Optimization of the magnetic horn for the nuSTORM non-conventional neutrino beam using the genetic algorithm

A. Liu\textsuperscript{a,1}, A. Bross\textsuperscript{a}, D. Neuffer\textsuperscript{a}

\textsuperscript{a}Fermilab, P.O.Box 500, Batavia, Illinois, 60510, USA

Abstract

This paper describes the strategy for optimizing the magnetic horn for the neutrinos from STORed Muons (nuSTORM) facility. The nuSTORM magnetic horn is the primary collection device for the secondary particles generated by bombarding a solid target with 120 GeV protons. As a consequence of the non-conventional beamline designed for nuSTORM, the requirements on the horn are different from those for a conventional neutrino beamline. At nuSTORM, muons decay while circulating in the storage ring, and the detectors are placed downstream of the production straight so as to be exposed to the neutrinos from muon decay. nuSTORM aims at precisely measuring the neutrino cross sections, and providing a definitive statement about the existence of sterile neutrinos. The nuSTORM horn aims at focusing the pions into a certain phase space so that more muons from pion decay can be accepted by the decay ring. The paper demonstrates a numerical method that was developed to optimize the horn design to gain higher neutrino flux from the circulating muons. A Genetic Algorithm (GA) was applied to the simultaneous optimization of the two objectives in this study. The application of the technique discussed in this paper is not limited to either the nuSTORM facility or muon based facilities, but can be used for other neutrino facilities that use magnetic horns as collection devices.

© 2014 Published by Elsevier Ltd.

Keywords: nuSTORM, Genetic algorithm, Magnetic horn, Neutrino beam

1. Introduction

The nuSTORM facility was designed to deliver $\nu_e$ and $\nu_\mu$ beams from the decay of $\mu^+$ beam stored in a racetrack ring. The facility, coupled with neutrino detectors, is capable of searching for sterile neutrinos with high sensitivity and can serve a future long-base neutrino oscillation program by providing high-precision measurements of neutrino cross sections. It can also serve as a critical step in muon accelerator development.

The nuSTORM facility is the simplest realization of the Neutrino Factory (NF) concept, of which the R&D challenges and cost are still actively being studied. At nuSTORM, a \textasciitilde 500 m racetrack-like decay ring is used to generate neutrinos with precisely known flux and flavor. The possibility of using such a muon racetrack decay ring with long straight sections to study low energy neutrino interactions has been recognized and was first proposed by Neuffer [1, 2, 3]. No fast kickers with large aperture are needed in the nuSTORM facility, owing to the fact that a new injection scheme was proposed for injecting pions in a muon storage ring. In this stochastic injection scenario, the pions are injected into the decay ring by a specially designed "Orbit Combination Section" (OCS) and then decay to muons in the \textasciitilde 150 m long production straight of the
ring. The feasibility of the stochastic injection used in nuSTORM was confirmed by Neuffer and Liu [4]. A schematic drawing of the facility is shown in Figure 1.

At nuSTORM, the pions are generated by bombarding a solid target with 120 GeV protons from the Fermilab Main Injector (MI) and collecting them with a magnetic horn. The pions are then transported in an injection beamline to the decay ring. The magnetic horn was first proposed by van der Meer in 1961 to focus pions forward in order to improve the beam in neutrino experiments [5]. Its success has been widely recognized, and it is now a standard element in all neutrino experiments worldwide. It is capable of focusing secondary particles with a wide angular divergence and momentum spectrum. Horn designs and the effectiveness of their collection are described in detail in [5, 6, 7, 8, 9].

The injection takes place at the OCS, in which a dispersive region is created with a combination of a dipole and a defocusing quadrupole. The orbits of the reference muon and the reference pion are combined after the OCS, where the dispersion function is suppressed to 0. The dispersion function describes the orbit displacement as a function of the momentum difference. The muons that are within the transverse and longitudinal phase space acceptances of the ring can effectively circulate. The stochastic injection is different from a conventional injection scenario in that it completely avoids the use of a fast kicker. This injection scenario relies on the decay kinematics of the pions and the difference between the reference momentum of the pion, $P_{0\pi}$, and that of the muon, $P_{0\mu}$, where $P_{0\pi} = 5$ GeV/c and $P_{0\mu} = 3.8$ GeV/c, respectively. $P_{0\mu}$ was determined as a balance among factors such as the decay ring size and the desired neutrino energy spectrum, and $P_{0\pi}$ was determined from maximizing the injection and particle production efficiency. The higher momentum pions decay to lower momentum muons that are accepted by the decay ring, which shares the production straight with the "pion beamline". From simulations using MARS [10], the number of pions within $5 \pm 1$ GeV/c after the horn is 0.29 per POT.

The nuSTORM decay ring is designed with a muon central momentum $P_{0\mu}$, aiming at achieving a transverse acceptance of 2000 µm · rad (denoted as $\Omega_x = \epsilon_x \Omega_p$), or expressed as 2 mm) and a momentum acceptance of $\pm 10\%$ ($\Omega_p$). From the simulations in G4Beamline (G4BL) [11], the number of muons within the acceptance of the ring is 0.013 per POT. Both FODO and FFAG ring lattice designs are currently being studied and optimized to achieve this acceptance. The goal of designing the pion beamline is to transport as many pions as possible that can decay to muons in both $\Omega_x$ and $\Omega_p$. Therefore, the nuSTORM horn is required to focus the pions into a certain phase space, rather than to focus them so that the beam motion is forward. The nuSTORM horn design in the project proposal [12] used the NuMI horn configuration [13], which was designed for the latter focusing purpose. Although recent studies have shown that a two-horn or even three-horn system can be more effective than a single-horn configuration for conventional neutrino beams [14, 15], only one horn was considered in the proposal and also in this optimization study due to the different constraints on the nuSTORM design. An example of the horn is shown in Figure 2.

2. nuSTORM horn and pion beamline

The magnetic field in the horn obeys Ampere’s Law such that $B_r = B_z = 0$, $B_{\phi} = \mu_0 I/2\pi r$, where $z$ is along the proton beam direction, $r$ is the radius from the center of the horn, and $\phi$ is the azimuthal angle. The parabolic shape of the horn inner conductor obtains a path length in the horn that is approximately $r^2$. With $B_{\phi} \propto 1/r$, the effective focusing strength of a particle passing through the horn is proportional to $r$. However, considering the physical length of the target, the initial longitudinal position of pions emerging from the target surface varies significantly (shown in Figure 3). The focusing effect for each particle is a function of the initial position, angle of the particle, and its momentum. Furthermore, the angular acceptance of the pion beamline (up to 20 mrad at the downstream end of the horn, regardless of the length of the pion beamline) can be much larger than that of a decay pipe used in a conventional neutrino beamline (10 mrad for a 200 m decay pipe with 2 m radius, or 15 mrad for one with 3 m radius, such as in LBNF). It is beneficial to re-optimize the horn specifically for the nuSTORM requirements. This paper discusses a numerical technique for this optimization, which can be adapted to the different goals of other experiments.

The pions emitted from the target surface are tracked in the magnetic field formed by the horn. The transverse phase space distribution of the pions at the downstream end of the horn is shown in Figure 4, where $x', y' = dx/dz, dy/dz$. In order to design the optics for the pion beamline, the transverse phase space distribution of the pions at the downstream end of the horn is fitted to obtain the Twiss parameters, which describe the phase space distribution of the particle beam. The Gauss-Newton (GN) algorithm [16] was used as the fundamental algorithm in the fitting method developed for this
Figure 1. The schematic beamline structure of the nuSTORM facility. The production straight contains 21 FODO cells, with a total length of \( \sim 150 \) meters. The length of the pion beamline is \( \sim 200 \) meters, in whose length approximately 50% of the 5 GeV/c pions will decay to muons.

Figure 2. An example of an NuMI magnetic horn design. The current direction in the example focuses positively charged secondary particles produced off the target.

study. An additional process is added to the regular GN method in order to iteratively remove pions with extra-large emittance (action) and shrink the beam admittance \( \Omega_{ad} \), conventionally defined as \( \Omega_{ad} = 6\epsilon_{rms}^2 \), to the \( \Omega \) acceptance of the pion beamline. More specifically, in the \( n \)th \((n = 0, 1, 2, \ldots, n_{\text{max}})\) iteration, the particles are sorted by their emittance defined by \( I_i = \gamma_n u_i^2 + 2\alpha_n u_i u_i' + \beta_n u_i'^2 \), where \( \alpha_n, \beta_n, \gamma_n \) are the Twiss parameters at the \( n \)th iteration, and \( u_i = x_i, y_i, u_i' = x_i', y_i' \) are the phase space coordinates for the \( i \)th particle. If \( \Omega_{ad} \) is larger than \( \Omega \), particles with the largest emittance are removed from the fitting. This process is repeated until \( \Omega_{ad} \) is reduced to be less than \( \Omega \). For a beam that occupies a diluted and distorted Gaussian phase space area, this new method works better in obtaining the most accurate Twiss values for optics matching, compared to a regular GN method or a direct fitting from the covariance matrix. The IGN method preserves the shape of the core phase space, and can be more accurate when fitting beam data with errors. The flowchart for the IGN method and an example of its application are shown in Figure 5. The IGN algorithm’s fitting (green ellipse) better preserves the core of the phase space over fitting using a covariance matrix approach (red ellipse).

The nuSTORM pion beamline consists of the pion delivery section, starting from the downstream end of the horn to the OCS, and the production straight section of the muon decay ring (see the red enclosure in Figure 1). The linear optics of the pion beamline from the downstream end of the horn into the production straight are shown in Figure 6. (Only one FODO cell is shown to omit the periodic Twiss values.) The design of the optics from the first dipole to the end of the production straight has been fixed for providing the best stochastic injection performance and a good accommodation of both the pion and the muon beams in the production straight [17]. The production straight contains 21 FODO cells, with a total length of \( \sim 150 \)
3. Optimization Objectives and Algorithm

3.1. Objectives

A straightforward optimization of the nuSTORM horn needs the complete tracking of pions collected by the horn and then tracked through the pion beamline, with muon decay enabled in the G4BL tracking code. The number of muons at the end of the pion beamline that are within the acceptance of both $\Omega_\pi$ and $\Omega_\mu$ is then compared for different horn configurations. The value can be used as the fitness value, or the figure of merit, of the optimization. This is thus essentially a single-objective optimization problem, referred to as the number of muons-in-acceptance. However, the full tracking simulation in the beamline with decay processes generally requires very large computing resources (for example, a full tracking of $10^5$ events using 24 cores takes several tens of minutes). Furthermore, for the pion phase space distribution after each horn, the optics of the pion beamline needs to be rematched in a design program like MAD-X [18], which generally adds another significant factor to the computing time. Overall, this algorithm saves approximately 80,000 CPU hours.

Considering the dynamical features of the pion beamline, and the well-known two-body decay kinematics of pions, the full tracking in the pion beamline can be eliminated. From the pion beam at the downstream end of the horn, it is possible to evaluate the number of muons in $\Omega_\pi$ ($N_{\mu,P}$) from the pion distribution data after the horn and maximize $N_{\mu,P}$ in the simulation. This estimation includes the muon decay kinematics, the momentum acceptance of the pion beamline, and the first-order assumption that the transverse phase space acceptance is the same for any $\Omega_\pi$, as long as the optics can be matched by the conventional capture quads. Simultaneously, the number of muons in $\Omega_\mu$ ($N_{\mu,e}$) can be increased via maximizing the number of pions in $\Omega_\pi$ ($N_{\pi,e}$). Maximizing both $N_{\mu,e}$ and $N_{\mu,P}$ simultaneously maximizes the number of muons-in-acceptance without full tracking. However, it also turns the single objective into two separate ones, which requires a different approach with an appropriate optimization algorithm. Accuracy is reduced by making these approximations and not doing full particle tracking for every individual case, since higher order nonlinearities of beam dynamics are not all included.

Constraints can be added to select the correct offspring parameters, which in this case, arise from the engineering limitations on the horn, and also from the range of optics that can be matched using the conventional capture quads in the pion beamline. As dis-
Fit the transverse distribution to a bivariate Gaussian function, calculate the covariance matrix to obtain the RMS beam emittance and Twiss functions.

\[ \varepsilon_{\text{rms}} > 2 \text{ mm?} \]

Incoming
Beam
To Fit

Sort the particles based on their individual emittance (action)
Remove a certain percentage of the ones with the largest values

Final Beam
Fitting ends

Figure 5. The flowchart for the Iterative Gauss-Newton (IGN) fitting algorithm (left) and an example of its application to fitting the transverse phase space distribution at the downstream end of the NuMI-like nuSTORM baseline horn.

Figure 6. The linear optics of the pion beamline. The whole FODO production straight was replaced by a single FODO cell to avoid repeating periodic Twiss functions. The pions move from the left to the right in the drawing.

yield unmatchable Twiss parameters are given very low priorities, to force the population to generate matchable optics.

3.2. Implementation of the Genetic Algorithm

The Genetic Algorithm (GA), especially the Multi-Objective GA (MOGA), has been widely recognized as an efficient multi-objective algorithm for multiple-criteria decision making problems. It has also been widely applied in simulation software and studies in high energy and accelerator physics [19, 20, 21]. The MOGA is more favorable in this study than the Single-Objective GA (SOGA) because more than one objective needs to be optimized. It provides an unbiased approach to the optimum values for each of the objectives, when the correlation between them can not be expressed analytically. A GA is a meta-heuristic Monte-Carlo method in which a new generation of individuals (different combinations of parameters) is constructed from the last generation of individuals through crossover and mutation. The parent individuals in each generation are selected based on the values they yield for the objectives. The MOGA is different from the SOGA in the way it handles the optimization objectives. MOGA treats more than one objective as equally important, and selects elite parents in each generation based on an evaluation using the “pareto front”. The pareto front describes the best individuals in the decision space [22, 23]. SOGA, as a contrast, selects the parents based on one and only one
fitness function, and has less complexity. The MOGA pushes the pareto front to the global optimum front in the decision space, and uses elites on the pareto front to improve the whole population.

In the case of the nuSTORM horn optimization, nine parameters are used as "genes" in the MOGA, eight of which are shown in Figure 7, and the last of which is the horn current $I$. The shape of the horn inner conductor was chosen to be parabolic, since the operation of NuMI horns with such configurations is already mature, and the engineering feasibility of such a horn is already confirmed. A linear inner conductor shape, although also feasible in manufacturing, does not provide comparably good results. Moreover, both the front and rear parabolic shapes are allowed to be changed to a straight shape, as the neck is, but the MOGA results shown later did not converge to that configuration. The length of the target was not included in the parameter set to reduce the computing cost of the Monte Carlo process in the target, but three different target lengths have been compared in the optimization runs. The parameters vary from one generation to the next by generating offspring values from a neighborhood function and two selected parent values. As two important features of the GA, the crossover and mutation operators for real-parameter optimization problems are well studied, in our case, the well-recognized Simulated Binary Crossover (SBX) and the parameter-based mutation [24] were used. In order to integrate the MOGA with G4BL, or any other SHELL callable programs, a Python-based, Message Passing Interface (MPI) implemented optimization toolkit called pyOPTmpi was developed to fulfill the tasks of analyzing the G4BL outputs from the MOGA inputs. Not only the GA, but also several other optimization algorithms including simulated annealing and Gauss-Newton fitting are built into the toolkit, which can be applied to other optimization problems. The flowchart for the algorithm is shown in Figure 8.

4. Optimization Results

A search for an optimal horn configuration was performed for three Inconel targets, which are 2.5, 3 and 4 interaction lengths, 38 cm, 46 cm, and 60 cm long, respectively. Pions produced from the three targets were used as the input beam in each case, which significantly reduces the simulation time. Fifty nodes, or 1200 computing cores were requested from NERSC in each iteration (shown in Figure 8) to model and compare 100 horns in each generation, and to construct the next generation of horns. Each iteration takes approximately 30 minutes of run time including the pyOPTmpi processing time and the G4Beamline tracking time. The number of generations is limited to 150 for each search. The algorithm terminates when the population ceases to improve, which can be detected when the objectives have not improved in 10 generations. The Python toolkit shows very good portability between platforms and has a smooth connection to G4BL.

The optimization results are shown by plotting the evolution of objective values in Figure 9 for the 38 cm and 46 cm targets. The arrows in the plots show the change of the fitness values $N_{\mu}P$ and $N_{r\mu}$ for the selected elite individual in each generation. Increasing the target length from 46 cm to 60 cm did not show any further improvement, thus consideration of the 60 cm case was abandoned. In order to confirm the optimization benefits, the single objective introduced in Section 3.1 is used. Full tracking was done in G4BL with rematched pion beamline optics to obtain the number of muons-in-acceptance. An increase of more than 8% was found for the horn used with the 38 cm Inconel target, compared with the pre-optimized model in the proposal. Redesigning the horn for the 46 cm Inconel target brings an increase of 16% to the number of muons-in-acceptance, compared with the pre-optimized horn-target configuration. Specifically, the number of muons within both the transverse phase space acceptance and momentum acceptance of the ring, obtained at the end of the production straight section, is increased from 0.013 POT to 0.015 POT. For comparison, increasing the target length to 46 cm from 38 cm without changing the horn only provides an increase of 5%.

The optimized horn for the 46 cm Inconel target is shown, and compared with the pre-optimized horn, in

![Figure 7. Schematic drawing of the horn showing the parameters that form the genes in the MOGA simulation for this horn optimization study. Eight of the nine are shown by the symbols L1, ..., ΔZ, and the ninth is the horn current I.](image-url)
GA starts with a population of random horn individuals, as the first generation. B-field in each horn is modeled in G4BL, then pions are tracked in the field. Calculate and find $N_\pi$ and $N_\mu$, end. Compare individuals based on the objectives; Record the best. Stopped improving? YES STOP. Perform Crossover and Mutation; Apply constraints on variables; Construct new generation. NO.

Figure 8. The flowchart for the MOGA applied to the horn optimization. The loop that has the MPI implemented is marked in blue.

Figure 10. The optimization stopped at the 81st generation, at which point the best fitness values for both the objectives stopped improving, or namely the pareto front stopped moving. The effective rear boundary of the horn conductor is shown with the dashed line considering that only the forward pions are useful. The optimized horn shows a simpler configuration. The optimized current is reduced from 230 kA to ~219 kA. It is therefore expected that the optimized horn is more cost effective. The corresponding pion phase space distributions and the fitted acceptance ellipses are shown in Figure 11. As mentioned earlier, in nuSTORM, the pion beamline is designed to transport and match the beam phase space into the ring. This is different from the simple point-to-parallel optimum of a conventional neutrino beam, such as NuMI, LBNE, etc. This can be seen in Figure 11 where the pions from the optimized horn have a larger angular divergence.

5. Summary

The optimization of the nuSTORM horn and the meta-heuristic algorithm and strategy for the study are described. The study aims at maximizing the number...
of muons that can be accepted by the nuSTORM decay ring, from pion decay during their transit in the pion beamline. The MOGA that is implemented in the numerical simulation avoids a full particle tracking for each horn candidate, by analyzing the pion distribution at the downstream end of the horn and maximizing two objective functions at the same time. This dramatically reduces the requirement on the computing resources by \( \sim 80,000 \) CPU hours. After the optimization, two new horn shapes were found for the 38 cm and 46 cm Inconel targets, which result in an increase of 8% and 16% in the number of muons-in-acceptance or the \( \bar{\nu}_e \) or \( \bar{\nu}_\mu \) flux at the detectors, respectively. The optimized 46 cm case has approximately 11% more muons in the ring acceptance than the optimized 38 cm case.

Figure 11. A comparison of the phase space distributions of pions focused by the pre-optimized and the optimized horn (left and right, respectively). The red ellipses are the fitted Gaussian acceptance ellipses. The number of pions included in the acceptance is \( 6.9 \times 10^5 \) and \( 8.2 \times 10^5 \) from \( 2.4 \times 10^6 \) 120 GeV POT. For reference, \( 1.2 \times 10^6 \) pions are emitted from the target surface in the momentum band of \( 5 \pm 1 \text{ GeV/c} \).
Acknowledgments

The authors thank David Adey and Prof. S.Y. Lee for offering valuable discussions, and the MAP collaboration for the useful comments. This work is supported by the Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy, and the Muon Accelerator Program within the United States Department of Energy.

References

[18] CERN, Mad-x. URL http://nad.web.cern.ch/nad/