

# Optimising the Active Muon Shield for the SHiP Experiment at CERN

A. Baranov<sup>1,2,\*</sup>, E. Burnaev<sup>3</sup>, D. Derkach<sup>1,\*</sup>, A. Filatov<sup>1,\*</sup>,  
N. Klyuchnikov<sup>3</sup>, O. Lantwin<sup>4,\*</sup>, F. Ratnikov<sup>1,2,\*</sup>, A. Ustyuzhanin<sup>1,2,\*</sup>  
and A. Zaitsev<sup>3</sup>

<sup>1</sup> National Research University Higher School of Economics, Moscow, Russia

<sup>2</sup> Yandex School of Data Analysis, Moscow, Russia

<sup>3</sup> Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>4</sup> Imperial College London, London, UK

\* on behalf of the SHiP Collaboration

E-mail: [filatovartm@gmail.com](mailto:filatovartm@gmail.com)

**Abstract.** The SHiP experiment is designed to search for very weakly interacting particles beyond the Standard Model which are produced in a 400 GeV/c proton beam dump at the CERN SPS. The critical challenge for this experiment is to keep the Standard Model background level negligible. In the beam dump, around  $10^{11}$  muons will be produced per second. The muon rate in the spectrometer has to be reduced by at least four orders of magnitude to avoid muon-induced backgrounds. It is demonstrated that new improved active muon shield may be used to magnetically deflect the muons out of the acceptance of the spectrometer.

## 1. Introduction

The Standard Model of elementary particles despite its enormous success has several places for improvement [1]. In particular, the baryon asymmetry of the Universe [2] and a Dark Matter existence are yet to be explained. Heavy Neutral Leptons (HNLs) [3], which are right-handed partners of the Standard Model neutrinos should they be discovered will give a boost in our understanding of both above mentioned riddles of Nature. The existence of such particles is strongly motivated by theory and were searched extensively previously. Cosmological constraints on the properties of HNLs now indicate that the majority of the interesting parameter space for such particles was beyond the reach of the previous searches at the PS191 [5], BEBC [7], CHARM [4], CCFR and NuTeV [6] experiments.

For this reasons, a new fixed-target experiment at the CERN SPS accelerator was proposed [8]. This experiment will use a 400 GeV proton beam on a fixed target to produce a large number of charm mesons. The HNLs from charm meson decays have a significant polar angle with respect to the beam direction, approximately 50 mrad on average. In order to maximise the geometric acceptance for a given transverse size of the detector, the detection volume must therefore be placed as close as possible to the target. The production of the charm mesons is accompanied by copious direct production of pions, kaons and short-lived light resonances. The subsequent decays of these particles would result in a large flux of muons and neutrinos. To minimise these decays, a combination of a target and a hadron absorber of a few metres length, both made of as dense a material as possible, is required. To reduce the detector occupancy and

backgrounds induced by the residual muon flux, a muon shield is required downstream of the hadron absorber. The experimental set-up must therefore balance the opposing requirements of locating the detector as close as possible to the target and of accommodating a sufficiently long muon shield upstream of the fiducial volume of the detector to reduce muon-induced backgrounds.

## 2. Bayesian Optimisation

The main goal of our research was to find a light and efficient magnetic shield. In order to achieve this, we applied a Bayesian optimisation algorithm. In this section we give an introduction to this method and motivation behind this approach.

Let  $f(x)$  be a black-box function on  $X \subset \mathbb{R}^d$ , which does not have an analytic expression and can be evaluated at a specific point only by conducting computationally intensive simulations, which in turn can yield noisy observations of the form  $y_i = f(x_i) + \epsilon_i$ , where  $\epsilon_i$  is the random variable with zero mean and constant variance. Our ultimate goal is to find a point  $x_*$  where  $f(x_*)$  is a true global optimum (henceforth without loss of generality we will consider a maximisation problem). In this problem setup, the general framework for optimum searching iteratively constructs the sample  $D_N = \{(x_i, y_i)\}_{i=1}^N$ , each iteration consists of three steps:

- (i) Choose a new point  $x_i$  for simulation
- (ii) Conduct simulations to obtain  $y_i = f(x_i) + \epsilon_i$
- (iii) If a stopping criterion is not met, add  $(x_i, y_i)$  to the sample  $D_N$  and proceed to iteration  $i + 1$

Convergence to the exact solution of the problem may require infinite amount of iterations, however, under some theoretical assumptions any degree of its approximation can be obtained within the corresponding finite horizon.

### Theorem 1

If  $f(x)$  is Lipschitz-continuous with a Lipschitz continuity constant  $C$ ,  $X$  is a  $d$ -dimensional unit hypercube, then in the noise-free case for any  $\varepsilon > 0$  some point  $x_+$  with near-optimal true value  $f(x_+) : f(x_+) \geq \max_{x \in X} f(x) - \varepsilon$  is guaranteed to be found within  $(C/2\varepsilon)^d$  iterations. (see [10])

Strong convergence guarantee is provided against the worst-case scenario, however the solution in this case will typically require an impractically large number of simulations for any reasonable accuracy of approximation. Since such scenarios are not plausible in practice and the number of simulations is very limited, one can come to the idea to relax guarantees in order to meet practical limitations, yet insure against average-cases. Bayesian approach to optimisation allows implementing this idea by providing specific choosing strategies (step i) in the framework for optimum search. In this approach, the objective function is approximated with a *surrogate model*, that is, a parametric function for which prior knowledge of parameters combined with likelihood of the observed sample defines a posterior distribution over the parametric family. Choice of new points to simulate is guided by the *principle of minimum expected risk* (its formalisation under the Bayesian framework has been provided in [9]) resulting in a secondary optimisation problem, which in turn depends not on the original function, but on its computationally cheaper surrogate-model.

The most popular choice for building surrogate models in the field of engineering design is based on taking *Gaussian processes* as a parametric family of priors along with assuming there is white noise in the observations  $\epsilon_i = \mathcal{N}(0, \sigma^2)$ . By definition [11, 22] Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. Consequently it is completely specified by its mean function  $m(x) = \mathbb{E}[f(x)]$  and the covariance function

$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$ . The most important property of Gaussian processes for Bayesian optimisation is that posterior takes also a form of Gaussian process, moreover its mean and covariance functions have analytic expressions:

- $\hat{f}_N(x) = \mathbb{E}[f(x)|D_N] = m(x) + \mathbf{k}(x)\mathbf{K}^{-1}(y_N - \mathbf{m})$ ,
- $\hat{\sigma}_N^2(x) = \mathbb{V}[f(x)|D_N] = k(x, x) - \mathbf{k}(x)\mathbf{K}^{-1}\mathbf{k}(x)^T$ ,

where  $\mathbf{m} = \{m(x_i)\}_{i=1}^N$ ,  $\mathbf{k}(x) = \{k(x, x_i) + \sigma^2\delta(x, x_i)\}_{i=1}^N$ ,  $\mathbf{K} = \{k(x_i, x_j) + \sigma^2\delta(x_i, x_j)\}_{i,j=1}^N$ . Several conditions on the family of priors that ensure convergence of the Bayesian optimisation to the optimum are specified in [12], they are satisfied by a Gaussian process with a constant mean function. Therefore, the convergence is guaranteed when the principle of minimum expected risk is used, as for instance an expected deviation from the global optimum:

$$x_{i+1} = \underset{x}{\operatorname{argmin}} \mathbb{E}[|f(x) - f(x_*)||D_i] \quad (2.1)$$

Expression 2.1 is computationally expensive since it requires estimation of  $f(x_*)$ , thus in present work we replace this criterion with its common approximation called *Expected Improvement* [13]:

$$x_{i+1} = \underset{x}{\operatorname{argmin}} \mathbb{E}[\max\{0, f(x) - f(x_+)\}|D_i], \quad (2.2)$$

where  $x_+ = \max_{j=1..i} y_j$ . Expected Improvement has also been shown to converge under additional mild assumptions [14], but its main advantage is a closed-form expression, that doesn't require numerical integration:

$$\mathbb{E}[\max\{0, f(x) - f(x_+)\}|D_i] = \hat{\sigma}_N^2(x)Z\Phi(Z) + \hat{\sigma}_N^2(x)\phi(Z), \quad (2.3)$$

where  $Z = \frac{\hat{f}_N(x) - f(x_+)}{\hat{\sigma}_N(x)}$ ,  $\phi$  and  $\Phi$  are PDF and CDF of the standard normal distribution respectively.

### 3. SHiP shield optimisation

The muon shield is a critical component of the SHiP experiment, which deflects the high flux of muons produced in the target, that would represent a very serious background for the particle searches, away from the detector. The shield consists of eight magnets and each magnet is parameterised by seven values: length, width, etc... Because the cost of the muon shield is significant we include the weight of the muon shield as a proxy for the cost. Therefore, our aim is to find the most efficient solution at a lowest cost possible. We apply Bayesian optimisation techniques to solve this problem. The optimisation is performed for already chosen material, thus the material balance is discussed elsewhere [15].

For the evaluation of the shield performance we use 18 million simulated events passed through the muon shield and measured at a scoring plane 64m downstream of the muon shield at the z of the first tracking station of the spectrometer. All subsystems downstream of the muon shield are removed.

Proton fixed target collisions and inelastic muon interactions are simulated using PYTHIA8 [16] and PYTHIA6 [18] respectively.

Heavy flavour cascade production is also taken into account [19]. The SHiP detector response is simulated in the GEANT4 [20] framework. The simulation is done within the FAIRROOT framework [21].

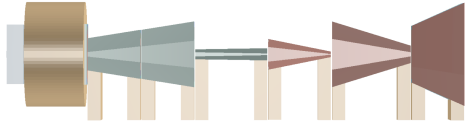
The transverse  $(x, y)$  position of muons is obtained at the first tracking station. For muons with  $|y| < 5\text{m}$ , their  $x$ -position is converted into "importance weight"  $\sigma_{\mu^\pm} = \sqrt{(260 \mp x_\mu/\text{cm})/560}$

for  $-300 < x_\mu < 260$  and  $-260 < x_\mu < 300\text{cm}$  for positive and negative muons respectively, else  $\sigma_{\mu^+} = 0$  [15].

Now we can introduce the complete loss function, that depends on the physical performance of the shield, its weight, and the length implicitly via the weight.

$$L(\Sigma, W) = (1 + \Sigma)(1 + \exp^{\frac{W-W_0}{W_0}}),$$

where  $\Sigma = \sum_\mu \sigma_\mu$ ,  $W$  is a weight of configuration and  $W_0$  is a weight of a baseline configuration [15] shown in Fig.1. The weight of the baseline is about 1900 tons and  $\Sigma$  is approximately equal to 32. As can be seen, weight is penalised exponentially because we are not interested in the heavy regions, as we will not be able to construct such configurations due to budget and engineering constraints.



**Figure 1:** Baseline configuration of the SHiP muon shield, as described in [15].

The following problems are addressed during the optimisation:

- (i) Optimisation is performed in 42-dimensional space
- (ii) Computation of the  $\Sigma$  is time consuming.
- (iii) Computation of the  $\Sigma$  is noisy due to limited statistics in Monte - Carlo simulations.

We choose only 500k muons with the biggest momentum and discard all the low-momentum muons.

To ensure a more even coverage of the muon phase space, the number of muons is capped in bins of momentum and transverse momentum, and augmented through resampling available muons in bins with low occupancy. This also reduces the overall number of muons to be simulated to about 500 thousand, and helps us to decrease the time of computations by factor of 8 times in average. We also introduce noise as a prior knowledge into Gaussian Process: in this setup GP tries to estimate the noise of the computation and incorporate it into final variance. The points are computed in batches of 100 points to optimise the usage of the available computing resources. The process of optimisation is illustrated in Fig. 2.

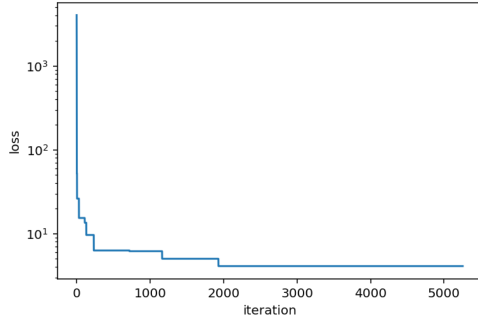
The optimisation procedure is stopped after 5000 iterations. The obtained configuration is found to be lighter by 25 % than the baseline while having about the same rejection capability as illustrated in Fig. 3. Fig. 4 illustrates the new found configuration.

## 4. Conclusion

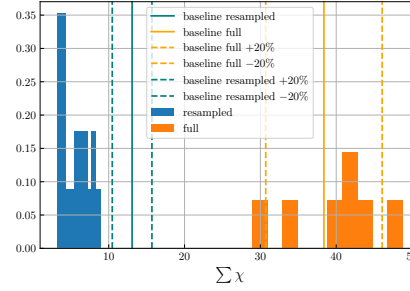
Bayesian optimisation is a powerful tool which can be applied for the optimisation of non-differentiable functions. We have demonstrated that this method can be successfully applied even to complicated optimisation problems in physics. However even though Gaussian processes can guarantee global optionality under ideal conditions, these are not satisfied for this particular problem.

## References

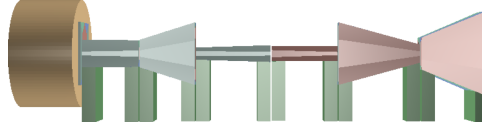
- [1] Ellis J *et. al* 2009 *Phys. Lett. B* **679** 369-375
- [2] Asaka *et. al* 2005 *Phys. Lett. B* **620** 17-26



**Figure 2:** Evolution of the best found configuration with GP iterations.



**Figure 3:** Distributions of rejection performance scores for re-sampled and full muon samples for the best found shield configuration, comparing with the baseline solution.



**Figure 4:** The best found muon shield configuration for the SHiP detector

- [3] Asaka *et. al* 2005 *Phys. Lett. B* **631** 151-156
- [4] Bergsma F *et. al* 1986 *Phys. Lett. B* **166** 437-478
- [5] Bernardi G *et. al* 1986 *Phys. Lett. B* **166** 479-483
- [6] Vaitaitis A *et. al* 1999 *Phys. Rev. Lett.* **83** 4943-4946
- [7] Cooper-Sarkar *et. al* 1985 *Phys. Lett. B* **160** 207-211
- [8] Bonivento W *et. al* 2013 Proposal to Search for Heavy Neutral Leptons at the SPS *Preprint 1310.1762*
- [9] Burnaev E and Panov M *et. al* 2015 *Proc. SLDS 2015* (Egham) 116-125
- [10] Betrò B 1991 *J. of Glob. Opt.* **1** 1-14
- [11] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (University Press Group Limited)
- [12] Mockus J 1994 *J. of Glob. Opt.* **4** 347-365
- [13] Mockus J, Tiesis V and Zilinskas A 1978 *Towards Global Optimization 2* (Elsevier Science Ltd.) 117-129
- [14] Vazquez E and Bect J 2010 *J. of Stat. Plan. and Inf.* **140** 3088-3095
- [15] Akmete A *et al.* 2017 *JINST* **12**
- [16] Sjöstrand T, Mrenna S, and Skands P 2008 *Comput.Phys.Commun.* **178** 852-867
- [17] Andreopoulos C, Bell A, Bhattacharya D, Cavanna F, Dobson J, *et al* 2010 *Nucl.Instrum.Meth.* **A614** 87-104
- [18] Sjöstrand T, Mrenna S, and Skands P 2006 *JHEP* **05** 026
- [19] Dijkstra H and Ruf T 2015 *SHiP-NOTE-2015-009*
- [20] Agostinelli S *et al.* 2003 *Nucl.Instrum.Meth.* **A506** 250-303
- [21] Al-Turany M Bertini D Karabowicz R Kresan D Malzacher P *et al.* 2012 *J.Phys.Conf.Ser.* **396**
- [22] Burnaev E Panov M and Zaytsev A 2016 *Journal of Communications Technology and Electronics* **61** 661-671.