WP2-2:	First application for the CMS detector in place	Erdmann	Haack, Kowalski
WP2-3:	Comparison to Symbolic Regression (WP3)	Erdmann	Haack, Kopper
WP2-4:	Implement Domain Shifts	Erdmann	Wiebusch
WP2-5:	Integration in Delphes	Erdmann	Wiebusch
Learning Physics for Ultra-Fast Simulations with Symbolic Regression			
WP3-1:	Proof of Concept: Symbolic Regression for PDFs	Kopper	Haack, Erdmann
WP3-2:	Photon propagation for a generic neutrino telescope	Kopper	Haack
WP3-3:	Comparison to KANs	Kopper	Erdmann, Kowalski
WP3-4:	Application to IceCube Photon Propagation	Kopper	Haack, Wiebusch
WP3-5:	Interface with WP5 to mitigate Domain Shifts	Kopper	Wiebusch
Active Control for Detector Simulations			
WP4-1:	Proof of Concept: Atmospheric Muon Background Simulation	Haack	Kopper, Wiebusch
WP4-2:	Development: Framework for Online Simulation Control	Haack	Erdmann, Kopper, Kowalski, Wiebusch
WP4-3:	Interface with Domain Shifts	Haack	Wiebusch
WP4-4:	Evaluation: Bias Quantification and Efficiency Gain	Haack	Kopper
WP4-5:	Application to IceCube	Haack	Kopper, Wiebusch
Domain Shifts for Enhanced Simulations			
WP5-1:	Development of Domain Shift Toolkit & Interface	Wiebusch	Kowalski
WP5-2:	Learn Domain Shifts related to simulated IceModels in IceCube	Wiebusch	Haack, Kopper, Kowalski
WP5-3:	Evaluate method performance and limitations of the IceCube test case	Wiebusch	Haack, Kopper
WP5-4:	CMS Interface, Application, and Evaluation	Wiebusch	Erdmann
WP5-5:	Evaluate Symbolic regression performances for different ice models	Wiebusch	Haack, Kopper

Table 2: Cooperation structure.

Context and Motivation Simulation models in engineering applications are often governed by parameters that are not known a priori and need to be calibrated, which means to identify the parameter that matches an observed process data best. In Bayesian parameter estimation, the parameter's posterior distribution is determined based on a prior, a likelihood function and the data's evidence. Evaluating the likelihood is computationally costly as soon as either the forward simulation is compute intense, or a large number of training data are considered. In order to render this a computationally feasible task one can therefore train a likelihood surrogate based on Gaussian processes (GPs) and utilize its built-in error estimate for active learning based on the information entropy. While this works well for calibrating isolated parameters, the calibration results do not automatically comply with known physical relations, e.g. incompressibility or invariants. Additional effort is needed, when we want to *calibrate* for the best built-in empirical model, e.g. a rheological or constitutive model. Model selection is then needed, which typically takes immense computational resources due to the need for high-dimensional integration.

Challenge: Model discovery based on novel constitutive neural networks, has been very successfully introduced in 2023 [23] as it enables the automatic identification of constitutive relations. To our knowledge, its application so far is restricted to cases, in which the network's target space corresponds to the observable space. In our case, we cannot directly observe in the target space, but need to conduct a simulation to relate both. In order to still learn from observations, we will employ Bayesian methods, potentially enhanced by Gaussian process surrogates and active learning. Combination of the two will yield an interpretable rheological or constitutive relation that is conditioned to indirect observations. Challenges will be associated with formulating the priors for predictors of the thermodynamically consistent network, and in evaluating the likelihood.

Tasks

- **WP 1.1 – Development: Bayesian Model Discovery** In a first step, thermodynamically consistent constitutive and rheological networks are build and tested against synthesized data in the network's target space. Next, Bayesian calibration is extended to learning hyperparameters of the network. We will also investigate computational feasibility enablers, such as GP-emulation of the likelihood and active learning strategies to minimize the necessary number of simulations.

- **WP 1.2 – Comparison with KANs** – We will investigate KANs (**WP2**) as an alternative to constitutive neural networks.
- **WP 1.3 – Comparison with SR** – We will investigate Symbolic Regression (**WP3**) as an alternative to constitutive neural networks.
- **WP 1.4 – Application: Moment propagation in heterogeneous background material** As a proof-of-concept, we will apply our novel Bayesian model discovery framework to calibrate for rheological relations in a moment based shallow flow setting [52] due to access to data. While the physical setting of the models differs from typical settings in astro-particle physics, the mathematical model structure of a propagating PDF shares commonalities with kinetic transport, and hence photon propagation. In a next step, we will extend our approach to other moment-based transport models in close collaboration with partners in this consortium.

3.2.2 WP2: Learning Physics for Ultra-Fast Simulations with KANs

The goal of this work package is to develop "ultra-fast simulations" by an automatized extraction of detector response parameterizations using KANs. The resulting simulations will be extremely fast to evaluate, as the particle responses are simple functions, and they will correspond much more closely to the real response of the fully simulated detector than existing parametrized simulations. We will apply the methods developed in this work in constitutive models (WP1) and the CMS (this WP), and IceCube experiments (WP 3). The active simulation control framework developed in WP4 will allow us to increase the efficiency of our simulation pipeline.

Deliverables We will provide a framework for parametrizing detector response functions with KANs (**Milestone 1**). The framework will be integrated in Delphes and applied to the CMS detector (**Milestone 5**). The resulting simulation efficiency gain will be quantified (**Milestone 4**).

Context and Motivation These ultra-fast simulations would be very useful for the design of analyses within the experimental collaborations. For example, the simulation of a new physics process would not need to be interfaced to the full software stack of the experiment. Still, the ultra-fast simulation can be used to get a first and reliable detector-level estimate. In addition, such ultra-fast simulations will be very useful for physicists outside of the experimental collaborations. For example, they will provide the possibility to account significantly better for detector effects in phenomenological studies that inspire new directions in the experimental data analysis. To our knowledge, no automated procedure for extracting response parametrizations exists in high-energy physics, despite their wide-spread use, for example, in the Delphes framework [6]. Kolmogorov-Arnold networks (KANs) [13] have recently been proposed as a more interpretable alternative to multilayer perceptrons (MLPs). Unlike MLPs, KANs have learnable activations on the edges, parametrized by splines. Given this recent development, only very few applications of KANs in particle physics exist, with the first proposed by the lead of this work package in Ref. [14]. In this work, small KANs have been shown to be interpretable. They can also be fit with a set of functions to obtain analytic expressions [13], which makes them attractive for the automated extraction of response parameterizations and their use in a Symbolic Regression context.

Tasks:

- **WP 2.1 – Proof of concept:** The concept of ultra-fast simulations will first be tested on a generic particle-physics detector. A small KAN will be used to learn analytic expressions of the detector response as a function of the relevent particle kinematics. These expressions are the parametrizations of the detector response for the ultra-fast simulation.

- **WP 2.2 – Application to CMS detector:** The concept will then be applied to full event simulations of the CMS experiment at the LHC. Ultra-fast simulations will be developed for different particle reconstruction and identification algorithms. The results will be compared with the current implement in the Delphes framework.
- **WP 2.3 – Symbolic regression comparison:** A detailed comparison of KAN-based symbolic regression and non-KAN-based symbolic regression methods from WP3 will be performed, assessing their precision and interpretability for detector simulations.
- **WP 2.4 – Domain shift implementation:** For several of the detector objects, such as electrons and muons, we will use experimental data from the CMS detector to learn the domain shift of experimental data to the ultra-fast simulations. For this sub-task, we will work closely together with researchers from WP5.
- **WP 2.5 – Integration in Delphes:** The detector response parametrizations will then be put into a detector card in the Delphes framework, due to its wide-spread use in particle physics. The expressions in the ultra-fast simulations, however, are independent of that particular framework and can also be used in other frameworks by the community.

3.2.3 WP3: Learning Physics for Ultra-Fast Simulations with Symbolic Regression

The goal of this work package is to utilize Symbolic Regression (SR) to obtain fast and interpretable analytic surrogate models for physics simulations. We will explore both the use of *vanilla* SR, which uses genetic programming or neural networks to explore the equation state space, and KANs (**WP2**), which provide a novel method of combinding symbolic regression and neural networks. **We will focus not only on classical regression problems, but aim to establish the use SR for density estimation**. The interpretability of the analytic expressions found by SR allows us to validate the models and explicitly include external inputs, such as medium properties, calibration constants, invariants, or conservation laws. The methods developed in this work package will be applied in constitutive models (**WP1**) and the CMS (**WP2**) and IceCube experiments (**this WP**).

Deliverables We will provide a framework for obtaining analytic surrogate models with SR (**Milestone 1**) and apply these surrogates in experiments (**Milestone 5**). The resulting simulation efficiency gain will be quantified (**Milestone 4**).

Context and Motivation Simulations for (astro-) particle physics experiments aim to produce samples from complex probability distributions, which encode the physical processes. As these distributions are generally not available in closed form, Monte Carlo simulation is used to transform random variates from simple distributions into the desired target distributions.

The computational complexity of these Monte-Carlo simulations is typically addressed by training generative models, such as diffusion models, variational auto-encoders, or normalizing flows. These implicitly learn the underlying distributions and allow for faster sampling. However, these models require large training datasets and specialized hardware (GPUs) and are not interpretable. **Invariants and conservation laws (such as energy or momentum conservation) typically have to be implicitly learned by the generative model**. Validation of these models can only be done statistically. Problems, such as distribution shifts, are challenging to detect and mitigate. Developing an SR framework that can generate high-quality approximations for probability distributions is a key goal of this work package. We will achieve this by combining SR with automatic differentiation to obtain cumulative distribution functions, which can then be differentiated to obtain the probability density functions.

Tasks

- **WP 3.1 – Development: Symbolic Regression for PDFs** The first goal of this work package is to develop a Symbolic Regression framework that can produce analytic expressions for probability distributions.
- **WP 3.2 – Proof of Concept: Photon propagation for a generic neutrino telescope** As a proof of concept, we will develop a surrogate model for photon propagation in a transparent medium (water or ice) as used in the simulation for neutrino telescopes. In this simulation, Cherenkov photons emitted by charged secondary particles produced in neutrino interactions are propagated through the detection medium by a first-principles physical model. The output of this simulation is a distribution of the number of detected photons and their arrival times. We will use SR to obtain analytic functions for the number of detected photons and the cumulative density function (CDF) of their arrival times. The resulting CDFs will be compared with MC simulations and existing simplified analytic models.
- **WP 3.3 – Comparison to KANs** Using the experiences and tools from **WP2** we will employ KANs as an alternative to *vanilla* SR for developing a surrogate model for photon propagation.
- **WP 3.4 – Application: Photon Propagation in Ice** Using the insights gained from the proof of concept, we will apply the SR framework to photon propagation in ice, as used in the simulation for the IceCube Neutrino Observatory. We aim to identify and extract physical parameters (such as the optical absorption and scattering lengths) from the SR-derived surrogate model and relate them to the known medium properties using the Bayesian Model Discovery framework developed in **WP1**.
- **WP 3.5 – Interface with WP5 to mitigate Domain Shifts** We utilize the IceCube interface to improve the performance of symbolic regression in WP3 and to enhance active control in WP4, assessing them on an event-by-event basis. Experimental data will be integrated into the evaluation process, allowing for a comprehensive analysis of systematic differences and their application to the fast simulations.

3.2.4 WP4: Active Control for Detector Simulations

This work package focuses on developing an *active simulation framework* to address the inefficiencies in simulation pipelines. The core idea is to integrate an adaptive decision-making mechanism within the simulation workflow to increase the relevance of simulated examples for downstream analysis workflows. **This work package provides a key synergy in our project, as it allows us to decrease the variance of Monte-Carlo estimators and model uncertainties for given computational resources**. The methods and framework developed here, will be applied in WPs 1, 2, 3 and 5.

Deliverables We will produce a software framework for implementing and integrating active simulation control for physics simulation pipelines (**Milestone 2**). The resulting simulation efficiency gain will be quantified (**Milestone 4**).

Context and Motivation Particle physics experiments generate vast quantities of raw detector data, which undergo multiple stages of processing and reduction to extract meaningful summary statistics. Accurate simulation of this data, including signals and backgrounds, is crucial for statistical analyses. The precision of the simulation results depends on minimizing MC statistical errors, which, if too large, introduce systematic uncertainties into the analysis.

The relevance of simulated events for downstream analysis tasks, such as measurements or the development of surrogate- and regression models, can vary dramatically. **The density of simulated events at the analysis stage is typically not uniform, resulting in regions of the parameter space with a large number of events (and thus low MC error) and regions that are poorly populated (and thus significant MC errors)**. Especially for

backgrounds, achieving low MC errors often conflicts with the data reduction processes designed to suppress such events. This leads to poor simulation efficiency, defined as the ratio of simulated events retained after reduction to the total number initially simulated. **For IceCube, this efficiency can be as low as** $\sim 10^{-7}$. Since the simulation process is resource-intensive, low efficiency translates to wasted computational resources and increased environmental and financial costs.

Proposed Solution: Active Simulation Framework Our approach is to train surrogate models to predict the relevance of simulated events for downstream analysis tasks early on in the simulation pipeline. Using predictions from the surrogate models, the framework evaluates whether an event is likely to contribute meaningfully to the final analysis. A heuristic, such as the current density of simulated events in specific regions of the observable space, determines whether an event is accepted or rejected. We can influence the tradeoff between computational complexity and bias by tuning the heuristic. An aggressive heuristic will drastically reduce the number of simulated events but might falsely reject events that have a significant impact on the downstream analysis. Studying and quantifying this tradeoff will be an essential aspect of this work package. Remaining biases will be addressed by the **Domain Adaptation Toolkit (WP5)**.

Tasks

- **WP 4.1 – Proof of Concept** – The initial proof-of-concept will focus on improving the simulation efficiency of the atmospheric muon background simulation pipeline. A surrogate model will be developed to predict the probability of an event surviving the selection process before the computationally expensive photon propagation step.
- **WP 4.2 – Development: Framework for Online Simulation Control** – Building on the proof-of-concept, we will generalize the active simulation framework for online control of simulations. This framework will integrate surrogate models in real-time decision-making to dynamically allocate resources during simulation runs. The framework will be designed to interface with common simulation production frameworks such as IceProd [74] or Gaudi [75]
- **WP 4.3 –Interface with Domain Shifts** – We will implement an interface with WP5 to mitigate biases introduced by the active simulation framework.
- **WP 4.4 –Quantify Bias and Efficiency Gain** – The framework will be rigorously evaluated to ensure no significant biases are introduced. Metrics such as simulation efficiency gains, resource usage, and accuracy of predictions will be quantified and compared to traditional MC simulations.
- **WP 4.5 –Application to IceCube** – Using the optimized simulation pipeline, we will produce a set of improved simulations for IceCube analyses. These simulations will serve as benchmarks to demonstrate the practical benefits of the active framework, including reduced computational cost and enhanced statistical precision.

3.2.5 WP5: Domain Shifts for Enhanced Simulations

This work package aims to develop AI-based tools to evaluate and quantify discrepancies between model simulations and experimental data, or between more accurate simulations. By employing domain adaptation techniques such as normalizing flows, GANs, and diffusion models, we will systematically learn and apply variations between datasets on an event-by-event basis, allowing us to investigate the root causes of these discrepancies. This iterative framework will enhance the accuracy of simulations while minimizing the need for computationally expensive resimulations. Additionally, these tools will offer valuable feedback for model development by identifying areas that require improvement. By doing this, **this work package provides support for WP2, WP3, and WP4 and enhances the precision of simulations and subsequent analyses**.

Deliverables We will produce a domain shift toolkit capable of identifying and mitigating discrepancies between datasets. This toolkit will be validated through applications to IceCube and CMS data (**Milestone 2 & 4**) and evaluated for its performance, scalability, adaptability and limitations (**Milestone 3**).

Context and Motivation Simulations in particle physics are essential for data analysis, but they still represent approximations of experimental conditions. This can introduce discrepancies that affect the accuracy and reliability of the results. These discrepancies often arise from simplifications made in the simulation models, such as assumptions regarding detector responses and environmental conditions. As these discrepancies accumulate over further approximation for surrogate models, they can lead to potential biases in subsequent analyses. Domain adaptation techniques, such as normalizing flow models, GANs, and diffusion models, have demonstrated success in other fields like computer vision [24, 25, 26, 27], where they are used to transfer features across domains with varying conditions. These methods enable the adaptation of datasets across different domains without needing direct pairs from each dataset. By utilizing these techniques, we can address systematic differences, enhance the accuracy of simulations, and optimize the use of existing computational resources, especially with legacy datasets. This approach enhances the accuracy of simulations and reveals systematic discrepancies, allowing for focused refinements.

Challenge Unlike grid-based datasets in computer vision, particle physics data involve irregular detector geometries and multidimensional features, such as timing, energy, and spatial information. Domain adaptation methods must be tailored to handle these complexities effectively. They need to integrate diverse data into a cohesive framework while maintaining the granularity required for detailed event-level analysis. Achieving this balance is critical to ensuring the robustness and accuracy of adapted simulations.

Tasks

- **WP 5.1 – Development and Proof of Concept of Domain Shift Toolkit** – Design and implement a flexible toolkit using state-of-the-art domain adaptation methods, including CycleGAN, normalizing flows, and diffusion models. This toolkit will be validated on an existing IceCube dataset, assessing its ability to detect and correct systematic differences. We will introduce artificial systematic deviations to the dataset to test the toolkit's accuracy in recovering the original inputs.
- **WP 5.2 – Learn Domain Shifts Related to Simulated Ice Models in IceCube** – Using an implemented interface to the IceCube experiment, we apply domain shifts to align older simulation datasets with current ice models. This approach reduces the need for new computationally expensive simulations while increasing the precision of analyses and enhancing data statistics.
- **WP 5.3 – Evaluate Method Performance and Limitations of the IceCube Test Case** – We evaluate the performance, scalability, and adaptability of the toolkit using the IceCube test case. This evaluation involves a thorough analysis of residual uncertainties, the ability to scale to larger datasets, and the adaptability to various calibration conditions. We will also identify potential limitations, such as biases or edge cases, to help guide future improvements
- **WP 5.4 – CMS Interface, Application, and Evaluation** – We are developing an interface for the CMS detector to ensure the toolkit's universal applicability. Next, we will apply domain shifts to CMS data to enhance the ultra-fast simulations developed in WP2, incorporating active feedback to refine the surrogate models.
- **WP 5.5 – Evaluate Symbolic Regression Performances for Ice Models, Including Experimental Data** – Leverage previously learned domain shifts from IceCube datasets to enhance symbolic regression performance on an event-by-event basis. Experimental data

will be incorporated into the evaluation, enabling a detailed examination of systematic differences and their application to fast simulations.

4 Exploitation plan

4.1 Commercialization prospects

The research topics addressed in this application are focused on applications in Astrophysics and Particle physics within the ErUM field. Though scientific progress is our driving motivation, our goal is the development of methods and tools that can be applied to other fields. This concept is also beneficial for our initial research goal because the validation of the applicability to problems in other domains enforces the generalization of approaches beyond the initially focused application. This concept is structurally embedded into the project by including partners from industry and engineering. ControlExpert, represented by Dr. Sebastian Schoenen, is a company that specializes in AI applications in automobile claim management. The chair of Methods for Model-based Development in Computational Engineering (MDB), represented by co-PI Julia Kowalski, focuses on innovative methods for engineering, particularly predictive process simulations for engineering science. For both fields, developing new innovative simulation methods is highly relevant, and the chances of commercial exploitation are very high. A relevant example for application is the generation of synthetic data of vehicles with different types of damages. For car damages, the amount of training data is very limited and the training of enhanced ML-models greatly benefits from a large sample of high-quality artificial data. The commercial interest is expressed in the attached letter of intent by *ControlExpert*. To foster this aspect, we plan for our doctoral researchers to conduct 2-month internships at the company to transfer our methods directly to industrial applications. The institute MDB is inherently entangled in the work plan and leads the WP-1.

4.2 Scientific and/or technological prospects

The proposed project team combines our long-year expertise in ML-based data science in the involved ErUM projects with expertise in model-based engineering, as well as an excellent industry partner within a closely entangled work plan. The work plan is based on clearly defined work packages and steps that can be achieved with tools and data that are currently available. Therefore, we consider the prospects of achieving our research goals high.

These results will directly impact the ErUM projects CMS and IceCube explicitly mentioned in the funding call. Furthermore, they can be applied to closely related projects such as ATLAS, other neutrino telescopes, and beyond. An example is the group in Erlangen, which is, beyond IceCube, involved in the neutrino telescopes KM3NeT and P-ONE. Common for most ErUM projects is that science results are immediately linked to the availability of high-quality and high-statistics of simulated data. Therefore, the central goal of improving both the quality and performance of simulations has very high prospects of immediate and long-term relevance in the ErUM research field and as a driver for innovation in the general community.

Throughout the project, our results will be continuously reported to the involved experimental collaborations as well as at national and international conferences. These are conferences in astro- and particle physics, conferences focused on AI-based methods in information science and engineering, and particularly meetings within the ErUM-Data and DIG-UM communities. For this, travel funds are requested in the application. Beyond that results will be published in journal papers with quality control. New methods will become technical papers authored by the collaborators of this project. Additionally, the successful application of science results will enter and improve publications by the involved experimental collaborations. Publications within our field, particularly if signed by full collaborations in view that first they need to be established in actual physics analyses of those collaborations, usually take longer time than few-author

technical papers. Therefore, we foresee that technical papers will be published mostly during the time frame of the project. Still, for the research results, we expect that the full exploitation of this project will continue for 2-5 years beyond the project's duration.

A further important exploitation aspect is the use of results for teaching and educating young researchers. Recently, requests for handing out research topics with the application of machine learning in data analysis for Bachelor and Master theses amount to more than 50% of requested thesis topics in our groups. Master and Bachelor students can be easily integrated into the workflow of this project. Based on the requests during recent years, we can estimate to supervise of the order of roughly 15-25 Bachelor and Master theses per year with direct or indirect connection to these projects. Furthermore, the involved PIs are very active in teaching, with dedicated lectures, seminars, and exercises for data science and ML applications. The topic of efficient simulation is ideally suited to be integrated into these courses. It will help improve the education of young students in data science and prepare them for later careers in the industry.

4.3 Scientific and commercial impact

The availability of high-quality and high-statistics simulation data is essential within the field of ErUM as discussed above. Therefore, the proposed research work is immediately connected to the full ErUM research field. Through explicit involvement in several experimental collaborations, the project team can directly apply the new research algorithms. The proposed methods have been planned as generic tools that can be easily adapted to different use cases and projects. We have planned our research program structurally in a multi-disciplinary framework. From this, we expect a high connectivity to other fields beyond ErUM.

4.4 Subsequent use

To successfully evaluate our developed methods and algorithms we need to implement them into the simulation software and data processing chain of the involved research facilities CMS and IceCube and potentially beyond. This guarantees the subsequent beneficial use within these facilities. Both projects have implemented thorough internal review procedures that evaluate each step of analyses, including verification of tools and excellent documentation. Furthermore, we plan to publish the code of our tools through platforms like GitHub with full open access. Our technical papers will be published in Open-Access journals and uploaded on preprint-servers like ArXiv. Also, our involved experimental collaborations usually publish all their results with open access and regularly release their data to the public. Examples are CMS Open data (https://cms-opendata-guide.web.cern.ch/) and data releases of IceCube (https://icecube.wisc.edu/science/data-releases/).

4.5 Sustainability

A more sustainable use of computing resources is put forward as a central goal of the funding call. This is targeted directly by this project. The generation of simulation data occupies roughly 50 % of the total computing budget of our facilities, see e.g. Figure 1 in [2]. The other half is shared by e.g. data processing, calibration, analysis, and statistical evaluation of confidence intervals. Therefore, simulation can be considered the hungriest consumer of energy for computing and correspondingly emission of CO_2.

Experience shows that improved computing —although more efficient — generally does not lead to switching off computing facilities and correspondingly to savings of CO_2. However, the exploitation of our methods within ErUM projects have the potential of dramatically improving the efficiency of used resources, resulting in a better ratio of resulting science per emitted CO_2. Beyond this, the estimation of needs for computing resources in future projects usually depends on the current state-of-the-art computing and the requirements of achieving the research goals. A detailed study for the HL phase of the LHC can be found in Figure 5 of [2].

Here, the implementation of our algorithms can lead to a significant reduction in the number of future computing facilities required for these projects. Furthermore, sustainability measures are implemented within our experimental collaborations. A prime example is the IceCube Neutrino Observatory which is hosted in one of the most fragile environments of our planet. For the planning of the next generation infrastructure, IceCube-Gen2, a dedicated *Sustainability & Environmental Impact working group* has been created to minimize the CO_2-footprint of the facility and significantly reduce the use of fossil fuels in comparison to the construction and operation of IceCube. Beyond power, transportation, and travel, computing has been identified as a key objectives for enhancing the sustainability (see e.g. the recently submitted proposal to the National Prioritization Process for Large Reserach Infrastructures of the BMBF).

Beyond the gain in efficiency, the transfer of methods to industry and engineering has the potential of a large multiplication of the gain within ErUM. In commercial industrial applications, improved modeling and more efficient simulation can immediately reduce computing demands on shorter timescales.

Beyond the central goal defined in the call, the Sustainability Guide for Research Processes of the DFG has also been taken into account. Within the applying universities, the involved groups at these universities can meet without the need of traveling. With only two applying universities also the traveling between them can be minimized by holding virtual phone meetings every month. In addition, we plan for a single in-person meeting per year only, but will utilize common meetings such as collaboration meetings or the DPG spring meeting for splinter meetings for in-person exchange. In addition, the proposed measures for reduced computing power are directly embedded in our work packages. Particular examples are WP-4 which directly steers the efficiency of simulations or WP-5 which has the potential of re-using existing simulations.

This proposal considers several of the sustainable development goals as defined by the Agenda 2030 of the United Nations. Our work program directly addresses goal 9 (resilient infrastructures and fostering innovation). Furthermore, our research is committed to the implementation of FAIR principles for exploitation and inclusive education based on equity and diversity, addressing not only goal 4 (Equity in quality education) but also goal 5 (Achieve gender equality). As our project structures are unbiased with respect to gender, they are thus ideally suited for empowering women and girls in science. Lastly, peace, as stated in the agenda, is a prerequisite for sustainability. The European Organization for Nuclear Research (CERN) is **the** European laboratory that represents peaceful cooperation between nations for collaborative research with an unprecedented history of meanwhile 70 years.

5 Work sharing/Cooperation with third parties

The work plan of the proposed project is structured according to closely entangled work packages (WPs) to ensure efficient collaboration and task distribution. The division of labor within the work packages and sub-tasks has been detailed in table 2. The close entanglement requires a well-organized cooperation in particular given the geographical distance between Aachen and Erlangen. Beneficial is that all project partners look back to a long-year history of joint collaborative work between each other. E.g. the groups of J. Kowalski and C. Wiebusch cooperate since more than a decade in space science projects funded by the DLR. The CMS and IceCube groups in Aachen have already successfully implemented a joint weekly meeting on ML-related topics in their research. The PI of our industry partner has a background in IceCube and many employees have been trained in particle physics. The overall project will be coordinated by the early career scientist C. Haack from Erlangen. As a former PhD graduate from Aachen, he provides excellent knowledge of both institutions as a key to implementing and supervising fruitful cooperation.

For internal cooperation and tracking the project's progress, we will run monthly virtual meetings and one dedicated in-person meeting per year — alternating between the sites. Additionally, researchers have the possibility of traveling between the institutions for dedicated work

meetings with colleagues at the respective other site. Furthermore, we can profit from existing structures, such as collaboration meetings and the possibility of a dedicated splinter meeting at the DPG spring meeting. Pre-defined milestones will be monitored continuously to ensure that issues and dependencies are resolved early and the project stays on track. During the later phase of the project, we have included internships at ControlExpert GmbH in the work plan, which will provide unique opportunities for hands-on collaboration with the industry.

Beyond internal collaboration, we will cooperate closely with all international collaboration partners from CMS and IceCube, as well as many other partners within the ErUM community. We will also cooperate closely with Dr. Martin Rongen from FAU Erlangen, an expert in modeling the optical properties of the Antarctic ice for IceCube. The German research structures provide excellent opportunities for collaborative work. That is, beyond the DPG spring meeting, frequent community meetings which are organized by the *Komitee für ElementarTeilchenphysik* (KET), the *Komitee für Astroteilchenphysik* (KAT), and last but not least within ErUM-Data, well-supported workshops organized by the ErUM-Data-Hub.

6 Necessity of the financial support

In this application we propose an interdisciplinary project that combines astroparticle physics with particle physics from within the ErUM program of the BMBF, engineering and an industrial partner for exploitation. The required resources are detailed in the work plan. The goals will be highly beneficial for the involved major research infrastructures IceCube and CMS, explicitly stated in the funding call. The proposed research plan is not covered by existing BMBF funding or other funding sources for these projects as it targets the development of cutting-edge AI methods for enhancing the digital competence within ErUM and beyond. The composition of the research group reflects this. The partners from the different research areas are not associated within another network or consortium nor can the work plan be mapped to existing project funding within the BMBF ErUM framework program. The necessary funds cannot be raised from the universities' basic funding. Therefore, funding within the framework of the "Software and Algorithms" program is essential for achieving the research objectives.

References

[1] C. Molnar, *Interpretable machine learning, 2 edn.(2022), Online Book: https://christophm. github. io/interpretable-ml-book* .

[2] ATLAS collaboration, *ATLAS HL-LHC Computing Conceptual Design Report*, CERN-LHCC-2020-015, https://cds.cern.ch/record/2729668.

[3] ATLAS collaboration, G. Aad et al., *AtlFast3: The Next Generation of Fast Simulation in ATLAS*, *Comput. Softw. Big Sci.* **6** (2022) 7, [2109.02551].

[4] M. Barbetti, *Lamarr: LHCb ultra-fast simulation based on machine learning models deployed within Gauss*, 21th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: AI meets Reality, 2303.11428.

[5] O. Amram et al., *CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation*, 2410.21611.

[6] DELPHES 3 collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [1307.6346].

[7] M. G. Aartsen, R. Abbasi, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers et al., *Energy reconstruction methods in the IceCube neutrino telescope*, *Journal of Instrumentation* **9** (Mar., 2014) P03009–P03009.

[8] Meagher, Kevin and van Santen, Jakob, *Parallelizing Air Shower Simulation for Background Characterization in IceCube*, *EPJ Web of Conf.* **295** (2024) 11016.

[9] P. Lemos, N. Jeffrey, M. Cranmer, S. Ho and P. Battaglia, *Rediscovering orbital mechanics with machine learning*, *Machine Learning: Science and Technology* **4** (oct, 2023) 045002.

[10] W. Tenachi, R. Ibata and F. I. Diakogiannis, *Deep symbolic regression for physics guided by units constraints: Toward the automated discovery of physical laws*, *The Astrophysical Journal* **959** (dec, 2023) 99.

[11] A. Butter, T. Plehn, N. Soybelman and J. Brehmer, *Back to the formula - LHC edition*, *SciPost Phys.* **16** (2024) 037.

[12] S.-M. Udrescu, A. Tan, J. Feng, O. Neto, T. Wu and M. Tegmark, *Ai feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity*, https://arxiv.org/abs/2006.10782.

[13] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić et al., *KAN: Kolmogorov-Arnold Networks*, 2404.19756.

[14] J. Erdmann, F. Mausolf and J. L. Späh, *KAN we improve on HEP classification tasks? Kolmogorov-Arnold Networks applied to an LHC physics example*, 2408.02743.

[15] E. Abasov, P. Volkov, G. Vorotnikov, L. Dudko, A. Zaborenko, E. Iudin et al., *Application of Kolmogorov-Arnold Networks in high energy physics*, 2409.01724.

[16] H. Bahl, N. Elmer, L. Favaro, M. Haußmann, T. Plehn and R. Winterhalder, *Accurate Surrogate Amplitudes with Calibrated Uncertainties*, 2412.12069.

[17] M. Raissi, P. Perdikaris and G. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, *Journal of Computational Physics* **378** (2019) 686–707.

[18] M. Seeger, *Gaussian processes for machine learning*, *International Journal of Neural Systems* **14** (2004) 69–106.

[19] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, vol. 2 of *Adaptive Computation and Machine Learning*. The MIT Press, Cambridge, MA, 2006.

[20] M. A. Alvarez and N. D. Lawrence, *Computationally efficient convolved multiple output gaussian processes*, *Journal of Machine Learning Research* **12** (2011) 1459–1500.

[21] A. Sauer, R. B. Gramacy and D. Higdon, *Active learning for deep gaussian process surrogates*, *Technometrics* **65** (2023) 4–18.

[22] K. Linka, A. Schäfer, X. Meng, Z. Zou, G. E. Karniadakis and E. Kuhl, *Bayesian physics informed neural networks for real-world nonlinear dynamical systems*, *Computer Methods in Applied Mechanics and Engineering* **402** (2022) 115346.

[23] K. Linka and E. Kuhl, *A new family of constitutive artificial neural networks towards automated model discovery*, *Computer Methods in Applied Mechanics and Engineering* **403** (2023) 115731.

[24] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, *Unpaired image-to-image translation using cycle-consistent adversarial networks*, Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.

[25] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh and M.-H. Yang, *Diverse image-to-image translation via disentangled representations*, Proceedings of the European conference on computer vision (ECCV), pp. 35–51.

[26] X. Huang, M.-Y. Liu, S. Belongie and J. Kautz, *Multimodal unsupervised image-to-image translation*, https://arxiv.org/abs/1804.04732.

[27] H.-Y. Lee, H.-Y. Tseng, Q. Mao, J.-B. Huang, Y.-D. Lu, M. Singh et al., *Drit++: Diverse image-to-image translation via disentangled representations*, https://arxiv.org/abs/1905.01270.

[28] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan and S. Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, https://arxiv.org/abs/1503.03585.

[29] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew et al., *GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models*, https://arxiv.org/abs/2112.10741.

[30] M. Baalouch, M. Defurne, J.-P. Poli and N. Cherrier, *Sim-to-real domain adaptation for high energy physics*, https://arxiv.org/abs/1912.08001.

[31] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman and J. Thaler, *Omnifold: A method to simultaneously unfold all observables*, *Physical Review Letters* **124** (May, 2020) .

[32] ATLAS collaboration, G. Aad et al., *Observation of Single-Top-Quark Production in Association with a Photon Using the ATLAS Detector*, *Phys. Rev. Lett.* **131** (2023) 181901, [2302.01283].

[33] ATLAS collaboration, M. Aaboud et al., *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, [1806.00425].

[34] ATLAS collaboration, G. Aad et al., *Observation of top-quark pair production in association with a photon and measurement of the $t\bar{t}\gamma$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Phys. Rev. D* **91** (2015) 072007, [1502.00586].

[35] CMS collaboration, *Measurements of inclusive and differential Higgs boson production cross sections at* 13.6 TeV *in the* $H \to \gamma\gamma$ *decay channel*, tech. rep., CMS-PAS-HIG-23-014, https://cds.cern.ch/record/2904882.

[36] C. C. Daumann, M. Donega, J. Erdmann, M. Galli, J. L. Späh and D. Valsecchi, *One Flow to Correct Them all: Improving Simulations in High-Energy Physics with a Single Normalising Flow and a Switch*, *Comput. Softw. Big Sci.* **8** (2024) 15, [2403.18582].

[37] J. Erdmann, A. van der Graaf, F. Mausolf and O. Nackenhorst, *SR-GAN for SR-gamma: super resolution of photon calorimeter images at collider experiments*, *Eur. Phys. J. C* **83** (2023) 1001, [2308.09025].

[38] J. Erdmann, T. Kallage, K. Kröninger and O. Nackenhorst, *From the bottom to the top—reconstruction of $t\bar{t}$ events with deep learning*, *JINST* **14** (2019) P11015, [1907.11181].

[39] J. Erdmann, *A tagger for strange jets based on tracking information using long short-term memory*, *JINST* **15** (2020) P01021, [1907.07505].

[40] P. Schillings and J. Erdmann, *Fighting gravity gradient noise with gradient-based optimization at the einstein telescope*, 2411.03251.

[41] M. G. Aartsen, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers, M. Ahrens et al., *Constraints on galactic neutrino emission with seven years of icecube data*, *The Astrophysical Journal* **849** (Oct., 2017) 67.

[42] A. Albert, M. André, M. Anghinolfi, M. Ardid, J.-J. Aubert, J. Aublin et al., *Joint constraints on galactic diffuse neutrino emission from the antares and icecube neutrino telescopes*, *The Astrophysical Journal Letters* **868** (Nov., 2018) L20.

[43] R. Abbasi, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers, M. Ahrens et al., *Observation of high-energy neutrinos from the galactic plane*, *Science* **380** (June, 2023) 1338–1343.

[44] M. G. Aartsen, R. Abbasi, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers et al., *Icecube-gen2: the window to the extreme universe*, *Journal of Physics G: Nuclear and Particle Physics* **48** (Apr., 2021) 060501.

[45] M. G. Aartsen, R. Abbasi, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers et al., *Detection of a particle shower at the Glashow resonance with IceCube*, *Nature* **591** (Mar., 2021) 220–224, Number: 7849 Publisher: Nature Publishing Group.

[46] C. Haack and L. J. Schumacher, *Machine-learning aided detector optimization of the Pacific Ocean Neutrino Experiment*, *PoS* **ICRC2023** (2023) 1059.

[47] D. Chirkin, J. C. Díaz-Vélez, C. Kopper, A. Olivas, B. Riedel, M. Rongen et al., *Photon propagation using gpus by the icecube neutrino observatory*, 2019 15th International Conference on eScience (eScience), pp. 388–393. DOI.

[48] C. Kopper, *clsim (public version)*, https://github.com/claudiok/clsim.

[49] IceCube collaboration, M. G. Aartsen et al., *Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector*, *Science* **342** (2013) 1242856, [1311.5238].

[50] IceCube collaboration, M. G. Aartsen et al., *Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data*, *Phys. Rev. Lett.* **113** (2014) 101101, [1405.5303].

[51] IceCube collaboration, M. G. Aartsen et al., *Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert*, *Science* **361** (2018) 147–151, [1807.08794].

[52] J. Kowalski and M. Torrilhon, *Moment approximations and model cascades for shallow flow*, *Communications in Computational Physics* **25** (2018) 669–702.

[53] L. Boledi, B. Terschanski, S. Elgeti and J. Kowalski, *A level-set based space-time finite element approach to the modelling of solidification and melting processes*, *Journal of Computational Physics* **457** (2022) 111047.

[54] M. T. I. Steldermann and J. Kowalski, *Shallow moments to capture vertical structure in open curved shallow flow*, *Journal of Computational and Theoretical Transport* **0** (2023) 1–31, [https://doi.org/10.1080/23324309.2023.2284202].

[55] U. Scholz, J. Kowalski and M. Torrilhon, *Dispersion in shallow moment equations*, *Communications on Applied Mathematics and Computation* (2023) 1–41.

[56] D. Weniger, P. L. Varghese, J. Kowalski and M. Torrilhon, *Unsteady stefan problem with kinetic interface conditions for rarefied gas deposition*, *International Journal of Heat and Mass Transfer* **217** (2023) 124696.

[57] H. Zhao, F. Amann and J. Kowalski, *Emulator-based global sensitivity analysis for flow-like landslide run-out models*, *Landslides* (2021) 3299–3314.

[58] H. Zhao and J. Kowalski, *Bayesian active learning for parameter calibration of landslide run-out models*, *Landslides* (2022) .

[59] A. Yildiz, I. Baselt, A.-K. Edrich, J.-T. Fischer, M. Mergili, H. Zhao et al., *Emulation techniques for rapid flow-like geohazards: A case study-based performance analysis*, EGU General Assembly 2021, online, 19–30 Apr 2021, pp. EGU2021–2453. DOI.

[60] J. Kowalski, A.-C. Plesa, M. Boxberg, J. Buffo, K. Kalousova, J. Kerch et al., *Compiling analysis-ready ice data across cryosphere disciplines*, EGU General Assembly 2024, pp. EGU24–21117. DOI.

[61] K. Kurgyis, P. Achtziger-Zupančič, M. Bjorge, M. S. Boxberg, M. Broggi, J. Buchwald et al., *Uncertainties and robustness with regard to the safety of a repository for high-level radioactive waste: introduction of a research initiative*, *Environmental Earth Sciences* **83** (2024) 82.

[62] A.-K. Edrich, A. Yildiz, R. Roscher and J. Kowalski, *A modular framework for fair shallow landslide susceptibility mapping based on machine learning*, *Natural Hazards* (2024) .

[63] J. Kowalski, L. Boledi, M. Boxberg, Q. Chen and A. Simson, *Cryotwin–digital infrastructure for virtually-assisted preparation and analysis of cryo-robotic exploration missions*, 84th EAGE Annual Conference & Exhibition, vol. 2023, pp. 1–5, European Association of Geoscientists & Engineers.

[64] IceCube collaboration, M. G. Aartsen et al., *The IceCube Neutrino Observatory: Instrumentation and Online Systems*, *JINST* **12** (2017) P03012, [1612.05093], [Erratum: JINST 19, E05001 (2024)].

[65] Double Chooz collaboration, H. de Kerret et al., *The Double Chooz antineutrino detectors*, *Eur. Phys. J. C* **82** (2022) 804, [2201.13285].

[66] JUNO collaboration, F. An et al., *Neutrino Physics with JUNO*, *J. Phys. G* **43** (2016) 030401, [1507.05613].

[67] D. Heinen, J. Audehm, F. Becker, G. Boeck, C. Espe, M. Feldmann et al., *The TRIPLE Melting Probe - an Electro-Thermal Drill with a Forefield Reconnaissance System to Access Subglacial Lakes and Oceans*, OCEANS 2021: San Diego – Porto, pp. 1–7. DOI.

[68] L. S. Weinstock, S. Zierke, D. Eliseev, P. Linder, C. Vollbrecht, D. Heinen et al., *The Autonomous Pinger Unit of the Acoustic Navigation Network in EnEx-RANGE: an autonomous in-ice melting probe with acoustic instrumentation*, *Annals of Glaciology* **62** (2021) 89–98.

[69] J. Kowalski, P. Linder, S. Zierke, B. von Wulfen, J. Clemens, K. Konstantinidis et al., *Navigation technology for exploration of glacier ice with maneuverable melting probes*, *Cold Regions Science and Technology* **123** (2016) 53–70.

[70] IceCube collaboration, M. G. Aartsen et al., *Observation and Characterization of a Cosmic Muon Neutrino Flux from the Northern Hemisphere using six years of IceCube data*, *Astrophys. J.* **833** (2016) 3, [1607.08006].

[71] M. Bachlechner, T. Birkenfeld, P. Soldin, A. Stahl and C. Wiebusch, *Partition pooling for convolutional graph network applications in particle physics*, *JINST* **17** (2022) P10004, [2208.05952].

[72] O. Janik et al., *MiniFool: Physics-Constraint-Aware Minimizer-Based Adversarial Attacks in Deep Neural Networks*, Journal paper in preparation, see also the thesis https://www.institut3b.physik.rwth-aachen.de/global/show_document.asp?id=aaaaaaaacgjxigu.

[73] IceCube collaboration, R. Abbasi et al., *Observation of Seven Astrophysical Tau Neutrino Candidates with IceCube*, *Phys. Rev. Lett.* **132** (2024) 151001, [2403.02516].

[74] J. C. Díaz-Vélez, *The iceprod (icecube production) framework*, *Journal of Physics: Conference Series* **513** (jun, 2014) 032026.

[75] G. Barrand et al., *GAUDI - The software architecture and framework for building LHCb data processing applications*, 11th International Conference on Computing in High-Energy and Nuclear Physics, pp. 92–95.

ControlExpert

ControlExpert GmbH | Marie-Curie-Straße 3 | 40764 Langenfeld

Dr. Sebastian Schoenen
Phone +49 2173 849 84-732
Mobile +49 151 527 032 73
s.schoenen@controlexpert.com

Tuesday, January 14, 2025

Letter of Intent – „ErUM-ARIA – Simulations: Active, Refined and Interpretable Approaches“

As an established market leader founded in 2002, ControlExpert (CE) has been processing over 20 million motor claims per year with 1000 employees worldwide. In our role as a digitalization expert, we combine AI-based processes and the latest technologies with the knowledge of our automotive experts. Together with CE's in-house research and development department with a team of more than 30 data scientists, we work to realize our vision: Car drivers all over the world will have their claims settled fairly at the very same day. Among others our research focuses on AI, Deep Learning, computer vision and telematics. Our customers include all the reputable insurance companies, leasing companies, car dealerships, and manufacturers in the automobile industry worldwide.

As an associated partner, CE would like to support the “ErUM-ARIA” research project to develop resource-saving physical simulations. The aim is to develop innovative AI methods in an interdisciplinary team that replace complex simulation chains with more efficient and more interpretable approximations. In addition, AI methods are to be developed to actively control the simulations and to precisely map systematic differences between data sets. CE considers this research project to be innovative and promising to offer significant added value for the sustainable use of AI simulation technologies in both academic and industrial contexts.

For CE, the generation of high-quality synthetic data, such as images of vehicles with different damage patterns, is a challenge. This data is valuable for training advanced damage assessment models, enabling models to be adapted more efficiently to changing market situations and minimizing resource consumption. Another specific problem for CE is minimizing the domain shift that occurs when models trained on synthetic data are applied to real data. Existing technologies do not offer satisfactory solutions to these problems. It is expected that the methodologies to be developed in the research project will offer promising solutions. In addition, the precise reconstruction of vehicle damage based on sensor data, which is often only available in small quantities, poses a further challenge. The methods developed in the research project, which make it possible to map systematic correlations between sensor data and vehicle damage using interpretable approximations, open up the possibility for CE to use these models profitably and effectively in the future.

CE will continuously support the developments in the research project. As part of the research project, the company will support doctoral students in the form of internships to transfer the methodologies developed on academic use cases to industrial use cases. Workstations, access to data, training and regular feedback will be provided. CE will cover the internal expenses by itself.

Best regards,

Sebastian Schoenen
Director of Innovation & Technology