Proposal for an experiment to search for dark sector particles weakly coupled to muon at the CERN SPS

D. Banerjee\textsuperscript{k}, J. Bernhard\textsuperscript{d}, V.E. Burtsev\textsuperscript{j}, A.G. Chumakov\textsuperscript{i}, P. Crivelli\textsuperscript{m}, E. Depero\textsuperscript{m}, A.V. Dermenev\textsuperscript{i}, S.V. Donskov\textsuperscript{i}, R. Dusaev\textsuperscript{j}, T. Enik\textsuperscript{b}, V. Frolov\textsuperscript{b}, A. Gardikiotis\textsuperscript{b}, S.N. Gninenko\textsuperscript{a}, M. H"{o}sgen\textsuperscript{a}, A. Karneyeu\textsuperscript{e}, G.D. Kekelidze\textsuperscript{b}, B. Ketzer\textsuperscript{a}, D. Kirpichnikov\textsuperscript{e}, M.M. Kirsanov\textsuperscript{c}, S. Kovalenko\textsuperscript{l}, L.V. Kravchuk\textsuperscript{c}, V.A. Kramarenko\textsuperscript{b,g}, N.V. Krasnikov\textsuperscript{e}, S.V. Kuleshov\textsuperscript{i}, V.E. Lyubovitskij\textsuperscript{b}, V.M. Lysan\textsuperscript{e}, V.A. Matveev\textsuperscript{b}, Yu.V. Mikhailov\textsuperscript{i}, L. Molina-Bueno\textsuperscript{m}, D.V. Peshekhonov\textsuperscript{b}, V.A. Polyakov\textsuperscript{i}, B. Radics\textsuperscript{m}, A. Rubbia\textsuperscript{m}, V. Samoylenko\textsuperscript{i}, D. Shchukin\textsuperscript{j}, V.O. Tikhomirov\textsuperscript{f}, D.A. Tlisov\textsuperscript{e}, A.N. Toropin\textsuperscript{e}, A. Yu.Trifonov\textsuperscript{j}, P. Ulloa\textsuperscript{l}, B.I. Vasilishin\textsuperscript{i}, B.M. Veit\textsuperscript{d}, P.V. Volkov\textsuperscript{b,g}, and V.Yu. Volkov\textsuperscript{g}

The NA64 Collaboration\textsuperscript{1}

\textsuperscript{a}Universit"{a}t Bonn, Helmholtz-Institut für Strahlen-und Kernphysik, 53115 Bonn, Germany
\textsuperscript{b}Joint Institute for Nuclear Research, 141980 Dubna, Russia
\textsuperscript{d}CERN, European Organization for Nuclear Research, CH-1211 Geneva, Switzerland
\textsuperscript{e}Institute for Nuclear Research, 117312 Moscow, Russia
\textsuperscript{g}Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
\textsuperscript{h}Physics Department, University of Patras, Patras, Greece
\textsuperscript{i}State Scientific Center of the Russian Federation Institute for High Energy Physics of National Research Center 'Kurchatov Institute' (IHEP), 142281 Protvino, Russia
\textsuperscript{j}Tomsk State Pedagogical University, 634061 Tomsk, Russia
\textsuperscript{k}University of Illinois, Urbana Champaign, Illinois, USA
\textsuperscript{l}Universidad Técnica Federico Santa María, 2390123 Valparaíso, Chile
\textsuperscript{m}ETH Zürich, Institute for Particle Physics, CH-8093 Zürich, Switzerland

March 27, 2019

\textsuperscript{1}http://na64.web.cern.ch
Abstract:
The NA64 collaboration proposes to carry out further searches for dark sector and other rare processes in missing energy events from high energy muon interactions in a hermetic detector at the CERN SPS for the 2021 pilot run.

A dark sector of particles predominantly weakly-coupled to the second and possibly third generations of the Standard Model is motivated by several theoretically interesting models. Additional to gravity this new very weak interaction between the visible and dark sector could be mediated either by a scalar ($S_{\mu}$) or $U'(1)$ gauge bosons ($Z_{\mu}$) interacting with ordinary muons. In a class of $L_{\mu} - L_{\tau}$ models the corresponding $Z_{\mu}$ could be light and have the coupling strength lying in the experimentally accessible region. If such $Z_{\mu}$ mediator exists it could also explain the muon $(g - 2)_{\mu}$ anomaly - the discrepancy between the predicted and measured values of the muon anomalous magnetic moment.

We propose an extension of the experiment called NA64$\mu$ to search for invisible decays of the $Z_{\mu}$ either to neutrinos or light Dark Matter particles. The primary goal of the experiment in the 2021 pilot run with the $\simeq 100 - 160$ GeV M2 beam is to commission the NA64$\mu$ detector and to probe for the first time the still unexplored area of the coupling strengths and masses $M_{Z_{\mu}} \lesssim 200$ MeV that could explain the muon $(g - 2)_{\mu}$ anomaly. Another strong point of NA64$\mu$ is its capability for a sensitive search for dark photon mediator ($A'$) of dark matter production in invisible decay mode in the mass range $m_{A'} \gtrsim m_{\mu}$, thus making the experiment extremely complementary to the ongoing NA64e and greatly increase the discovery potential of sub-GeV dark matter. Other searches for $S_{\mu}$’s decaying invisibly to dark-sector particles, and millicharged particles will probe a still unexplored parameter areas.
## Contents

1 Introduction and Motivation 4

2 Phenomenology and existing experimental bounds 5
   2.1 The vector $Z_\mu$ case 5
   2.2 The scalar $S_\mu$ case 8
   2.3 Dark matter in the $L_\mu - L_\tau$ model 8

3 The $Z_\mu(S_\mu)$ production and decays 9

4 Phase I: The experiment to search for the $Z_\mu \rightarrow$ invisible decay 12
   4.1 The detector 12
   4.2 The method of the search 14
   4.3 The signal simulations 15
   4.4 The M2 beam 18
   4.5 The simulation of the beam and NA64 $\mu$ setup 20
      4.5.1 The incoming and outgoing muon momentum measurements 20
   4.6 The trigger rate 23
      4.6.1 The trigger option 1 24
      4.6.2 The trigger option 2 29
   4.7 DAQ and electronics upgrades 30
   4.8 Background 31
      4.8.1 The purity of the incoming muon momentum reconstruction 32
      4.8.2 The purity in the muon final state 33
      4.8.3 The empty target measurements 35
      4.8.4 Hadron contamination and decays in the M2 beam 36
      4.8.5 The detector hermeticity 39
      4.8.6 Dimuon events from the reaction $\mu^-Z \rightarrow \mu^-\mu^+\mu^-Z$ 43
      4.8.7 Other background sources 44
   4.9 Expected sensitivity for the scalar and vector cases 44

5 The Phase II physics program 47
   5.1 Search for dark photon mediator: complementarity of NA64e and NA64$\mu$ 47
      5.1.1 NA64$\mu$ projections for the $\gamma - A'$ mixing strength 48
      5.1.2 Combined LTM sensitivity of NA64e and NA64$\mu$ 50
   5.2 Constraints on scalar and $L_\mu - L_\tau$ sub-GeV Dark Matter 52
   5.3 Search for millicharged particles 55
      5.3.1 Missing energy signature of millicharged particles at NA64 57
      5.3.2 Millicharge reach estimate 59
6 The 2021 pilot run
   6.1 The main goals of the run
   6.2 The requirements for the run
   6.3 Beam time request
   6.4 Possible scenarios for the 2021 run

7 The detector cost, manpower, and schedule

8 Conclusion
Executive Summary

We propose an extension of the NA64 experiment called NA64\(\mu\) dedicated to a sensitive search for dark sectors via the invisible decays \(Z_\mu \rightarrow \text{invisible}\) of a new dark boson \(Z_\mu\) with a mass in the sub-GeV range, which is coupled predominantly to the second and third lepton generations through the \(L_\mu - L_\tau\) current. If the new dark boson \(Z_\mu\) exists, it could explain the long-standing 3.6 \(\sigma\) discrepancy between the predicted and measured values of the anomalous magnetic moment of positive muons. Another strong motivation for \(Z_\mu\) comes from the observation that the \(Z_\mu\) could mediate the interaction between the visible and dark sectors and decay invisibly either into \(\nu\nu\) or Dark Matter (DM) particle pairs. We show that if the \(Z_\mu\) exists, it could be observed in the reaction \(\mu + Z \rightarrow \mu + Z + Z_\mu\) of a muon scattering off a nuclei by looking for an excess of events with a large missing muon energy in a detector due to the prompt bremsstrahlung \(Z_\mu\) decay \(Z_\mu \rightarrow \text{invisible}\). We describe the experimental technique and the preliminary study of the feasibility of the proposed search. We show that this specific signal allows a search for the \(Z_\mu\) with a sensitivity to the coupling constant \(\alpha_\mu \gtrsim 10^{-11}\), which is three orders of magnitude lower than the value required to explain the (g-2)\(\mu\) discrepancy. We point out that the availability of high-energy and high-intensity muon beams at the CERN SPS provides a unique opportunity either to discover or rule out the \(Z_\mu\) in the near future. The NA64\(\mu\) is based on the missing-energy approach we developed for the searches for invisible decays of dark photons and (pseudo)scalar mesons at CERN and is complementary to these experiments. The combined NA64\(e\) and NA64\(\mu\) results with \(\approx 10^{13}\) EOT and a few \(10^{13}\) MOT, respectively, will allow to cover almost fully the parameter space of the most interesting light DM models. This makes NA64\(e\) and NA64\(\mu\) extremely complementary to each other and greatly increases the discovery potential of sub-GeV DM.

The experiment has a capability to search for invisible decays \(S_\mu \rightarrow \text{invisible}\) of light scalars with a high sensitivity. The feasibility study shows that the sensitivity in the search for the \(S_\mu \rightarrow \text{invisible}\) decay mode in branching fraction \(\text{Br}(S_\mu) = \frac{\sigma(\mu^- Z \rightarrow \mu^- Z S_\mu), S_\mu \rightarrow \text{invisible}}{\sigma(\mu^- Z \rightarrow \mu^- Z \gamma)}\) could be achieved at the level below a few times \(10^{-11} - 10^{-12}\). The intrinsic background due to the presence of low energy muons in the beam can be suppressed by using a tagging system based on the combined incoming muon detection with the BMS and MS1 magnetic spectrometers. It has also a capability of a sensitive search for millicharged particles and lepton flavour violation process of the \(\mu - \tau\) conversion and cover a significant part of the still unexplored part of the parameter space complementary to what is intended to be probed by other searches.

The main goal of NA64\(\mu\) in 2021 is to probe for the first time the muon (g-2)\(\mu\) parameter space. To achieve this goal we propose a pilot run focusing on the tests of the detector components and obtaining the first background estimate, which might
require further detector upgrade. This can be done by using the M2 muon beam line of the SPS that provide enough intensity in the given energy range of measurements.
1 Introduction and Motivation

Despite the intensive searches at the LHC and in non-accelerator experiments Dark Matter (DM) is still a great puzzle. Though stringent constraints obtained on DM coupling to Standard Model (SM) particles ruled out many DM models, little is known about the origin and dynamics of the dark sector itself. The main difficulty so far is that the only established way to probe DM is through its gravitational interaction. An exciting possibility is that in addition to gravity, a new force between the dark sector and visible matter transmitted by a new vector boson $A'$ (dark photon) might exist. Such $A'$ could have a mass $m_{A'} \lesssim 1$ GeV - associated with a spontaneously broken gauged $U(1)_D$ symmetry- and couple to the SM through kinetic mixing with the ordinary photon, $-\frac{1}{2} \epsilon F_{\mu\nu} A_{\mu\nu}'$, parametrized by the mixing strength $\epsilon \ll 1$. This has motivated a worldwide theoretical and experimental effort towards searches for dark forces and other portals between the visible and dark sectors, shifting the strategy from the high energy to the high intensity frontier [1].

An additional motivation for the existence of the $A'$ has been provided by hints of the astrophysical signals of dark matter, as well as the $3.6 \sigma$ deviation from the SM prediction of the muon anomalous magnetic moment $(g-2)_\mu$. The precise measurement of the anomalous magnetic moment of the positive muon $a_\mu = (g-2)/2$ from the Brookhaven AGS experiment 821 [2] gives a result which is about $3.6 \sigma$ higher than the Standard Model (SM) prediction

$$a_\mu^{exp} - a_\mu^{SM} = 288(80) \times 10^{-11}$$ (1.1)

This result may signal the existence of new physics beyond the Standard Model. At present the most popular explanation of this discrepancy is supersymmetry with a chargino and sneutrino lighter than 800 GeV [3]. Other possible explanations include leptoquarks [4] or some exotic flavor-changing interactions [5]. All of these explanations assume the existence of new heavy particles with masses $\geq O(100)$ GeV. Another explanation of the $(g-2)_\mu$ anomaly is related to the existence of a new light (with a mass $m_{Z'} \leq O(1)$ GeV) vector boson (dark photon) which couples very weakly with the muon with $\alpha_{Z'} \sim O(10^{-8})$ [6]-[11], see also Ref. [12].

In this proposal we consider the muon $(g-2)_\mu$ anomaly as an indication of the existence of the new light vector boson $Z_\mu$, which is coupled predominantly to the second and third lepton generations [13]. We propose an experiment to search for the $Z_\mu$ in the high-energy muon beam at the CERN SPS. If the $Z_\mu$ exists, it could be observed in the reaction $\mu + Z \rightarrow \mu + Z + Z_\mu$ of a high-energy muon scattering off a nuclei by looking for an excess of events with a specific signature, namely large missing muon beam energy in the detector. The experiment uses the missing-energy approach developed for the search for invisible decays of dark photons and (pseudo)scalar mesons at CERN [20–22] and is complementary to these searches.
2 Phenomenology and existing experimental bounds

2.1 The vector $Z_\mu$ case

As discussed in the Introduction one of the possible explanations of the $(g-2)_\mu$ anomaly assumes the existence of a new light vector boson $Z'$ which interacts with muons with a coupling $e_\mu$ like a photon [13], namely

$$L_{Z'} = e_\mu \bar{\mu} \gamma_\nu \mu Z'^\nu.$$  \hspace{1cm}(2.1)$$

The interaction 2.1 gives additional contribution to the muon anomalous magnetic moment $a_\mu \equiv \frac{g_\mu - 2}{2}$

$$a'_\mu = \frac{\alpha_\mu}{\pi} \int_0^1 \frac{x^2(1-x)}{x^2 + (1-x)M_{Z'}^2/m_\mu^2},$$  \hspace{1cm}(2.2)$$

where $\alpha_\mu = (e_\mu)^2/4\pi$ and $M_{Z'}$ is the mass of the $Z'$ boson. Equation 2.2 allows one to determine the $\alpha_\mu$ which explains the $(g-2)_\mu$ anomaly. For $M_{Z'} \ll m_\mu$ we find from Eq 1.1 that

$$\alpha_\mu = (1.8 \pm 0.5) \times 10^{-8}$$  \hspace{1cm}(2.3)$$

For another limiting case $M_{Z'} \gg m_\mu$ Eq 1.1 leads to

$$\alpha_\mu \frac{m_\mu^2}{M_{Z'}^2} = (2.7 \pm 0.8) \times 10^{-8}$$  \hspace{1cm}(2.4)$$

Figure 1. Diagram illustrating the massive $Z_\mu$ production in the reaction $\mu + Z \rightarrow \mu + Z + Z_\mu$ of muons scattering off a nuclei (A,Z). The $Z_\mu$ is either stable or it decays invisibly if its mass $M_{Z_\mu} \leq 2m_\mu$, or it could subsequently decay into a $\mu^+\mu^-$ pair if $M_{Z_\mu} > 2m_\mu$.

But the postulation of the interaction (2.1) of the $Z_\mu$ boson with a muon is not the end of the story. The main question is, what about the interaction of the $Z'$ boson with other quarks and leptons? There is a lot of possibilities here. For instance, a very popular scenario involves the interaction of the $Z'$ boson with quarks and leptons that is proportional to the electromagnetic current $J_\nu$ of the SM, namely

$$L_{Z'} = e_\mu J_\nu Z'^\nu.$$  \hspace{1cm}(2.5)$$
The natural realization of this scenario is the existence of a new gauge boson \( Z' \) which interacts with the SM fields through the mixing with the SM hypercharge [23–25]

\[
\Delta L = \frac{\epsilon}{2} F_{\alpha\beta} F_{\alpha\beta}^{Z'}. \tag{2.6}
\]

For the scenario with the interaction (2.5) of the \( Z_\mu \) boson with the electromagnetic current of the SM there are several interesting constraints. Bounds [26, 27] from the electron magnetic moment value

\[
\Delta a_e = a_e^{exp} - a_e^{SM} = -1.06(0.82) \times 10^{-12} \tag{2.7}
\]

exclude the region \( M_{Z'} < 30 \text{ MeV} \). Other experiments use dilepton resonance searches, \( Z' \rightarrow l^+l^- \). First we consider the bounds obtained under the assumption that the \( Z' \) boson decays mainly into charged leptons, i.e. \( \text{Br}(Z' \rightarrow l^+l^-) = 1 \), with \( l = e, \mu \). The Phenix Collaboration looked for the \( Z' \) boson in the \( \pi^0, \eta \rightarrow (Z' \rightarrow e^+e^-)\gamma \) decays and excluded the masses \( 36 < M_{Z'} < 90 \text{ MeV} \) [28]. The A1 Collaboration used the reaction \( eZ \rightarrow eZ'Z; Z' \rightarrow e^+e^- \) to search for the \( Z' \) boson and excluded the masses \( 40 < M_{Z'} < 300 \text{ MeV} \) [29]. The \( BABAR \) Collaboration looked for the \( Z' \) boson in the reaction \( e^+e^- \rightarrow \gamma Z', Z' \rightarrow e^+e^-; \mu^+\mu^- \) and excluded the masses \( 30 \text{ MeV} < M_{Z'} < 10.2 \text{ GeV} \) [30]. Finally, taking into account the recent results from the \( K^- \) decay experiments [31], the possibility of the \( (g-2)_\mu \) explanation in the model with the interaction 2.6 of the \( Z' \) boson with the assumption that \( \text{Br}(Z' \rightarrow l^+l^-) = 1 \) is excluded, see, e.g. Ref.[32] for a discussion.

For the model with the interaction (2.5) there is the possibility that the \( Z' \) boson decays dominantly invisibly into new light particles \( \chi \) with the branching \( \text{Br}(Z' \rightarrow \chi\tilde{\chi}) = 1 \). For this scenario the \( K^+ \rightarrow \pi^+ + \text{missing energy} \) bound [33] and the off-resonance \( BABAR \) result [34] exclude a sizable parameter space except \( 30 < M_{Z'} < 50 \text{ MeV} \) and the narrow region around \( M_{Z'} = 140 \text{ MeV} \) [35], [36].

Another interesting scenario is that from Ref. [36], where the light gauge boson (the dark leptonic gauge boson) interacts with the leptonic current, namely

\[
L_{Z'} = \epsilon'[\bar{e}\gamma_\nu e + \bar{\nu}_{eL}\gamma_\nu \nu_{eL} + \bar{\mu}\gamma_\nu \mu + \bar{\nu}_{\mu L}\gamma_\nu \nu_{\mu L} + \bar{\tau}\gamma_\nu \tau + \bar{\nu}_{\tau L}\gamma_\nu \nu_{\tau L}]Z'^\nu \tag{2.8}
\]

This interaction does not contain quarks and as a consequence the corresponding model escapes many quarkonium-decay constraints [36]. The relevant searches for the dark leptonic gauge boson include fixed-target [29] and neutrino trident experiments [37, 38], the \( BABAR \) search for \( e^+e^- \rightarrow \gamma + \text{missing energy} \) [30], beam-dump experiments [39–41], and last but not least the Borexino experiment [42]. It appears that for the model with the dark leptonic gauge boson the most restrictive bound comes from the latter. The presence of a new vector boson would alter the charged-current interaction between solar \( \nu_e \) neutrinos and target electrons in the detector.
The bounds from the 862 KeV $^7$Be solar neutrino flux measurement at the Borexino experiment exclude the possibility that the leptonic gauge boson can explain the (g-2)$_\mu$ anomaly [43], see Fig. 3 in Ref.[36] where the constraints on the parameter space of the dark lepton gauge boson model were presented 1.

In Refs. [6] - [8], an explanation of the (g-2)$_\mu$ anomaly was given by a model where the new light gauge boson (hereafter denoted as $Z_\mu$) interacts with the $L_\mu - L_\tau$ current as

$$L_{Z_\mu} = e_\mu [\bar{\mu} \gamma_\nu \mu + \bar{\nu}_L \gamma_\nu \nu_\mu L - \bar{\tau} \gamma_\nu \tau - \bar{\nu}_\tau L \gamma_\nu \nu_\tau L] Z_\mu^\nu$$ (2.9)

which is anomaly free and corresponds to the global flavor symmetry $U(1)_{L_\mu - L_\tau}$ which commutes with with the SM $SU_c(3) \otimes SU(2)_L \otimes U(1)_Y$ gauge group [44]. In addition, it was recently shown that the $Z_\mu$ with a mass $\approx 2$ MeV can explain the gap in the cosmic neutrino spectrum observed by the IceCube Collaboration [45].

As the $Z_\mu$ does not couple to quarks, electrons and $\nu_e$ neutrinos, it escapes most of the current experimental constraints. The most restrictive bound comes from the results of experiments on neutrino trident production $\nu_\mu N \to \nu_\mu N + \mu^+ \mu^-$. As was shown in Ref. [46], the CCFR data [38] on $\nu_\mu N \to \nu_\mu N + \mu^+ \mu^-$ production exclude the $g_\mu - 2$ explanation for a $Z_\mu$-boson mass $m_{Z_\mu} \geq 400$ MeV.

We note that at the one-loop level the $Z_\mu$ and the photon are kinetically mixed. The effective coupling of $Z_\mu$ to electrons (or quarks) due to the muon or $\tau$-lepton loop is $\simeq O(\frac{\alpha}{\pi}) e_\mu$, i.e. it is suppressed by at least a factor $\approx 3 \cdot 10^{-3}$. This results in rather modest constraints on the invisible decays of $Z_\mu$ from dark-photon and other experiments. For example, the bound on the coupling $\alpha_\mu$ from the $K^+ \to \pi^+ + \text{missing energy}$ decay is on the level $\alpha_\mu \leq O(10^{-3})$, which is several orders of magnitude below the value from Eq.(4). The visible decay $Z_\mu \to e^+ e^-$ can also occur at the one-loop level. Its branching fraction is estimated to be $\text{Br}(Z_\mu \to e^+ e^-) = O((\frac{\alpha}{\pi})^3 \alpha_\mu) = O(10^{-5}) \alpha_\mu$. As a consequence, in any experiment using electrons or quarks as a source of $Z_\mu$'s, the number of $Z_\mu \to e^+ e^-$ signal events is suppressed by a factor $O((\frac{\alpha}{\pi})^4) \approx 10^{-10}$, resulting in a very weak bound on $\alpha_\mu$ (here the factor $(\frac{\alpha}{\pi})^2$ comes from the $Z_\mu$ production). Finally, we note that if the $Z_\mu$ couples to light dark matter, then the additional contribution from the invisible decay mode $Z_\mu \to \text{dark matter}$ that increases the $Z_\mu \to \text{invisible}$ decay rate is possible. Such a scenario requires additional study, which is beyond the scope of this project. To conclude, let us stress that the existing experimental data restrict the explanation of the (g-2)$_\mu$ anomaly by the existence of a new light gauge boson rather strongly, but they do not completely eliminate it. For the model with the interaction (2.5) the realization with invisible $Z_\mu$-boson decays into new light $\chi$-particles for $M_{Z_\mu} = 30 - 50$ MeV and around $M_{Z_\mu} = 140$ MeV is still possible. Moreover, for the interaction of the $Z_\mu$

---

1For instance, for $m_{Z_\mu} = 1$ MeV the experimental bounds on $e'$ is $e' \leq 3 \times 10^{-6}$ [43]. Note that in the early proposals [60]-[62] (see also Ref.[63]) similar bounds on the $Z_\mu$-boson coupling from neutrino reactions were obtained.
boson with $L_\mu - L_\tau$ current bounds are rather weak, and a light $Z_\mu$- boson with a mass $M_{Z_\mu} \leq 400$ MeV is not excluded as a source of the $(g-2)_\mu$ discrepancy.

2.2 The scalar $S_\mu$ case

In addition we note, that relevant $(g - 2)_\mu$ parameter space may be probed by the scalar particles $S_\mu$ from Dark Sector interacting with SM muons as well as with Dirac DM fermions $\chi$. The simplified Yukawa-like Lagrangian of $S_\mu$ has the following form [15–17]

$$L_{S_\mu} \supset \frac{1}{2} (\partial_\nu S_\mu)^2 - \frac{1}{2} m_{S_\mu}^2 S_\mu^2 + g_S S_\mu \bar{\mu} \mu + g_\chi S_\mu \bar{\chi} \chi $$

where $g_S$ is the scalar coupling to be constrained by NA64$\mu$ in the reaction $\mu Z \rightarrow \mu Z S_\mu$, followed by invisible decay, $S_\mu \rightarrow \bar{\chi} \chi$. The Yukawa-like interaction provides the leading order contribution to the muon anomalous magnetic momentum

$$\Delta a_\mu^S = \frac{g_S^2}{16\pi^2} \int_0^1 \frac{dx}{m_{S_\mu}^2 (1-x)^2 + m_S^2 x}. $$

The differential cross-section of the process $\mu Z \rightarrow \mu Z S_\mu$ can be calculated in the generalized Weizsacker-Williams approximation. Namely, the scalar production rate has the following form

$$\frac{d\sigma}{dx} = \frac{g_S^2 \alpha^2}{12\pi} \chi \beta_\mu \beta_S x^3 (m_{S_\mu}^2 (3x^2 - 4x + 4) + 2m_\mu^2 (1-x)) \left( \frac{m_{S_\mu}^2 (1-x)}{m_{S_\mu}^2 + m_\mu^2 x^2} \right)^2$$

where $x = E_S/E_\mu$ and the boost factors for muon and scalar are respectively defined by

$$\beta_\mu = \sqrt{1 - m_\mu^2/E_\mu^2}, \quad \beta_S = \sqrt{1 - m_{S_\mu}^2/E_S^2}$$

We consider the benchmark case, when $m_S > 2m_\chi$ and $g_\chi \gtrsim g_\mu$. This means that $S_\mu$ primarily decays invisibly to $\bar{\chi} \chi$.

2.3 Dark matter in the $L_\mu - L_\tau$ model

Recently, it has been shown that an extension of the $L_\mu - L_\tau$ model is able to explain the observed relic Dark Matter density in the Universe for $m_{Z_\mu} \simeq O(10)$ MeV [14], see also [15–17]. This strengthen motivation for the experimental search of the $L_\mu - L_\tau$ mediator of the DM production in invisible decay mode. We also note that if the $Z_\mu$ couples to light DM, then an additional contribution from the invisible decay mode $Z_\mu \rightarrow \text{dark matter}$ increases the $Z_\mu \rightarrow \text{invisible}$ decay rate. Thus, for $m_{Z_\mu} > 2m_\mu$ the visible decay $Z_\mu \rightarrow \mu^+ \mu^-$ is more suppressed.
Figure 2. Calculated distributions of the $Z_\mu$ fractional energy $x = E_{Z_\mu}/E_\mu$ from the reaction $\mu + Z \rightarrow \mu + Z + Z_\mu$ at a muon beam energy $E_\mu = 160$ GeV for different $Z_\mu$ masses indicated near the curves. The spectra are normalized to a common maximum.

3 The $Z_\mu(S_\mu)$ production and decays

In this section we consider the production and decays of the $Z_\mu$, however the results obtained are also applicable to the $S_\mu$ case. As a source of $Z_\mu$s we use the bremsstrahlung $Z_\mu$'s produced in the reaction

$$\mu(p) + Z(P) \rightarrow Z(P') + \mu(p') + Z_\mu(k)$$

(3.1)

of high-energy muons scattering off nuclei in the target, as shown in Fig. 1. Here $p, P, P', p', k$ are the four-momenta of the incoming muon, incoming nucleus $Z$, outgoing nucleus $Z$, outgoing muon and outgoing $Z_\mu$ boson, respectively. In this section we give the main formulas for the production of the $Z_\mu$ boson in the reaction of Eq.(3.1). In the Weizsacker-Williams approximation [47], in the rest frame of the nucleus ($P = (M, 0)$, $p = (E_0, \vec{p})$, $p' = (E', \vec{p}')$, and $k = (E_{Z_\mu}, \vec{k})$, the $Z_\mu$-production cross section on the nucleus

$$\frac{d\sigma(\mu + Z \rightarrow \mu + Z_\mu + Z)}{dE_{Z_\mu} d\cos\theta_{Z_\mu}}$$

(3.2)
Figure 3. The branching fraction of the decays $\text{Br}(Z_\mu \rightarrow \nu \nu)$ and $\text{Br}(Z_\mu \rightarrow \mu^+\mu^-)$ as a function of the $Z_\mu$ mass.

is related to the cross section for real photon scattering, $\mu(p)\gamma(q) \rightarrow \mu(p')Z_\mu(k)$ with $q = P' - P$; namely, the following formula applies:

$$\frac{d\sigma(\mu + Z \rightarrow \mu + Z_\mu + Z)}{dE_{Z_\mu} d\cos \theta_{Z_\mu}} = \frac{\alpha \chi E_0 x \beta_{Z_\mu}}{\pi} \frac{1}{1 - x} \times \frac{d\sigma(p + q \rightarrow p' + k)}{d(pk)} \bigg|_{t=\min}$$

(3.3)

Here

$$x \equiv E_{Z_\mu}/E_0,$$

(3.4)

$$t \equiv -q^2,$$

(3.5)

$$\beta_{Z_\mu} = \sqrt{(1 - m^2_{Z_\mu}/E_0^2)}$$

(3.6)

and $\chi$ is the effective flux of photons integrated from $t = \min$ to $t = \max$ [47]. The kinematics is determined at $t = \min$. For a given $Z_\mu$ momentum the virtuality $t$ has its minimum value $t_{\min}$ when $\vec{k}$ is collinear with the three-vector $\vec{k} - \vec{p}$ [48]. One can find that in the Weizsacker-Williams approximation the cross section of the $\mu(p) + Z(P) \rightarrow Z(P') + \mu(p') + Z_\mu(k)$ reaction is given by [13]

$$\frac{1}{E_0^2 x} \frac{d\sigma}{dx d\cos \theta_{Z_\mu}} = 4 \left( \frac{\alpha^2 \alpha_{\mu} \chi \beta_{Z_\mu}}{1 - x} \right) \left[ \frac{C_2}{U^2} + \frac{C_3}{U^3} + \frac{C_4}{U^4} \right],$$

(3.7)
where
\begin{equation}
C_2 = (1 - x) + (1 - x)^3, \quad (3.8)
\end{equation}
\begin{equation}
C_3 = -2x(1 - x)^2m_Z^2 - 4m_{\mu}^2x(1 - x)^2, \quad (3.9)
\end{equation}
\begin{equation}
C_4 = 2m_{Z_{\mu}}^4(1 - x)^3 + (1 - x)^2[4m_{\mu}^4x^2 + 2m_{\mu}^2m_{Z_{\mu}}^2(x^2 + (1 - x)^2)]. \quad (3.10)
\end{equation}

By integrating with respect to \(\theta_{Z_{\mu}}\), we find that
\begin{equation}
\frac{d\sigma}{dx} = 2\left( \frac{\alpha^2 \alpha_{\mu} \beta_{Z_{\mu}}}{1 - x} \right) \left[ \frac{C_2}{V} + \frac{C_3}{2V^2} + \frac{C_4}{3V^3} \right], \quad (3.11)
\end{equation}
where
\begin{equation}
V = U(x, \theta_{Z_{\mu}} = 0) = m_{Z_{\mu}}^2 \frac{1 - x}{x} + m_{\mu}^2x \quad (3.12)
\end{equation}
For the general electrical form factor \(G_2(t)\) [47], the effective flux of photons \(\chi\) is
\begin{equation}
\chi = \int_{t_{min}}^{t_{max}} \frac{dt}{t^2} \frac{(t - t_{min})}{G_2(t)}. \quad (3.13)
\end{equation}

Note that for heavy atomic nuclei \(A\) we also have to take into account the inelastic nuclear form factor. Numerically, \(\chi = Z^2 \cdot \text{Log}\), where the function Log depends weakly on atomic screening, nuclear size effects and kinematics [48]. Numerically, \(\text{Log} \approx (5-10)\) for \(m_{Z_{\mu}} \leq 500\) MeV [47, 48]. One can see that compared to the photon bremsstrahlung rate, the \(Z_{\mu}\) production rate is suppressed by a factor \(\approx \alpha_{\mu}m_{\mu}^2/\alpha M_{Z_{\mu}}^2\).

In Fig. 2 an example of the expected distributions of the energy of a \(Z_{\mu}\) produced by a 160 GeV muon impinging on the Pb target is shown for different \(Z_{\mu}\) masses. The spectra are calculated for the coupling \(\alpha_{\mu} = \alpha\). One can see that for masses \(M_{Z_{\mu}} \gtrsim 100\) MeV the \(Z_{\mu}\) bremsstrahlung distribution is peaked at the maximal beam energy.

For \(M_{Z_{\mu}} < 2m_{\mu}\), the decays \(Z_{\mu} \to \mu\bar{\mu}\) are prohibited and the \(Z_{\mu}\) decays mainly into \(Z_{\mu} \to \nu_{\mu}\bar{\nu}_{\mu}, \nu_{\tau}\bar{\nu}_{\tau}\). For \(2m_{\mu} < M_{Z_{\mu}} < 2m_{\tau}\), in addition to decays into neutrino pairs \(Z_{\mu}\) also decays into \(\mu^+\mu^-\) pairs with the decay width
\begin{equation}
\Gamma(Z_{\mu} \to \mu^-\mu^+) = \frac{\alpha_{\mu}M_{Z_{\mu}}}{3} \left( 1 + \frac{2m_{\mu}^2}{M_{Z_{\mu}}^2} \right) \sqrt{1 - 4 \frac{m_{\mu}^2}{M_{Z_{\mu}}^2}}, \quad (3.14)
\end{equation}
The branching ratio into \(\mu^-\mu^+\) pairs is determined by the formula
\begin{equation}
Br(Z_{\mu} \to \mu^-\mu^+) = \frac{K \left( \frac{m_{\mu}}{M_{Z_{\mu}}} \right)}{1 + K \left( \frac{m_{\mu}}{M_{Z_{\mu}}} \right)}, \quad (3.15)
\end{equation}
where
\begin{equation}
K \left( \frac{m_{\mu}}{M_{Z_{\mu}}} \right) = \left( 1 + \frac{2m_{\mu}^2}{M_{Z_{\mu}}^2} \right) \sqrt{1 - 4 \frac{m_{\mu}^2}{M_{Z_{\mu}}^2}}. \quad (3.16)
\end{equation}
For the coupling of Eqs.(4) and (5), the \( Z_\mu \) with the mass \( M_{Z_\mu} \gtrsim 100 \text{ MeV} \) is a short-lived particle with the lifetime \( \tau_{Z_\mu} \lesssim 10^{-15} \text{ s} \). In Fig. 3 the branching fractions of the decays \( Z_\mu \rightarrow \nu \nu \) and \( Z_\mu \rightarrow \mu^+\mu^- \) are shown as functions of the \( Z_\mu \) mass. One can see that for \( M_{Z_\mu} \gtrsim 2m_\mu \), 50% of the \( Z_\mu \)'s decay invisibly into a couple of neutrinos, while another 50% decay into a \( \mu^+\mu^- \) pair. The latter would result in the muon trident signature in the detector. For \( Z_\mu \) energies \( E_{Z_\mu} \simeq 100 \text{ GeV} \), the opening angle \( \Theta_{\mu^+\mu^-} \simeq M_{Z_\mu}/E_{Z_\mu} \) of the \( \mu^+\mu^- \) pair from the decay is still big enough, and the muons could be resolved as two separate tracks, so the pairs would be mostly detected as double-track events. However, the main problem in the search for the \( Z_\mu \rightarrow \mu^+\mu^- \) decay is the background from the muon trident events from the QED reaction \( \mu Z \rightarrow \mu Z \mu^+\mu^- \), whose rate substantially exceeds the rate of the reaction (3.1). An additional study, which is beyond the scope of this work, is required for this decay channel. Here we mostly focus on the case when the reaction (3.1) is accompanied by the decay \( Z_\mu \rightarrow \nu \nu \), resulting in the invisible final state.

4 Phase I: The experiment to search for the \( Z_\mu \rightarrow \text{invisible} \) decay

The experiment can to perform in two Phases:

- The main goal of Phase I is to probe the muon \((g-2)\mu\) parameter space, which would require accumulation of \( \lesssim 10^{11} \) muons on target (MOT)

- The main goal of Phase II is to cover the DM parameter space by collecting \( \gtrsim 10^{13} \) MOT.

In this document we focus specifically on the 2021 run.

4.1 The detector

The experimental setup specifically designed to search for the \( Z_\mu \) production and subsequent decay \( Z_\mu \rightarrow \nu \nu \) from the reaction of Eq. (3.1) of high-energy muon scattering off nuclei in a high density target is schematically shown in Fig. 4. The detector shown in Fig. 4 utilizes two magnetic spectrometers upstream (MS1) and downstream (MS2) of the target. Each of these spectrometers, in turn, consists of two magnetic spectrometers aiming at two independent and consistent measurements of the muon momentum and high purity reconstruction of the initial and final muon state, respectively. The momentum of the incident muon is measured by the upstream spectrometer, which consists of the Beam Momentum Stations (BMS) and MS1 spectrometer. The MS1 includes an MBPL dipole magnet and the corresponding tracker stations. The momentum of the scattered muon is measured by the downstream spectrometer MS2, which is composed of two spectrometers, MS2-1 and MS2-2, each similar to the MS1. The tracker system is a set of low-material
Figure 4. Schematic illustration of the setup to search for dark $Z_\mu$. The bremsstrahlung $Z_\mu$s are produced in the forward direction in the reaction $\mu + Z \rightarrow \mu + Z + Z_\mu$ of a high-energy muon scattering off nuclei of an active target (see text for definition of colors).

budget Micromegas, GEM and straw-tube chambers allowing for the reconstruction and precise measurements of momenta for incident and outgoing muons. It also uses scintillating hodoscopes $H_{0-7}$. The scintillator counters $S_0$ and $S_1$, and veto counters ($V_{0-1}$) are used to define the small size and divergence of the primary muon beam, while the counter $S_2$ defines the scattered muons. The active target is surrounded by a high-efficiency electromagnetic calorimeter (ECAL) serving against photons and other secondaries emitted from the target at large angles. Downstream of the target the detector is equipped with a high-efficiency veto HCAL ($VHCAL$) with a small entrance hole, and a veto system $V_2$ consisting of counters $V_{21}$ and $V_{22}$, located inside the MBPLs of the MS2 and a massive, hermetic hadronic calorimeter (HCAL) located at the end of the setup all serving against charged and neutral secondaries produced from the muon interaction in the target. The HCAL consists of several modules, each with lateral and longitudinal segmentation. The central part of the first (last) module is a cell with the lateral size $\simeq 100 \times 100$ mm$^2$ ($\simeq 400 \times 400$ mm$^2$), used to detect scattered muons and secondaries emitted in the very forward direction. The rest of each HCAL module serves as a filter to completely absorb and detect the energy of secondary hadrons, electrons, and photons produced in the muon interactions $\mu^- A \rightarrow anything$ in the target. The last downstream counter $M$ is used for the final-state muon identification. The size of the central HCAL cells, straw-tube chambers and the counter $M$ is determined from simulations by the requirement to keep the acceptance for deflected scattered muons with momentum in the range 20-50 GeV $\gtrsim 90\%$. For example, the lateral size of the $M$ counter should
be at least $50 \times 50 \text{ cm}^2$ and is determined mostly by the deflection angle in the second magnet and multiple scattering in the HCAL modules of scattered muons. In order to suppress background due to the detection inefficiency, the HCAL must be longitudinally completely hermetic. To enhance its hermeticity, the HCAL thickness is chosen to be in the range $\simeq 20 - 30 \lambda_{\text{int}}$ (nuclear interaction lengths).

4.2 The method of the search

The method of the search is as follows. The bremsstrahlung $Z\mu$s are produced in the reaction (3.1) which occurs uniformly over the length of the target (T). A fraction $(f)$ of the primary beam energy $E'_{\mu} = f E_{\mu}$ is carried away by the scattered muon which is detected by the second magnetic spectrometer, as shown in Fig. 4, tuned for the scattered muon momentum $p'_{\mu} \lesssim f p_{\mu}$. The remaining part of the primary muon energy $(1 - f)E_{\mu}$ is carried away beyond all the subdetectors by the neutrinos from the prompt $Z_{\mu} \rightarrow \nu \nu$ decay resulting in the missing energy $E_{\text{miss}} = E_{\mu} - E'_{\mu}$.

The occurrence of $Z_{\mu}$ produced in $\mu^{-}Z$ interactions would appear as an excess of events with a single scattered muon accompanied by zero-energy deposition in the detector, as shown in Fig. 4, above those expected from the background sources. The signal candidate events have the signature:

$$S_{Z_{\mu}} = S_0 \cdot S_1 \cdot T \cdot \mu_{\text{out}} \cdot V_0 \cdot V_1 \cdot \text{VHCAL} \cdot V_2 \cdot \text{HCAL} \quad (4.1)$$

and should satisfy the following selection criteria:

(i) $S_0 \cdot S_1 \cdot T$: The presence of an incoming muon with momentum within $2\sigma$ of the central peak from the nominal beam momentum measured by both the BMS and MS1 spectrometers. The energy deposited in the target is consistent with that expected from a minimum ionizing particle (MIP).

(ii) $\mu_{\text{out}}$: The presence of a single scattered muon with energy $E'_{\mu} \lesssim 0.5 \times E_{\text{beam}}$ after the target, the presence of a single muon track in the straw-tube chambers between the HCAL modules and in the last counter $M$.

(iii) $V_0 \cdot V_1 \cdot \text{VHCAL} \cdot V_2 \cdot \text{HCAL}$: No energy deposition in the veto counters $V_0$, $V_1$, VHCAL and $V_2$ located inside the MS2 magnets, no energy deposition in the central HCAL cells above those expected from the scattered muon, and no energy in the rest of the HCAL modules.

The ”zero-energy” signal is defined as:

(i) The presence of the energy $E_{\text{ECAL}}^\mu$ deposited in the target, which is consistent with that deposited by a MIP.

(ii) The presence of energy $E_{\text{HCAL}}^\mu$ deposited in the HCAL cell crossing by the scattered muon compatible with that expected from a MIP, $E_{\text{HCAL}}^\mu \simeq E_{\text{mip}}$. 

- 14 -
Figure 5. The distribution of the energy deposited in the central cell of the first HCAL module (red histogram) and in all four HCAL central cells (shaded histogram) by traversing muons with energy $E_{\mu} = 80$ GeV. The peak of the pedestal sum over the rest of the HCAL in units of hadronic energy is also shown.

(iii) No energy deposition in the veto counters $V_0$, $V_1$ and $V_2$, and in the rest of the HCAL, $E_{HCAL} \lesssim 100$ MeV, presented by the sum of pedestals of the readout system, as shown in Fig. 5. The effective width of the signal in the rest of the HCAL is $\simeq 100$ MeV.

(iv) The total energy deposited in the ECAL and HCAL is $E_{tot} = E_{ECAL} + E_{HCAL} \lesssim 12$ GeV.

The optimal primary beam energy is selected using the following considerations. First, it has to be high enough to provide the highest rate for the production of $Z_\mu$'s in the sub-GeV mass range; second, it should correspond to as large absolute value of $E_{miss}$ as possible; and third, it should allow one to operate muon beam at high intensity. Taking these considerations into account, a beam energy in the region 100-160 GeV is chosen.

4.3 The signal simulations

In Fig. 6, the simulated distribution of the events from the SM reactions [59]: bremsstrahlung, knock-on, pair-production and photonuclear muon interactions in the tar-
get in the \((E'_\mu; E_{\text{tot}})\) plane is shown for the primary muon beam energy \(E_\mu \sim 160\) GeV and a total number of incident muons \(n_\mu \simeq 3 \times 10^8\). The events are selected with the requirement of no energy deposition in the veto counters \(V_1\) and \(V_2\). The signal of the reaction \(\mu + Z \rightarrow \mu + Z + Z_\mu, Z_\mu \rightarrow \nu\nu\) is defined by a scattered muon energy \(15 \lesssim E'_\mu \lesssim 100\) GeV and a total energy \(E_{\text{tot}} = E_{\text{ECAL}} + E_{\text{HCAL}} \lesssim 12\) GeV. The width of the signal region along the \(E_{\text{tot}}\) axis corresponds to an energy deposition around 2.4 GeV in each central cell of four consecutive HCAL modules, as illustrated in Fig. 5, plus about 0.5 GeV deposited in the ECAL, i.e. around 10 GeV for the total energy.

The production of light particle \(Z_\mu\) off nuclei by muons, which is the signal that we search for, was performed by the code similar to one described in [53], compiled as a part of the Geant4 application. There are three main differences, listed below.

- Different beam particle mass.
- It turned out that the difference between WW and exact tree level cross section
calculations [54] is small, within a factor $\lesssim 1.5$, in the interesting range of $Z_\mu$ masses. For this reason we use simply the former without K-factors.

- In the case of muon beam the change of direction of the muon can be important because we have to detect it. This required additional calculations.

We assumed that the target is the ECAL used in our previous searches, it contains $\simeq 40 X_0$ of Pb (here $X_0$ is the radiation length).

**Figure 7.** Left panel: The distribution of the $Z_\mu$ energy, nominal beam momentum 160 GeV. Right panel: The distribution of the scattered muon angle after the $Z_\mu$ emission calculated for $m_{Z_\mu} = 100$ MeV for the primary beam energy of 160 GeV.

**Figure 8.** The 2D distribution of the recoil muon energy after the target $E'_\mu$ and the total registered energy $E_f = E'_\mu + E_{tot}$, for the signal events calculated for the mass $m_{Z_\mu} = 100$ MeV and the primary beam energy of 160 GeV.

The distribution of $E_{Z_\mu}$ for the 160 GeV beam energy and the mass $m_{Z_\mu} = 100$ MeV is shown in the left panel of Fig. 7. An example of the muon angle distribution after the $Z_\mu$ emission is shown in the right panel of Fig. 7. It scales as the mass of $Z_\mu$ and as $1/E_\mu$. Taking into account the beam size, the edge of the hole in the VHCAL
located before the MS2-1 magnet, is seen from the upstream interaction vertex in the ECAL at the angle starting from $\approx 8$ mrad. The inefficiency due to the scattered muon angle becomes sizeable for the masses of $Z_\mu$ higher than 300 MeV. This figure also represents the distribution of missing energy. The 2D distribution of the recoil muon energy after the target, $E_{\mu}^{ECAL}$ and the total detected energy $E_f = E_{\mu} + E_{\text{tot}}$, is shown for the signal events in Fig. 8. The signal events are located in the region $E_{\text{tot}} \lesssim 20$ GeV. The signal efficiency lost due to recoil muon interactions in the HCAL is at the level of a few %. No cuts on $V_2$ are used in this case.

Figure 9. Beam distribution at the entrance to the NA64 setup predicted by EN-EA-LE section with HALO software to be $\sigma_x \approx \sigma_y \approx 20$ mm.

4.4 The M2 beam

The experiment is supposed to employ the upgraded M2 muon beam at the CERN SPS described in details in Ref.[49]. The beam was designed to transport high fluxes of muons of maximum momenta in the range between 100 and 225 GeV/c that could be derived from a primary proton beam of 400 GeV/c with intensity between $10^{12}$ and $10^{13}$ protons per SPS spill. The beam is produced by protons impinging on a primary beryllium target and transported to the detector in an evacuated beam-line [50]. The typical maximal intensity for a beam energy $\simeq 160$ GeV, is $\gtrsim 5 \times 10^7 \mu^-$ for a SPS spill with $10^{12}$ protons on target. The typical SPS cycle for fixed-target (FT) operation lasts 14.8 s, including 4.8 s spill duration. The maximal number of FT cycles is four per minute. Our plan is to use $\lesssim 10^7 \mu^-$/spill for the Phase I, and the maximal intensity at the second phase. The hadron contamination in the muon beam is remarkably low ($\pi/\mu \lesssim 10^{-6}$) and the size of the beam at the detector position is of the order of $\sigma_x \sim \sigma_y \sim 20$ mm with a angular divergence of $\sigma_x' \approx \sigma_y' \approx 0.2$ mrad as studied by the CERN EN-EA-LE section (D. Banerjee and J. Bernhard). Fig. 9 and 10 shows the beam distribution obtained with HALO, a muon beam simulation software [51] which simulated the entire M2 beam line. The halo component in a
muon beam is particularly large as can be seen in Fig. 11 for the region $|X| < 3$ m and $|Y| < 3$ m and $|\rho| = |\sqrt{X^2 + Y^2}| > 15$ cm and is estimated to be $\sim 27\%$ of the full beam intensity. The increased halo flux at the position of the experiment ($Z = 1078.772$ m according to the M2 co-ordinate system) compared to the detected flux of 20% at COMPASS target ($Z = 1131.824$ m) is due to the upstream location and the shorter distance from the Magnetized Iron Blocks (at $Z = 1069.559$ m) which results in an insufficient lever arm for optimal halo suppression. However, because of the small transverse acceptance of all NA64 detectors of maximum $X \times Y = 10$ cm $\times$ 10 cm the major part of this halo component is not within the experimental acceptance and can be controlled as explained by the trigger setup and signal selection procedure described in the following sections.
4.5 The simulation of the beam and NA64\(\mu\) setup

The simulation of the M2 beam muons in the setup is performed in two ways. The first is to parameterize the distributions of coordinates, angles and momentum of muons at some Z coordinate, usually at the exit of MS1 which are obtained from HALO and to let Geant4 simulate primary particles according to this parameterization (the Geant4 subpackage gps). The second way is to transform the beam files, which are the lists of muons with their coordinates etc. as obtained from the beam simulation program, HALO, to HepMC ASCII format and read them one by one using the Geant4 HepMC interface.

![Figure 12. Trigger Setup.](image)

For the simulation of the trigger conditions the geometry of setup shown in Fig. 12 was used. Two cases with a single and the two MBPL magnets for the MS2 spectrometer were considered as shown. The muons at the exit of MS1 magnet were shot as primary particles. The muon distributions at this depth are shown in Fig. 13 and Fig. 14. The distribution includes the main beam core profile and a big halo (\(\simeq 20\%\) of the full beam intensity). In the momentum distribution of triggered muons (right lower plots in figures) the beam is defined by two round counters (diameter 4.2 cm) separated by 2.7 meters. The energy in the trigger counters that define the beam is deposited mainly by the primary muons, but sometimes also by secondary particles (for example delta electrons). We estimated the corresponding fraction for the muons with momentum below 120 GeV in the 160 GeV beam to be \(\simeq 0.2\%\). A fraction of 87\% of them pass through the trigger counters.

4.5.1 The incoming and outgoing muon momentum measurements

As can be seen from Fig. 13 - Fig. 14, there are incoming triggered muons with initial momentum significantly lower than the nominal beam momentum. For this reason the measurement of the initial momentum is necessary. The momentum of
Figure 13. Muon distributions at the exit of MS1, nominal beam momentum 100 GeV

Figure 14. Muon distributions at the exit of MS1, nominal beam momentum 160 GeV
the incoming muon is defined by the measurements performed by the BMS and MS1 spectrometers as shown in Fig. 15. Simulation is performed to estimate the resolution of momentum reconstruction and check the level of purity of this combined measurement. The BMS consists of beam defining hodoscopes labelled BM01-06. The entire beamline is simulated with HALO and the particle hits are recorded at the BMS hodoscope positions and the entrance to the NA64 setup after Q33 as shown in Fig. 15. The BMS hodoscopes have a detector resolution of 1.3 mm for BM01-4 and 0.7 mm and 0.4 mm for BM05 and BM06 respectively. The hodoscopes can define the beam only in the vertical (Y) co-ordinate. The system of bending magnets (Bend 6) consists of three 5 m vertical bends with $\sim 3.3$ T.m each. The momentum reconstruction with the BMS is estimated using the TMVA analysis of ROOT with Boosted Decision Tree method taking the vertical hit positions and the direction of the particle in the upstream and downstream part as the input variables. The detector resolution is taken into account. To estimate the momentum reconstruction with MS1 the incident distribution of the beam at the entrance of NA64 is obtained from HALO and propagated through the NA64 setup using Geant4. The momentum reconstruction is done taking into account two tracker modules upstream and three modules downstream of the 2 m MS1 MBPL magnet of 3.6 T.m. The detector resolution $\sim 0.1$ mm is taken into account. The resolution of momentum reconstruction is shown in Fig 16 with $\sigma_{BMS} \sim 1.3$ GeV/c and $\sigma_{MS1} \sim 2.3$ GeV/c for the 160 GeV/c incoming beam. The results of the reconstruction for BMS and MS1 is shown in Fig. 17 as $P_{\text{True}}$ vs $P_{\text{BMS}}$ and $P_{\text{True}}$ vs $P_{\text{MS1}}$.
Figure 16.  Resolution of momentum reconstruction with BMS and MS1.  $\sigma_{BMS} \sim 1.3$ GeV/$c$ and $\sigma_{MS1} \sim 2.3$ GeV/$c$ for 160 GeV/$c$ incoming beam.

Figure 17.  Distribution of the incoming muon momentum $P_{in}$ reconstruction by the BMS and MS1 spectrometers as a function of the True Momentum for the beam energy 160 GeV/$c$.  The initial muon momentum is defined from the consistency of these two independent measurements of $P_{in}$ with the BMS and MS1.

4.6 The trigger rate

The current NA64 DAQ acquisition rate is limited to 6 kHz ($\sim$ 30 kEvents/spill).  In order to run NA64e at $\sim 10^7$ e$^{-}$/s/spill, an energy threshold of $E_{th} \sim 80$ GeV in the ECAL was applied to suppress the trigger rate by a factor of $\lesssim 10^3$.  As discussed in Sec. 4.2, the signal for the NA64$\mu$ experiment is a muon with momentum below $0.5 \times P_0$ reconstructed with the MS2 spectrometer.  In analogy to NA64e, the muon missing momentum/energy will be used to keep the trigger rate at the required level.  For the minimum missing energy to be triggered on, which is about one half of the beam energy, the deflection angle in the magnetic spectrometer, e.g. $\gtrsim 20$ mrad for the 80-100 GeV signal muon, is much larger then the muon scattering angle from the reaction (3.1) and the multiple-scattering angle of the muon in the target.  Given these facts and taking into account that for the initial phase we can run experiment at a relatively low beam intensity, $\lesssim 10^7 \mu$/spill, a simplified solution has been first considered at the cost of the acceptable signal efficiency drop.
4.6.1 The trigger option 1

This trigger option is based on the selection of the well defined primary muon beam within the small lateral size, divergency, and momentum spread by using small size beam defining counters and the MS1 magnet. To do this in NA64µ, the deflection of the scattered muon in the transverse plane will be used in the trigger condition. The trigger setup used in simulations is composed of the following parts that are shown in Fig. 12:

**Figure 18.** Simulation of a muon monoenergetic beam traversing the setup. Expected signal at the trigger scintillator after traversing the MS2.

1. $S_0$ and $S_1$: two scintillator counters (42 mm diameter, 3 mm thick) before the ECAL spaced by 3 meters will define the incoming beam.

2. $VHCAL$: A veto hadronic calorimeter, with a hole in the middle of 120x60 mm$^2$ placed before the MS2 magnet will be used to suppress events with large angle scattered muons and large angle secondary particles, see e.g. Figs. 4,12.

3. $S_2$: A scintillator counter before the HCAL shifted from the beam axis (see Fig. 12) to trigger on deflected muons.

4. $V_2$: A scintillator base veto system (see Sect.4.1 and Fig.4) which will be used to veto the primary beam if $S_2$ is triggered by a secondary particle.

5. Muon counter $M$: a scintillator counter at the end of the HCAL used to identify muons.

The trigger condition will thus be:

$$TRIGGER = S_0 \times S_1 \times V_0 \times V_1 \times S_2 \times V_2 \times M \quad (4.2)$$
A detailed Geant4 simulation is used in order to estimate the rate from multiple-scattering (MS) muons in the ECAL that can trigger $S_2$. The expected MS divergence ($\theta_{MS}$) and spread ($\sigma_{MS}$) for a muon crossing the ECAL with a corresponding length of $40X_0$ can be estimated with

$$\theta_{MS} = \frac{13.6\text{MeV}}{\beta_c p} \cdot z \cdot \sqrt{\frac{X}{X_0}} \left(1 + 0.038 \ln \left(\frac{X}{X_0}\right)\right) = 9.8 \text{ mrad} \quad (4.3)$$

$$\sigma_{MS} = 0.00098 \times 4950 = 4.851 \text{ mm} \quad (4.4)$$

As shown in Fig. 18, the MC done by simulating a mono-energetic beam starting on beam axis with no divergence reproduces this properly as well as the expected deflection ($X$) after MS2

$$\phi = \frac{L[m]}{\rho[m]} = \frac{0.3LB[T \cdot m]}{p_T[GeV]} = 0.3 \frac{3.91}{100} = 0.012 \quad (4.5)$$

$$X = 0.012 \times 2100 = 25.2 \text{ mm} \quad (4.6)$$

Figs. 19-20, show the projection on the deflection axis ($x$-axis) of the expected hit positions for both signal and primary beam at the location of $S_2$ and $V_1$ for a beam energy of 160 GeV and a MS2 configuration with one MBPL. The orange area are
the events that would be detected if the edge of $S_2$ would be placed at $x=-50$ mm while the white area are the events hitting $V_2$. The plots for the beam momentum 160 GeV are shown in Fig. 21. On the right plots in this figure one can see how the muon ID counter after the HCAL removes very soft muons. The trigger rate due to the MS muons from the primary beam and the efficiency to detect signal events as a function of $S_2$ position are summarized in the Tables of Fig. 22 for two different experimental setups and two different beam energies. The figure in the top corresponds to the values obtained for a beam energy of 100 GeV, and the one in the bottom for 160 GeV. Both cases are also calculated defining MS2 as only one MBPL, case 1 MBPL, or two magnets separated 1 m, case 2 MBPL. The red values correspond to the number of signal events with a projection on the deflection value below the value specified in each column. The orange numbers between parenthesis correspond to this number multiplied by the ratio of signal events reaching the last muon counter in the setup. In all cases, signal is defined as those events with a $Z_\mu$ candidate and an energy, $E_{Z_\mu} > 0.5 \times E_0$.

The signal was simulated for different $Z_\mu$ masses, namely 10, 100, 1000 MeV. The efficiency for the signal has been calculated for the different cases described in Fig.22. For 100 GeV the signal efficiency to be detected $S_2$ when its edge is at $X_{S_2} = -70$ mm ranges from $\epsilon_{SIG} = 46\% - 36\%$ for 1 MBPL while the trigger rate from MS muons

Figure 20. Projection on the deflection axis (x-axis) of the expected hit positions in $S_2$ (orange area) and $V_1$ (white area) for signal events using the profiles calculated with HALO. The signal events are defined as those with a $Z_\mu$ candidate and an energy, $E_{Z_\mu} > 0.5 \times E_{beam}$.
Figure 21. The distributions of the muon X coordinate at the front face of HCAL. Upper row: no signal. Lower row: signal events with $Z_\mu$ energy higher than $0.5 \times E_{beam}$. Nominal beam momentum is 160 GeV.

from the primary beam is below $10^{-3}$. In the case of 2 MBPL the efficiency after cutting at $X_{S_2} = -130$ mm ranges between $\epsilon_{SIG}=17-50\%$, while the rate of muons from the primary beam is still below $10^{-3}$. Therefore, one can conclude that at a rate of $N_\mu = 5 \times 10^6$ muons/spill the DAQ will be capable to operate without any significant deadtime. To acquire $\simeq 10^{11}$ MOT in order to cover the (g-2)$_\mu$ muon favoured parameter space assuming a conservative number of $N_{SP}=3000$ spills/day will thus require about:

$$N_{days} = \frac{\propto 10^{11}}{\epsilon_{S_0S_1} \times \epsilon_{SIG} \times N_{SP} \times N_\mu} \simeq 50$$  \hspace{1cm} (4.7)

Including setting up, debugging and calibration this would be feasible in a couple of months of beamtime.

The final step in the trigger rate study was the estimation of the accidental trigger rate to check that the experiment is feasible. For this estimate we assume for the incoming muons the maximal counting core beam rate $R_b = S_0 \cdot S_1 \simeq 10^7 \mu$/spill, or the rate $R_b \simeq 2 \times 10^6 \mu/s$ given by the coincidence of two beam defined counters $S_0 \times S_1$. The simulation shows that for this condition the counting rate of the trigger counter $S_2$ with the X-Y size $150 \times 100$ mm$^2$ and with an offset $X < 50$ mm is $1.7 \times 10^6 \mu$/spill, or $R_b \simeq 3 \times 10^5 \mu/s$. This value is mostly defined by the muon beam halo rate. Assuming conservatively the time width of the coincidence gate to
Figure 22. Summary tables: an example of the trigger rate and signal efficiency versus $S_2$ position for different MS2 configurations and beam energies. The red values correspond to the number of signal events with a projection on the deflection value below the value specified in each column. The orange numbers between parenthesis correspond to this number multiplied by the ratio of signal events reaching the last muon counter in the setup. In all cases, signal is defined as those events with a $Z_\mu$ candidate and an energy, $E_{Z_\mu} > 0.5 \times E_0$.

be $\delta_t \lesssim 10$ ns, one can get for the trigger condition, see l.h.s. of Fig.23:

$$Tr_1 = S_0 \cdot S_1 \cdot S_2 \cdot M$$ (4.8)

the corresponding accidental rate

$$R_{tr}^{acc} \simeq R_b \cdot R_h \cdot \delta_t \simeq 6 \times 10^3 \mu/s$$ (4.9)
or $R_{tr1}^{acc} \simeq 3 \cdot 10^4 \mu$/spill, which is at the current DAQ rate limit $\gtrsim 3 \cdot 10^4$/spill. This estimate shows that further improvement of the trigger is required.

1) 

2) 

Figure 23. Two trigger options for NA64$\mu$. See text.

4.6.2 The trigger option 2

To reduce the accidental trigger rate, one could try i) to implement the hodoscope based trigger system, similar to the one used for the muon scattering experiments, see e.g. [55], and ii) to optimise the incoming beam properties such as divergency and the core/halo beam intensity ratio. Below we report on design of a new trigger system, while the simulations on the beam optimisation are still in progress.

As discussed previously, for the minimum missing energy $E_{\text{miss}} \gtrsim E_0/2$ to be triggered on the deflection angle in the magnetic spectrometer, e.g. $\gtrsim 20$ mrad for the 100 GeV signal muon, is much larger then the muon scattering angle from the reaction of (3.1) and the multiple-scattering angle of the muon in the target. Given the fact that the minimal deflection angle for signal events is relatively large, the muon trigger system is designed to select the muon scattering events with the help of dedicated trigger hodoscopes. The possibility to use in the trigger the hadron calorimeter modules HCAL1 - HCAL3, shown in r.h.s. of Fig. 23, where the energy deposition should not to significantly exceed the energy deposited by a MIP is also under study. The hodoscope trigger selection is based on the directionality properties of the track of the signal muon candidate which should pointed back to the target.
In this design the counter $S_2$ consists of four vertical scintillator strips $T_{1-4}$ as shown in r.h.s. Fig. 23, each covering the angle of the scattered signal muon $\Delta \Theta \simeq 5$ mrad. The counter $M$ consists of a four partly overlapped vertical counters $M_{1-4}$. The trigger requires the coincidence of any pair $T_i \cdot M_i$, with $i = 1, 2, 3, 4$. The coincidence between the corresponding strips $T_i$ and $M_i$ corresponds to the track which being extrapolated back originates from the target region. The simulation shows that for this condition the counting rate of the trigger $\sum_i T_i \cdot M_i$ from (4.10) is reduced down to $\simeq 0.58 \cdot 10^6 \mu$/spill, resulting in the accidental trigger rate to $R_{acc}^{tr} \simeq 1.2 \cdot 10^4 \mu$/spill, which is a factor 3 less that the current DAQ rate limit $\gtrsim 3 \cdot 10^4$/spill thus making the experiment feasible.

4.7 DAQ and electronics upgrades

The following steps are foreseen to improve the DAQ efficiency,

- to increase the DAQ rate by at least a factor 2-3 allowing running at a higher beam intensity of the order $S_0 \times S_1 \simeq 5 \times 10^7 \mu$/spill
- to decrease the width of the coincidence gate to $\delta t \lesssim 3$ ns

This would result in the ratio $R_{acc}^{tr}/R_{tr} \lesssim 0.1$.

(i) DAQ: The DAQ was successfully commissioned during NA64 runs at H4 beam. For the NA64$\mu$ experiment it will be upgraded to read out the following detectors: - $\sim 500$ channels of the ECAL target and HCAL calorimeter; - $\sim 5000$ channels of Straw detector; - $\sim 200$ channels of hodoscopes; - up to $\sim 5000$ channels of MM and GEM detectors using APV ASIC;

(ii) The ECAL and HCAL Calorimeters: The readout chain of the electromagnetic calorimeter consists of a Shaper, MSADC and multiplexer. No shaper modules are planned to be used, while new MSADC modules operating at $\sim 200$MHz are supposed to be developed and produced. The delivery is expected in the middle of 2020.

(iii) Straw front-end electronics:

The Straw detector will be equipped with new preamplifier-discriminator cards developed for CBM experiment and based on OKA-1 ASIC. The cards already exist. A dedicated TDC card for NA64 experiment based on an FPGA has been developed. The first prototype will be tested and production of cards is scheduled for 2019. It’s expected to start commissioning of the straw readout electronics in the second half of 2020.
(iv) Hodoscope TDC and beam counter TDC: A TDC for hodoscope detectors as well as for beam counters is a faster version of the straw TDC. It will be produced simultaneously with the straw TDC.

(v) MM and GEM detectors: MM and APV detectors are read out by COMPASS electronics. The electronics exists and could be provided for the test beam.

4.8 Background

To estimate the background and sensitivity of the proposed experiment, a simplified feasibility study based on GEANT4 [52] Monte Carlo simulations has been performed for incoming muons with an energy of 160 GeV. In these simulations the target is the radiation-hard shashlik module ($X_0 \simeq 1.3$ cm) with a total thickness of about 40 $X_0$, surrounded by the ECAL, which is a hodoscope array of the lead-scintillator counters that are also of the shashlik type, each with a size of $38 \times 38 \times 490$ mm$^3$, allowing for accurate measurements of the lateral energy leak from the target. The shashlik calorimeter is a sampling calorimeter in which scintillation light is read out though wavelength-shifting fibers running perpendicular to the absorber plate; see, e.g. Ref. [56]. The target module consists of 150 layers of 1.5 mm thick lead and 1.5 mm thick plastic scintillator plates and has longitudinal segmentation. The veto counters are assumed to be $\simeq 1$ cm thick, high-sensitivity scintillator arrays with a high light yield of $\gtrsim 10^2$ photoelectrons per 1 MeV of deposited energy. It is also assumed that the veto inefficiency for the MIP detection is, conservatively, $\lesssim 10^{-4}$. The hadronic calorimeter HCAL0 is a set of three modules. Each module is a sandwich of alternating layers of iron and scintillator with a thickness of 25 mm and 4 mm, respectively, and with a lateral size simulated up to $150 \times 150$ cm$^2$. Each module consists of 48 such layers and has a total thickness of $\simeq 7\lambda_{int}$. The number of photoelectrons produced by a MIP crossing the module is in the range $\simeq 150$-200. The HCAL1 and HCAL2 are a single module calorimeters with the similar longitudinal structure. The energy resolution of the whole HCAL calorimeter as a function of the beam energy is taken to be $\frac{\sigma_E}{E} \simeq \frac{60\%}{\sqrt{E}}$ [58]. The energy threshold for zero energy in the HCAL for the case of the subtraction of the muon energy deposition is $\simeq 1$ GeV. We assume that the momenta of the in- and outgoing muons are measured with a precision of a few percent. The scattered muon produced in the target is defined as a single track crossing the HCAL and the straw-tube stations ST9-ST12 and accompanied by no activity in the HCAL modules. The background reactions resulting in the signature of Eq. (4.1) can be classified as being due to physical-, detector-, and beam-related sources. To investigate these backgrounds down to the level $\lesssim 10^{-10}$ with the full detector simulation would require a prohibitively large amount of computer time. Consequently, only the following background sources - identified as the most dangerous- are considered and evaluated with reasonable statistics combined with numerical calculations:
4.8.1 The purity of the incoming muon momentum reconstruction

One of the main background sources is related to the low-energy tail in the energy distribution of beam muons. The muon energy is lost due to the interaction of the particles with passive material, such as, entrance windows and the residual gas of beam lines. Another source of low-energy muons is due to the in-flight decays of pions and kaons that contaminate the beam. The uncertainty arising from the lack of knowledge of the dead material composition in the beam line is potentially the largest source of systematic uncertainty for accurate calculations of the fraction and energy distribution of these events. An estimation shows that the fraction of events with energy below $\lesssim 80$ GeV in the muon beam tuned to 160 GeV could be as large as $10^{-6}$. Hence, the sensitivity of the experiment could be determined by the presence of such muons in the beam, unless one takes special measures to suppress this background.

![Figure 24](image)

**Figure 24.** Simulated 2D-distribution of the reconstructed momentum with MS1 as a function of the True Momentum for a selected sample of $P_{BMS}$ between 140 GeV/c and 180 GeV/c to estimate the momentum reconstruction impurity for the incoming 160 GeV/c muons. The red ellipse indicates the signal region.

To improve the high-energy muon selection and suppress the background from possible admixture of low-energy muons, an additional tagging system utilizing a magnetic spectrometer MS1 is used along with the BMS spectrometer as schematically shown in Figs. 4, 15. The precision of the muon momentum measurement with MS1 is dominated by the track measurement errors, $\sigma_x \simeq 100 \, \mu$m for the MS1 trackers and $\sigma_y \simeq 1.3 \, \text{mm}$ for BM01-BM04 and $\sigma_y \simeq 0.7 \, \text{mm}$ and $0.4 \, \text{mm}$ for BM05 and BM06 respectively. The momentum resolution, $\sigma_{p_{BMS}} \sim 1.3 \, \text{GeV/c}$ for BMS and $\sigma_{p_{MS1}} \sim 2.3 \, \text{GeV/c}$ for MS1 for a 160 GeV/c beam, obtained from simulation is explained in Section 4.5.1. The level of purity of the measurement was calculated to estimate the background level.
By impurity \( I_{in} \) of incoming muon momentum \( (P_{true-in}) \) distribution we mean the variable

\[
I_{in} = \frac{n(P_{true-in} \lesssim P_{th}^{in})}{n(P_{recon-in} \simeq P_0)}
\]  

(4.11)

which is the fraction of muons with the true momentum \( P_{true-in} \) below the threshold \( P_{th}^{in} \) in the sample selected with BMS and MS1 where \( P_{BMS} \sim P_{MS1} = P_{recon-in} \sim P_0 \). \( P_0 \) is the nominal beam momentum within the momentum spread. Typically, \( P_{th}^{in} \) is selected as \( P_0/2 \). Fig. 24 shows the distribution with the ellipse indicating the region consistent with potential background (signal region). Extrapolation of the double Gaussian-like function into the signal region \( P_{BMS} \simeq P_{MS1} \simeq P_0 \) and \( P_{true-in} \lesssim 80 \text{ GeV/c} \) results in the corresponding impurity conservatively estimated as

\[
I_{in} \lesssim 10^{-12} - 10^{-11}
\]  

(4.12)

The estimated value is defined from the extrapolation of the observed shape shown in Fig. 24 to the signal region and by the number of accumulated events in this simulation. More precisely, to estimate the level of purity of the measurement with simulation a sample of simulated events in a momentum window of 140-180 GeV/c was chosen for both the MS1 and BMS reconstructions and the distribution of the difference of the True Momentum from the reconstructed MS1 momentum was plotted as shown in the left panel of Fig. 25. In this plot, the events with the difference value \(< -80 \text{ GeV/c} \) will be the background for a 160 GeV/c nominal beam and a threshold of 80 GeV/c for signal detection. The only cut used for this distribution was the requirement of only one hit on each tracking module. The distribution was then fitted with a Crystal Ball Function to account for the lower energy tail and extrapolated to find the level of background. The right panel shows the extrapolated value and the level of background for a threshold of 80 GeV/c < \( 10^{-13} \). Few events with difference \(< -20 \text{ GeV/c} \) was found wherein a part of the primary muon energy was carried by secondary particles from muon interactions in the detector materials which were successfully rejected with the hit requirement cut.

4.8.2 The purity in the muon final state

One of the sources of background is expected from the inaccurate measurements of the incoming muon momenta in the MS2 spectrometer for those muons that passed the target without interactions. We can define the impurity for such muon momentum \( (P_{recon-out}) \) measurements as the ratio

\[
I_{out} = \frac{n(P_{recon-out} \lesssim P_{th}^{out})}{n(P_{out} \simeq P_0)}
\]  

(4.13)

which is the fraction of muons reconstructed as the scattered muons with the momentum \( P_{recon-out} \lesssim P_{out}^{th} \simeq 80 \text{ GeV/c} \) for the sample of \( P_0 \simeq 160 \text{ GeV/c} \) incoming
**Figure 25.** Distribution of True Momentum - Reconstructed MS1 momentum for a reconstructed momentum window of 140-180 GeV with both MS1 and BMS. The distribution is fitted with a Crystal Ball function with parameters $x = 0$, $\sigma = 2.3$, $\alpha = 2.5$ and $N = 15$ (left panel). Fitted Crystal Ball Function with parameters $x = 0$, $\sigma = 2.3$, $\alpha = 2.5$ and $N = 15$ extrapolated till Difference $< -80$ GeV/c to get level of background $\lesssim 10^{-13}$ (right panel).

muons passing the detector without interactions with momentum $P_{out} \sim 160$ GeV/c. For the case of MS2 consisting of two MBPL magnets an example of evaluation of the $I_{out}$ value is shown in Fig. 26. The extrapolation of the event reconstructed with the

**Figure 26.** Measured 2D-distribution of the reconstructed momenta with the BMS (MS2-1) and COMPASS Spectrometer (MS2-2) magnets used for the estimation of the momentum reconstruction impurity for incoming 160 GeV muons that passed the detector without interactions.

BMS and SM2 spectrometers for the data sample collected by the COMPASS experiment with the empty target into the signal region was done in the following way. The left panel of Fig. 27 shows the distributions of the incoming muon momenta $P_{SM2}$
reconstructed in the COMPASS SM2 magnetic spectrometer for the event sample selected with the requirement to have the incoming momentum $P_{BMS}$ reconstructed with the BMS in the region $P_{BMS} \simeq 160$ GeV. Assuming that the performance of

\begin{align*}
\begin{array}{c}
\text{Entries} & 145209 \\
\text{Mean} & 159.3 \\
p_0 & 9.1 \pm 200.4 \\
p_1 & 0.204 \pm 3.697 \\
p_2 & 0.0341 \pm 0.5902 \\
p_3 & 0.0205 \pm 0.1822 \\
\end{array}
\end{align*}

the NA64$\mu$ reconstruction of the incoming muon momentum is comparable to the one obtained by COMPASS, allows us to get an estimate of the muon momentum mis-measurement rate for 1-MBPL and 2-MBPL cases. More precisely, this estimate can also be interpreted as a lower limit on the background level due to the muon momentum mis-measurement rate for 1-MBPL and 2-MBPL cases. The extrapolation of the fitted distribution shape to the signal region $P_{SM2} \lesssim P_{BMS}/2 = 80$ GeV shown in the right panel in Fig. 27 together with the parameters of the fit results gives rather crude, order of magnitude estimate of the background level to be

$$I_{out} \lesssim 10^{-12} - 10^{-11}$$  \hfill (4.14)

Although the background level estimates for the mis-measurement rate from simulations and real data agreed within an order of magnitude, the true values of (4.13) and (4.14) are defined mostly by the uncertainty in the extrapolation of the observed shape in Figs. 24, 26 to the signal region, the number of accumulated events in this test and mostly by the yet unknown low momentum tails in the distributions shown in Figs. 25, 27. We plan to perform the corresponding measurements of the level of impurity (mis-measurement) for the incoming muons in the pilot run.

4.8.3 The empty target measurements

The important cross-check of the performance of the tagging system for the incoming and outgoing muons is the measurement with the empty target, when the muons pass the detector without interactions. In this case the performance of the upstream
and downstream spectrometers, as well as of each of the four magnetic spectrometers BMS, MS1, MS20-1 and MS2-2 can be evaluated by the cross-checking of the measurement results of one spectrometer with respect to another. For example, the purity of the incoming momentum reconstruction by the BMS-MS1 system can be evaluated by measuring the ratio

\[ I_{in} = \frac{n_{MS2}(P_{\text{down}} \lesssim 80 \text{ GeV})}{n_{BMS-MS1}(P_{\text{up}} \simeq 160 \text{ GeV})} \] (4.15)

where \( n_{MS2}(P_{\text{down}} \lesssim 80 \text{ GeV}) \) is the number of outgoing muons with the momentum \( P_{\text{down}} \lesssim 80 \text{ GeV} \) obtained from the combined and consistent measurements from the downstream spectrometers MS2-1 and MS2-2, and \( n_{BMS-MS1}(P_{\text{up}} \simeq 160 \text{ GeV}) \) is the number of incoming muons with the momentum \( P_{\text{up}} \simeq 160 \text{ GeV} \), as defined from two independent and consistent measurements performed with the BMS and MS1. These are the crucial benchmark measurements for the experiment, which require accumulation of \( \simeq 10^{11} \) muons.

### 4.8.4 Hadron contamination and decays in the M2 beam

The low-energy muons could appear in the beam after the target due to the in-flight \( \pi \to \mu \nu \) decay of the punch-through 160 GeV pions in the region between the tracker stations ST3 and ST4. In this case the pion could mimic the primary muon, while the decay muon could be taken as a fake scattered muon. Taking into account that the admixture of the pion in the beam is at the level \( P_\pi = \pi/\mu \lesssim 10^{-6} \) [49] and the probability for the pion to decay at a distance of 4 m between the two spectrometers \( P_d \sim 5 \times 10^{-4} \) results in an overall expected background at the level of \( 5 \times 10^{-10} \) per incoming muon. To suppress this background further, one can use a cut that requires the maximal scattered muon energy to be below the minimal kinematically allowed decay muon energy \( E_{\mu, \text{min}}^\nu \approx 86 \text{ GeV} \). The combined probability for the \( \pi \to \mu \nu \) decay background is then \( P_\pi P_{\text{dec}} P_{\text{cut}} \lesssim 10^{-12} \).

![Figure 28](image)

**Figure 28.** Simulated distribution of energy deposited in the HCAL2 by \( 10^6 \), 160 GeV muons and \( 10^3 \), 160 GeV pions.

In order to estimate and cross-check the above \( \pi/\mu \) ratio, we have performed simulations and measurements of the hadron contamination in the M2 muon beam.
in 2017, which are describe below. The method used to estimate the ratio is based on the analysis of the far region of the energy loss spectra of 160 GeV muons which is expected to have an excess of events due to beam hadron absorption in the hadronic calorimeter installed in the M2 beam. More precisely, it is the following. In Fig.28 the simulated distributions of the energy deposited by 160 GeV muons and pion beam is shown. The integral probability of the fractional energy loss \( f = \Delta E_\mu / E_\mu \gtrsim 0.9 \) for the HCAL module of about 7.5\( \lambda_0 \) is \( \simeq 5 \times 10^{-4} \) according to simulations. This probability can be also well estimated by using extrapolation of the muon energy deposition spectrum in the HCAL2 shown in Fig.28 from the low energy side region to the hadron signal region as shown in Fig. 29. In this case the number of true muon events with energy deposition in the range corresponding to the signal from hadronic events can be well predicted from the measurements itself. Therefore, any excess event in the hadron region will signal on the presence of hadron admixture in the beam. In our estimate, the low energy region from 60 to 120 GeV was used to predict the number of true muon events in the energy region 140 -180 GeV, defining as a hadron signal region. The simulation are performed with Geant4, assuming a resolution of \( \sigma / E \simeq 25\% / \sqrt{E (\text{GeV})} \) for electromagnetic showers, and \( \sigma / E \simeq 60\% / \sqrt{E (\text{GeV})} \) for hadronic shower. The muon beam was simulated with the Gaussian energy spread \( \pm 10 \text{ GeV} \), assuming that its momentum is measured with a good precision by the SM2. The hadronic energy deposited in the HCAL2 by muon with energy \( E_\mu \) was normalised to the average beam energy of 160 GeV, assuming no significant difference in hadronic shower development for hadrons in the narrow energy range 140 - 180 GeV. We first looked at the energy distribution for the muon energy deposition in the region \( 60 \lesssim E_{hc} \lesssim 120 \text{ GeV} \) regions and estimated the contamination level into the hadron signal range \( 140 \lesssim E_{hc} \lesssim 180 \text{ GeV} \) by fitting the HCAL energy distribution shape with exponential function \( f(E_{hc}) = \)

![Figure 29. The sum of two simulated distribution of energy deposited in the HCAL2 by 10^6, 160 GeV muons and 10^3, 160 GeV pions shown above 60 GeV. The red curve shows single exponential fit to the spectrum, extrapolated from the true muon region 60-120 GeV to the hadron signal region 140-180 GeV allowing to predict the number of background muon events there.](image)
\( \exp[p_1 + p_2 \times E_{hc}(GeV)] \), where \( p_1 \) is a constant, and \( p_2 \) is the slope, as shown in Fig. 29. As a result, from this extrapolation the expected number of \( 514 \pm 26.7 \) muon events in the hadron signal region \( 140 < E_{hc} < 180 \) GeV have been found and shown in Fig. 29. Thus number agrees well with the simulated 492 events in this region. Taking 3 \( \sigma \) level for the observation of an excess event, the precision of the measurements of the hadron admixture is scaled down as, roughly \( \sim 6 \times 10^{-5}/\sqrt{n_{\mu}/10^6} \). Therefore for the number of muon \( n_{\mu} \gtrsim 10^8 \) the sensitivity to the hadron contamination is expected to be at the level a few \( 10^{-6} \), which is suitable for our goals.

**Figure 30.** Left panel: Distribution of energy deposited in the COMPASS HCAL from the 160 GeV muon beam taken with 6 absorbers in the 2017 test run. Right panel: End point distributions of energy deposited in the COMPASS HCAL from the 160 GeV muon beam taken with 6 and 9 absorbers and 160 GeV pions in the 2017 test run.

In 2017 on agreement with the COMPASS Collaboration, we irradiated the HCAL2 of the COMPASS detector with the 160 GeV muon beam and accumulated a data sample with a few \( 10^8 \) muon events. For the purpose of the measurement, we used the following steps:

- Removal of Hydrogen target from the beam.
- Calibration of a few HCAL2 modules with the 160 GeV hadron beam with SM2 ON.
- Accumulation of a few \( 10^8 \) 160 GeV muons with SM2 ON with Beam Trigger and air target.

The distributions of energy deposited in the COMPASS HCAL2 from the 160 GeV muon beam taken with 6 and 9 absorbers in the 2017 test run are shown in Fig. 30. To define the hadron contamination in the muon beam we performed the fit of the amplitude spectrum in hadron calorimeter module using extended likelihood method (R.Barlow, C.Beeston, Comp.Phys.Comm. V. 77 219 (1993) ). For the 6-absorbers case for the beam composition we got an estimate for the fraction \( f \) of particles: \( f_{\mu} = 0.99987 \pm 0.18905 \times 10^{-3}, f_{\pi} = (0.98715 \pm 0.13659) \times 10^{-4} \). As a cross-check, the 6-absorber distribution shape was also analysed with the previously described
method, by fitting it to the sum of the 9-absorber shape plus 160 GeV pion HCAL distribution shape obtained directly from the calibration data. This results in the estimate of the pion contamination in the beam $R(\pi/\mu) \simeq (0.97 \pm 0.14) \times 10^{-4}$ for 6 absorbers. (See, the NA64++ talk, PBC workshop, November, 2017). The both methods work reliably at the level $f_\pi \simeq 10^{-5}$. Finally, the first estimate of the hadron contamination in the M2 beam for the 9-absorbers case results in the $\pi/\mu$ ratio which is expected to be at the level $\pi/\mu \simeq 10^{-6}$. More accurate measurements of the hadronic admixture are planned in the NA64$\mu$ pilot run. We also plan to use more detailed fit of the data with a more accurately measured HCAL response function in the muon energy range 100-160 GeV.

4.8.5 The detector hermeticity

The fake signature of Eq.(4.1) could also arise when a high-energy muon loses energy through hard bremsstrahlung (BR), knock-on electrons (KN), pair-production (PP) or photonuclear (PN) interactions in the target. The fraction of these reactions, compared with the total muon energy losses including ionization losses, depends on the ratio $E'_\mu/E_\mu$ and for the Pb target is in the range $\simeq 10^{-3} - 10^{-5}$ per $X_0$ for $0.1 \lesssim E'_\mu/E_\mu \lesssim 0.9$ [59]. Such reactions could yield a low-energy scattered muon accompanied by neutral penetrating particles in the final state (e.g. photons, neutrons, $K^0_L$, etc.), which then could escape detection in the rest of the detector. Simulations show that in this case, the background is dominated by the photonuclear reactions accompanied by the emission of hadrons or a leading hadron $h$ from the muon-induced reactions $\mu A \rightarrow \mu h X$ which could escape detection due to incomplete hermeticity of the HCAL. For the energy range discussed the muon photonuclear cross section is $\sigma_{PN}(\mu N \rightarrow \mu X) \simeq 10^{-2}\sigma_{tot}$ of the total interaction cross section $\sigma_{tot} = \sigma_{BR} + \sigma_{KN} + \sigma_{PP} + \sigma_{PN}$ [59].

In Fig. 31 the 2D distribution of the scattered muon energy after the target vs the total detected energy $E_{tot}$ registered in the calorimeters (ECAL + HCAL) is shown for the SM events for the single MBPL option. Extrapolation of the background to the signal region gives an estimate of its level of $\lesssim 10^{-7}$ and $\lesssim 10^{-11}$ for event selected without (right panel) and with (left panel) the cut on the VHCAL signal, respectively. A clear difference between these event distributions is seen, thus illustrating the need for the use of the veto system for the background suppression. The two-MBPL case is illustrated in Fig. 32, where the similar 2D distribution in the plane ($E_{tot}$ vs $E'_\mu$), is shown for the SM events. Extrapolation of the background to the signal region gives an estimate of its level of $\lesssim 10^{-7}$ and $\lesssim 10^{-10}$ for event selected without (right panel) and with (left panel) the cut on the $V_{21}$ calorimeter signal, respectively, while the cut on the VHCAL signal is always on. A difference between the event distributions illustrating the need for the use of the veto system for the background suppression is clearly seen. The detector hermeticity for the two-MBPL option is worser because of the smaller event acceptance due to increased distance between the target and the
Figure 31. The distributions of the events from the SM muon interactions in the ECAL target in the plane \( (E_{\text{tot}} \text{ vs } E_{\mu}') \), i.e. the total detected energy \( E_{\text{tot}} \) registered in the calorimeters (ECAL + HCAL) vs the scattered muon energy after the target for the one-MBPL case. The events are selected without \( V_2 \) cuts (left panel), and with the cuts on the \( V_2 \) (right panel). The beam momentum is 160 GeV, the total number of simulated events corresponds to \( \gtrsim 10^7 \) MOT.

Figure 32. The distributions of the events from the SM muon interactions in the ECAL target in the plane \( (E_{\text{tot}} \text{ vs } E_{\mu}') \), i.e. the total detected energy \( E_{\text{tot}} \) registered in the calorimeters (ECAL + HCAL) vs the scattered muon energy after the target for the two-MBPL case. The events are selected without VHCAL cuts (left panel), and with the cuts on the VHCAL (right panel). The beam momentum is 160 GeV, the total number of simulated events corresponds to \( \gtrsim 10^7 \) MOT.
Figure 33. The l.h.s. shows the simulated distribution of the energy deposited in the ECAL+HCAL by secondaries from the bremsstrahlung, pair-production, knock-on and photonuclear muon interactions in the target. The events are selected by requiring the presence of a scattered muon crossing the central HCAL cells with initial energy $E'_\mu \lesssim 100$ GeV and no energy deposition in the VHCAL for the single MBPL case. The energy $(E_{ECAL} + E_{HCAL})$ deposited by the scattered muon in the ECAL and the HCAL central cells is subtracted. On the r.h.s. the same distribution (dots) is shown on a logarithmic scale. Other plots correspond to the energy distribution in the ECAL+HCAL for the HCAL with half of the lateral size, i.e., $60\times60$ cm$^2$ (triangles) and $150\times150$ cm$^2$ (open circles). The curves are the fit of the low-energy tail of the distributions by a smooth polynomial function extrapolated to the signal region $E_{tot} = E_{ECAL} + E_{HCAL} - E_{ECAL} - E_{HCAL} \lesssim 12$ GeV (MIP energy deposited in the HCAL1), indicated by the arrow, in order to conservatively evaluate the expected number of background events. The importance of the HCAL transverse size for the minimization of the lateral leak of the energy and the reduction of the number of background events is clearly seen.

HCAL. Adding the second VHCAL calorimeter (catcher) between the MBPL2-1 and MBPL2-2 (not shown in Fig. 4) improve HCAL hermeticity to the level $\lesssim 10^{-11}$.

The presented above results are based on the study of the background sources due to the HCAL non-hermeticity, which we conditionally divided into two categories: i) lateral energy leak, and ii) longitudinal energy leak. In Fig. 33 we show the simulated distribution of the energy deposited in the (ECAL+HCAL0) by secondaries from the bremsstrahlung, knock-on, pair-production or photonuclear muon interactions in the target for the single MBPL case. The events are selected by requiring additionally the presence of a scattered muon crossing the central HCAL0 cells with an energy $E'_\mu \lesssim 100$ GeV and no energy deposition in the VHCAL. The energy deposited by the scattered muon in the ECAL and the HCAL0 central cell is subtracted. The low-energy tail of this distribution was fitted by a smooth polynomial function and extrapolated to the energy region to evaluate the number of background events in the signal region. Those events with an energy deposition below the typical
Figure 34. Expected distributions of energy deposited by $\simeq 10^6 K^0$ with energy $\simeq 95$ GeV in two (a) and four (b) consecutive HCAL modules. The peak at zero energy in spectrum (a) is due to the punch-through neutral kaons.

energy deposited by the MIP could mix with the muon signal resulting in the fake signal. In the same plot the distribution of the energy in the ECAL+HCAL0 for the HCAL0 with half the lateral size, i.e. $60 \times 60$ cm$^2$ and $150 \times 150$ cm$^2$ are shown for comparison. We also tried to answer the question what should be the transverse size of the HCAL0 to optimise its acceptance for the the secondary particles in particular for the two-MBPL case. The effect of the HCAL0 transverse size on the lateral leak of the energy deposition from the bremsstrahlung, pair-production, knock-on or photo-nuclear muon interactions in the target and the corresponding number of background events is clearly seen on the right panel in Fig. 33. Using this rough estimate we find that for the transverse HCAL size increased up to $150 \times 150$ cm$^2$ the background level can be reduced down to $\lesssim 10^{-12}$ per incoming muon, which however substantially increase the cost of the project. Finally, the HCAL transverse size of $120 \times (60 - 80)$ cm$^2$ was chosen, resulting in the background level $\lesssim 10^{-11}$ for both one- and two-MBPL options.

Further improvement of the detector hermeticity is possible by using, in addition to the larger transverse size HCAL calorimeter, an algorithm for the subtraction of the outgoing muon energy deposition in the HCAL modules. The later results in an effective reduction of the energy threshold for the variable $E_{tot}$ shown in the right panel in Fig. 33. For that purpose, the central cell of the HCAL0 modules (see Fig.4) should be represented by a matrix with $3 \times 3$ cells each of $\simeq 70 \times 70$ mm$^2$ with its own signal photoreadout.

Another important question remains: What is the contribution to the estimated above background level from the possible longitudinal HCAL nonhermeticity, e.g. due the punch-through effect which could result in large missing energy in the detector. Indeed, the energy could leak, e.g. when the leading neutron or $K^0_L$ punches
through the HCAL without depositing energy above a certain threshold $E_{th}$. In this case, if the sum of the energy released in the HCAL is below $E_{th}$, the event is considered as a "zero-energy" event. The punch-through probability $P_{pth}$ is defined roughly by $P_{pth} \approx \exp(-L_{tot}/\lambda_{int})$, where $L_{tot}$ is the HCAL length. As discussed previously, it can be suppressed by using the HCAL with a thickness of $\approx 30\lambda_{int}$, resulting in a $P_{pth}$ of $\ll 10^{-10}$. This value should be multiplied by a conservative factor $\lesssim 10^{-4}$, which is the probability of a single leading hadron production in the target, resulting in the final estimate of $\lesssim 10^{-13}$ for the level of this background per incoming muon. For completeness, the HCAL0 nonhermeticity and corresponding background for high-energy secondary hadrons were cross-checked with GEANT4-based simulations in the following way. In Fig. 34 we show the expected distributions of energy deposited by $\approx 10^6 K_0$ with energy $\approx 95$ GeV in two (a) and four (b) consecutive HCAL0 modules. The peak at zero energy in the spectrum (a) is due to the punch-through neutral kaons, while for the full HCAL thickness there are no such missing energy events in distribution (b). Due to the difference in the nuclear interaction length for the neutral $K$-mesons and neutrons, we have also simulated another sample of $\approx 10^7$ 160 GeV neutrons events in the HCAL0. For this sample, the low-energy tail in the distribution of energy deposited by neutrons was fitted by a smooth polynomial function and extrapolated to the lower-energy region in order to evaluate the number of events below a certain threshold $E_{th}$. This procedure results in an estimate for the HCAL nonhermeticity, defined as the ratio of the number of events below the threshold $E_{th}$ to the total number of incoming particles: $\eta = n(E < E_{th})/n_{tot}$. For the energy threshold $E_{th} \approx 0.1$ GeV the nonhermeticity is expected to be at the level $\eta \lesssim 10^{-9}$. Taking into account a probability of producing a single leading hadron per incoming muon of $P_h \lesssim 10^{-4}$, results in an overall background level from the longitudinal HCAL nonhermeticity of the order $\lesssim 10^{-13}$, in agreement with the previous rough estimate.

### 4.8.6 Dimuon events from the reaction $\mu^- Z \to \mu^- \mu^+ \mu^- Z$

The dimuon production by 100 GeV $e^-$ beam from the reaction

$$e^- Z \to e^- Z \gamma; \gamma \to \mu^+ \mu^- \quad (4.16)$$

has been previously studied by NA64e in order to estimate additional uncertainty in the $A'$ yield prediction, and to make a cross-check between a clean sample of observed and MC predicted $\mu^+ \mu^-$ events with the energy depositions $E_{ECAL} \lesssim 60-70$ GeV [84]. The number of dimuon events from $e^-$ (or $\mu$) reactions are both proportional to the square of the Pb nuclear form factor $F(q^2)$ and are sensitive to its shape. As the mass $(m_{A'} \approx m_{\mu})$ and $q^2$ $(q \approx m_{A'}^2/E_{A'} \approx m_{\mu}^2/E_{\mu})$ ranges for both reactions are similar, the observed difference can be interpreted as due to the accuracy of the dimuon yield calculation for heavy nuclei and, thus can be conservatively accounted for as
additional systematic uncertainty in $n_{A'}(\epsilon, m_{A'}, \Delta E_{A'})$. In that way, the performance of NA64 was monitored online, and the sensitivity of the experiment was gauged by recording the number of dimuon events reconstructed online from the two-muon trigger. These events were from reaction (4.16) in which high-energy bremsstrahlung photons converted in the reaction $\gamma Z \rightarrow \mu^+ \mu^-$ to two muons. The overall probability to produce dimuon pair in this way with $E_{ECAL} \lesssim 60$ GeV is $\lesssim 10^{-5}$. During the NA64 running time, over $10^5$ of these events were reconstructed using loose cuts on the presence of muons in the HCAL. The production mechanisms of the $A'$ and $\mu^+ \mu^-$ pair is different, however there are some similarities. The cross section in both cases is $\propto F(q^2)$ and the final state is in the same (sub-GeV) mass range. The dimuons can produce a background if the pair is poorly detected in the ECAL or HCAL, e.g. due to a very asymmetric energy distribution between the primary muons and muon pairs.

4.8.7 Other background sources

The fake signature of Eq. (3.1) could be due to the QED production of muon trident events, $\mu Z \rightarrow \mu Z \mu^+ \mu^-$, with asymmetrical muon momenta in the muon pair. In this case, the lower-energy muon could be poorly detected, and another one could admix to the scattered muon in the HCAL central cell. A preliminary simulation study shows that this background can be suppressed down to the $10^{-12}$ level, provided the inefficiency of veto counters $V_1$ and $V_2$ is below $10^{-4}$ and the two tracks separated by a distance $\simeq 1$ mm are resolved by the ST5-ST8 and ST9-ST12 trackers.

In Table 1 contributions from the all dominant background processes are summarized for a primary muon beam energy of 160 GeV. The total background is expected to be at the level $\lesssim 10^{-11} - 10^{10}$. The contribution of additional subdominant background sources (e.g., such as very asymmetric $\mu \rightarrow e\nu\nu$ decays accompanied by low-energy muon production in the HCAL by the decay electron, cosmic muons, etc.) is negligible. This means that up to $\simeq 10^{12} \mu^-$ accumulated events this search is expected to be background free.

4.9 Expected sensitivity for the scalar and vector cases

To estimate the expected sensitivities we used simulations of the process (3.1) occurring in the detector shown in Fig. 4. The calculations of the production rate and energy distributions of muons produced in the SM reactions in the target are based on the results of the publication [59]. The calculated fluxes and energy distributions of the scattered muons produced in the target are used to predict the number of signal events in the detector. For a given total number of primary muons $n_\mu$, the expected number of events from the reaction $\mu + Z \rightarrow \mu + Z + Z_\mu, Z_\mu \rightarrow \nu\nu$ occurring
Table 1. Expected contributions to the total level of background from different background sources estimated for a beam energy of 160 GeV (see text for details).

<table>
<thead>
<tr>
<th>Source of background</th>
<th>Expected level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impurity of incoming muons</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>Impurity of outgoing muons</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>Pion contamination (9 absorber case)</td>
<td>(&lt; 10^{-12})</td>
</tr>
<tr>
<td>HCAL lateral nonhermeticity</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>- single-MBPL case</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>- two- MBPL case with additional VHCAL</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>(\mu) induced reactions in the target</td>
<td>(&lt; 10^{-13})</td>
</tr>
<tr>
<td>(\mu) trident events</td>
<td>(&lt; 10^{-12})</td>
</tr>
<tr>
<td>Total (preliminary)</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>- single-MBPL case</td>
<td>(&lt; 10^{-11})</td>
</tr>
<tr>
<td>- two- MBPL case</td>
<td>(&lt; 10^{-11})</td>
</tr>
</tbody>
</table>

within the decay length \(L\) of the detector is given by

\[
n_{Z_{\mu}} = k n_{\mu} Br(Z_{\mu} \rightarrow \nu \nu) \frac{\rho N_A}{A} \int \frac{\sigma(\mu+Z \rightarrow \mu+Z+Z_{\mu})}{dx} d\zeta(M_{Z_{\mu}}) dx \tag{4.17}
\]

with \(d = 1\) for \(M_{Z_{\mu}} < 2m_{\mu}\), and \(d = \left[1 - \exp\left(-\frac{LM_{Z_{\mu}}}{\tau_{Z_{\mu}}}\right)\right]\) for \(M_{Z_{\mu}} > 2m_{\mu}\). Here, the coefficient \(k\) is a normalization factor that was tuned to obtain the total cross sections of meson production, \(P_{Z_{\mu}}\) and \(\tau_{Z_{\mu}}\) are the produced \(Z_{\mu}\) momentum and lifetime at rest, respectively, \(\zeta(M_{Z_{\mu}})\) is the overall signal reconstruction efficiency, \(\rho\) is the density of the target, \(L\) is the decay length in the detector, and \(N_A\) is the Avogadro number. In this estimate we neglect the scattered \(\mu\) interactions in the target, the momentum of the incoming muons is \(< p_{\mu} > \approx 160\) GeV, and the efficiency \(\zeta(M_{Z_{\mu}})\) is in the range \(\simeq 0.1 - 0.5\) for the masses \(1\) MeV \(< M_{Z_{\mu}} \leq 1\) GeV.

The obtained results can be used to impose constraints on the previously discussed coupling strength \(\alpha_{\mu}\) as a function of the \(Z_{\mu}\) mass. Using the relation \(n^{90\%}_{Z_{\mu}} > n_{Z_{\mu}}\), where \(n^{90\%}_{Z_{\mu}} (= 2.3\) events\) is the 90\% C.L. upper limit for the number of signal events and Eq. (4.17), one can determine the expected 90\% C.L. upper limits from the results of the proposed experiment, which are shown in Fig. 35 together with values of the coupling \(g_{V}\) required to explain the muon \(g_{\mu} - 2\) anomaly. Here, the coupling \(g_{V} = \epsilon_{\mu}\) introduced in Section 2 and hence, \(\alpha_{\mu} = g_{V}^2 / 4\pi\). These bounds are calculated for a scattered muon energy \(10 \lesssim E'_{\mu} \lesssim 100\) GeV and a total of \(10^{12}\) incident muons in the background-free case. Here we assume an exposure to the muon beam with a nominal rate is a few months.

The statistical limit on the sensitivity of the proposed experiment is mostly set
by the number of accumulated events. However, there is a limitation factor related to the HCAL signal duration ($\tau_{HCAL} \approx 100$ ns) resulting in a maximally allowed muon counting rate of $\lesssim 1/\tau_{HCAL} \approx 10^7 \mu^-/s$ in order to avoid significant loss of the signal efficiency due to the pileup effect. To evade this limitation, one could implement a special muon pileup removal algorithm to allow for high-efficiency reconstruction of the zero-energy signal properties and the shape in high muon pileup environments, and run the experiment at the muon beam rate $\approx 1/\tau_{HCAL} \approx 10^8 \mu^-/s$. Thus, in the background-free experiment one could expect a sensitivity in the process $\mu + Z \rightarrow \mu + Z + Z_\mu, Z_\mu \rightarrow \nu\nu$ that is even higher than those presented above, assuming an exposure to the high-intensity muon beam of a few months. In the case of the $Z_\mu$ signal observation, several methods could be used to cross-check the result. For instance, to test whether the signal is due to the HCAL non-hermeticity or not, one could perform measurements with different HCAL thicknesses, i.e., with one, two, three, and four consecutive HCAL modules. In this case the background level can be evaluated by extrapolating the results to an infinite HCAL thickness. To ensure that there is no additional background due to the HCAL transverse hermeticity one could perform measurements for different distances between the target and the HCAL. The evaluation of the signal and background could also be obtained from the

![Figure 35](image-url)  

**Figure 35.** Exclusion region in the ($m_{Z_\mu}, g_V$) plane that might be expected from the results of the proposed experiment in the 2021 run provided that the total number of good $\approx 10^{10}$ MOT is collected, and for $10^{12}$ incident muons at the energy $E_\mu = 160$ GeV. The red line represents the value of $\alpha_\mu$ required to explain the muon $(g-2)_\mu$ discrepancy as a function of the $Z_\mu$ mass.
results of measurements at different muon beam energies. Finally, we note that the presented analysis gives an illustrative order of magnitude for the sensitivity of the proposed experiment and may be strengthened by more detailed simulations of the experimental setup.

5 The Phase II physics program

The NA64µ collaboration is currently updating the full physics program (Phase II) of the experiment, which aims to start running after the LS3. The Phase II program includes searches for the invisible decay $A' \rightarrow \text{invisible}$ of dark photons and sub-GeV Dark Matter [13, 14, 85], millicharged particles [86], and $\mu - \tau$ conversion in flight, which is under the preliminary developing stage [87].

5.1 Search for dark photon mediator: complementarity of NA64e and NA64µ

Among several renormalizable light DM extensions of the SM, the model with dark photon, where dark sector includes an abelian gauge field $A'_\mu$ (dark photon) is the most popular now. In these dark photon models, dark sector interacts with the SM particles only through nonzero kinetic mixing of the ordinary photon and dark photon, $-\frac{\epsilon}{2} F'_{\mu\nu} F^\mu\nu$. In renormalizable models the DM particles interacting with the $A'$ have spin 0 or 1/2. Spin 1/2 DM particles can be Majorana or pseudo-Dirac particles. The annihilation cross section for scalar or Majorana DM has $p$-wave suppression that allows to escape the GMB bound [18] while for Dirac fermions the annihilation cross section is $s$-wave that contradicts to the GMB bound. For the model with pseudo Dirac fermions it is also possible to avoid the GMB bound.

Let us consider, as an example, charged scalar dark matter interacting with dark photons. The charged dark matter field $\phi_d$ interaction with the $A'$ dark photon field is

$$L_{\phi Z'} = (\partial^\mu \phi - ie_d Z'^\mu\phi)^* (\partial_\mu \phi - ie_d Z'_\mu \phi) - m_{DM}^2 \phi^* \phi - \lambda_\phi (\phi^* \phi)^2$$  \hspace{0.5cm} (5.1)

The nonrelativistic DM annihilation cross section $\phi_d \bar{\phi}_d \rightarrow e^- e^+$ has the form\(^2\)

$$\sigma_{an} v_{\text{rel}} = \frac{8\pi}{3} \frac{\epsilon^2 \alpha_D m_{DM}^2 v_{\text{rel}}^2}{(m_{A'}^2 - 4m_{DM}^2)^2}.$$  \hspace{0.5cm} (5.2)

Here $\alpha_D = \frac{e_D^2}{4\pi}$ is an analog of the fine-structure constant $\alpha = 1/137$ for the DM particles interacting with DM photon. We shall use standard assumption that in the hot early Universe DM is in equilibrium with ordinary matter, see e.g. Ref. [1]. During the Universe expansion the temperature decreases and at some $T_d$ the thermal

---

\(^2\)Here we consider the case $m_{A'} > 2m_{DM}, m_{A'} \gg m_e$. 

- 47 -
decoupling of the DM starts to work. Namely, at some freeze-out temperature $T_d$ the cross-section of annihilation $\text{DM particles} \rightarrow \text{SM particles}$ becomes too small to obey the equilibrium of DM particles with the SM particles and DM decouples. The experimental data are in favour of scenario with cold relic for which the freeze-out temperature $T_d$ is much lower than the mass of the DM particle. In other words DM particles decouple in the non-relativistic regime. The value of the DM annihilation cross section at the decoupling temperature determines the value of the current DM density in the Universe. Too big annihilation cross section leads to small DM density and vice versa too small annihilation cross section leads to DM overproduction. The observed value of the DM density $\rho_{DM} \approx 0.23$ ( here $\rho_{DM}, \rho_c$ is the dark matter density, and the total energy density of the Universe, respectively) allows to estimate the DM annihilation cross-section into the SM particles and hence to estimate the discovery potential of light dark matter both in direct underground and accelerator experiments. Very crude estimate for the DM annihilation cross section is

$$<\sigma_{\text{ann}} v_{\text{rel}} > = O(1) \text{ pb} \cdot \text{c}.$$ (5.3)

As a consequence of the formulae (2,3) for fixed values $m_{A'}$ and $m_{DM}$ we can estimate the product $\epsilon^2 \alpha_D$. Note that fixed target NA64 experiment [83, 84] uses the reaction of the dark photon electroproduction on nuclei allowing to obtain only upper bounds on $\epsilon^2$ vs $m_{A'}$. Therefore, to test the prediction for the $\epsilon^2 \alpha_D$ we have to know either the $\alpha_D$ value or at least an upper bound $\alpha_D \leq \alpha_o$ on the $\alpha_D$. The arguments based on the use of the renormalization group and the assumption of the absence of the Landau pole singularity up to some scale $\Lambda$ allow to obtain upper limit on the coupling constant $\alpha_D$ [25] in the formula (2) for the annihilation cross section. The bound on $\alpha_D$ depends on the scale $\Lambda$ logarithmically. Moreover, the scale $\Lambda$ has to be larger than 1 TeV [25]. So for fixed values of $m_{A'}$ and $m_{DM}$ the knowledge of the upper bound on $\alpha_D$ together with the requirement that the dark photon model correctly reproduces the observed Dark matter density allows to obtain lower bound on $\epsilon^2$ as a function of $m_{A'}$ or $m_{DM}$.

5.1.1 NA64$\mu$ projections for the $\gamma - A'$ mixing strength

The NA64$\mu$ experiment is proposed to search for dark sector particles weakly coupled to the muon [13], which could explain the muon $(g-2)_\mu$ anomaly [2]. One of the good examples of such a particle, is a new light vector $L_{\mu} - L_{\tau} Z'$ boson, which interacts predominantly with the $L_{\mu} - L_{\tau}$ current, [13]$^3$. Interestingly, the $Z'$ could also serve as a new leptophilic mediator of dark force between the ordinary and dark matter, which is charged with respect to $U(1)_{L_{\mu}-L_{\tau}}$, and explain the relic DM abundance [13, 15]. Another interesting possibility involves muon-specific scalar mediator which

$^3$One loop corrections lead to nonzero interactions with electron, and other quarks and leptons [13]
could connect the visible and dark sectors and also account for the \((g-2)_{\mu}\) anomaly [16, 17].

The NA64\(\mu\) plans to perform a sensitive search for \(L_{\mu} - L_{\tau} Z'\) as a mediator of sub-GeV dark matter particle \((\chi)\) production in missing energy events from the reaction of 100-160 GeV muon scattering on heavy nuclei:

\[
\mu^- + Z \rightarrow \mu^- + Z + Z'; Z' \rightarrow \nu\nu, \chi\chi \tag{5.4}
\]

at the CERN SPS [13].

In the \(A'\) models the interaction of dark photon with the leptons and quarks is given by \(L_{A'} = e\epsilon A'_\mu J_{SM}^\mu\). Here, \(J_{SM}^\mu\) is the electromagnetic current. So, we see that muons and electrons interact with the dark photon universaly, with the same coupling constant. Hence, similar to the reaction \(e^- Z \rightarrow e^- ZA'; A' \rightarrow invisible\), the dark photons will be also produced in the reaction of Eq.(5.4) with the same experimental signature of the missing energy. For the \(A'\) mass region \(m_{A'} \gg m_e\),

![Diagram](image.png)

**Figure 36.** The NA64e 90% C.L. current [84] and expected exclusion bounds obtained with \(4.3 \times 10^{10}\) EOT and \(5 \cdot 10^{12}\) EOT, respectively, in the \((m_{A'}, \epsilon)\) plane. The NA64\(\mu\) projected bounds calculated for \(n_{MOT} = 5 \cdot 10^{12}\) and \(5 \cdot 10^{13}\) are also shown [85].

the total cross section of the dark photon electroproduction \(eZ \rightarrow eZA'\) scales as \(\sigma_{A'} \sim \epsilon_e^2/m_{A'}^2\). On the other hand, for the dark photon masses, \(m_{A'} \lesssim m_\mu\), the similar \(\mu Z \rightarrow \mu ZA'\) cross section can be approximated in the bremsstrahlung-like limit as \(\sigma_{A'} \sim \epsilon_\mu^2/m_\mu^2\). Let us now compare expected sensitivities of the \(A'\) searches with NA64e and NA64\(\mu\) experiments for the same number \(\approx 5 \times 10^{12}\) particles on target. Assuming the same signal efficiency the number of \(A'\) produced by the 100
GeV electron and muon beam can approximated, respectively, as follows

\[ N_{A'}^e \approx n_{EOT} L^e \sigma_{A'}^e, \quad N_{A'}^\mu \approx n_{MOT} L^\mu \sigma_{A'}^\mu, \]

(5.5)

where \( L^e \simeq X_0 \) and \( L^\mu \simeq 40X_0 \) are the typical distances that are passed by an electron and muon, respectively, before producing the \( A' \) with the energy \( E_{A'} \gtrsim 50 \) GeV in the NA64 active Pb target of the total thickness of \( \simeq 40 \) radiation length \((X_0)\) [13, 84]. The detail comparison of the calculated \( A' \) sensitivities of NA64e and NA64\( \mu \) is shown in Fig.36, where the 90\% C.L. limits on the mixing \( \epsilon \) are shown for different number of particles on target for both the NA64e and NA64\( \mu \) experiments. The limits were obtained for the background free case by using exact-tree-level (ETL) cross-sections rather than the Weizsacker-Williams (WW) ones calculated for NA64e in Ref.[54], and for the NA64\( \mu \) case in the work of Ref.[85]. The later are shown in Fig. 37 as a function of \( E_{A'}/E_\mu \) for the Pb target and mixing value \( \epsilon = 1 \). One can see that in a wide range of masses, 20 MeV \( \lesssim m_{A'} \lesssim 1 \) GeV, the total WW cross-sections are larger by a factor \( \simeq 2 \) compared to the ETL ones. As the result, the typical limits on \( \epsilon \) for the ETL case are worse by about a factor \( \simeq 1.4 \) compared to the WW case. One can see that, e.g. for \( n_{EOT} = n_{MOT} = 5 \cdot 10^{12} \) the sensitivity ob NA64e is enhanced for the mass range \( m_e \ll m_{A'} \simeq 100 \) MeV . While for the \( A' \) masses \( m_{A'} \gtrsim 100 \) MeV NA64\( \mu \) allows to obtain more stringent limits on \( \epsilon^2 \) compared to NA64e.

5.1.2 Combined LTM sensitivity of NA64e and NA64\( \mu \)

The reported previously limits on the \( \gamma - A' \) mixing strength, allow us to set the combined NA64e and NA64\( \mu \) constraints on the LDM models, which are shown in

![Figure 37. Cross-section of Dark Photon production by muons as a function of \( x = E_{A'}/E_\mu \) for various masses \( m_{A'} \) and \( \epsilon = 1 \). Solid lines represent ETL cross-sections and dashed lines show the cross-sections calculated in WW approach [54, 85, 86].](image)
Figure 38. The NA64e current (solid) [84] and 90% C.L. expected (dashed light blue) exclusion bounds for $5 \times 10^{11}$ and $5 \times 10^{12}$ EOT in the $(m_\chi, y)$ and $(m_\chi, \alpha_D)$ planes. The combined limits from NA64e and NA64\(\mu\) are shown for $10^{13}$ EOT plus $2 \times 10^{13}$ MOT (dashed green), and $10^{13}$ EOT plus $5 \times 10^{13}$ MOT (dashed dark blue), respectively. The limits are calculated for $\alpha_D = 0.1$ and 0.005 and assuming $m_{A'} = 3m_\chi$. The results are also shown in comparison with bounds obtained from the results of the LSND, E137, BaBar and MiniBooNE experiments. For explanation of the limit curves, see Ref.[1].

the $(y; m_\chi)$ plane in Fig.38. As discussed in Sec. I, as a result of the $\gamma - A'$ mixing the cross-section of DM particle annihilation into SM particles, which determines the relic DM density, is proportional to $\epsilon^2$. Hence using constraints on the cross section of the DM annihilation freeze out (resulting in Eq.(5.2), and obtained limits on mixing strength of Fig. 36 one can derive constraints in the $(y; m_\chi)$ plane, which can also be used to restrict models predicting existence of LTDM for the masses $m_\chi \lesssim 1$ GeV. Here, the variable $y = \epsilon^2 \alpha_D (m_\chi/m_{A'})^4$.

These limits obtained from the data sample of the 2016 run [84] and expected from the run after the LS2 are shown in the top panels of Fig. 38 together with
combined limits from NA64 and NA64µ for $10^{13}$ EOT and $10^{13}$ MOT (dashed green curve) and $10^{13}$ EOT and $2 \times 10^{13}$ MOT (dashed dark blue curve), respectively. The favoured parameter curves for scalar, pseudo-Dirac (with a small splitting) and Majorana scenario of LDM taking into account the observed relic DM density. The limits are calculated by using Eq.(5) from Ref.[84] under the assumption $\alpha_D = 0.1, 0.005$, and $m_{A'} = 3 m_\chi$, here $m_\chi$ stands for the LDM particle’s masses, either scalars or fermions. The plot shows also the comparison of our results with limits from other experiments. The choice of $\alpha_D = 0.1$ is based on arguments for the running of the dark gauge coupling, presented in Sec. I. It should be noted that the $\chi$-yield in the NA64 case scales as $\epsilon^2$, not as $\epsilon^4 \alpha_D$. Therefore, for sufficiently small values of $\alpha_D$ the NA64 limits will be much stronger. This is illustrated in the upper right panel of Fig. 38, where the NA64 limits are shown for $\alpha_D = 0.005$. One can see, that for this, or smaller, values of $\alpha_D$, the direct search for LDM in NA64 excludes models of LDM production via vector mediator for the full mass region $m_\chi \lesssim 0.05$ GeV. While being combined with the NA64µ limit, the NA64 results for the coupling $\alpha_D = 0.1$ exclude the models for the entire mass region $m_\chi \lesssim 1$ GeV.

The upper bounds on $\epsilon$ also allow to obtain lower bounds on coupling constant $\alpha_D$ by using a relation among the parameters derived from the requirement of the thermal freeze-out of DM annihilation into visible matter through $\gamma - A'$ kinetic mixing [25]:

$$\alpha_D \simeq 0.02 f \left( \frac{10^{-3}}{\epsilon} \right)^2 \left( \frac{m_{A'}}{100 \, \text{MeV}} \right)^4 \left( \frac{10 \, \text{MeV}}{m_\chi} \right)^2 (5.6)$$

which are shown in lower panels of Fig. 38 in the $(\alpha_D;m_\chi)$ plane. For the case of pseudo-Dirac fermions and small splitting, the limits in the lower left panel of Fig. 38 were calculated by taking the value $f = 0.25$. One can see that for the full mass range $m_\chi \lesssim 0.05$ GeV the obtained combined NA64e and NA64µ bounds are more stringent than the limits obtained from the results of NA64e allowing to probe almost the full parameter space. The limits for the Majorana case shown in the lower right panel of Fig. 38 were calculated by setting $f = 3$, see [84]. Similar to pseudo-Dirac case the the combined NA64e and NA64µ limits exclude the full remained parameter area. Note, that new constraints for the large pseudo-Dirac fermion splitting can also be derived. They will be more stringent than for the case of the small splitting and similar to the one obtained for the Majorana case.

5.2 Constraints on scalar and $L_\mu - L_\tau$ sub-GeV Dark Matter

Models with light dark matter ($m_{DM} \leq 1$ GeV) can be classified by the spins and masses of the dark matter particles and mediator. The scalar dark matter mediator models are severely restricted or even excluded by non-observation of rare B-meson decays so we consider here only the case of vector mediator. The most popular vector mediator model is the model with additional massive photon $A'$ which couples with dark matter particles via interaction $L_I = g_D A'_\mu J_D^\mu$ with coupling constant $g_D(\alpha_D =
The mixing $L_{mix} = -\frac{1}{2} \epsilon F^{\nu \nu} F_{\mu \nu}$ between photon field $A_\mu$ and dark photon field $A'_\mu$ leads to nonzero interaction of dark photon $A'_\mu$ with the electrically charges SM particles with the charges $e = e e_{SM}$. So as a result of mixing the annihilation cross-section of DM particles is proportional to $\epsilon^2$. The dark matter annihilation cross-section into SM particles determines the dark matter density. Consider at first the case of scalar dark matter. For $m_{A'} \geq m_\phi$ the rate of annihilation $\phi \phi^* \rightarrow f \bar{f}$ determines the relic density. For $m_e \leq m_\phi \leq m_\mu$ the main annihilation channel is into electron-positron pair $\phi \phi^* \rightarrow e^- e^+$ determines the relic density. Neglecting the $m_f$ mass, the tree level cross section at relative velocity $v_{rel} \ll c$ is

$$\sigma v_{rel} = \frac{8 \pi \alpha_D m_\phi^2 v_{rel}^2}{3 (m_{A'}^2 - 4 m_\phi^2)^2 + m_{A'}^2 \Gamma^2}, \quad (5.7)$$

where $\Gamma$ is the $A'$ width. In the limit $m_{A'} \gg m_\phi, \Gamma$, this cross-section depends on dark-sector parameters only through the dark matter mass $m_\phi$ and the dimensionless variable

$$y \equiv \frac{\epsilon^2 \alpha_D (m_\phi / m_{A'})^4}, \quad (5.8)$$

---

The currents $J_{DM}^\mu = \bar{\psi}_{DM} \gamma^\mu \psi_{DM}$ and $J_{DM}^\mu = i(\phi_{DM} \partial^\mu \phi_{DM} - \phi_{DM} \partial^\mu \phi_{DM})$ for spin 1/2 and 0.

---

**Figure 39.** Parameter space for a muon-philic scalar $S$ (left) or vector $V$ (right) particle as described in Ref.[13, 15–17]. Left panel: the expected sensitivity for the search for dark scalar $S$ with the fermion coupling $L_S \supset g_S f \bar{f}$ in the experiments NA64 $\mu$ [13] and $M^3$ [15] Right panel: the projected sensitivity to probe vector mediator with SM fermion coupling $L_V \supset g_v V_\mu f \gamma_\mu f$. It is assimed that both the $S$ and $V$ decay predominantly invisibly. The green bands represent the parameter space for which such particles can reconcile the $(g-2)_\mu$ anomaly to within 2$\sigma$ of the measured value.
For fermion dark matter $\phi$ with vector interaction $L_I = \epsilon_D \bar{\phi} \gamma^\mu \phi A'_\mu$, the dark matter annihilation cross section is

$$\sigma_{v_{rel}} = \frac{8\pi \epsilon^2 \alpha_D \alpha_D m^2_{\phi}}{(m^2_{A'} - 4m^2_{\phi})^2 + m^2_{A'} \Gamma^2}.$$  \hspace{0.5cm} (5.9)

Note that the absence of $v_{rel}^2$ factor is due to the fact that the annihilation of fermions takes place in the s-wave. Numerically for $v_{rel} \sim 1/3$ the annihilation cross-section for scalar dark matter is suppressed by factor $\sim 25$ in comparison with fermion dark matter (for the same mass and $\alpha_D, \epsilon$). For the axial-vector interaction we have p-wave suppressed annihilation with $\sigma_{v_{rel}} \sim v_{rel}^2$ as for the scalar case. If the global $U(1)$ symmetry under which the Weyl components $\phi_{1,2}$ of Dirac fermion $\phi = (\phi_1, \phi_2)$ have opposite charges is broken (by a Higgs field that gives mass to the $A'$) the interaction $L_{break} = \delta \phi_1 \phi_2$ yield mass eigenstates $\phi_{\pm} = \frac{1}{\sqrt{2}} (\phi_1 \pm \phi_2)$ split in mass by $\delta$. This corresponds to the inelastic or pseudo-dirac scenario. Analogously inelastic interaction can also arise in scalar case. The dark matter annihilations freeze out before the era of recombinations, however residual annihilations can re-ionize hydrogen and distort the high $l$-CMB power spectrum. The data on the high $l$-CMB power spectrum exclude thermal-relic Dirac fermion dark matter, but not other scenarios like scalar dark matter in which p-wave suppression of annihilation at late times leads to more weak bound on the parameter $y$. Also this bound becomes weak for inelastic dark matter. So scalar dark matter or inelastic fermion dark matter

![Figure 40](image.png)

**Figure 40.** Parameter space for predictive thermal DM charged under $U(1)_{L_\mu - L_\tau}$, for DM charges near the pertubativity limit (left), based on the running of the dark gauge coupling arguments Ref. [25], or smaller such that the $(g-2)_\mu$ region overlaps with the thermal relic curves (right) [15]. Here the relic abundance arises through direct annihilation to SM particles via $s$ -channel $Z_\mu$ exchange. Left panel represents enhanced DM coupling $g_\chi = 1$ to vector mediator $Z_\mu$, and right panel corresponds to suppressed DM coupling $g_\chi = 5 \cdot 10^{-2}$. 


models survive. From the equations (5.8,5.9) we can estimate the value of \( y \). Namely, for fermion dark matter we find

\[
y = \frac{< \sigma v >}{8 \pi \alpha} \cdot \left( 1 - \frac{4m_{\psi}^2}{m_A^2} \right)^2 \quad (5.10)
\]

Numerically, for \( < \sigma v > = 0.3 \cdot 10^{-8} \text{GeV}^{-2} \) we find that

\[
y = 1.6 \cdot 10^{-12} (m_\psi/10 \text{ MeV})^2 \cdot \left( 1 - \frac{4m_{\psi}^2}{m_A^2} \right)^2 \quad (5.11)
\]

For scalar dark matter the corresponding estimate reads

\[
y = 0.4 \cdot 10^{-10} (m_\psi/10 \text{ MeV})^2 \cdot \left( 1 - \frac{4m_{\psi}^2}{m_A^2} \right)^2 \quad (5.12)
\]

In Fig. 39 and 40 the NA64 constraints on the scalar \((g_S)\) and vector \((g_V)\) couplings to muon as well as limits on the variable \( y \) are shown. The limits are calculated with the simulation described in details in Ref.[13] assuming background free case and 50\% efficiency for the signal.

### 5.3 Search for millicharged particles

The mechanism of the electric charge quantization remains a fundamental puzzle of the particle physics. In the context of grand unification models this mechanism may be linked to magnetic monopole, which quantizes the electric charge of the fermions [64]. However, the magnetic monopoles have not been observed yet, and the underlying mechanism for charge quantization remains unknown.

The particles with fractional electric charge have been proposed in various scenarios of the Standard Model extension. In particular, for the Dark Photon scenario in the Holdom phase [65, 66], the mixing between hidden sector and SM is described by a term \( \frac{i}{2} F'_{\mu\nu} F^{\mu\nu} \). After a redefinition of the hidden vector field \( A'_\mu \rightarrow A'_\mu + \epsilon A_\mu \) one obtains the electromagnetic-like interactions \((A'_\mu + \epsilon A_\mu) J'^{\mu}_D \) with \( J'^{\mu}_D \) being hidden current of fermions, \( J'^{\mu}_D = g_D \chi \gamma^{\mu} \chi \). This means that the millicharged particles interact with the photon via the effective coupling \( Q_\chi = \epsilon g_D \). In this scenario the Dark Photon remains massless and interacts only with dark sector particles, thus the emission of the hidden fermions by \( A' \) is forbidden. Nevertheless, the millicharged particles can be effectively produced via electromagnetic process. Furthermore, the possible existence of the particles with fractional electric charge has stimulated interest in the investigation of missing momentum processes [67]

\[
e^+ e^- \rightarrow \gamma \chi \bar{\chi}, \quad \mu N \rightarrow \mu N \chi \bar{\chi}, \quad (5.13)
\]

We emphasize that in the leading order the millicharge production rate of these processes is scaled as \( Q_\chi^2 \). These millicharged particles could easily escape the detection
in experiment, which unable to observe them directly. The Lagrangian of $\chi$ can be written as follows

$$L \supset i \bar{\chi} \gamma^\mu \partial_\mu \chi - m_\chi \bar{\chi} \chi + Q_\chi A_\mu \bar{\chi} \gamma^\mu \chi$$  \hspace{1cm} (5.14)$$

here we suppose that $Q_\chi$ is a small fraction of electron charge, $Q_\chi/e \ll 1$.

This muon missing momentum experiments [13] and the one recently proposed at FNAL [15, 16], both aiming at the sensitive probe of the $(g - 2)_\mu$ anomaly [70] and light dark sector [82]. Moreover, the beyond collider experimental activity has been stimulated widely during last years due to the LHC null result in the field of new physics. We emphasize, that LHC is insensitive to probe sub-GeV dark sector scenario [72]. In particular, the millicharge parameter space in the ranges $0.1 \mathrm{GeV} \lesssim m_\chi \lesssim 1 \mathrm{GeV}$ and $10^{-4} \lesssim Q_\chi/e \lesssim 10^{-3}$ has not been constrained yet by dedicated experiments. On the other hand, these beyond standard model scenarios of sub-GeV hidden particles can be probed at SHIP proton beam dump facility as well as at MiniBoone, DUNE and LSND neutrino detectors. In particular, due to the significant meson flux in these experiments the dominant millicharge production signatures are the missing energy decays

$$\pi^0/\eta \rightarrow \gamma \chi \bar{\chi}, \quad J/\psi, \ U \rightarrow \chi \bar{\chi}. \hspace{1cm} (5.15)$$

Once being produced in meson decay, the millicharges elastically scatter on an atomic electrons in the dump, $\chi e \rightarrow \chi e$. So that, the detection of millicharges is based on measuring low electron energy recoils. However, the millicharge signal yield from hadrons is suppressed by both the production term $\sim Q_\chi^2$ and the interaction factor $\sim Q_\chi^2$, such that $N_{\chi \chi} \sim Q_\chi^4$. On the other hand, muon missing momentum signature of $\chi \bar{\chi}$ production at NA64$\mu$ scales as $\sim Q_\chi^2$. Therefore, a significant gains in millicharge sensitivity may be achieved by optimizing the active target design of NA64$\mu$ even for fewer events accumulated on target. In particular, $10^{13}$ muons on target are expected to be accumulated at NA64$\mu$ during couple of months running. For comparison, one assumes to accumulate $10^{20}$ POT and $10^{22}$ POT at SHIP and DUNE respectively; LSND and MiniBoone collected about $10^{23}$ POT and $10^{21}$ POT respectively. So the millicharge constraints for these neutrino fixed target experiments are performed in [73] by taking into account the background, the detector acceptance and efficiency of the regarded facilities.

In addition, we note that unlike the electron beam-dump NA64$e$ setup, the muon beam-dump missing energy facility of NA64$\mu$ has increased sensitivity to the millicharge searches. Indeed, 100-GeV electron beam degrades significantly even in the relatively thin lead target of $40X_0$ ($\approx 20 \mathrm{cm}$). Therefore, the electron missing momentum yield is suppressed by the electron beam struggling factor $X_0/(40X_0)$, such that $N_{\chi \chi} \sim X_0$. On the other hand, the relativistic 100 GeV muons pass through the dump minimally interacting with the target material. This implies, that millicharge production signal in the muon missing energy process scales as the length

– 56 –
of the target, \(N_{\chi} \sim 40X_0\). Moreover, one can improve the millicharge sensitivity for the muon beam by varying the effective interacting length of the active lead target in the range \(40X_0-80X_0\). This provides an excellent opportunity for NA64\(\mu\) to probe wider range of millicharge parameter space.

### 5.3.1 Missing energy signature of millicharged particles at NA64

![Feynman diagrams of milli-charge pair production]

**Figure 41.** Feynman diagrams of milli-charge pair production

In this section we calculate the differential cross-section of the missing energy process for the NA64 facility with the active target and incident muon beam

\[
\mu N \rightarrow \mu N\gamma^* \rightarrow \mu N\chi\bar{\chi}
\]  

(5.16)

with \(\chi\) being a milli-charged Dirac fermions from hidden \(U(1)_D\) sector. The most interesting signature for NA64 is the milli-charge emission process (5.16), which scales at tree level as \(O(\alpha^3Q_\chi^2)\). We neglect \(O(\alpha^2Q_\chi^4)\) trident milli-charge production signature and relevant \(O(\alpha^{5/2}Q_\chi^3)\) interference terms in our calculation. The underlying processes are shown in Fig. 41. Given that one can factorize the current tensor of the milli-charged particles \(\chi_{\alpha\beta}\) in the regarding cross-section

\[
d\sigma_{\mu N \rightarrow \mu N\chi\bar{\chi}} = d\text{Lips}_{2\rightarrow 3}|\mathcal{M}_{2\rightarrow 3}|^2 \frac{dk_{\gamma^*}^2}{(2\pi)} \times \chi_{\alpha\beta}
\]  

(5.17)

where \(d\text{Lips}_{2\rightarrow 3}\) is Lorentz invariant phase space for a process \(\mu N \rightarrow \mu N\gamma^*\) with off-shell photon in the final state. The external \(\gamma^*\)-line is replaced by the photon propagator in the amplitude relevant to this process, such that

\[
|\mathcal{M}_{2\rightarrow 3}|^2_{\alpha\beta} = \sum_{\text{span}} \mathcal{M}^\mu \mathcal{M}^{\nu\dagger} g_{\mu\alpha} g_{\nu\beta} \frac{1}{k_{\gamma^*}^4},
\]  

(5.18)
where an averaging over initial muon spin and summation over outgoing muon state is performed. The milli-charged tensor $\chi_{\alpha\beta}$ has the following form

$$
\chi_{\alpha\beta} = \int \frac{d^3k_1}{(2\pi)^3 2E_1} \frac{d^3k_2}{(2\pi)^3 2E_2} (2\pi)^4 \delta(4) (k - k_1 - k_2) \sum_{+,-} j_{\alpha} j_{\beta},
$$

(5.19)

with $j_{\alpha} = Q_{\chi} \bar{\chi} \gamma_\alpha \chi$ being a milli-charged current, $k \equiv k_{\gamma^*}$ is a total four-momentum of the milli-charged pair, $k_{\gamma^*} = k_1 + k_2$. The integration over the variables of the millicharged pair in (5.19) can be performed analytically. One easily finds [74]

$$
\chi_{\alpha\beta} = \frac{k_{\gamma^*}^2}{6\pi} \sqrt{1 - \frac{4m_{\chi}^2}{k_{\gamma^*}^2}} \left( 1 + \frac{2m_{\chi}^2}{k_{\gamma^*}^2} \right) \times \left( -g_{\alpha\beta} + \frac{k_{\alpha} k_{\beta}}{k_{\gamma^*}^2} \right).
$$

(5.20)

We note that term $\sim k_{\alpha} k_{\beta}$ in (5.20) vanishes after contracting with (5.18). This means that millicharge production cross-section has the following form

$$
d\sigma_{\mu N \rightarrow \mu N \bar{\chi}} = d\sigma_{\mu N \rightarrow \mu N \gamma^*} \times \frac{Q_{\chi}^2}{12\pi^2} \frac{d\sigma_{\mu N \rightarrow \mu N \gamma^*}}{dk_{\gamma^*}^2} \sqrt{1 - \frac{4m_{\chi}^2}{k_{\gamma^*}^2}} \left( 1 + \frac{2m_{\chi}^2}{k_{\gamma^*}^2} \right).
$$

(5.21)

The first factor in (5.21) can be calculated in the equivalent photon approach [75, 76],

![Figure 42](image_url)

**Figure 42.** Upper limits on the fractional electric charge $Q_{\chi}/e$ of the hypothetical millilicharged fermions of mass $m_{\chi}$. The areas with the grey shading are the bounds excluded by SLAC [77], collider [78, 79] and EDGES experiment [80, 81]. The projected limits are shown by solid lines. In particular, the expected reaches for SHIP and MilliQuan are taken from [73]. The sensitivity of LDMX is based on MadGraph missing momentum simulation with 16 GeV electron beam on aluminium target [82].
corresponding differential cross-section is
\[
\frac{d}{dx}\sigma_{2\to3} \approx \frac{2}{3}\alpha^3\zeta|k_{\gamma^*}| \left[ m^2_\mu x(-2 + 2x + x^2) - 2(3 - 3x + x^2)\tilde{u} \right], \quad x = E_{\gamma^*}/E_0
\]
(5.22)

with \( \zeta \) being the photon flux from nucleus
\[
\zeta = \frac{k^2_{\gamma^*} + m^2_\mu}{k^2_{\gamma^*}/(4E^2_0)} \cdot \left| k_{\gamma^*} \right| = \left( x^2 E^2_0 - k^2_{\gamma^*} \right)^{1/2},
\]
(5.23)

where \( a = 111Z^{-1/3}/m_e \) parametrizes the electron screening effect and \( d = 0.164A^{-2/3} \) GeV\(^2\) stands to account the finite nuclear size. Such form-factor parametrization (5.23) accounts for elastic scattering effects only. The inelastic form-factor is proportional to \( \sim Z \) and thus can be neglected in high-\( Z \) target experiment. The quantities \( \tilde{u} \) and \( |k_{\gamma^*}| \) in (5.22) are defined by
\[
\tilde{u} = -\frac{k^2_{\gamma^*}(1 - x)}{x - m^2_\mu x} \quad \text{and} \quad |k_{\gamma^*}| = \left( x^2 E^2_0 - k^2_{\gamma^*} \right)^{1/2},
\]
respectively. Therefore, one can estimate the \( \chi\bar{\chi} \)-production rate by integrating (5.21) over \( \gamma^* \) invariant mass
\[
\sigma_{\mu N \to \mu N\chi\bar{\chi}} \approx \frac{4}{3} \frac{\alpha^3 \zeta}{(2m_\chi)^2} \ln \left[ \frac{1}{2} \left( \frac{2m_\chi}{m_\mu} \right)^2 + \frac{11}{12} \right] \times \frac{Q^2_\chi}{12\pi^2}, \quad \kappa = 0.8.
\]
(5.25)

with a reasonable accuracy. In addition, we note that (5.25) generally resembles, up to the numerical factor \( \sim Q^2_\chi \) and additive correction to the logarithm, the total cross-section of Dark Photon (DP) production [76], in which the DP mass is redefined as \( m_{A'} \to 2m_\chi \). This observation allows one to estimate the expected constraints for the parameter space of the millicharged particles directly in NA64\( \mu \) and \( M_3 \) muon missing momentum experiments.

5.3.2 Millicharge reach estimate

Using the above cross-section, one can estimate the millicharge yield for given muon beam and target properties. As the experimental signatures of the \( Z_\mu \) and the millicharged particles are identical, in simulations we use the Phase 1 setup shown in Fig. 4 and the experimental criteria for the signal candidates selection as described
in Section 4.2. Namely, the muon beam energy is 100 GeV and the total number of accumulated muons is about $N_{MOT} = 5 \cdot 10^{13}$ MOT. We consider a single shashlyk ECAL cell as target with the thickness of $L_T = 40X_0$.

The millicharge yield which originates from the elementary $\mu N \rightarrow \mu N\chi\chi$ process may be expressed as

$$N_{\chi\chi} = N_{MOT} \times \frac{\rho \times N_A}{A} \times L_T \times \sigma_{\chi\chi},$$

(5.26)

where $A$ is the atomic weight, $N_A$ is Avogadro’s number, $\rho$ denotes the target density and $\sigma_{\chi\chi}$ is the the cross-section of milli-charged pair production discussed above. For the sake of simplicity, we neglect muon stopping lost in the lead target. This seems to be reasonable, because the muon energy struggling reported in [16] is rather small for the beam energy range, $\langle dE_\mu/dz \rangle \approx 12.7 \cdot 10^{-3}$ GeV/cm. In addition, the muon missing energy signature $\mu N \rightarrow \mu N\overline{\chi}\chi$ can be used for the signal process of millicharge pair production at fixed target experiment, as long as an initial and final states of muon are measured accurately. In NA64$\mu$ facility one assumes to utilize two, upstream and downstream, magnetic spectrometers allowing for precise measurements of momenta for incident and recoiled muons, respectively [13].

Integrating numerically (5.24), we obtain the sensitivity in fraction charge for the background free case. In particular, we require $N_{\chi\chi} > 2.3$ to obtain 95\%CL exclusion limit on $Q_\chi/e$ and $m_\chi$ from (5.26). In Fig. 42 we show the expected reach of NA64$\mu$ and NA64$e$ detectors for $N_{MOT} = 5 \cdot 10^{13}$ and $N_{EOT} = 5 \cdot 10^{12}$ respectively, the corresponding beam energy is $E_0 = 100$ GeV for both $e$- and $\mu$- modes.

It instructive to compare millicharge limits for muon and electron beam in order to show that the bound from muon setup is enhanced at $m_\chi \gtrsim m_\mu$. Indeed, the ratio of the reaches can be naively approximated as follows

$$\frac{Q_{\chi}^{(e)}}{Q_{\chi}^{(\mu)}} \approx \left( \frac{L_{eff}^{(e)}}{L_{eff}^{(\mu)}} \cdot \frac{\sigma_{\chi\chi}^{(e)}}{\sigma_{\chi\chi}^{(\mu)}} \cdot \frac{N_{MOT}}{N_{EOT}} \right)^{1/2},$$

(5.27)

where labels $(\mu/e)$ specify a beam type and $L_{eff}$ is the effective length of millicharge lepto-production in the lead target, namely $L_{eff}^{(\mu)} \approx 40X_0$ and $L_{eff}^{(e)} \approx X_0$. The latter means that the millicharges are essentially produced in a few radiation of the target because of the large electron energy loss. While the muons produce millicharges uniformly over the whole length of the target. For beam energy $E_0 = 100$ GeV and millicharge masses $m_\chi \gtrsim 200$ MeV the electron and muon cross-sections scale respectively as [76]

$$\sigma_{\chi\chi}^{(e)} \sim \frac{1}{(2m_\chi)^2} \left( \ln \frac{1}{2} \left[ \frac{E_0}{2m_\chi} \right]^2 + \mathcal{O}(1) \right), \quad \sigma_{\chi\chi}^{(\mu)} \sim \frac{1}{(2m_\mu)^2} \left( \ln \frac{1}{2} \left[ \frac{2m_\chi}{m_\mu} \right]^2 + \mathcal{O}(1) \right),$$

(5.28)
where factor $1/2$ under the logarithms comes from the integration of production cross-section over the missing energy range, $1/2 < E_{\gamma^*}/E_0 < 1$, here the left bound corresponds to $E_{\text{miss}} \equiv E_{\gamma^*} > 50$ GeV for $E_0 = 100$ GeV. Finally for $N_{\text{MOT}} = 5 \cdot 10^{13}$ and $N_{\text{EOT}} = 5 \cdot 10^{12}$ one gets $Q_{\chi}^{(e)}/Q_{\chi}^{(\mu)} \approx 9$ at $m_\chi = 200$ MeV. This result can be illustrated from Fig. 42. Indeed, for $m_\chi = 200$ MeV we have $Q_{\chi}^{(e)}/e \approx 10^{-2}$ and $Q_{\chi}^{(\mu)}/e \approx 10^{-3}$. One can also estimate the fraction (5.27) for the relatively light millicharges $m_\chi \ll m_\mu$. In this case the muons produce the millicharges in the bremsstrahlung-like limit. Such that the regarding cross-sections for both electron and muon beam can be naively approximated as

\[
\sigma_{\chi\chi}^{(e)} \sim \frac{1}{(2m_\chi)^2} \left( \ln \frac{1}{2} \left[ \frac{2m_\chi}{m_e} \right]^2 + \mathcal{O}(1) \right), \quad \sigma_{\chi\chi}^{(\mu)} \sim \frac{1}{m_\mu^2} \left( \ln \frac{1}{2} \left[ \frac{m_\mu}{2m_\chi} \right]^2 + \mathcal{O}(1) \right). \tag{5.29}
\]

Finally one gets $Q_{\chi}^{(e)} \gtrsim Q_{\chi}^{(\mu)}$ from (5.27) and (5.29) for $m_\chi \gtrsim 2$ MeV. This is illustrated in Fig. 42, such that the electron and muon curves approach closely each other in this mass region for the benchmark incident fluxes $N_{\text{EOT}} = 5 \cdot 10^{12}$ and $N_{\text{MOT}} = 5 \cdot 10^{13}$.

6 The 2021 pilot run

6.1 The main goals of the run

Our main goals for the one week pre-beam time and two weeks of the 2021 pilot run are:

(i) assembly and commissioning of the beamline and the detector

(ii) study and optimisation of the M2 beam and beam halo.

(iii) study of the trigger and the accidental trigger rate.

(iv) study of the background level from the hadron contamination in the M2 beam at different beam energies.

(v) hermeticity: the study of the background level with at least $\lesssim 10^9$ MOT.

(vi) first data taking with $\simeq 5 \times 10^6 \mu$/spill

(vii) analysis of the data sample

It should be noted, that even with $\lesssim 10^{10}$ collected $\mu^-$s, which could be collected during a few days of running, we can already set new constraints for the $Z_\mu \to \text{invisible}$ decay mode, and explore a part of the muon $(\text{g-2})_\mu$ favoured parameter space. These results will be very timely and of great importance given that the new results of the FNAL experiment on $(\text{g-2})_\mu$ can support or even strengthen the anomaly.
6.2 The requirements for the run

For the initial stage of the experiment, the proposed plan is to be setup 45 m upstream of the entrance to the COMPASS experiment at the position of the present CEDARS.

(i) Since the experiment is extremely sensitive to the incoming beam energy it will be defined by the Beam Momentum station (BMS) consisting of three 5 m vertical bending magnet of 3.3 T.m each which is currently used by COMPASS, in addition to a 2 m MBPL magnet of 140 mm aperture (MS1).

(ii) For MS1 a MBPL magnet currently placed downstream in the M2 beamline (Bend 7) at Z=1100.228 m, according to the M2 co-ordinate system, shall be moved around 20 m upstream at Z = 1080.722 m.

(iii) Two additional MBPL magnets shall be placed 3 m downstream of that for defining the outgoing beam (MS2), which is crucial for the signal detection.

(iv) The beam requirements for the physics run of the experiment is consistent with the M2 beam capabilities of a typical maximal intensity for $\simeq 160$ GeV of the order of $10^7 \mu^-/sec$ with $10^{12}$ protons on target per SPS spill. The requirement of a parallel beam is also met by the M2 beam of size of the order of $\sigma_x \sim \sigma_y \sim 20$ mm with a angular divergence of $\sigma_x' \sim \sigma_y' \sim 0.2$ mrad at the detector position as studied by the EN-EA-LE Group [68].

(v) Additional $e^-$, $\pi$ and $\mu$ beams of energies between 80 and 160 GeV/c shall be required for detector commissioning and checking their response. More precisely, we would need

- an electron beam for the electromagnetic calorimeter calibration, the existing 40-50 GeV electron beam suits well for this purpose.
- a hadron beam for the hadron calorimeters calibration. The beam energy could be in the range $\simeq 50 - 160$ GeV.

(vi) For the pilot run the experiment also requires a control room for data taking.

(vii) The trackers used during the run requires gas mixture Ar-CO$_2$ (93-7% pre-mixed) at a flow rate of 3-5 l/hr. For this reason gas panels shall be installed around the experimental setup for 15-20 tracker modules.

(viii) Patch panel for high voltage and signal cables and ethernet connection shall also be installed for the DAQ of the experiment. Details of the DAQ system is given in Section 4.7.

(ix) For the ECAL calibration a DESY table shall be installed.

(x) A calibration table for the HCAL or scanning with the beam shall be requested for HCAL cell calibration.
Table 2. The required number of MOT.

<table>
<thead>
<tr>
<th>Main goal</th>
<th>Energy, GeV</th>
<th>(\epsilon_{\text{tot}} \cdot n_{\text{MOT}})</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>search for the (Z_\mu \rightarrow \text{invisible}) decay</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1. Coverage of the ((g - 2)_\mu) parameter space</td>
<td>100-160</td>
<td>(\approx 10^{10})</td>
</tr>
<tr>
<td>Phase 2. Coverage of the Dark Matter models parameter space</td>
<td>100-160</td>
<td>(\approx 10^{12} - 10^{13})</td>
</tr>
<tr>
<td><strong>search for millicharged particles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improvement of limits in the mass range (m_{mQ} \gtrsim 100) MeV</td>
<td>100-160</td>
<td>(\gtrsim 10^{13})</td>
</tr>
</tbody>
</table>

6.3 Beam time request

We anticipate requiring one week for the initial commissioning tasks - such as moving the detector into position at the M2 beam line in front of COMPASS, connecting and testing cables, electronics and the various sub-detectors. At this time we would need to contact the NA transport group regarding using the NA Hall crane to move detectors into the M2 experimental area. This will require various safety and technical documentation which will be provided by us if necessary.

In addition to one week for pre-beam commissioning, we require two weeks of beam time - one week for the detector commissioning and one week for the first physics run. During these two-weeks beam period the detector assembly, DAQ and corresponding electronics, the realistic testing of sub-detectors and their response to \(e^-, \pi\) and \(\mu\) of different energies (80 - 160 GeV/c), and the measurements related to detector performance, noise level, etc... will be performed with the goal to debug and test the whole setup. The first preliminary results on the search for the \(Z_\mu \rightarrow \text{invisible}\) decay mode could also be expected. During these three weeks a good quality M2 beam cannot be expected downstream for any experiment since we will be located in the upstream area of the beam line.

We could request to be the first user of the M2 beam in 2021 - i.e. we could begin commissioning the detector before the scheduled start of beam time. However, as we plan also to have the first run at the new location at the H4 beam, it would be helpful to combine both beam period into a single one towards the end of the 2021 SPS beam run. This will allow using the required time to commission the detector fully, without wasting beam time, resources and otherwise impacting on the COMPASS program and other users. The three-weeks test run during the year 2021 will be followed by another \(\sim 1.5\) months physics run in 2022.
For the Phase 2 of the signal search, a more detailed study of the apparatus design might be required. The experience and knowledge gathered in the first phase of this experiment will certainly help to improve the design and the sensitivity of the experiment. The experimental Phase 2 after the LS3 stop aims to perform a more sensitive search for the $Z_{\mu}$, provided that the pilot-run and Phase I show encouraging results. It would also require the average muon rate of $\gtrsim 5 \times 10^7$ per SPS spill. The full running time of the proposed measurements is estimated to be 12 months, and the total number of accumulated good MOT including efficiency $\epsilon_{tot} \approx 0.3$ is estimated to be $\approx 2 \times 10^{13}$ MOT, which is enough to achieve the main goals of the experiment summarised in Table 2.

### 6.4 Possible scenarios for the 2021 run

From our previous experience on the $A' \rightarrow invisible$ decay search at H4 line, it is clear that detailed and careful detector and background studies and understanding are crucial for the sensitive searching for the reaction $\mu Z \rightarrow \mu Z Z_{\mu}$. If the background level in the 2021 data sample motivates more study and running we would propose additional running time at M2 to investigate the issue.

To this end we would like to point out that the three week beam request is the very lower limit required for our run and is extremely optimistic with minimum interference with the COMPASS beam time. It does not take into account any possible beam loss time or unforeseen detector issues that one may encounter. For example, to have one more week of running would allow us to have the first collected data set preliminary analysed before starting the data taking for the final result analysis. Therefore, a total of four-six week beam time would be a more realistic option and provide additional insurance for any encountered issues during the beam time period. However, if the four-six week option is not possible taking into account the COMPASS beam run, even with the three week beam time we believe the 2021 run data sample will be extremely useful in order to perform the successful search for the $Z_{\mu}$ as an explanation of the muon g-2 anomaly with complete coverage of the $Z_{\mu}$ parameter space in the near future.

### 7 The detector cost, manpower, and schedule

A breakdown of the cost estimates for the items described earlier and the institutional responsibilities for detector construction are shown in Table 3. Although the cost estimate is based on our experience in NA64e running at the H4 electron beam, the estimate is still very preliminary and will be updated over time. Considering the importance of the 2021 run for the success of the NA64$\mu$ project we tried for these costs to be reasonably justified. In order to complete the NA64$\mu$ detector construction in time for the 2021 run, timely receiving of funding is of critical importance. Significant fraction of the construction funds must be made available by the middle
Table 3. Cost projection and tentative responsibilities for NA64 experiment on muon beam in the 2018-2020.

<table>
<thead>
<tr>
<th>Item</th>
<th>Institute</th>
<th>Invest, kCHF 2019</th>
<th>Invest, kCHF 2020</th>
<th>Invest, kCHF total</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM-calorimeter</td>
<td>IHEP, INR</td>
<td></td>
<td></td>
<td></td>
<td>from NA64e</td>
</tr>
<tr>
<td>BMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CERN, Compass</td>
</tr>
<tr>
<td>Hadron calorimeter</td>
<td>IHEP, INR, JINR,</td>
<td>85</td>
<td>90</td>
<td>175</td>
<td>new</td>
</tr>
<tr>
<td></td>
<td>UTFSM, ETH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hadron modules</td>
<td>IHEP, INR</td>
<td></td>
<td></td>
<td></td>
<td>from NA64e</td>
</tr>
<tr>
<td>Veto hadron calorimeter</td>
<td>IHEP, INR, JINR,</td>
<td>45</td>
<td>50</td>
<td>95</td>
<td>new</td>
</tr>
<tr>
<td></td>
<td>UTFSM, ETH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scintillator hodoscopes</td>
<td>UTFSM, IHEP, INR</td>
<td>35</td>
<td>35</td>
<td>70</td>
<td>3 hodoscopes from NA64e</td>
</tr>
<tr>
<td>8 GEM station</td>
<td>Bonn</td>
<td>40</td>
<td>40</td>
<td>80</td>
<td>4 GEM from NA64e</td>
</tr>
<tr>
<td>8 MM station</td>
<td>ETH</td>
<td>40</td>
<td>40</td>
<td>80</td>
<td>6 MM from NA64e</td>
</tr>
<tr>
<td>Veto counters</td>
<td>IHEP, INR, JINR,</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>new</td>
</tr>
<tr>
<td></td>
<td>UTFSM, ETH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Straw</td>
<td>JINR</td>
<td>70</td>
<td>80</td>
<td>150</td>
<td>new</td>
</tr>
<tr>
<td>DAQ, electronics</td>
<td>All</td>
<td>70</td>
<td>70</td>
<td>140</td>
<td>part of electronics for calorimeters, MM, Straw and GEM from NA64e</td>
</tr>
<tr>
<td>M2 line</td>
<td>CERN</td>
<td>170</td>
<td>180</td>
<td>350</td>
<td>new</td>
</tr>
</tbody>
</table>

| Total                    |                   | 1160              |                   |                   |                               |

The list of equipments which can be used from NA64e and which has to be built for the 2021 run is shown in Table 3. The goals of the 2021 run outlined previously in Sec. 6, could be partially achieved even without using the full NA64μ detector. For example, the Veto HCAL, some of the HCAL modules, or tracker chambers, scintillator hodoscopes etc... might not be ready for the 2021 run. Instead the

Table 4. Milestones for the NA64μ 2021 run.

<table>
<thead>
<tr>
<th>date</th>
<th>milestone</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2019</td>
<td>Submission of the Proposal</td>
</tr>
<tr>
<td>May 2019</td>
<td>Begin design and upgrade of DAQ and electronics</td>
</tr>
<tr>
<td>June 2019</td>
<td>Complete design of the detector, beamline and infrastructure</td>
</tr>
<tr>
<td>July 2019</td>
<td>Begin construction of sub-detectors and electronics</td>
</tr>
<tr>
<td>July 2020</td>
<td>Complete components construction of calorimeters</td>
</tr>
<tr>
<td>August 2020</td>
<td>Deliver calorimeter modules and begin assembly at CERN</td>
</tr>
<tr>
<td>January 2021</td>
<td>Begin commissioning of sub-detectors</td>
</tr>
<tr>
<td>October 2021</td>
<td>Commissioning of the NA64μ detector</td>
</tr>
<tr>
<td>October 2021</td>
<td>First data taking 2021 run</td>
</tr>
</tbody>
</table>
corresponding detectors from the NA64e setup, e.g. the ECAL, a few modules of the HCAL and the tracker might be used allowing to reduce time for the commissioning and calibration of the NA64µ detectors. We believe the 2021 beam period would be more efficient if the NA64µ 2021 run is scheduled after the NA64 2021 run at H4 beam. It will allow us to optimise the funding for the presence of 25-30 people, including detector and DAQ experts, at CERN.

8 Conclusion

The availability of high-energy and -intensity muon beams at CERN SPS provides a unique opportunity to search for a new light gauge boson $Z_\mu$ with a predominant coupling to the second and third generations which could serve as a mediator of a new interaction between the visible and dark sector. If the $Z_\mu$ exists, it could be produced in the reaction $\mu + Z \rightarrow \mu + Z + Z_\mu$ and be observed by looking for events with a specific signature, namely those missing a large fraction of the beam energy in the detector. We propose an experiment NA64µ at the M2 muon beamline aiming at either discover or rule out the $Z_\mu$ in the near future. The experiment is based on the missing-energy approach developed for the searches for invisible decays of dark photons and (pseudo)scalar mesons at CERN [20–22] and is timely and complementary to these experiments. It also provides interesting motivations for further muon studies and fits well with the present muon physics program at CERN.

Moreover, the experiment will be able to check if the discrepancy between the measured and predicted values of the muon $(g-2)_\mu$ could be explained by the existence of the $Z_\mu$ predominantly coupled to the second and third generations. A feasibility study of the experimental setup shows that this specific signal of the $Z_\mu$ allows for searches for the $Z_\mu$ with a sensitivity in the coupling constant $\alpha_\mu \gtrsim 10^{-11}$, i.e., 3 orders of magnitude stronger than the value $\alpha_\mu \sim 10^{-8}$ explaining the 3.6$\sigma$ muon $g_\mu - 2$ discrepancy for the $Z_\mu$ mass range $M_{Z_\mu} < O(5)$ GeV [6].

The experiment has the capability for a sensitive search for dark scalars $S_\mu$s decaying invisibly to dark-sector particles, such as dark matter. Due to their specific properties, dark scalars $S_\mu$ are also an interesting probe of dark matter well motivated by physics beyond the standard model both from the theoretical and experimental point of view. We propose to perform an experiment dedicated to the sensitive search for dark $S_\mu$ in the still unexplored area of the coupling strength and masses $m_S \lesssim 200$ MeV by using $\approx 100$ GeV muon beams from the CERN SPS. If $S_\mu$s exist, their invisible decays $S_\mu \rightarrow$ invisible could be observed by looking for events with large energy deposition in the detector. Our feasibility study shows that a sensitivity for the search of the $S_\mu \rightarrow$ invisible decay mode in branching fraction $Br(S_\mu) = \frac{\sigma(\mu^- Z \rightarrow \mu^- ZS_\mu), S_\mu \rightarrow \text{invisible}}{\sigma(\mu^- Z \rightarrow \mu^- Z + \text{all})}$ at the level below a few parts in $10^{12}$ is in reach. The intrinsic background due to the presence of low energy muons in the beam can be suppressed by using a tagging system, which is based on the double measurements...
of incoming muons by the BMS and MS1 spectrometers. The search would allow to cover a significant fraction of the yet unexplored parameters space for the $S_\mu \rightarrow \text{invisible}$ decay mode.

These results could be obtained with a detector optimised for several of its properties, namely, i) the intensity and purity of the primary muon beam, ii) the high efficiency of the veto system, and iii) the high level of hermeticity for the downstream veto system and hadronic calorimeters. Availability of a large amount of high-energy muons and high background suppression are crucial for improving the sensitivity of the search. To obtain the best limits, the choice of the energy and intensity of the beam, as well as the background level should be compromised.

We believe the data sample collected in the 2021 pilot run will be extremely useful in order to perform the optimisation of the beam and detector components, background measurements and the first $Z_\mu$ coverage of the muon $(g-2)_\mu$ parameter space. The successful search for the $Z_\mu$ as an explanation of the muon g-2 anomaly with complete coverage of the $Z_\mu$ parameter space could be then performed in the near future by collecting $\sim 10^{10}$ MOT. In the second phase, after the LS3, the NA64$\mu$ goal is to reach the previously mentioned sensitivity to light dark matter, millicharged particles, and other rare processes, or better by exploiting a possible upgrade of the detector, which might be necessary given the results of Phase I with $\sim 10^{12} - 10^{13}$ MOT. In particular, we would like to underline the NA64 perspectives for discovery of sub-GeV dark matter by running the experiment in electron and muon modes at the CERN SPS. While with $5 \times 10^{12}$ EOT NA64e is able to test the scalar and Majorana LDM scenarios, the combined NA64e and NA64$\mu$ results with $\sim 10^{13}$ EOT and a few $10^{13}$ MOT, respectively, will allow to cover almost fully the parameter space of the all most interesting LDM models. This makes NA64e and NA64$\mu$ extremely complementary to each other, and greatly increases the NA64 discovery potential of sub-GeV DM.

Acknowledgments We gratefully acknowledge the support of the CERN management and staff and the technical staffs of the participating institutions for their vital contributions to this proposal preparation. This work was supported by the HISKP, University of Bonn (Germany), JINR (Dubna), MON, RAS and RF program “Nauka” (Contract No. 0.1764.GZB.2017) (Russia), ETH Zurich and SNSF Grant No. 169133 (Switzerland), and grants FONDECYT 1140471, 1150792, and 3170852, Ring ACT1406 and Basal FB0821 CONICYT (Chile). Part of the work on MC simulations was supported by the RSF grant 14-22-00161. We thank the COMPASS collaboration, in particular O. Denisov and G. Mallot, for helping us with the hadronic contamination measurements in the M2 beam, and the COMPASS DAQ group and the Institute for Hadronic Structure and Fundamental Symmetries of TU Munich for the technical support.
References


[34] B. Aubert et al. [BABAR collaboration], arXiv:0808.0017.


[48] J.D.Bjorken, R.Essig, P.Schuster and N.Toro, 
Phys. Rev. D 80,075028 (2009); 
S.Andreas, G.Niebuhr and A.Ringwald, Phys. Rev. D 86, 095019; 
[50] See, for example, http://sba.web.cern.ch/sba/ 
[56] G.S. Atoian, V.A. Gladyshev, S.N. Gninenko, V.V. Isakov, A.V. Kovzelev, et al., 
[68] D. Banerjee, J. Bernhard, Beam Studies for Physics Beyond Collider Projects at EHN2 (ATS Note under preparation)


