HAL open science

Testing the standard model via the electromagnetic coupling and going beyond with axion-mediated dark matter

Sophie Mutzel

▶ To cite this version:

Sophie Mutzel. Testing the standard model via the electromagnetic coupling and going beyond with axion-mediated dark matter. High Energy Physics - Lattice [hep-lat]. Aix-Marseille Université, 2023. English. NNT : . tel-04524810

HAL Id: tel-04524810

https://cnrs.hal.science/tel-04524810v1

Submitted on 28 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Aix*Marseille université
Socialement engagée

NNT : 2023AIXM0001

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 14 décembre 2023 par

Sophie MUTZEL

Testing the standard model via the electromagnetic coupling and going beyond with axion-mediated dark matter

Discipline
Physique et sciences de la matière

Spécialité
Physique Théorique et Mathématique

École doctorale
École doctorale 352 - Physique et sciences de la matière

Laboratoire/Partenaires de recherche
Centre de Physique Théorique (CPT)
UMR 7332
A*MIDEX
Institut Physique de l'Univers (iPhu)

Composition du jury

Geneviève BÉLANGER Directrice de recherche, CNRS, LAPTh, Annecy-le-Vieux	Rapporteure
Jonathan FLYNN Professeur, University of Southampton	Rapporteur, Président du jury
Christian HÖLBLING Professeur, Universität Wuppertal	Examinateur
Aoife BHARUCHA Chargée de recherche, CNRS, CPT Marseille	Directrice de thèse
Laurent LELLOUCH Directeur de recherche, CNRS, CPT Marseille	Co-directeur de thèse

CPT
A*Midex Initiative d'excellence Aix-Marseille
Institut Physique de l'Univers Aix*Marseille Université

So, you are keen on physics?

Aiveen Bharucha

Affidavit

I, undersigned, Sophie Mutzel, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Aoife Bharucha and Laurent Lellouch, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the french national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Marseille, October 18, 2023

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et participation aux conférences

Liste des publications réalisées dans le cadre du projet de thèse

Liste des publications avec comité de lecture:

1. A. Bharucha, F. Brümmer, N. Desai, and S. Mutzel. Axion-like particles as mediators for dark matter: beyond freeze-out. Journal of High Energy Physics, 02:141, 2023, `[arXiv:2209.03932 [hep-ph]]`

Liste des proceedings:

1. A. Bharucha, S.Mutzel. ALPs as dark matter mediators: beyond freeze-out. Published in: *PoS(DISCRETE2022)023*, avec lecture du comité de conférence
2. S. Mutzel for the BMW collaboration. Hadronic contributions to the running of α_{QED} from lattice QCD: challenges at short distances. Published in: *PoS LATTICE2022*, avec lecture du comité de conférence
3. S. Mutzel. Axion-like dark matter mediators. Contribution to the 2022 electroweak session of the 56th Rencontres de Moriond. Published in: *56th Rencontres de Moriond on Electroweak Interactions and Unified Theories*
4. G. Brooijmans et al. Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report, 2020. Published in *In 11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*

Participation aux conférences et écoles d'été au cours de la période de thèse:

1. **European Tensor Network school**, Abingdon, September 2023

2. **NePsi 2023***, New Physics Signals 2023, Pisa, February 2023
3. **39th international conference on lattice field theory (Lattice 2022)**$^{\dagger *}$, Bonn, August 2022
4. **58th Course on subnuclear physics***, Erice, June 2022
5. **GDR QCD annual workshop***, Saint-Pierre d'Oléron, May 2022
6. **56th Rencontres de Moriond conference**$^{\dagger *}$, La Thuile, March 2022
7. **GDR-InF annual workshop***, Paris, November 2021
8. **EuroPLEx Summer School on lattice field theory and applications**, August 2021, online
9. **Les Houches Summer School on dark matter**, École des Houches, France, July 2021
10. **16th Patras workshop on axions, WIMPs and WISPs***, online, June 2021
11. **Muon g-2 theory initiative workshop in memoriam Simon Eidelman**, online, June/July 2021
12. **The hadronic vacuum polarization from lattice QCD at high precision**, online, November 2019
13. **GDR-InF annual workshop**, Sommières, November 2019
14. **Mini-Workshop on search for Dark Matter**, Marseille, October 2019
15. **GDR-Intensity Frontier lectures: Determination of V_{ub}**, Paris, October 2019

* = Une présentation orale a été effectuée par l'auteur.

† = Un proceeding a été preparé par l'auteur.

Acknowledgements

First and foremost I thank my two advisors, Aoife Bharucha and Laurent Lellouch for their constant support during these four years. Thank you Aoife, for pushing me to grow scientifically, for unwaveringly guiding me in questions related and unrelated to physics, for considering me a colleague from the very first day and finally for being a friend. Thank you Laurent, for your trust, for always taking even the most basic question seriously, for staying in the institute until 8 pm and later to never leave a question unanswered, *bref*, for teaching me so much about physics during all the discussions, which I truly enjoyed. You both also made it possible for me to attend numerous international conferences and schools which I enjoyed enormously and where I found many new friends and hopefully future colleagues – in particular the four weeks long Les Houches summer school on dark matter which seemed like a resurrection after the Covid lockdowns. Finally, thank you both for all your input, work and corrections that went into this manuscript and simply for everything that you both have done for me.

I thank Geneviève Belanger and Jonathan Flynn for serving as a committee member and reviewer of this (too?) long thesis, and Christian Hölbling for being a member of my thesis committee.

Collaboration is the key to scientific advancement; no one person has all the answers (Max Planck). I would like to thank all members of the BMW collaboration, particularly Andrey Kotov, Kalman Szabo and Gen Wang for the work and discussions on the project discussed in the first part of this thesis. Andrey has provided the results on the charm contribution and has taken the heroic task of implementing the perturbative calculation of the running of α, which unfortunately cannot yet be presented here; and Gen cross-checked many steps in my calculations and provided some of the plots that went into this thesis. I thank my collaborators Felix Brümmer and Nishita Desai for their invaluable expertise on and their contributions to the project discussed in the second part of this thesis.

A big thanks goes to all the members of the particle physics group for their (not only scientific) support; especially Antoine Gerardin for answering many questions and Savvas Zafeiropoulos for his help with applications for postdoc positions. I warmly thank Leti(zia) for the great time we had together, but also for always wanting to understand every single step in a calculation and then explaining it to me. I thank Willem simply for all the (physics and non-physics) fun we had together and Malaka Eduardo for always making me laugh!

Science and everyday life cannot and should not be separated (Rosalind Franklin). I consider it a privilege to study at a place where coming to work means meeting (and working with) friends. Moving to the old library turned out to be extremely fruitful, not least in making many more friends, and even mathematicians! I would like to thank all the CPT and I2M PhD students and

postdocs who are too numerous to be all mentioned here. Thank you for all the rooftop breaks, barbecues and apéros, for the pleasantly endless hikes in the calanques, for the weekly seminars where we tried to teach you some real science and you tried to teach us some real maths, for the countless volleyball tournaments (Willem, we are unbeatable!), for the ski and camping trips and simply for your friendship. A special thanks goes to the members of the office of the *fragiles* – with whom I shared many laughters and a few tears – for all their support, especially Kenza, who has always been there for me during the last four years. *Vous allez me manquer.*

I thank all regular CPT and I2M users of the old library for creating an *atmosphère sympathique*, and the CPT administration, especially Stephanie Suciu, for forgiving me that I was often late with my paperwork.

Finally, I would like to thank my family: My parents, Helga and Rupert, for the way they raised me and for teaching me to be curious; Verena for being an unerring big sister to look up to and for being my best friend; Simon and now also Felix for always being there for me; my late Oma Sofie, for having taught me many (useful?) life lessons I try to apply and finally Antoine, whom I would have never met if he was not keen on physics, too.

Abstract

The electromagnetic coupling, α, is a cornerstone of the standard model (SM) of particle physics. If left unimproved, the precision of its value at the scale of the Z boson mass (M_Z) will limit the discovery potential of future collider experiments. This precision hinges on our understanding of the hadronic contribution to the running of α at low energies, a region where the SM's description is hampered by non-linear effects in quantum chromodynamics (QCD). In this study, we take a first step in reducing this uncertainty by calculating the hadronic running of α from the Thomson limit (at a scale $Q^2 = 0$) to spacelike $Q^2 = 8\,\mathrm{GeV}^2$. We employ large-scale simulations in lattice QCD, which are performed using $N_f = 2 + 1 + 1$ flavours of staggered quarks with masses tuned to their physical values, and which also account for leading-order isospin-breaking corrections. To improve control over systematic uncertainties, especially those associated with the necessary continuum extrapolations, we divide the running into sub-intervals of Q^2, optimizing our analyses within each one. In particular, we have developed and implemented a series of improvement strategies. These strategies are based on lattice perturbation theory at short distances, effective theories addressing taste breaking at larger distances, and the utilization of 48 discretizations of the charm electromagnetic current. Our results already match the level of uncertainty achieved via the reference data-driven approach, and we outline a roadmap for a significant reduction of these uncertainties in the future.

In the second part of this thesis we investigate a scenario to go beyond the standard model of particle physics by considering the possibility of an axion-like particle (ALP) acting as a mediator between SM particles and dark matter (DM) particles. Experimental advances have significantly constrained the standard electroweak-scale WIMP produced via thermal freeze-out. This has led to a paradigm shift towards alternative scenarios. We focus on the case where the couplings are too small to allow for DM generation via freeze-out and the DM is thermally decoupled from the SM particles. However, alternative DM genesis mechanisms like freeze-in and decoupled freeze-out (DFO) can still reproduce the observed DM relic density. Having determined the region of parameter space for these scenarios, we then revisit experimental constraints on ALPs, in particular we (re)calculate and improve beam dump, flavour and supernova constraints. We find that in the DFO region, ALP-fermion couplings are tiny and are probably only accessible via cosmological constraints. The freeze-in scenario is possible in a wide parameter space and subject to other, more direct constraints. While certain regions of this parameter space are already ruled out, a significant part should be accessible to forthcoming collider experiments.

Résumé

Le couplage de l'interaction électromagnétique, α, est une pierre angulaire du modèle standard (SM) de la physique des particules. Si elle n'est pas améliorée, la précision de sa valeur à l'échelle de la masse du boson Z (M_Z) limitera le potentiel de découverte d'expériences auprès de futurs collisionneurs. Cette précision dépend de notre compréhension de la contribution hadronique à l'évolution de α aux basses énergies, une région où la description du SM est entravée par les effets non linéaires de la chromodynamique quantique (QCD). Dans cette étude, nous faisons un premier pas vers la réduction de cette incertitude en calculant l'évolution hadronicque de α depuis la limite de Thomson (correspondant une l'échelle $Q^2 = 0$) jusqu'à l'échelle de genre espace, $Q^2 = 8\,\mathrm{GeV}^2$. Nous utilisons des simulations à grande échelle en QCD sur réseau, qui sont réalisées avec $N_f = 2 + 1 + 1$ saveurs de quarks "staggered" avec des masses ajustées à leurs valeurs physiques, et qui prennent également en compte les corrections de brisure d'isospin d'ordre dominant. Afin d'améliorer le contrôle des incertitudes systématiques, en particulier celles associées aux extrapolations nécessaires vers la limite du continu, nous divisons cette évolution en sous-intervalles de Q^2, en optimisant nos analyses dans chacun d'entre eux. En particulier, nous avons développé et mis en œuvre une série de stratégies d'amélioration. Ces stratégies sont basées sur la théorie des perturbations sur réseau à courte distance, sur des théories effectives traitant de la brisure de la symétrie de "taste" à plus grande distance, et sur l'utilisation de 48 discrétisations du courant électromagnétique du charme. Nos résultats ont déjà des incertitudes comparables à celles obtenues dans l'approche de référence basée sur les données, et nous esquissons une feuille de route pour une réduction significative de ces incertitudes à l'avenir.

Dans la deuxième partie de cette thèse, nous étudions un scénario permettant d'aller au-delà du modèle standard de la physique des particules en envisageant la possibilité qu'une particule de type axion (ALP) agisse comme médiateur entre les particules du modèle standard et les particules de la matière sombre. Les progrès expérimentaux ont considérablement limité l'existence d'un WIMP standard à l'échelle électrofaible qui serait produit par *freeze-out*. Cela a conduit à un changement de paradigme vers des scénarios alternatifs. Nous nous concentrons sur le cas où les couplages sont trop faibles pour permettre la génération de matière sombre par freeze-out et où la matière sombre est découplée thermiquement des particules du modèle standard. Cependant, des mécanismes alternatifs de génération de matière sombre comme le freeze-in et le freeze-out découplé peuvent toujours reproduire la densité relique de matière sombre observée. Après avoir déterminé la région de l'espace des paramètres pour ces scénarios, nous revisitons les contraintes expérimentales sur les axions, en particulier nous (re)calculons et améliorons les contraintes de beam dump, de saveur et de supernova. Nous constatons que dans la région du

freeze-out découplé, les couplages axion-fermion sont minuscules et ne sont probablement accessibles que par des contraintes cosmologiques. Le scénario du freeze-in est possible dans un large espace de paramètres et est soumis à d'autres contraintes plus directes. Alors que certaines régions de cet espace de paramètres sont déjà exclues, une partie importante devrait être accessible à des expériences à venir, auprès de collisionneurs.

Résumé étendu

Introduction

Le modèle standard (SM) de la physique des particules est une théorie quantique des champs au succès spectaculaire. Il fournit un cadre mathématique permettant de comprendre les propriétés des particules fondamentales connues et de trois des quatre interactions connues, et prédit avec succès l'existence de particules élémentaires telles que le boson de Higgs. En particulier, il existe des preuves expérimentales solides que la théorie de la chromodynamique quantique (QCD) fournit une description précise des interactions fortes. Bien que le modèle standard ait résisté à de multiples tests expérimentaux au cours des cinquante dernières années, il ne fournit pas d'explication à plusieurs phénomènes observés, tels que les masses des neutrinos, le problème de la hiérarchie, l'origine de la saveur et l'asymétrie baryonique et l'origine des "composantes sombres" (énergie noire et matière sombre) de l'univers.

Sur le plan théorique, il existe généralement deux grands axes de recherche : le premier consiste à examiner le modèle standard en effectuant des mesures et des calculs précis dans le cadre de ce modèle. Une autre approche consiste à construire des modèles pour les phénomènes inexpliqués et à étudier les conséquences de ceux-ci. Dans cette thèse, nous explorerons ces deux approches.

Première partie : Tester le modèle standard via le couplage électromagnétique

L'une des approches pour rechercher de la physique au-delà du modèle standard consiste à se placer à la frontière de l'intensité, c'est-à-dire à augmenter la luminosité plutôt que l'échelle d'énergie des expériences. Étant donné qu'il n'existe pas encore de mesures directes ou indirectes d'un effet de nouvelle physique aupres d'éxperience de physique des particules (à part la masse des neutrinos), nous nous attendons à ce que ses effets soient très faibles. La détection d'un écart par rapport au modèle standard exige que la prédiction théorique et la mesure expérimentale soient déterminées avec une extrême précision. Du côté théorique, les effets hadroniques limitent souvent la précision, ce qui nécessite une description fiable de ces processus de chromodynamique quantique (QCD) à basse énergie.

L'une de ces quantités de précision est la constante de couplage électromagnétique, α, dont la valeur à l'échelle de la masse Z est une donnée importante dans les tests de précision électro-faibles du modèle standard [1]. Alors que la moyenne mondiale de α à l'échelle de la masse de

l'électron est connue avec une précision étonnante d'une partie par milliard [2], ce couplage électromagnetique perd cinq ordres de grandeur en précision lorsqu'il est évalué à M_Z [3]. Cela fait de $\alpha(M_Z)$ l'un des paramètres les moins bien déterminés du modèle standard. La plus grande source d'erreur provient des effets hadroniques non perturbatifs dans la région de basse énergie.

Les futurs collisionneurs réduiront considérablement les incertitudes sur $\alpha(M_Z)$ obtenues indirectement via des ajustements globaux d'observables à l'échelle électrofaible. En fait, les incertitudes sur le calcul direct de $\alpha(M_Z)$ devront être réduites d'un facteur deux pour tirer pleinement parti de ces futures mesures dans la recherche d'une nouvelle physique fondamentale [3].

L'évolution du couplage électromagnétique en fonction de l'énergie Le propagateur complet d'un photon d'impulsion q est donné par des insertions irréductibles à une particule du tenseur d'auto-énergie, noté $\Pi_{\mu\nu}(q)$, qui est la valeur moyenne dans le vide du produit ordonné dans le temps du corrélateur de deux courants électromagnétiques. Par invariance de Lorentz et de jauge, ce tenseur de polarisation du vide peut être décomposé en une fonction invariante de Lorentz et une structure de Lorentz qui est transverse et proportionnelle à q^2,

$$\Pi_{\mu\nu}(q) = i \int d^4x \; e^{iq\cdot x} \left\langle 0|T\{J_\mu(x)J_\nu(0)\}|0\right\rangle = \left(q_\mu q_\nu - g_{\mu\nu}q^2\right)\Pi\left(q^2\right) \; , \tag{1}$$

avec $J_\mu(x) \equiv \sum_f q_f \bar{f}(x)\gamma_\mu f(x)$, où $f = \{e, \mu, \tau, u, d, s, c, b, t\}$, et avec une charge éléctrique q_f. Après resommation des diagrammes irréductibles à une particule il est possible de définir le couplage effectif éléctromagnétique :

$$\alpha(q^2) = \frac{\alpha}{1 - \Delta\alpha(q^2)} \; , \qquad \text{avec} \quad \Delta\alpha(q^2) = 4\,\pi\,\alpha\hat{\Pi}(q^2) = 4\,\pi\,\alpha\left(\Pi(q^2) - \Pi(0)\right) \; , \tag{2}$$

avec $\hat{\Pi}(q^2)$ la fonction de polarisation du vide renormalisé. Il est pratique de diviser les contributions au *running* en une partie leptonique, la contribution hadronique des cinq quarks les plus légers et la contribution du quark top,

$$\Delta\alpha(q^2) = \Delta\alpha_{\text{lep}}(q^2) + \Delta\alpha^{(5)}_{\text{had}}(q^2) + \Delta\alpha_{\text{top}}(q^2) \; . \tag{3}$$

Alors que la contribution leptonique et la contribution du quark top peuvent être calculées de manière fiable en théorie des perturbations, la fonction de polarisation du vide hadronique (HVP) reçoit d'importantes contributions non perturbatives en dessous d'une échelle de quelques GeV, ce qui la rend inaccessible aux méthodes analytiques connues. En raison des complications liées au calcul de cette contribution hadronique, elle domine l'incertitude sur le *running* de α. La manière traditionnelle d'obtenir la HVP, dans ce régime non perturbatif, consiste à utiliser les mesures de la section efficace $e^+e^- \to$ hadrons en fonction de l'énergie dans le centre de masse et une relation de dispersion [4, 5]. Cependant, pour des impulsions de genre espace, la HVP est accessible aux calculs ab-initio en QCD sur réseau qui fournit donc une approche complémentaire qui ne dépend pas des données de section efficace [6–8]. Ceci est devenu particulièrement important depuis la récente détermination du $g-2$ du muon par la collaboration BMW, pour laquelle les prédictions obtenues via l'approche basée sur les données et celle à partir du réseau diffèrent [7].

La contribution hadronique à l'évolution de α en fonction de l'énergie en QCD sur réseau Sur le réseau, la quantité qui nous intéresse est le tenseur de polarisation du vide dans

un espace temps euclidien $Q^2 = -q^2$,

$$\Pi_{\mu\nu}(Q) = \int d^4x \; e^{iQ\cdot x} \left\langle J_\mu(x) J_\nu(0) \right\rangle = \left(Q_\mu Q_\nu - \delta_{\mu\nu} Q^2\right) \Pi\left(Q^2\right) \; , \tag{4}$$

où la dernière égalité découle de l'utilisation de la symétrie O(4) et de l'invariance de jauge. Cependant, l'équation (4) n'est pas valable dans un volume fini - le tenseur de polarisation du vide n'est totalement transverse que dans un volume et un temps infini. Comme le souligne [9], $\hat{\Pi}(Q^2)$ peut être obtenu dans la représentation temps-impulsion par un corrélateur projeté sur une impulsion zéro et multiplié par une fonction qui dépend de Q^2. Nous définissons donc

$$\hat{\Pi}(Q^2) \equiv \Pi(Q^2) - \Pi(0) = 2a \sum_t \mathrm{Re}\left[\frac{e^{iQt} - 1}{Q^2} + \frac{t^2}{2}\right] \mathrm{Re}\, C(t) \qquad \forall Q \in \mathbb{R} \; , \tag{5}$$

avec

$$C(t) = \frac{a^3}{3} \sum_{i=1}^{3} \sum_{\vec{x}} \left\langle J_i(x) J_i(0) \right\rangle = C^{ud}(t) + C^s(t) + C^c(t) + C^{\mathrm{disc}}(t) \; , \tag{6}$$

avec a l'espacement du réseau et où nous avons décomposé en saveur le corrélateur du courant électromagnétique J_μ. Cela présente l'avantage de pouvoir traiter séparément les incertitudes statistiques et systématiques très différentes des diverses contributions.

Par conséquent, pour notre analyse finale, nous avons décidé de calculer les différences finies de la HVP, que nous appelons la fonction d'Adler discrète

$$\Delta\Pi(Q^2) \equiv \Pi(Q^2) - \Pi(Q^2/2) \; . \tag{7}$$

La HVP est obtenue à partir d'une somme télescopique

$$\hat{\Pi}(Q_n^2) = \Pi(Q_n^2) - \Pi(0) = \sum_{i=1}^{n-1} \Delta\Pi(2^i Q_0^2) - \hat{\Pi}(Q_0^2) \tag{8}$$

et nous choisissons $Q_0^2 = 1$ GeV^2. Pour la contribution légère et la contribution déconnectée, nous divisons en outre la contribution à faible énergie $\hat{\Pi}(Q_0^2 = 1 \text{ GeV}^2)$ en une fenêtre à courte distance et une fenêtre à longue distance,

$$\begin{aligned} \hat{\Pi}(Q_0^2) = \hat{\Pi}_{00-04}(Q_0^2) + \hat{\Pi}_{04-\infty}(Q_0^2) = & \int \mathrm{d}t \; w(t; 0, t_2 = 0.4 \text{ fm}) \; C(t) \; k(t, Q^2) + \\ & \int \mathrm{d}t \; w(t; t_1 = 0.4 \text{ fm}, \infty) \; C(t) \; k(t, Q^2) \; , \end{aligned} \tag{9}$$

où nous avons convolué l'intégrande par une fonction lisse, dite de fenêtre [10]

$$w(t; t_1, t_2) \equiv \Theta(t; t_1, \Delta) - \Theta(t; t_2, \Delta) \quad \text{avec} \quad \Theta(t; t', \Delta) \equiv \tfrac{1}{2} + \tfrac{1}{2} \tanh[(t - t')/\Delta] \; . \tag{10}$$

Δ est la largeur de la fonction en escalier lissée et nous utilisons $\Delta = 0,15$ fm. Comme nous le montrerons dans les sections suivantes, à courte distance, la HVP subit des effets de discrétisation augmentés par un logarithme tandis qu'à longue distance, le spectre de la fonction de corrélation à deux points est déformé par des effets de brisure de la symétrie de "goût". Un mélange de ces effets compromettrait une extrapolation contrôlée vers le continu et la séparation des quantités à courte et à longue distance est donc nécessaire. Le résultat est que le comportement de nos

observables est amélioré à des échelles très différentes, ce qui permet un contrôle optimal des sources d'incertitudes systématiques.

Strategies d'analyse

- **Défis à courte distance: effets de discrétisation.** Bien que l'équation (5) puisse être utilisée pour définir $\hat{\Pi}(Q^2)$ pour tout $Q \in \mathbb{R}$, aux grandes valeurs de Q nous sommes limités par la coupure ultraviolette sur le réseau, quant aux petites valeurs de Q, l'observable sentira la taille finie du réseau. À ces grandes distances donc, les effets de volume fini et, puisque nous utilisons des quarks *staggered*, les effets de brisure de la symétrie de "goût" joueront un rôle important. En raison de la soustraction de $\Pi(0)$, la HVP mélange, pour de grands Q^2, des échelles très différentes.

  Pour de grandes valeurs de Q^2, les effets de discrétisation deviennent importants et finissent par détériorer l'extrapolation vers le continu de nos observables. En outre, comme cela a été montré récemment dans [8, 11], les contributions des quarks de valence légers de masse m_l à nos observables subissent des erreurs de discrétisation d'$\mathcal{O}(a^2)$ augmentés par un logarithme, même à l'ordre dominant (LO) dans la théorie des perturbations sur réseau,

$$O\left(Q^2, a\right) = O\left(Q^2\right)\left\{1 + \Gamma_0 (aQ)^2 \ln(aQ)^2 + \mathcal{O}\left[(aQ)^2, (am_l)^2\right]\right\} , \tag{11}$$

  où $O\left(Q^2\right)$ est la valeur de nos observables dans le continu et Γ_0 est une constante. Ces effets de coupure augmentés par un logarithme proviennent de petites séparations entre les deux courants : nos observables ne sont pas des quantités *on-shell*. Notons également que les corrections perturbatives au terme logarithmiquement renforcé, de la forme $\alpha_s^n (1/a)(aQ)^2 \ln(aQ)^2$, sont d'ordre $\alpha_s^{n-1}(1/a)(aQ)^2$ parce que $\alpha_s \sim -1/\ln(a\Lambda_{\mathrm{QCD}})$ et ne sont donc, en fait, pas logarithmiquement renforcées.

  Une façon d'éliminer certaines des erreurs de discrétisation est d'améliorer les données du réseau à l'aide de la théorie des perturbations sur réseau.

- **Effets de brisure de la symétrie d'isospin.** Travailler dans la limite de symétrie d'isospin n'est plus suffisant si l'on vise une précision $< 1\%$. Nous incluons donc, dans nos calculs, les effets de brisure de symétrie isospin induits par des effets d'électrodynamique quantique et de différence de masse du quark u et d. Ces effets sont pris en compte au premier ordre dans les paramètres de brisure de la symétrie isospin $\delta m = m_u - m_d$ et α en développant l'intégrale de chemin QCD+QED dans ces paramètres et en mesurant ces observables sur des configurations symétriques à l'isospin, où nous distinguons également les charges des quarks de la mer et de valence.

- **Défis à longue distance: effets de la brisure de symétrie de "goût".** Pour de faibles valeurs de Q^2, la contribution des quarks légers est dominée par l'état à deux pions et la résonance ρ, et est donc affectée par des effets de discrétisation supplémentaires provenant de la brisure de symétrie de goût dans la formulation *staggered*. À basse énergie, ces violations de symétrie de goût deviennent en fait les effets de discretisation les plus importants des fermions *staggered*, puisqu'elles déforment le spectre des pions. Sur la base de l'étude approfondie réalisée dans la Réf. [7] concernant les effets de brisure de la

symétrie de goût, nous modélisons ces erreurs de discrétisation en utilisant le modèle rho-pion-gamma dans sa version *staggered* (que nous appelons SRHO). Nous constatons que le modèle SRHO décrit bien les erreurs de discrétisation observées dans nos simulations pour $t \geq 0.4$ fm.

- **Extrapolation à la limite de volume infini.** Alors que les effets de volume fini deviennent faibles pour des Q^2 élevés, à longue distance, l'observable ressent la taille finie du réseau. Ainsi, en particulier pour la contribution d'isospin $I = 1$ à $\hat{\Pi}(1\ \mathrm{GeV}^2)$, les effets de taille finie sont non négligeables. Pour déterminer la limite de volume infini de la partie isospin-symétrique, $I = 1$, de la HVP et de la fonction d'Adler discrète, nous avons réalisé une étude dédiée des effets de taille finie, similaire à la procédure de Réf. [7].

- **Extrapolation à la limite du continu.** Afin d'effectuer une extrapolation contrôlée vers le continu, nous incorporons dans notre analyse les diverses améliorations mentionnées précédemment. Pour obtenir le résultat physique pour nos observables, nous effectuons un ajustement global qui comprend une extrapolation vers le continu, une interpolation au point physique, où X_l et X_s paramètrisent la petite différence des masses des quarks simulés par rapport à leurs valeurs physiques, et la détermination des corrections de la brisure de la symétrie d'isospin,

$$O(a, Q^2) = O(0, Q^2) + \underbrace{A(a)}_{\substack{\text{extrap.}\\ \text{continu}}} + \underbrace{BX_l + CX_s}_{\substack{\text{interpolation au}\\ \text{point physique}}} + \underbrace{DX_{\delta m} + Ee_v^2 + Fe_v e_s + Ge_s^2}_{\substack{\text{détermination des}\\ \text{corrections à } \mathcal{O}(\delta m, e^2)}} , \tag{12}$$

avec

$$X_l = \left[\frac{(M_{\pi_0}/M_{\Omega^-})^2}{(M_{\pi_0}/M_{\Omega^-})_*^2}\right] - 1, \quad X_s = \left[\frac{(M_{ss}/M_{\Omega^-})^2}{(M_{ss}/M_{\Omega^-})_*^2}\right] - 1 , \tag{13}$$

et

$$A(a) = A_2[a^2\alpha_s^n(1/a)] + A_{2l}a^2\log(a^2) + A_4[a^2\alpha_s^n(1/a)]^2 , \tag{14}$$

et où $A_2, A_{2l}, A_4, B, C, E, F, G, H$ sont des paramètres d'ajustement. M_{Π_0} est la masse du pion de Goldstone, M_{Ω^-} la masse du baryon Ω^-, M_{ss} la masse du méson construit à partir d'un quark étrange et d'un anti-quark étrange et les indices $*$ indiquent la valeur physique. Notons que nous distinguons également les charges électriques des quarks de valence, e_v, et de la mer, e_s, comme dans [7, 12].

Résultats. Sur la base de la procédure d'échantillonnage qui sera décrite à la Sec. 4.3.2.4, nous calculons la somme, en préservant les corrélations, des contributions connectées légères, étranges, du charme et des contributions déconnectées à l'évolution du couplage électromagnétique jusqu'à $Q^2 = 8\ \mathrm{GeV}^2$. Le Tab. 1 contient les résultats des différentes contributions de saveur, ainsi que les corrections de volume fini. Nous avons également inclus la contribution du quark *bottom*, qui a été obtenue en utilisant les résultats de Réf. [13]. La somme des contributions connectées légères, étranges, charm, déconnectées, bottom et de taille finie donne la polarisation du vide hadronique finale à cinq saveurs $\hat{\Pi}(Q^2)$, donnée dans la dernière colonne du Tab. 1. Alors que la somme des contributions connectées légères, étranges, charmantes et déconnectées a été obtenue en effectuant la procédure d'échantillonnage, les contributions de taille finie et de fond ont été

$O(Q^2)\cdot 10^4$	O^{light}	O^{strange}	O^{charm}	O^{disc}	$\infty -$ ref	O^{bottom}	$O^{(5)}$
$\hat{\Pi}(1\,\text{GeV}^2)$	355[5]	41.6[4]	16.30[15]	−4.6[7]	3.10[22]	0.34[5]	411[5]
$\hat{\Pi}(2\,\text{GeV}^2)$	465[5]	60.4[4]	30.79[26]	−4.8[7]	3.24[25]	0.7[1]	555[5]
$\hat{\Pi}(4\,\text{GeV}^2)$	574[5]	80.8[4]	55.6[4]	−4.8[7]	3.31[25]	1.33[19]	710[5]
$\hat{\Pi}(8\,\text{GeV}^2)$	681[5]	101.70[35]	94.2[6]	−4.8[7]	3.36[25]	2.6[4]	878[5]
$\Delta\Pi(2\,\text{GeV}^2)$	110.12[32]	18.81[8]	14.49[12]	−0.165[12]	0.135[12]	0.34[5]	143.75[35]
$\Delta\Pi(4\,\text{GeV}^2)$	109.06[34]	20.45[8]	24.85[18]	−0.046[4]	0.08[1]	0.7[1]	155.1[4]
$\Delta\Pi(8\,\text{GeV}^2)$	107.0[4]	20.48[14]	38.57[25]	−0.0076[12]	0.04[1]	1.27[19]	167.7[5]

Table 1: Évolution de $\hat{\Pi}(Q^2)$ jusqu'à $Q^2 = 8\,\text{GeV}^2$, par pas de puissances de deux, ainsi que de la fonction d'Adler discrète. Nous donnons séparément les contributions légères, étranges, *charm* et déconnectées qui sont les résultats finaux présentés dans la Sec. 4.3.3. Nous énumérons également les corrections de taille finie calculées dans la Sec. 4.3.2.3. La contribution du quark *bottom* a été obtenue à partir des travaux de la collaboration HPQCD [13]. La dernière colonne contient la somme de toutes les contributions. Notons que la somme des contributions légères, étranges, *charm* et déconnectées a été obtenue en utilisant une procédure d'échantillonnage que nous présentons dans la Sec. 4.3.2.4 tandis que les corrections bottom et de volume fini ont été simplement ajoutées, en combinant leurs erreurs en quadrature. Les erreurs doivent être comprises comme portant les derniers chiffres de la valeur centrale et nous appliquons les conventions du PDG pour l'arrondi [2].

simplement ajoutées, et les erreurs ont été ajoutées en quadrature puisque ces contributions sont complètement indépendantes de notre résultat de réseau à quatre saveurs $\hat{\Pi}(Q^2)$.

Les valeurs de la fonction d'Adler discrète dans nos intervalles de $Q^2 \in [2^n, 2^{n+1}]$, $n = 0, 1, 2$ sont plus ou moins de la même taille puisque le *running* est logarithmique à ordre dominant en théorie des perturbations en QCD. Nos résultats deviennent plus précis lorsque nous considérons des valeurs plus élevées de Q^2. À basse énergie, nous sommes actuellement limités par l'erreur systématique provenant de l'extrapolation vers le continu de la fenêtre à longue distance de la contribution légère à $\hat{\Pi}(1\,\text{GeV}^2)$. Bien que les violations de la symétrie de goût déforment de manière significative le spectre des pions et conduisent à des effets de discretisation importants, nous avons néanmoins aussi inclus les ajustements aux données non corrigées par le modèle SRHO pour une estimation prudente de l'erreur, ce qui entraîne une incertitude systématique relativement importante à ces basses énergies. Une étude plus approfondie des violations de la symétrie de goût et l'ajout d'un ensemble avec une maille de réseau plus fine sont en cours.

Dans la Fig. 1, nous comparons nos résultats pour la contribution hadronique à l'évolution en fonction de l'énergie du couplage électromagnétique $\Delta\alpha^{(5)}_{\text{had}}(Q^2)$ avec la récente détermination du réseau par la collaboration de Mayence [8], avec deux déterminations antérieures de notre collaboration [6, 7] et avec une détermination guidée par les données de section efficace $e^+e^- \to$ hadrons de la collaboration KNT [14]. Nos résultats incluent la contribution du quark bottom de la collaboration HPQCD [13]. Dans la Réf. [8], les auteurs fournissent une paramétrisation de leurs résultats basée sur des approximants de Padé incluant une matrice de corrélation. Aux énergies élevées, nous sommes en général en assez bon accord : la différence sur chacun des pas finis de la fonction d'Adler discrète se situe entre un et deux écart types.

Cependant, nos résultats sont systématiquement inférieurs à ceux obtenus par Mayence, ce qui se traduit par une valeur plus faible pour $\hat{\Pi}(8\,\text{GeV}^2)$. Ces différences peuvent être dues au fait que, contrairement à Mayence, nous incluons, dans nos paramétrisations pour les extrapolations vers le continu des données non corrigées, les effets de discrétisation augmentés logarithmiquement qui sont présents, ou nous améliorons les données, avant de les extrapoler, avec des corrections calculées à l'ordre supérieur dans la théorie des perturbations *staggered*.

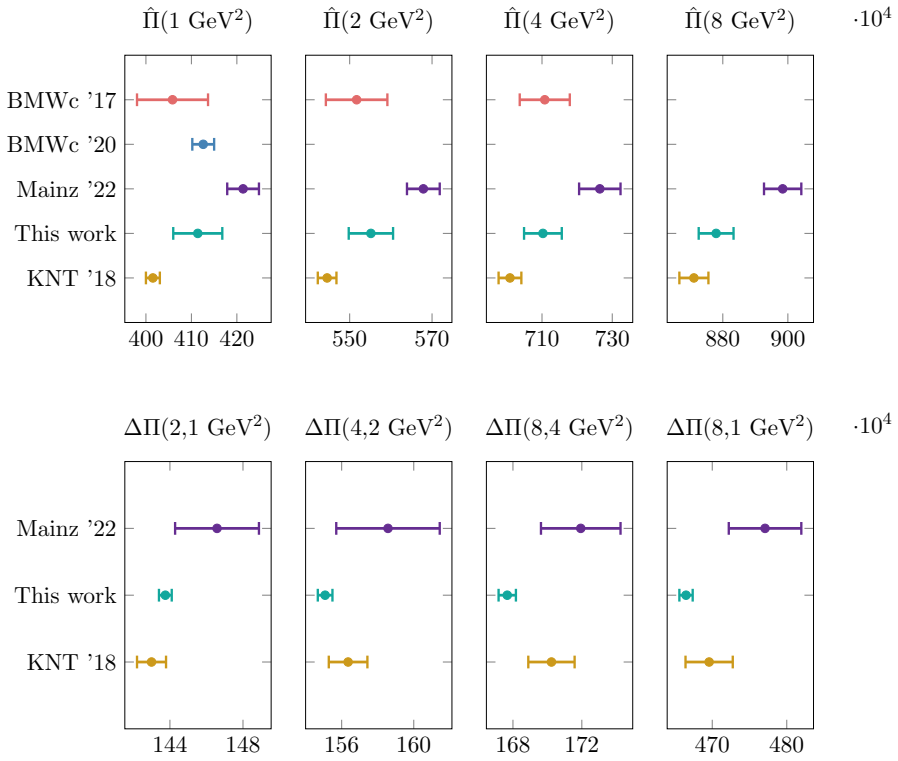


FIGURE 1: Compilation des résultats pour la fonction de polarisation du vide hadronique $\hat{\Pi}(Q^2)$ et son échelonnement $\Delta\Pi(2Q^2, Q^2)$ en unités de 10^4 pour diverses valeurs de Q^2 en GeV^2. Les résultats en violet sont une détermination récente en QCD sur réseau par la collaboration de Mayence [8]. Les auteurs dans Réf. [8] fournissent une paramétrisation de leurs résultats basée sur des approximations de Padé, y compris une matrice de corrélation. Les marques jaunes sont les résultats d'une détermination basée sur les données du rapport R fournies par KNT [14] du seuil à $Q = 11,199$ GeV et de la QCD perturbative à $\mathcal{O}(\alpha_s^4)$ à l'aide du programme rhad [15] pour l'intégrale restante. Pour $\hat{\Pi}(1\,\mathrm{GeV}^2)$ nous comparons également à une détermination antérieure par notre collaboration [7].

Le résultat guidé par les données est obtenu en intégrant les données de section efficace $e^+e^- \to$ hadrons de la Réf. [14] du seuil à $Q = 11.199$ GeV, en tenant compte des corrélations via la matrice de corrélation qui a également été fournie par les auteurs. Pour l'intégrale restante de $Q = 11.199$ GeV à ∞, nous intégrons la section efficace $e^+e^- \to$ hadrons obtenues en QCD perturbative à $\mathcal{O}(\alpha_s^4)$ que nous calculons à l'aide du programme rhad [15].

Bien que nous trouvions un léger excès par rapport à l'approche guidée par les données pour $\hat{\Pi}(1\,\mathrm{GeV}^2)$ de $\sim 2.5\%$, nous sommes d'accord pour les étapes discrètes au-dessus de $1\,\mathrm{GeV}^2$. Par conséquent, pour $\hat{\Pi}(8\,\mathrm{GeV}^2)$, nous sommes déjà d'accord, à quelques erreurs près, avec l'approche basée sur les données. Remarquons également que si notre incertitude est encore plus grande que l'incertitude sur le résultat obtenu pour $\hat{\Pi}(1\,\mathrm{GeV}^2)$, nos erreurs totales sur la fonction d'Adler discrète à des énergies plus grandes sont deux fois moins importantes.

Conclusions La valeur du couplage électromagnétique à l'échelle Z, $\alpha(M_Z^2)$, est une donnée importante pour les tests de précision du modèle standard. La précision de la détermination théorique de cette quantité est actuellement limitée par les incertitudes de la contribution hadronique dans la région de basse énergie. Alors que l'approche traditionnelle pour calculer la contribution hadronique à l'évolution en fonction de l'énergie de α dans ce régime non-perturbatif repose sur une relation de dispersion et nécessite des données expérimentales associées à la section efficace $e^+e^- \to$ hadrons, la QCD sur réseau fournit un outil idéal pour accéder à ce régime

non-perturbatif à partir des principes fondamentaux. Dans la première partie de cette thèse, nous présentons un calcul ab-initio de QCD sur réseau de l'évolution en fonction de l'énergie de la contribution hadronique de α jusqu'à $8\,\mathrm{GeV}^2$ qui est obtenue avec $2+1+1$ saveurs de fermions *staggered* dont les masses sont ajustées à leurs valeurs physiques.

Pour notre résultat final, l'évolution en fonction de l'énergie de α jusqu'à $8\,\mathrm{GeV}^2$, $\Delta\alpha^{(5)}_{\mathrm{had}}(Q^2 = 8\,\mathrm{GeV}^2) = (80.0 \pm 0.5) \cdot 10^{-4}$, nous avons déjà atteint le même niveau de précision que les déterminations phénoménologiques, voir par exemple le Tab. VIII de Réf. [16]. Pour la contribution provenant de la partie basse énergie de l'évolution, $\hat{\Pi}(1\ \mathrm{GeV}^2)$, nous sommes actuellement limités par la précision sur la contribution légère, qui constitue également la contribution dominante à cette échelle. À grande distance, les violations de la symétrie de goût dans le spectre des pions introduisent de grands effets de discrétisation dans cette quantité. Pour les éliminer, nous utilisons le modèle dit SRHO au-dessus de $t \geq 1.0$ fm, qui décrit bien les erreurs de discrétisation observées dans nos simulations, même jusqu'à $t = 0.4$ fm. Néanmoins, comme nous aimerions confirmer notre compréhension des effets de brisure de symétrie de goût avant de citer de très petites barres d'erreur, nous estimons l'incertitude sur $\hat{\Pi}(1\ \mathrm{GeV}^2)$ de manière très conservatrice en incluant également les ajustements aux données non corrigées dans notre estimation finale. Nous prévoyons de réduire l'incertitude sur cette quantité en effectuant une analyse plus approfondie des effets de brisure de symétrie de goût et en incluant au moins deux ensembles supplémentaires proche du point physique à un espacement de réseau plus fin $a \approx 0.048$ fm.

Partie II : Sonder au-delà du modèle standard avec un axion comme médiateur pour la matière sombre

Introduction Les récentes avancées expérimentales limitent désormais fortement les modèles dans lesquels la particule de matière sombre (DM) est directement chargée par rapport au groupe de jauge du modèle standard via les interactions renormalisables. Le standard *weakly interacting massive particle* (WIMP) produit par *freeze-out* est, par exemple, remis en question par le fait que, malgré des recherches expérimentales depuis des décennies, il n'a pas été détecté jusqu'à présent. On peut donc se demander si d'autres scénarios et d'autres candidats ne sont pas également possibles. En fait, l'axion QCD est lui-même un exemple pionnier de candidat pour la matière sombre produit de manière non thermique. Dans ce travail, l'axion ne constitue pas la matière sombre mais nous explorons la possibilité d'un *axion-like particle* (ALP) de type DFSZ agissant comme médiateur entre la particule de matière sombre et le modèle standard. L'intensité des nouveaux couplages est déterminée par la nouvelle échelle de brisure de symétrie f_a qui devrait être suffisamment grande pour être en accord avec les expériences actuelles des collisionneurs. Par conséquent, les nouveaux couplages sont faibles et nous nous concentrons sur les régions où les particules de matière sombre sont en déséquilibre thermodynamique par rapport aux particules du modèle standard.

Le modèle Dans notre modèle, nous étendons le modèle standard par un fermion de Dirac χ et un ALP a de type DFSZ, c'est-à-dire le pseudo-boson de Goldstone d'une symétrie globale approximative supplémentaire $U(1)$ qui elle est spontanément brisée à une échelle élevée f_a. L'ALP agit comme un médiateur entre le modèle standard et le secteur sombre. En dessous de

l'échelle électrofaible, l'ALP ne se couple efficacement qu'aux fermions du modèle standard au *tree-level* via des interactions de type Yukawa et interagit avec les bosons de jauge/Higgs via des boucles fermioniques. S'il n'y a pas de fermions vectoriels lourds qui ont été intégrés, c'est-à-dire si le contenu en fermions provient uniquement du modèle standard, les conditions d'annulation des anomalies impliquent automatiquement que les contributions des termes de dimension 5 du type $aF\tilde{F}$ s'annulent pour ne laisser que des termes de dimension 4, dépendant de la masse [17]. Cependant, comme l'ALP peut émerger naturellement d'un secteur de Higgs étendu (ous fournissons une complétion UV explicite dans l'appendice F), le couplage de dimension 5 entre le Higgs, l'ALP et les fermions du modèle standard doit également être pris en compte. Ces couplages donnent des contributions à la densité relique de la matière sombre qui sont sensibles au régime ultraviolet. Nous considérons donc les nouveaux termes suivants dans le lagrangien effectif :

$$\begin{aligned}\mathcal{L} \supset & \frac{1}{2}\partial_\mu a \partial^\mu a + \bar{\chi}(i\partial\!\!\!/ - m_\chi)\chi - \frac{1}{2}m_a^2 a^2 + ia\sum_f \frac{m_f}{f_a} C_f \bar{f}\gamma_5 f \\ & + ia\frac{m_\chi}{f_a} C_\chi \bar{\chi}\gamma_5\chi + a\sum_f C_f \frac{y_f}{\sqrt{2}f_a} h\, \bar{f} i\gamma_5 f + \dots \end{aligned} \tag{15}$$

Ici, f est n'importe quel fermion du Modèle Standard avec une masse m_f, les masses du χ et de l'ALP sont m_χ et m_a, respectivement. Dans la suite, nous travaillerons cependant avec les couplages redéfinis : $g_{a\chi\chi} \equiv m_\chi C_\chi / f_a$ sera appelé *couplage du secteur caché* et $g_{aff} \equiv C_f/f_a$ comme *couplage de connecteurs*. Ce lagrangien est valide en dessous de l'échelle de brisure de la symétrie électrofaible.

Notons que si a était l'axion QCD (de type DFSZ), de multiples contraintes observationelles et experimentales s'appliqueraient (voir [18] pour une vue d'ensemble), bien qu'il puisse être possible de contourner ces contraintes par une modélisation poussée. Nous supposons donc que la masse m_a est principalement due à un effet de brisure explicite de $\mathrm{U}(1)_{\mathrm{PQ}}$ autre que l'anomalie, ce qui nous permet de considérer m_a et g_{aff} indépendamment et, par conséquent, d'échapper aux contraintes.

Scenarios de la genèse de matière sombre Comme l'échelle f_a doit être grande, les couplages sont en général trop petits pour donner la densité relique correcte par le mécanisme *freeze-out*. Une alternative au scénario standard de *freeze-out* est que les abondances initiales des particules du secteur sombre, c'est-à-dire la matière sombre et l'ALP, sont négligeables par rapport au bain thermique des particules du modèle standard après rechauffage, et qu'elles n'atteignent pas l'équilibre avec les particules du modèle standard avant le découplage. Par exemple, dans le mécanisme de *freeze-in* : les particules du secteur sombre sont progressivement produites par des processus de scattering dans le plasma thermique sans atteindre l'équilibre, jusqu'à ce que la température diminue suffisamment pour que les processus se découplent et que l'abondance de matière sombre reste constante (voir par exemple [19]). Notons qu'il y a deux possibilités, le *freeze-in* infrarouge où la densité rélique finale est insensible à la température de rechauffage, ou le *freeze-in* ultraviolet où le choix de la température de rechauffage influence le résultat, en fonction de la dépendance de T de la section efficace thermique. Dans notre travail, nous nous concentrons sur les mécanismes suivants :

- Dans le cas où $g_{a\chi\chi}$ et g_{aff} sont faibles, la matière sombre est produite par *freeze-in*. Soit, lorsque $T_{RH} \lesssim 1$ TeV directement via $f\bar{f} \to \chi\bar{\chi}$ ou, par exemple, $f\bar{f} \to ag$ suivi de $aa \to \chi\bar{\chi}$, où l'ALP peut ou non être en équilibre avec le modèle standard [20]. En revanche, pour des T_{RH} plus grandes, la contribution dominante provient de *freeze-in* UV via le processus 2→3 $f\bar{f} \to h\bar{\chi}$ ainsi que $f\bar{f} \to ha$ suivi de $aa \to \chi\bar{\chi}$.
- Au niveau intermédiaire de g_{aff}, les ALPs n'atteignent pas l'équilibre avec les particules du modèle standard, mais si $g_{a\chi\chi}$ est suffisamment grand, les ALP et les particules de matière sombre s'équilibrent. Si $g_{a\chi\chi}$ est suffisamment grand, les ALP et les particules du modèle standard peuvent entrer en équilibre, ce qui conduit au *freeze-out* d'un secteur sombre thermiquement découplé ou, pour le dire plus simplement, au *freeze-out découplé* (DFO), voir [21], et plus récemment [22].

Étudions la résolution des équations de Boltzmann dans ces différents cas. La déduction des interactions entre les particules du modèle standard, de la matière sombre et des ALP à partir du lagrangien de l'equation (15) nous permet d'écrire l'ensemble le plus général et le plus complet d'équations différentielles couplées régissant l'évolution des densités de a et χ

$$\begin{aligned}
\frac{\mathrm{d}n_\chi}{\mathrm{d}t} + 3Hn_\chi &= \sum_f \left\langle \sigma_{\chi\bar{\chi}\to f\bar{f}} v \right\rangle \left(\left(n_\chi^{\mathrm{eq}}\right)^2 - n_\chi^2 \right) + \left\langle \sigma_{aa\to\chi\bar{\chi}} v \right\rangle n_a^2 - \left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle n_\chi^2 \\
&\quad + \sum_{i,j,k} \left(\left\langle \sigma_{\chi\bar{\chi}\to ijk} v \right\rangle \left(\left(n_\chi^{\mathrm{eq}}\right)^2 - n_\chi^2 \right) + \left\langle \sigma_{\chi\bar{\chi} i\to jk} v^2 \right\rangle n_i^{\mathrm{eq}} \left(\left(n_\chi^{\mathrm{eq}}\right)^2 - n_\chi^2 \right) \right), \\
\frac{\mathrm{d}n_a}{\mathrm{d}t} + 3Hn_a &= \sum_{i,j,k} \left\langle \sigma_{ia\to jk} v \right\rangle \left(n_a^{\mathrm{eq}} n_i^{\mathrm{eq}} - n_a n_i^{\mathrm{eq}} \right) + \left\langle \Gamma_a \right\rangle \left(n_a^{\mathrm{eq}} - n_a \right) \\
&\quad - \left\langle \sigma_{aa\to\chi\bar{\chi}} v \right\rangle n_a^2 + \left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle n_\chi^2 .
\end{aligned} \tag{16}$$

avec i, j, k des particules du modèle standard qui participent aux processus de changement de nombre de l'ALP (la contribution dominante provenant du processus *t-channel* $i = g$, $j = \bar{k} = t$). Les processus $2 \to 2$ ou $2 \to 3$ entrent par la section efficace typique moyennée thermiquement $\langle \sigma v \rangle$ et la désintégration (inverse) du pseudoscalaire par le taux de désintégration moyenné thermiquement, $\langle \Gamma_a \rangle$ [23]. Nous avons appliqué le principe d'équilibre lorsque cela était approprié. La résolution de cet ensemble d'équations pour une gamme de couplages $g_{a\chi\chi}$ et $g_{a\chi\chi} \cdot g_{aff}$ donne le diagramme de phase de la figure 2, la ligne en pointillé indiquant la combinaison de couplages qui conduit à la densité relique de matirère sombre observée pour des masses de valeur $m_\chi = 10$ GeV et $m_a = 1$ GeV. Pour une théorie particulière avec un certain contenu en particules et diverses interactions, il est générique que différents types de mécanismes puissent fixer l'abondance de matière sombre en fonction du type d'interaction le plus important ; notre modèle donne donc lieu à une riche phénoménologie. En fait, comme il a été montré dans [21] et plus tard dans [22], le cas du médiateur massif conduit à cinq mécanismes différents de production dynamique de matière sombre. Dans ce travail, nous nous concentrerons sur les scénarios où les particules de matière sombre sont thermiquement découplées du bain du modèle standard. En bas à gauche du diagramme de phase de la figure 2, correspondant à un petit $g_{a\chi\chi}$ et un petit produit $g_{aff} \cdot g_{a\chi\chi}$, les particules de matière sombre ne sont pas en équilibre, ni avec les particules du modèle standard, ni avec les ALPs. Cependant, le couplage fermion-ALP est suffisamment grand pour que les ALPs et les particules du modèle standard s'équilibrent, $a \overset{T}{\longleftrightarrow}$ SM. La matière sombre est produite directement à partir de collisions de fermions du modèle standard, $f\bar{f} \to \chi\bar{\chi}$. Ce régime est donc appelé *freeze-in via le modèle standard*. En augmentant $g_{a\chi\chi}$ tout en gardant

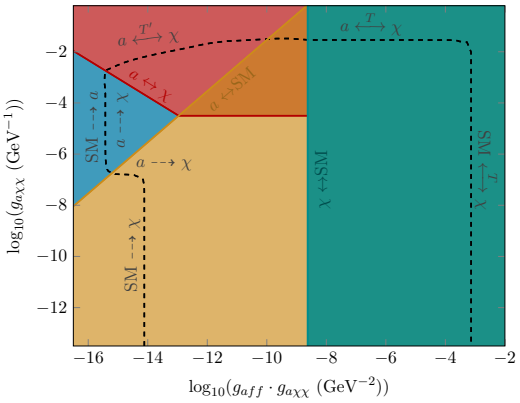


FIGURE 2: Couplage ALP-DM $g_{a\chi\chi}$ en fonction du produit du couplage ALP-DM et du couplage ALP-fermion, $g_{a\chi\chi} \cdot g_{aff}$, pour m_χ = 10 GeV et m_a = 1 GeV. La ligne en pointillés correspond à la combinaison de couplages qui donne la densité de reliques DM correcte $\Omega_{\rm DM} h^2$ = 0.12 telle que mesurée par le télescope Planck.

le produit $g_{aff} \cdot g_{a\chi\chi}$ fixe, $aa \dashrightarrow \chi\bar{\chi}$ devient le processus dominant. Etant conceptuellement très similaire au mécanisme précédent, ce mécanisme est appelé *freeze-in via le médiateur*. Notons qu'il existe deux possibilités, le *freeze-in* IR pour lequel la densité relique finale est insensible à la température de rechauffage, ou le *freeze-in* UV pour lequel le choix de la température de rechauffage influence le résultat, en fonction de la dépendance T de la section efficace thermique. Si nous continuons à diminuer le couplage fermion ALP g_{aff}, les interactions ALP-modèle standard deviennent trop rares pour que les deux secteurs se thermalisent. Pourtant, la quantité d'ALP produite est suffisante pour obtenir la densité de relique de matière sombre correcte via la chaîne de productions séquentielles hors équilibre SM $\rightarrow$ ALP et $aa \rightarrow \chi\bar{\chi}$; la matière sombre est produite via *le freeze-in séquentiel*. Cependant, comme les ALPs ne sont pas en équilibre cinétique, les équations de Boltzmann non intégrées décrivant l'évolution de la distribution de l'espace des phases des ALPs doivent être étudiées pour une analyse quantitative fiable [20], ce qui rend le calcul de l'abondance finale de matière sombre non trivial. En augmentant encore le couplage du secteur caché $g_{a\chi\chi}$, à un moment donné, l'interaction DM-ALP devient si forte qu'ils se thermalisent l'un l'autre. En particulier, la matière sombre et les ALPs partageront une température commune T' qui est différente et en général beaucoup plus petite que la température des photons, T. Nous devons donc évaluer leur section efficace d'interaction à leur température commune. Plus précisément, nous calculons la température du secteur caché T' à partir de sa densité d'énergie ρ' qui peut être obtenue en résolvant une équation de Boltzmann supplémentaire décrivant la quantité de transfert d'énergie du bain des particules du modèle standard vers le secteur caché. Nous devons donc résoudre numériquement un système de trois équations différentielles couplées (rigides). La densité relique de la matière sombre est déterminée par les interactions $aa \leftrightarrow \chi\bar{\chi}$ et le mécanisme ressemble au *freeze-out* ordinaire mais se produit à une température différente. Nous appelons donc ce régime *freeze-out découplé (DFO)*. Il convient de noter que, comme pour le *freeze-in*, les contributions ultraviolettes aux interactions SM-DM et SM-ALP peuvent indirectement introduire une dépendance de la température de rechauffage via le transfert d'énergie du modèle standard. Nous avons vérifié explicitement que pour des températures de rechauffage inférieures à $\mathcal{O}$(TeV), les effets sur l'abondance finale de matière sombre sont faibles. Plus précisément, la variation de $T_{\rm RH}$ de 200 GeV à 2000 GeV conduit à un changement de ~ 22% dans l'abondance de matière sombre et de ~ 7% dans la valeur requise de $g_{a\chi\chi}$. Nous les négligeons donc. Cependant, nous ajoutons des corrections à température finie à

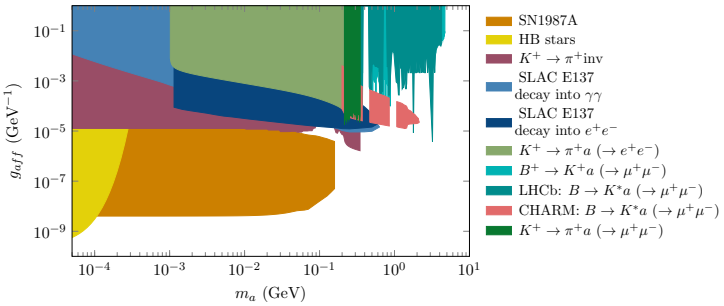


FIGURE 3: Tracé de la contrainte complète sur le couplage ALP-fermion g_{aff} en fonction de la masse de l'ALP m_a pour un ALP qui se couple uniquement aux fermions du modèle standard au tree-level.

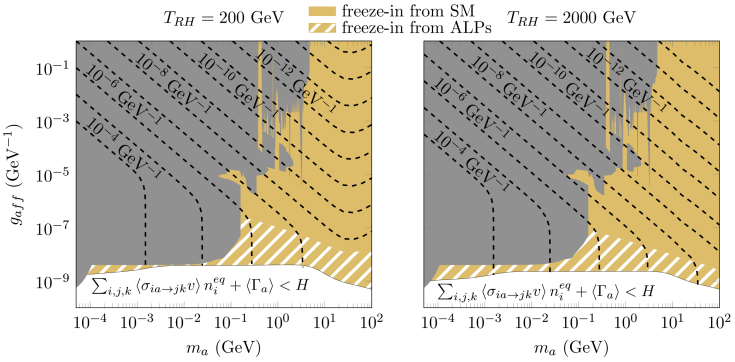


FIGURE 4: Couplage ALP-fermion g_{aff} en fonction de m_a pour les lignes de couplage constant du secteur caché $g_{a\chi\chi}$ (en noir pointillé) qui reproduisent la densité relique de matière sombre observée via le *freeze-in via le modèle standard* et le *freeze-in via l'ALP*, comme indiqué pour les températures de rechauffage T_{RH} = 200 GeV (à gauche) et T_{RH} = 2000 GeV (à droite). Nous avons fixé le rapport m_χ/m_a = 10. La ligne inférieure indique la valeur de g_{aff} pour laquelle les ALP et le SM atteignent l'équilibre. En gris, nous avons inclus les contraintes relevantes pour notre modèle ALP sur le couplage g_{aff} dans cette région de paramètres (cf. figure 3).

la production d'ALP et de matière sombre à partir des particules du modèle standard, ce qui se traduit par des corrections d'ordre un.

Enfin, dans la partie droite du diagramme de phase, les trois secteurs partagent une température commune et la matière sombre est produit par *freeze-out* thermique.

Implications phénoménologiques Enfin, nous explorons les implications phénoménologiques de notre modèle en le comparant aux expériences. La Fig. 3 résume les limites de *beam dump*, astrophysiques et de physique des saveurs sur les ALPs que nous avons calculées pour notre modèle spécifique, à savoir lorsque l'ALP se couple uniquement aux fermions du modèle standard au *tree-level* et que les couplages de bosons ALP-gauge sont induits par des boucles de fermions. Les contraintes astrophysiques provenant du sursaut de neutrinos de SN1987A ou du rapport entre le nombre d'étoiles à branche horizontale et de géantes rouges dans les amas globulaires reposent sur des arguments de perte d'énergie [24, 25]. L'expérience de *beam dump* d'électrons SLAC a recherché des axions et n'a observé aucun événement [26]. En ce qui concerne les contraintes de saveur, nous dérivons des bornes pour les canaux les plus contraignants : $K^+ \to \pi^+ a(\to \text{inv.})$ de NA62, $K^+ \to \pi^+ a(\to \mu^+\mu^-)$ et $K^+ \to \pi^+ a(\to e^+e^-)$ de NA48/2, $B^+ \to K^+ a(\to \mu^+\mu^-)$ de LHCb et $B \to K^* a(\to \mu^+\mu^-)$ de LHCb et CHARM.

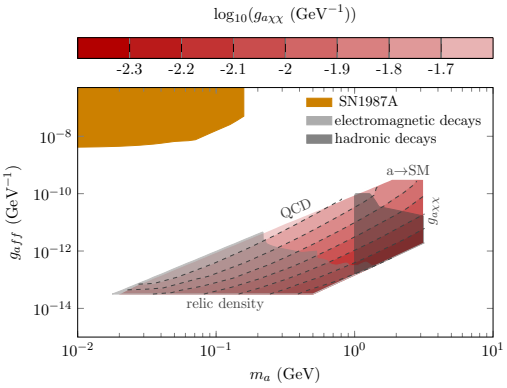


FIGURE 5: Courbe de niveau des couplages du secteur caché $g_{a\chi\chi}$ dans le plan (m_a, g_{aff}) qui donnent la densité relique de matière sombre observée via le *freeze-out* découplé (DFO). Nous avons fixé le rapport m_χ/m_a = 10. Nous avons inclus les contraintes relevantes pour notre modèle ALP sur le couplage de connecteur g_{aff} dans cette région de paramètres (cf. figure 3). Nous indiquons également la région exclue par les contraintes cosmologiques des particules supplémentaires se désintégrant électromagnétiquement (gris clair) et se désintégrant hadroniquement (gris foncé). Pour des explications sur les contraintes des désintégrations électromagnétiques et hadroniques, voir le texte principal.

Comme nous l'avons vu précédemment, les contributions UV à la densité relique de matière sombre provenant des processus $2 \to 3$ médiés par l'ALP $f\bar{f} \to h\chi\bar{\chi}$, $fh \to f\chi\bar{\chi}$ et $\bar{f}h \to \bar{f}\chi\bar{\chi}$ deviennent importantes pour des températures de rachauffage supérieures à quelques centaines de GeV. Ces contributions introduisent une dépendance en la température de rechauffage puisque l'abondance relique final s'échelonne avec T_{RH}, de telle sorte que nous considérons deux scénarios représentatifs : T_{RH} = 200 GeV, où les contributions UV sont presque négligeables ; l'abondance de matière sombre sera donc approximativement indépendante de T_{RH} et $T_{RH} \sim 2000$ GeV, où le résultat est dominé par les contributions UV. Dans la figure 4, les régions jaune et jaune hachurée représentent la plage de g_{aff} et m_a où la densité relique correcte est obtenue, pour différentes valeurs de $g_{a\chi\chi}$, montrant également les contours pour différentes valeurs du couplage du secteur caché (lignes noires en pointillé) pour ces deux scénarios. Dans la région jaune, la production de matière sombre par diffusion de fermions du modèle standard est plus efficace (*freeze-in via le modèle standard*), et dans la région jaune hachurée, c'est la production par diffusion d'ALP qui est dominante (*freeze-in via l'ALP*). Notons que pour T_{RH} = 200 GeV et $m_\chi > m_t$, la suppression par un facteur de Boltzmann des particules produisant principalement la matière sombre via $t\bar{t} \to \chi\bar{\chi}$ se traduit par une courbure vers le haut des lignes de couplage constant du secteur caché. Dans le cas dominé par des proccesus ultraviolets, les lignes de couplage constant $g_{a\chi\chi}$ sont déplacées vers des valeurs plus petites de g_{aff} puisque plus de matière sombre est produits à des moments plus précoces. On trouvera plus de détails et une discussion dans le Chap. 7 ou dans la Réf. [27].

Dans le régime DFO, contrairement aux régimes de *freeze-in*, $g_{a\chi\chi}$ est grand pour assurer l'équilibre entre les particules du secteur caché. Ceci est visible dans la figure 5 où nous montrons la région de l'espace des paramètres où la densité relique correcte est obtenue dans le plan (m_a, g_{aff}), en gardant le rapport m_χ/m_a = 10 fixe, ainsi que les contraintes. Notons que les limites de cette région sont décrites dans le Chap. 7 ou dans [27]. On observe qu'il est difficile de tester cette région avec des recherches d'ALP dans les collisionneurs : en particulier, la limite SN et les limites des désintégrations rares B et K devraient être améliorées de plusieurs ordres de grandeur. D'autre part, la cosmologie pourrait tester cette région car l'ALP a une longue durée de vie en raison du faible couplage avec les particules du modèle standard. Comme les

ALPs sont produits en abondance avec les particules du modèle standard, leur désintégration a des implications importantes pour l'histoire cosmologique au moment de la nucléosynthèse du big bang. Pour les ALPs de masse $m_a \lesssim 2m_\mu$, nous appliquons les limites de la Réf. [28] sur les ALPs à très longue durée de vie dans la gamme de masse inférieure au GeV, à l'exclusion des durées de vie $\tau_a \sim 10^3 - 10^5$ s (limite étiquetée "electromagnetic decays"). Pour des masses de $2m_\mu \lesssim m_a \lesssim 1$ GeV, l'ALP se désintègre principalement en muons, et ici la durée de vie est suffisamment courte pour qu'ils n'aient pas d'importance. Pour les désintégrations hadroniques, les limites deviennent plus sévères et des durées de vie supérieures à $\tau \sim 0,1$ s peuvent être exclues [28] (limite étiquetée "hadronic decays"). On trouvera plus de détails dans le Chap. 7 ou dans la Réf. [27].

Conclusion Notre modèle d'un ALP agissant comme médiateur entre les particules du modèle standard et les particules de matière sombre engendre une riche phénoménologie : En fonction des masses et des couplages de l'ALP et de la matière sombre, divers mécanismes sont possibles pour reproduire la densité relique de matière sombre mesurée aujourd'hui. Le calcul de la densité relique dans la région du *freeze-out découplé* est particulièrement complexe. Nous effectuons un calcul détaillé de cette région dans ce modèle de matière sombre médié par l'ALP. Dans le cas du *freeze-in*, la densité relique est insensible à la température de rechauffage si elle est inférieure à quelques centaines de GeV. Nous présentons également de tout nouveaux calculs des contraintes sur l'ALP, en particulier pour les SN, le beam-dump et les contraintes de saveur qui sont cruciales pour l'espace des paramètres d'intérêt. De vastes régions dans les régimes de *freeze-in* et de DFO restent inexplorées, et alors que le premier pourrait être sondé par de futures expériences de collisionneurs, le second serait difficile d'accès. Cependant, la cosmologie peut nous aider à contraindre cette région, et nous réalisons une première estimation de ces limites, qui devrait être améliorée à l'avenir. Plusieurs autres améliorations de notre analyse seraient également avantageuses : Tout d'abord, nous aimerions améliorer la précision de nos calculs dans la région de *freeze-in séquentiel*, en résolvant l'équation de Boltzmann non intégrée. Ensuite, nous avons l'intention d'analyser plus en détail les corrections thermiques. Enfin, nous aimerions procéder à une évaluation approfondie de la sensibilité potentielle des expériences futures à la région concernée et nous attendons avec impatience de voir l'impact des futurs résultats expérimentaux sur notre modèle.

Abbreviations

AIC	Akaike Information Criterion
ALP	Axion-like Particle
BSM	Beyond the Standard Model
CM	Centre-of-mass
DFO	Decoupled Freeze-Out
DOF	Degree Of Freedom
DM	Dark Matter
EW	Electroweak
FIMP	Feebly Interacting Massive Particle
FV	Finite Volume
FP	Fixed Point
HMC	Hybrid Monte Carlo
HVP	Hadronic Vacuum Polarization
IB	Isospin Breaking
LPT	Lattice Perturbation Theory
MCMC	Markov Chain Monte Carlo
MS	Minimal Subtraction (scheme)
NLO	Next-to-leading order
NNLO	Next-to-next-to-leading order
ppb	part-per-billion
ppm	part-per-million
PT	Perturbation Theory
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
RHMC	Rational Hybrid Monte Carlo
RG	Renormalization Group
RT	Renormalized Trajectory
Sup.Inf.	Supplementary Information
SIB	Strong Isospin Breaking
SM	Standard Model
SSB	Spontaneous Symmetry Breaking
SXPT	Staggered Chiral Perturbation Theory
VEV	Vacuum Expectation Value
WIMP	Weakly Interacting Massive Particle
XPT	Chiral Perturbation Theory

Contents

Affidavit i

Liste de publications et participation aux conférences iii

Acknowledgements v

Abstract vii

Résumé ix

Résumé étendu xi

I General introduction 1

1 Motivation 3

2 QED and QCD in the standard model 9

2.1 Electroweak theory . . . 10

2.1.1 Construction of the Lagrangian . . . 10

2.1.2 The Higgs mechanism: recovering quantum electrodynamics . . . 11

2.2 Quantum Electrodynamics . . . 13

2.2.1 Prologue: Normalizing Renormalization . . . 13

2.2.2 Vacuum polarization and charge renormalization . . . 15

2.3 (Low energy) Quantum Chromodynamics . . . 21

2.3.1 The QCD Lagrangian . . . 21

2.3.1.1 Non-Abelian $SU(N_c)$ pure gauge theory . . . 22

2.3.2 The perturbative QCD β-function: asymptotic freedom . . . 22

2.3.3 QCD at low energies: The $U(1)_A$ problem, instantons and θ-vacua . . . 23

2.3.3.1 The heavy η' . . . 23

2.3.3.2 Instantons . . . 25

2.3.4 The strong CP problem . . . 27

2.3.4.1 The chiral anomaly . . . 27

2.3.4.2 Adding massive fermions: the strong CP problem . . . 28

II Testing the standard model via the electromagnetic coupling 31

3 Lattice Quantum Chromodynamics 33

3.1 Non-perturbative regularization . . . 34

3.2 Lattice fermion action . . . 35

3.2.1 Fermion doubling problem . . . 36

3.2.2 Kogut–Susskind (staggered) quarks . . . 39

3.3 Lattice gauge action . . . 44

3.4 Quantization and Euclidean path integral of lattice QCD . . . 46
3.4.1 Gauge measure . . . 46
3.4.2 (Fermionic) path integral . . . 46
3.4.3 Relation to statistical mechanics . . . 47
3.5 Example: Hadron Spectroscopy . . . 48
3.6 Continuum limit . . . 49
3.6.1 Renormalization group transformation, fixed point and behaviour near the fixed point . . . 50
3.6.2 Universality, asymptotic scaling, taking the continuum limit . . . 51
3.6.3 Symanzik effective theory . . . 54
3.6.4 Recovering full QCD . . . 56
3.7 Electromagnetism on the lattice . . . 57
3.7.1 QED in finite volume . . . 57
3.8 Numerical methods . . . 60
3.9 Analysis . . . 64
3.9.1 Statistical error . . . 64
3.9.2 Systematic error . . . 65

4 The running of the electromagnetic coupling 67
4.1 Introduction . . . 67
4.2 The dispersive approach: relating the HVP to $e^+e^- \to$ hadrons cross-section data 70
4.2.1 Analyticity: dispersion relations and vacuum polarization . . . 70
4.2.2 Unitarity: the optical theorem and the R-ratio . . . 72
4.3 Lattice QCD calculation of the hadronic running of α . . . 75
4.3.1 Lattice details . . . 76
4.3.1.1 Ensembles . . . 76
4.3.1.2 Staggered electromagnetic currents of quarks . . . 77
4.3.1.3 Definition of the hadronic vacuum polarization on the lattice . . 81
4.3.1.4 Definition of observables . . . 83
4.3.1.5 Renormalization of non-conserved currents . . . 86
4.3.1.6 QED and strong isospin breaking corrections . . . 87
4.3.1.7 Calculation of various isospin-breaking components . . . 90
4.3.1.8 Blinding . . . 91
4.3.2 Analysis strategy . . . 91
4.3.2.1 Challenges at short distances: discretization errors . . . 92
4.3.2.2 Long-distance taste-breaking corrections . . . 99
4.3.2.3 Finite-volume corrections . . . 101
4.3.2.4 Global fit procedure and definition of the physical point . . . 104
4.3.3 Lattice results . . . 114
4.3.3.1 Connected light contribution . . . 114
4.3.3.2 Strange contribution . . . 120
4.3.3.3 Connected charm contribution . . . 123
4.3.3.4 Disconnected contribution . . . 126
4.3.3.5 Hadronic running of α up to 8 GeV^2: lattice results . . . 127
4.3.4 Comparison with previous results . . . 130
4.4 Conclusion . . . 133

III Going beyond with axion-mediated dark matter 137

5 Dark Matter 139
5.1 A Brief Overview of Cosmology . . . 139
5.2 A brief history of the universe: gathering evidence for Dark Matter . . . 141

5.2.1 What we know about the missing mass . . . 148
5.2.2 Why the standard model is not enough . . . 149
5.3 Detection strategies: shake it, make it, break it, brake it? . . . 149
5.4 Bake it! . . . 155
5.4.1 Boltzmann equation . . . 155
5.4.2 Freeze-out and WIMPs . . . 160
5.4.3 Freeze-in . . . 162
5.4.3.1 Infrared dominated freeze-in . . . 162
5.4.3.2 Ultraviolet dominated freeze-in . . . 163

6 Axions and axion-like particles 167
6.1 The axion solution to the strong CP problem . . . 167
6.2 DFSZ model . . . 169
6.3 A note on anomalies . . . 170
6.4 Properties of the axion at low energies . . . 171
6.5 Axion-like Particles . . . 173

7 Axion-like particles as mediators for dark matter 175
7.1 Introduction . . . 175
7.2 Particle content, interactions and relevant processes . . . 176
7.3 The phase diagram of production mechanisms . . . 178
7.3.1 General Boltzmann equations . . . 178
7.3.2 Freeze-in . . . 180
7.3.2.1 Freeze-in from SM particles . . . 181
7.3.2.2 Freeze-in from ALPs . . . 182
7.3.2.3 Sequential freeze-in . . . 182
7.3.3 Decoupled freeze-out (DFO) . . . 183
7.4 Finite temperature effects . . . 188
7.5 Testing the model: current constraints on ALPs . . . 189
7.5.1 Astrophysical constraints . . . 189
7.5.1.1 Horizontal branch stars . . . 189
7.5.1.2 Supernova SN1987A . . . 191
7.5.2 Collider constraints . . . 192
7.5.2.1 SLAC E137 electron beam dump experiment . . . 192
7.5.2.2 Exotic Higgs decays at the LHC . . . 195
7.5.2.3 Production in decays of mesons . . . 196
7.5.3 Cosmological constraints . . . 197
7.5.4 Constraints on the DM . . . 198
7.5.4.1 DM relic density constraint . . . 198
7.5.4.2 (No) constraints from dark matter phenomenology . . . 198
7.6 Results . . . 200
7.6.1 Collider and astrophysical constraints . . . 201
7.6.2 Phase diagram . . . 203
7.6.3 Freeze-in region . . . 205
7.6.4 DFO region and cosmological constraints . . . 208
7.7 Conclusion . . . 212

IV Outlook 215

Appendices 219
A Compilation of current precision . . . 223

B Staggered leading order lattice perturbation theory . . . 224
B.1 Current correlators . . . 224
B.2 HVP and (discrete) Adler function . . . 227
B.3 Coefficients of the leading logarithmically-enhanced cutoff effects . . . 230
C Continuum results in leading order perturbation theory . . . 232
C.1 Current correlator . . . 232
C.2 HVP and (discrete) Adler function . . . 233
D Removal of an additional discretization error . . . 233
E Continuum extrapolations and distributions . . . 234
F Effective ALP-DM Lagrangian from a two Higgs doublet model . . . 243
G Cross sections and input for Boltzmann equations . . . 247
G.1 Hidden sector process . . . 248
G.2 Connector processes . . . 248
G.3 SM-ALP processes . . . 248
G.4 Decay width of the ALP . . . 249
G.5 DM self-interaction cross sections . . . 251
H Axion production cross sections in SLAC E137 experiment . . . 251
H.1 Primakoff production . . . 251
H.2 Bremsstrahlung . . . 252
H.3 Positron annihilation . . . 253
I Details on the ALP's detection probability . . . 254
J Details on constraints from exotic Higgs decays . . . 256
J.1 Loop diagrams relevant for constraint calculations . . . 257

Part I

General introduction

1 Motivation

The standard model (SM) of particle physics is a quantum field theory that has been spectacularly successful. It has provided a mathematical framework for understanding the properties of the known fundamental particles and three among the four known interactions. It also successfully predicted the existence of elementary particles such as the top quark [29, 30] and the Higgs boson [31]. In addition, there is strong experimental evidence that the SM's quantum chromodynamics (QCD) provides an accurate description of the strong interactions. While the SM has resisted multiple experimental tests over the last fifty decades, it does not provide an explanation for several observed phenomena, such as neutrino masses, the hierarchy problem, the origin of flavour, baryon asymmetry and the origin of the "dark components" in the universe, dark energy and dark matter.

One way to search for phenomena which go beyond the current paradigm is of course by building bigger accelerators in order to directly produce yet unknown particles.

On the theoretical side, there are usually two main lines of research. One approach to uncover new physics consists in trying to fault the SM by comparing very precise predictions of this model to equally precise measurement. Alternatively, one can construct models for unexplained phenomena and explore the consequences of these new theories. In this thesis we will explore both lines of research.

The search for new physics "at the intensity frontier" relies on increasing the luminosity rather than the energy scale of experiments. Since there is still no conclusive direct or indirect evidence for new physics in such experiments, we expect its effects to be very small. Thus, detecting a deviation from the SM requires both the theoretical prediction and the experimental measurement to be extremely precise. The first approach hence relies on providing these highly accurate theoretical predictions. Their precision is often limited by hadronic effects due to the non-linearity of QCD at low energies, thus requiring a reliable description of these non-perturbative processes.

A prime example of the search for new physics at the intensity frontier is the anomalous magnetic moment of the muon, a_μ. Its value has both been measured and calculated to unprecedented precision. The theoretical prediction includes up to five-loop corrections for the dominant contribution coming from quantum electrodynamics (QED) [32]. While the hadronic contribution is about 17.000 times smaller in size, its uncertainty contributes 87% to the error-squared on a_μ. The traditional approach to calculating this contribution is based on the use of a dispersion relation and low-energy measurements of hadronic production in e^+e^--annihilation [4, 5]. Already the measurement of a_μ at Brookhaven National Laboratory (BNL) published in 2006 [33]

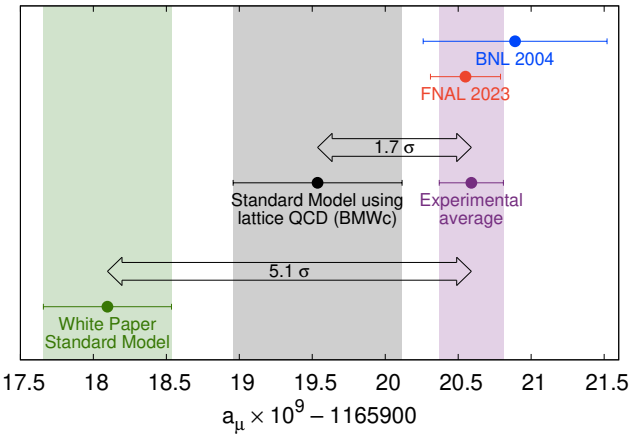


FIGURE 1.1: Comparison between experimental measurements [33, 34, 36] and standard model predictions of the muon magnetic anomaly. The calculation using the traditional approach (White Paper standard model) [35] gives a significantly lower value (5.1 standard deviations) than the weighted combination of the experimental measurements from Brookhaven National Laboratory (BNL 2004) and Fermilab (FNAL 2023). The results of this calculation were published in a journal article by the Muon g-2 Theory Initiative in 2020. The difference is reduced, to just 1.7 standard deviations, when the result of the lattice QCD calculation on the hadronic contribution (BMWc) replaces that of the traditional approach.

pointed to a disagreement with the SM prediction, and over the last twenty years the muon $g-2$ has become one of the most promising new physics signals.

During the course of this thesis, in April 2021, the first result from a new experiment at Fermilab [34] confirmed the BNL measurement and increased the deviation with the SM reference prediction to 4.2σ [34, 35]. The Fermilab release was however accompanied by the simultaneous publication, on April 7, of the ab-initio result from a lattice QCD calculation of the leading hadronic contribution to the muon $g-2$ by the BMW collaboration with comparable accuracy to the data-driven approach [7]. This result predicted a much smaller disagreement, suggesting that new physics might not be needed to explain the difference. However, it is at odds with the prediction from the traditional approach.

This summer Fermilab released their runs 2 & 3 results [36], significantly reducing their uncertainties and aggravating the tension with the prediction of the traditional approach to more than 5σ, which is the threshold commonly quoted to claim the discovery of as yet unknown fundamental physics [36]. However, several other independent lattice calculations now confirm the BMW result for a sub-quantity where lattice methods are expected to work particularly well [7, 37–40]. The current situation is summarized in Fig. 1.1. To make the situation even more puzzling, the CMD-3 collaboration very recently published new measurements of the hadronic cross section in the two-pion region, indicating that its prediction could be close to that of the lattice calculations. However, it disagrees with other (and also their own) previous measurements [41]. Therefore, the issue of whether new physics is necessary to account for the muon $g-2$ discrepancy remains rather ambiguous. Notably, the focus has shifted to the disagreement between the two standard model predictions [16].

In this thesis we will use lattice QCD simulations to calculate the hadronic contribution to a closely related SM precision observable, the running of the electromagnetic coupling α. It is determined by the same hadronic vacuum polarization (HVP) function that enters the leading strong-interaction contribution to the muon anomaly. This function is related to the hadronic

contribution to the photon self-energy that quantifies how much a photon of virtuality q^2 polarizes quarks and gluons in the vacuum.

Lattice field theories provide a powerful tool to access the non-linear low-energy regime of QCD. The lattice approach relies on a numerical evaluation of the Euclidean QCD path integral. This is made possible by discretizing space-time in a finite volume, hence reducing the problem to a finite (yet extremely large) number of variables. The numerical evaluation requires the use of large-scale supercomputers. Lattice calculations suffer from systematic errors but these can be quantified and controlled. In particular, one main focus of this thesis is to ensure a reliable continuum extrapolation and determination of the associated systematic errors.

Besides its importance for improving our understanding of the current disagreement with the dispersive approach in the low energy hadronic region, the running of the electromagnetic coupling plays an important role in electroweak precision tests of the SM. As all SM input parameters have experimentally been measured since the discovery of the Higgs boson in 2012 [31], global comparisons of electroweak measurements to theoretical predictions allow for consistency checks between experiment and theory [1, 42, 43].

While there are already slight tensions at the EW scale, for instance between the experimentally measured value for the Higgs mass and that indirectly determined from global fits to the remaining EW precision observables [44], it has been argued that the increase in the SM prediction for a_μ obtained from the lattice calculation of the HVP would drastically augment this discrepancy [45–49]. While this is disfavoured by both a first investigation of the BMW collaboration [7] and a recent result by the Mainz collaboration [8], consequences for electroweak fits remain to be understood. Moreover, future colliders are expected to significantly reduce the uncertainty on the indirect determination of the running of α and the precision in the theory prediction of the hadronic contribution to this running will have to be halved to fully harness the discovery potential of those instruments [3].

The second approach to search for new physics consists in extending the established model, based on well-motivated scenarios, and in exploring the extension's phenomenological implications and possible signatures in experiments. This is the strategy we will pursue in the second part of this thesis. We propose and investigate a particle physics solution to the dark matter (DM) puzzle, one of the major enigmas in physics. While there is compelling gravitational evidence that it exists, coming from a wide range of scales, no non-gravitational evidence of DM has been found so far. Accounting for about 27% of the total energy budget in the universe (dark energy making up about 68%), we have not yet been able to answer the question of what our universe is mainly made of. Observations support the idea that dark matter (DM) has a particle nature, that it is weakly coupled to SM particles and that it exhibits a small velocity dispersion. The cold DM paradigm forms one of the building blocks of the framework of the ΛCD model, the standard model of cosmology, which has accurately described cosmological and astrophysical observations over the past decades [2].

However, none of the known particles within the SM is a suitable candidate, thereby necessitating beyond the standard model (BSM) physics. Numerous extensions of the standard model have been proposed. One of the simplest scenarios, where the DM particle directly couples to SM states via renormalizable interactions, is one in which a so-called weakly interacting massive particle (WIMP) is produced via thermal freeze-out, with a mass at the electroweak scale and

a coupling on the order of that of the weak interaction. In the freeze-out scenario, one assumes that DM and SM particles were in equilibrium in the early universe. Once the expansion rate of the universe becomes bigger than the reaction rate, fewer and fewer particles will collide and interact until the number density stays unchanged. However the WIMP scenario is challenged by the fact that – despite experimental searches for decades – it has not been detected so far. On the contrary, a large parameter space is nowadays excluded by LHC data, indirect and direct detection experiments [50, 51]. Hence, one is led to wonder if different scenarios and different candidates are also possible, in particular non-thermal models which require weaker couplings and which are still compatible with the current null results. A very simple non-thermal scenario is the freeze-in mechanism and the associated DM candidate is a feebly interacting massive particle (FIMP) [19, 52]. Contrary to the freeze-out case, the DM particles interact so weakly with the SM particles that they never reach thermal equilibrium during the cosmic evolution. The initial abundance of DM particles is assumed to be negligibly small, and DM is produced from SM scatterings or decays in the thermal plasma. At a certain point, the abundance of SM particles gets suppressed and the DM freezes in [19, 52]. This mechanism can occur in various well-motivated extensions of the standard model. Alternative scenarios, in particular the freeze-in mechanism, have received much attention in recent years but are still less explored than the standard thermal freeze-out case. In this thesis we explore different DM production mechanisms in a simple extension of the SM.

We investigate the possibility for the interactions between SM and DM particles to be mediated by an additional particle allowing the DM particle to be a singlet of the SM gauge group. As a mediator we consider a light pseudoscalar, an axion-like particle (ALP), by which we mean the pseudo-Goldstone boson of an additional approximate U(1) global symmetry spontaneously broken at a high scale f_a. The axion was originally proposed to solve a problem within the theory of quantum chromodynamics and constitutes itself a pioneering example of a non-thermally produced DM candidate [53–56]. As is generically the case for a particular theory with a variety of interactions, different types of mechanisms can set the DM relic abundance, depending on which interactions are more important. Hence our model gives rise to a rich phenomenology. The size of the couplings is driven by the new symmetry breaking scale f_a which should be large enough to be in accord with current collider limits. Therefore the new couplings are small and we focus on parameter regions where the DM particles are not in thermal equilibrium with the SM bath, such as freeze-in and decoupled freeze-out. The pseudoscalar mediator case is largely insensitive to direct or indirect DM searches. However, the axion/ALP mediators can have many implications for astrophysical, cosmological and collider processes and we constrain our model via measurements of those processes.

Besides, axions or axion-like particles have been put forward as a beyond-the-standard-model (BSM) explanation for the muon $(g-2)$ puzzle [57–62] and the existence of such a particle might even reconcile the predictions for the muon $g-2$ based on the R-ratio and the lattice [63].

This thesis treats therefore two very different subjects and is divided into two distinct parts. It is structured as follows:

- In Chap. 2 I give, as an introduction common to both topics, a broad overview of quantum electrodynamics and quantum chromodynamics in the SM. We briefly discuss the perturbative running of the electromagnetic coupling in QED, whose hadronic contibution

we address non-perturbatively in Chap. 4. We then focus on the low energy properties of quantum chromodynamics and discuss how the solution of the so-called U(1) problem introduces a new problem, the strong CP problem for which axions were proposed as a solution.

- In the first part of this thesis we make use of ab-initio lattice QCD calculations to access the non-linear low-energy regime of QCD. In Chap. 3 I therefore give an introduction to lattice quantum chromodynamics which, until now, is the only known way to make physical predictions from first principles in that regime. I discuss the lattice discretization, the intrinsic difficulties it holds and put an emphasis on the importance the renormalization group played in its development. Since we include leading order isospin breaking effects in our calculations I also present different possible formulations of QED in a finite volume.

- Chap. 4 presents the main results of the first part of this thesis. I first provide a summary of the phenomenological status and importance of the running of the electromagnetic coupling and discuss the reference method based on a dispersive integral and $e^+e^- \to$ hadrons cross-section data. I then present the methodology employed to compute the running with lattice QCD, including leading-order isospin-breaking effects. Finally, in Sec. 4.3.3 I present our lattice results that I also compare to previous results in Sec. 4.3.4 and conclude in Sec. 4.4.

- Chap. 5 gives an overview of the theoretical aspects discussed in the second part of this thesis. We start with a brief introduction to the gravitational evidence for DM and proceed by presenting the basic principles of DM detection modes in Sec. 5.3. We then discuss the formalism/framework to calculate the evolution of the DM relic density where we focus on freeze-out, which is the standard DM production mechanism, and freeze-in, an alternative which requires weaker couplings to the SM particles.

- Since the interactions between the SM and the DM particles in our model are mediated by axion-like particles I briefly introduce in Chap. 6 the theoretical properties of axions. I also make a brief comment about anomalies which motivates the UV completion of our particular model.

- Chap. 7 is dedicated to the main results of the second part of this thesis. We start by discussing the theoretical basis for our DM model in Sec. 7.2, followed by a discussion of a selection of collider and astrophysical constraints on axion-like particles in Sec. 7.5. Sec. 7.6 summarizes our results and we conclude in Sec. 7.7.

- Finally, in part IV we conclude this thesis and discuss future prospects for both topics treated in this work.

2 QED and QCD in the standard model

The standard model of particle physics describes the dynamics of all known particles and their interactions, except for gravity. The gauge symmetry of the SM is the product of the colour group (the symmetry group of QCD) and the electroweak symmetry group which breaks spontaneously via the Higgs mechanism

$$SU(3)_C \times SU(2)_L \times U(1)_Y \to SU(3)_C \times U(1)_{\text{em}} \,. \tag{2.1}$$

Adding all fermion fields, a scalar Lagrangian for the Higgs mechanism in the electroweak sector and Yukawa couplings to endow the fermions with masses, the SM Lagrangian can be summarized as [64]

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_H + \mathcal{L}_Y \,, \tag{2.2}$$

where $\mathcal{L}_g$ includes the dynamics of all gauge bosons, $\mathcal{L}_H$ contains the Higgs scalar field, $\mathcal{L}_f$ includes the Dirac Lagrangians for leptons and quarks and $\mathcal{L}_Y$ their Yukawa interactions with the Higgs field.

The standard model postulates the existence of six leptons and six quarks and their corresponding antiparticles (fermions), four gauge bosons which are the mediators of the electroweak and the strong interactions, and the Higgs boson. The particle content and the particles' quantum numbers are summarized in Tab. 2.1. Since the discovery of the Higgs in 2012, the existence of the known SM particles has experimentally been confirmed [31].

Particle(s)	Field(s)	Content	Spin	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	$q = I^3 + Y/2$
Quarks	Q_i	$(u,d)_L$	1/2	$\mathbf{3}$	$\mathbf{2}$	1/3	$(2/3, -1/3)$
(Three generations)	u_{Ri}	u_R	1/2	$\overline{\mathbf{3}}$	$\mathbf{1}$	4/3	2/3
	d_{Ri}	d_R	1/2	$\overline{\mathbf{3}}$	$\mathbf{1}$	$-2/3$	$-1/3$
Leptons	L_i	$(\nu_e, e)_L$	1/2	$\mathbf{1}$	$\mathbf{2}$	-1	$(0, -1)$
(Three generations)	l_{Ri}	e_R	1/2	$\mathbf{1}$	$\mathbf{1}$	-2	-1
Gluons	G^a_μ	g	1	$\mathbf{8}$	$\mathbf{1}$	0	0
W bosons	$W^{1,2}_\mu$	$W^\pm$	1	$\mathbf{1}$	$\mathbf{3}$	0	± 1
Photon, Z boson	W^3_μ, B_μ	γ, Z^0	1	$\mathbf{1}$	$\mathbf{3}, \mathbf{1}$	0	0
Higgs boson	ϕ	H	0	$\mathbf{1}$	$\mathbf{2}$	1	0

Table 2.1: Field content of the standard model with the corresponding particles, their spin, and the field's representation. $SU(N)$ representations are labelled by their dimension n, with a bar for the conjugate representation. U(1) representations are labelled by the charge. The electric charge $Q = eq$ (in units of e) is given by the combination of the third component of the weak isospin I^3 and hypercharge Y of each of the fermions via the Gell-Mann-Nishijima formula in the last column.

2.1 Electroweak theory

This chapter is intended to give a brief idea of the construction of the standard model. We focus here on the electroweak sector and the Higgs mechanism to recover quantum electrodynamics (QED). In Sec. 2.3 we will discuss quantum chromodynamics, the theory of strong interactions.

2.1.1 Construction of the Lagrangian

For a gauge theory, the theory should be symmetric under a *local* gauge group $\mathcal{G}$ which acts independently at all locations in space-time. We will focus on a non-abelian part of the SM gauge group, namely $\mathrm{SU}(2)_L$, but try to keep the discussion general. The concept works analogously for the gauge group of the strong interactions, $SU(N_c)$, and the abelian $\mathrm{U}(1)_Y$ is then a special case of this. Let us start from the Lagrangian for a free massive fermion field (we only consider one flavour for this discussion)

$$\mathcal{L} = \overline{\psi}(x)(i\gamma_\mu\partial^\mu - m_f)\psi(x) \,, \tag{2.3}$$

with m_f the mass of the quark and where two Dirac fermions are grouped into the doublet ψ

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix} . \tag{2.4}$$

Eq. (2.3) is the most general Lagrangian up to dimension 4 which is invariant under Lorentz transformation and global $\mathrm{SU}(2)_L$ transformations. If we now require our theory to be invariant under local, space-time dependent $\mathrm{SU}(2)_L$ transformations

$$\psi(x) \to \Omega(x)\psi(x) \,, \qquad \Omega(x) \in \mathrm{SU}(2)_L \,, \tag{2.5}$$

the derivative in Eq. (2.3) is no longer well defined under a $\mathrm{SU}(2)_L$ transformation. Indeed,

$$\overline{\psi}\,\partial_\mu\psi \to \overline{\psi}\,\partial_\mu\psi + \overline{\psi}\,\Omega^\dagger(x)\left[\partial_\mu\Omega(x)\right]\psi \tag{2.6}$$

is not invariant under a gauge transformation since $\partial_\mu\Omega(x) \neq 0$ for a generic $\Omega(x)$. Since we want to keep our field dynamical we cannot get rid of the derivative. Instead, we can introduce an auxiliary field, a so-called *connection*, and replace the derivative by a covariant derivative containing this auxiliary field. From Eq. (2.6) we see that, in order for the Lagrangian to be invariant we need

$$D_\mu(\Omega\psi) \overset{!}{=} \Omega(D_\mu\psi) \,, \tag{2.7}$$

hence the covariant derivative of the group element has to vanish, $D_\mu\Omega = 0$, such that $D_\mu\psi$ transforms in the same way as ψ. Any element of $\mathrm{SU}(2)_L$ can be obtained by $\Omega = e^{ig\theta^a T_a}$ with θ^a real-valued and $T_a = T_a^\dagger$ the generators of the associated algebra which are hermitian and traceless and we have added a factor g. They satisfy the $\mathfrak{su}(2)$ Lie-algebra (the so-called Lie-bracket),

$$[T_a, T_b] = i f_{ab}{}^c T_c \,, \tag{2.8}$$

where f_{abc} are the structure constants of $\mathfrak{su}(2)$. The doublet ψ transforms in the fundamental representation and the usual choice for the SU(2) generators are the Pauli spin matrices $T_a = \sigma_a/2$. Hence $\partial_\mu \Omega = ig(\partial_\mu \theta^a) T_a \Omega$ and the covariant derivative can be constructed as

$$D_\mu \equiv \partial_\mu - ig\partial_\mu \theta^a T_a \ , \tag{2.9}$$

expressed only in terms of local data. It is common to define $\partial_\mu \theta^a \equiv A^a_\mu(x)$. $D_\mu \psi$ transforms as ψ under $\mathrm{SU}(2)_L$ transformations provided $A_\mu(x) \equiv T_a A^a_\mu$ transforms in a specific way, namely

$$A_\mu(x) \to \Omega(x) A_\mu(x) \Omega^{-1}(x) + \frac{i}{g} \Omega(x)(\partial_\mu \Omega^{-1}(x)) \ , \qquad \Omega \in \mathrm{SU}(2)_L \ . \tag{2.10}$$

Since the doublet transforms in the fundamental representation, from the transformation law we see that A_μ is in the adjoint representation. However, A_μ is now a field and as such should be quantized, too. Therefore, we need to give it dynamics. To do so, we notice that

$$[D_\mu, D_\nu]\psi(x) \to \Omega(x)[D_\mu, D_\nu]\psi(x) \tag{2.11}$$

under a group transformation and the commutator of two covariant derivatives $[D_\mu, D_\nu]$ is in fact a function and not an operator. It is a rank-2 tensor,

$$[D_\mu, D_\nu] = -igT_a F^a_{\mu\nu} \ , \tag{2.12}$$

where we have defined $F_{\mu\nu} = F^a_{\mu\nu} T_a \equiv \partial_\mu A_\nu - \partial_\nu A_\mu - ig\left[A_\mu, A_\nu\right]$. $F_{\mu\nu}$ transforms as $F_{\mu\nu} \to \Omega F_{\mu\nu} \Omega^\dagger$, hence its trace is gauge invariant and

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2} \operatorname{Tr}\left[F_{\mu\nu} F^{\mu\nu}\right] = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} \tag{2.13}$$

is both gauge and Lorentz invariant (we have used $\operatorname{Tr}[\sigma^a \sigma^b] = 2\delta^{ab}$). Finally, the Lagrangian is

$$\mathcal{L} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \overline{\psi}(i\not{D} - m)\psi \ . \tag{2.14}$$

Based on this derivation, the gauge Lagrangians of the SM can be built, by substituting the required gauge group. We will see in the next section that the abelian U(1) is a special case of the $SU(2)$ case.

2.1.2 The Higgs mechanism: recovering quantum electrodynamics

The first part of this thesis treats the hadronic contribution to the running of the electromagnetic coupling. This section is therefore intended to set the notation and introduce the parameters in the electroweak sector of the SM. The Lagrangian containing a complex doublet H reads

$$\mathcal{L} = -\frac{1}{4}\left(W^a_{\mu\nu}\right)^2 - \frac{1}{4} B^2_{\mu\nu} + \left(D_\mu H\right)^\dagger \left(D_\mu H\right) + m^2 H^\dagger H - \lambda \left(H^\dagger H\right)^2 \ , \tag{2.15}$$

where B_μ is the hypercharge gauge boson, $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$, and $W^a_{\mu\nu}$ the field strengths of the $\mathrm{SU}(2)_L$ gauge bosons (see previous section). The gauge symmetries in the theory are

$SU(2)_L \times \mathrm{U}(1)_Y$. Then, the covariant derivative is given by

$$D_\mu = \partial_\mu - igW_\mu^a T^a - \frac{1}{2} ig' B_\mu \tag{2.16}$$

with g and g' $SU(2)$ and $U(1)_Y$ couplings, respectively and $T^a = \sigma^a/2$ normalized as in the previous section. Choosing λ and m in the Higgs potential appropriately H gets a vacuum expectation value (vev) $v = m/\sqrt{\lambda}$ which we take real and require only the lower component of the doublet to develop the vev

$$H = \exp\left(2i\frac{\pi^a T^a}{v}\right)\begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + \frac{h}{\sqrt{2}} \end{pmatrix} . \tag{2.17}$$

where $h(x)$ and $\pi^a(x)$ are real fields and the factor $\sqrt{2}$ is a canonical normalization factor. In particular, this breaks the symmetry down to

$$SU(2)_L \times U(1)_Y \to U(1)_{\mathrm{em}} .$$

In the unitary gauge ($\xi = \infty$) we have $\pi = 0$ and with the parametrization for H (2.17) we obtain

$$|D_\mu H|^2 \supset g^2 \frac{v^2}{8}\left[\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2 + \left(\frac{g'}{g}B_\mu - W_\mu^3\right)^2\right] , \tag{2.18}$$

and we see that the gauge bosons have acquired mass terms. We now rotate the fields to diagonalize them to the mass basis

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & -\sin\theta_{\mathrm{W}} \\ \sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix}\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \quad \cos\theta_{\mathrm{W}} = \frac{g}{\sqrt{g^2+g'^2}}, \quad \sin\theta_{\mathrm{W}} = \frac{g'}{\sqrt{g^2+g'^2}} \tag{2.19}$$

where we have defined the weak mixing angle θ_W. We also define the linear combinations

$$W_\mu^\pm \equiv \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right)$$

Plugging these definitions into Eq. (2.18), the theory now contains two massive gauge bosons with $m_W = gv/2$ and a third gauge boson with mass

$$m_Z = 1/(2\cos\theta_W)gv = v/2\sqrt{g^2+g'^2} = m_W/\cos\theta_W .$$

The fourth gauge boson, the photon, remains massless, $m_A = 0$. Furthermore, one identifies the electric charge as

$$e = g\sin\theta_W = g'\cos\theta_W = \frac{gg'}{\sqrt{g^2+g'^2}} ,$$

and the physical fields $W^\pm$ have charges ± 1. Using these normalizations the electron will indeed have charge -1 since the generator of the remaining $U(1)_{\mathrm{em}}$ gauge symmetry is given by $q = T^3 + Y/2$. We won't go into the details of the fermionic sector of the standard model. Let us just mention that there are three generations of fermion fields. $\mathrm{SU}(2)_L$ is a chiral theory and only couples to left handed fermions which are grouped into weak doublets. The right handed fields are $\mathrm{SU}(2)_L$ singlets. Coupling the gauge bosons to the fermion content of the SM we

recover the QED Lagrangian after electroweak symmetry breaking,

$$\mathcal{L}_{\mathrm{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \sum_f \overline{\psi}_f \left(\not{D}_f - m_f\right)\psi_f - \frac{\xi}{2}(\partial^\mu A_\mu)^2\,, \tag{2.20}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $\not{D}_f = \gamma^\mu(\partial_\mu - ieq_f A_\mu)$, with q_f the electric charge in units of e, see Tab. 2.1, and the sum runs over all fermions charged under $\mathrm{U}(1)_{\mathrm{em}}$. These are Dirac fermions built from the left- and right-handed fields $\psi = (\psi_L, \psi_R)^T$. The mass term for the fermions comes from the Yukawa coupling with the Higgs, $m_f = y_f v/\sqrt{2}$. The leptons e, μ, τ have charges $q_f = -1$ formed from the hypercharge and $\mathrm{SU}(2)_L$ charge $T^3 = -1/2$ of the left handed fields, cf. Tab. 2.1. We have included an explicit gauge fixing term.

2.2 Quantum Electrodynamics

2.2.1 Prologue: Normalizing Renormalization

> *It is hard to underestimate the importance of the renormalization group in shaping the way we think about quantum field theory.*
> Matthew D. Schwartz

Learning about the renormalization group has definitely shaped my way of thinking about quantum field theories and science in general. Further, one certainly cannot underestimate the importance of the renormalization group in the development of lattice field theory. Incidentally, it is also the reason for the scale dependence, or "runnning", of the electromagnetic coupling, which is the main observable under study in the first part of this thesis. I would first like to give a brief overview of the conceptual importance of the renormalization group.

QED was developed in the 1930's by Dirac, Fermi, Heisenberg, Pauli, Jordan, Wigner etc. and the solutions were worked out as perturbative series in the "charge" (and the "mass") of the electron. It was then quickly found that higher-order corrections, such as the diagram depicted in Fig. 2.2 are infinite and lead to infinities in the prediction of physical observables. This caused Dirac, Bohr, Oppenheimer and others to conclude that QED should be thrown away [1]. However, Kramers, Feynman, Bethe, Schwinger, Tomonaga and Dyson managed to develop an algorithm – called *renormalization* – where these infinities could be eliminated by redefining the parameters in the theory. To appreciate that this redefinition is in fact not "neglecting it just because it is infinitely great", let us review our algorithm of "doing physics" (or science in general) and let us apply this algorithm to quantum field theory [65]:

1. *Observe.* Observations provide empirical data. We collide particles in particle colliders and measure e.g. the scattering cross section σ.

[1] *I must say that I am very dissatisfied with the situation, because this so-called "good theory" does involve neglecting infinities which appear in its equations, neglecting them in an arbitrary way. This is just not sensible mathematics. Sensible mathematics involves neglecting a quantity when it is small - not neglecting it just because it is infinitely great and you do not want it!* — P. A. M. Dirac, Directions in Physics

2. *Construct a model.* In the framework of quantum field theory, choose some fields Φ_n to build Lagrangians for "simple" models, dependent on a finite number of parameters, e.g. the Lagrangian in Eq. (2.20) – or even the SM Lagrangian – to explain the data.

3. *Fix the model's parameters.* Use the model to obtain mathematical expressions for at least as many observables as the model has parameters and adjust the values of these parameters to make the numerical values of these expressions agree with the corresponding measurements.

4. *Predict.* Use the values of the parameters determined in 3 and use the model to make predictions for other physical processes.

These four steps are the same for any mathematical description of physical processes, be it quantum or classical. The steps can be iterated at will to check the model against ever more measurements. Of course, we can never confirm hypotheses (our models in step 2), we can only falsify them. If falsified, we have to iterate through the steps 1-4, formulating new hypotheses, and rejecting them if they are falsified. As more measurements are made, the list of models which can explain all observations shrinks.

The prediction is a map from the model, dependent on some parameters, to observation. Usually, this map is not injective, i.e. we can have multiple explanations consistent with observation. In particular, this map is therefore clearly not invertible. An important remark to make is that the coupling constants (the parameters) are not directly measurable and are not operationally well defined. Thus we make a list of predictions based on a given model and compare these predictions to observations. Then we invert the map given by this particular model to obtain the parameters. As we will see, it is possible to make finite predictions with "infinite" parameters.

In the framework of our "simple" QFTs it turns out that, as we go to higher orders in perturbative quantum corrections, the models in step 2 produce infinities in steps of the calculations that yield the predictions for physical observables for which we measure finite values. This has led to a paradigm shift: instead of attempting to explain physics up to infinite energies one may step back from this very ambitious task and try to explain a restricted number of degrees of freedom, but more than we are ever likely going to observe. This can be achieved by imposing a cutoff, where we can symbolically regard the cutoff as the number of degrees of freedom we try to explain with our simple theory. Of course, we would be concerned if the predictions depended on this arbitrary cutoff. However, we might be able to invert the prediction $F(g_i, \Lambda)$ to get not a constant coupling (e.g. a mass in Newtonian mechanics) but $g_i = g_i(\Lambda)$, i.e. allow the parameters (the "masses" and coupling "constants") to depend on the number of degrees of freedom we are attempting to explain,

$$F_j(g_1(\Lambda), \ldots, g_n(\Lambda), \Lambda) = \langle O_j \rangle \quad \forall \Lambda \, .$$

It is however still far from obvious whether it is possible to change the cutoff while keeping the number of parameters fixed (intuitively, to explain more degrees of freedom, one has to add more and more parameters). Theories in which this is possible are called "renormalizable" and are very remarkable objects in the space of explanations. That this is indeed possible is related to the fact that in order to explain physics at a given long distance scale, only a few parameters are important. This is the reason we can do physics after all. We will discuss this in more detail in Sec. 3.6.

QED is a perturbatively renormalizable theory, i.e. all amplitudes can be written in terms of a finite number of parameters (e, m) to all loop orders and they are finite after renormalization, even when the cutoff Λ is sent to infinity.

2.2.2 Vacuum polarization and charge renormalization

A diagram which will keep us occupied during the first part of the thesis is the self energy of the photon. In this section we will explain why the electromagnetic coupling "runs" with energy. For illustrative purposes and to introduce the notation we will be concerned with the 1-loop running in pure QED and explain how the electric charge can be renormalized. We will also define an effective electromagnetic coupling based on the resummation of 1-particle irreducible vacuum bubbles. This will be the basis of our discussion in Chap. 4 where we calculate the contribution from QCD to the vacuum polarization non-perturbatively.

The (bare) Lagrangian of QED neglecting all fermions except the electron reads

$$\begin{aligned}\mathcal{L}_{\mathrm{QED}} &= \bar{\psi}_0 \left(i \not{D} - m_0 \right) \psi_0 - \frac{1}{4} F_{\mu\nu,0} F_0^{\mu\nu} - \frac{\xi}{2} (\partial_\mu A_0^\mu)^2 \,, \\ &= \bar{\psi}_0 \left(i \not{\partial} - m_0 \right) \psi_0 - \frac{1}{4} F_{\mu\nu,0} F_0^{\mu\nu} - e_0 \bar{\psi}_0 \gamma_\mu \psi_0 A_0^\mu - \frac{\xi}{2} (\partial_\mu A_0^\mu)^2 \,,\end{aligned} \tag{2.21}$$

with ξ the gauge-fixing parameter. The full photon propagator is defined as

$$i D^\gamma_{\mu\nu}(x-y) = \langle 0 | T \left\{ A_\mu(x) A_\nu(y) \right\} | 0 \rangle \,. \tag{2.22}$$

It can be obtained by insertions of one-particle-irreducible (1PI) blobs, called the vacuum polarization tensor, which itself is given by the vacuum expectation value of the time-ordered product of two electromagnetic currents

$$\Pi_{\mu\nu}(q) = i \int \mathrm{d}^4 x e^{iqx} \langle 0 | T \left\{ j^{\mathrm{em}}_\mu(x) j^{\mathrm{em}}_\nu(0) \right\} | 0 \rangle \,, \tag{2.23}$$

with q_μ the four-momentum of the photon and $j_\mu(x) = \overline{\psi}(x)\gamma_\mu\psi(x)$. Due to Lorentz invariance and current conservation $\partial_\mu j^\mu_{\mathrm{em}}(x) = 0$ which follows from Noether's theorem from the invariance of the QED Lagrangian under a global U(1) symmetry, the vacuum polarisation tensor can be decomposed into a Lorentz invariant function and a Lorentz structure which is completely transverse,

$$\Pi_{\mu\nu}(q) = \left(q_\mu q_\nu - g_{\mu\nu} q^2 \right) \Pi \left(q^2 \right) \,, \tag{2.24}$$

from which it follows directly that indeed $q_\mu \Pi^{\mu\nu}(q) = 0$. The free photon propagator is given by

$$i e_0^2 D_{\mu\nu}(q) = -i \frac{e_0^2}{q^2} \left[g_{\mu\nu} - (1-\xi) \frac{q_\mu q_\nu}{q^2} \right] \,, \tag{2.25}$$

where we have introduced an overall factor of e_0^2 for reasons that will be made clear below. Finally, the full, resummed photon propagator will be given by multiple (infinite) insertions of

the vacuum polarization tensor (so-called Dyson series of self-energy insertions), see Fig. 2.1

$$
\begin{aligned}
ie_0^2 D^{\mathrm{F}}_{\mu\nu} &= e_0^2 \left(iD_{\mu\nu}(q) + iD_{\mu\lambda}(q) ie_0^2 \Pi^{\lambda\rho} iD_{\rho\nu}(q) + iD_{\mu\lambda}(q) ie_0^2 \Pi^{\lambda\rho} iD_{\rho\sigma}(q) ie_0^2 \Pi^{\sigma\kappa} iD_{\kappa\nu}(q) + \dots \right) \\
&= -i \frac{e_0^2}{q^2} \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) \left(\sum_{n=0}^{\infty} \left[e_0^2 \Pi\left(q^2\right) \right]^n \right) - i\xi \frac{e_0^2}{q^2} \frac{q_\mu q_\nu}{q^2} \\
&= -i \frac{1}{q^2} \frac{e_0^2}{1 - e_0^2 \Pi\left(q^2\right)} \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) - i\xi \frac{e_0^2}{q^2} \frac{q_\mu q_\nu}{q^2}
\end{aligned}
\tag{2.26}
$$

for which we used the transverse nature of the VP tensor to cancel almost all terms containing the gauge-fixing parameter. Here we restrict our discussion to the lowest order contribution

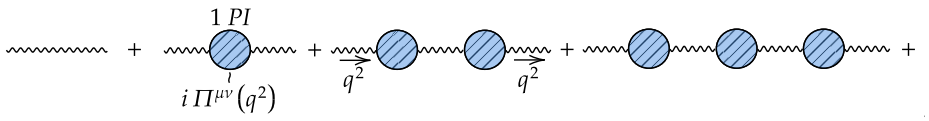


FIGURE 2.1: The full photon propagator is given by infinite insertions of the vacuum polarization tensor, $\Pi_{\mu\nu}$.

($\mathcal{O}(e_0^2)$) to the 1PI diagram. Restricting ourselves to pure QED with only one lepton of charge e and mass m (e.g. the electron), the one-loop correction to the photon propagator is simply given by an electron-positron loop

$$
ie_0^2 \Pi^{\mu\nu}_{\text{1-loop}}(q) = (-1)(-ie_0)^2 \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \operatorname{Tr} \left\{ \gamma^\mu \frac{\not{k} + m_0}{k^2 - m_0^2 + i\epsilon} \gamma^\nu \frac{\not{k} + \not{q} + m_0}{(k+q)^2 - m_0^2 + i\epsilon} \right\} , \tag{2.27}
$$

which superficially diverges quadratically in the loop momentum k. However, due to gauge invariance, Eq. (2.24), the divergence is actually only logarithmic. Nevertheless, this infinity appears to get in the way of making sensible predictions for physical observables which are necessarily finite. As we discussed in the prologue, this led to a paradigm shift: to analyze what is really going on, one must first quantify these infinities and study their structure. Since they arise from the fact that intermediate/virtual states can have infinite four-momentum components, one has to place an upper limit on how large those four-momenta can be and study how the corresponding contributions depend on this upper limit. Imposing such a limit is called "regularization". There are various ways of regularizing a theory. In QED, we require a regulator which preserves gauge invariance. In this specific example we will use dimensional regularization which is the standard regularization procedure used in perturbative QED calculations. While it is somewhat less intuitive physically it exactly preserves gauge and Lorentz invariance. In Chap. 4 we will use a lattice to regularize our theory non-perturbatively. The calculation is a standard textbook example, see for instance [66–69]. The goal is to temporarily make the theoretical predictions finite and isolate the divergent parts. Then these divergent parts can be subtracted by a redefinition of the bare parameters in the theory to be left with a well-defined finite result. Hence, the bare Lagrangian parameters will ultimately depend on the choice of regulator and on the cutoff. The many ways to regularize QFTs should however give the same physical predictions, a consequence of universality which we discuss in Sec. 3.6.

The idea of dimensional regularization is to Wick-rotate to Euclidean spacetime and regulate the integral in Eq. (2.27) by analytically continuing it to $d = 4 - 2\eta$ dimensions, where $\eta > 0$ is

an infinitesimal parameter. By reducing the dimension of spacetime in this way, logarithmically divergent integrals become convergent. Then by taking the limit $d \to 4$ the UV divergences appear as isolated singularities of the analytic continuation, as $1/\eta^n$ poles. In d-dimensions, the electromagnetic coupling acquires a mass dimension $4-d$. In order to keep a dimensionless coupling we replace

$$e_0^2 = \mu^{4-d} g^2(\mu)$$

where μ is an arbitrary mass scale $[\mu] = 1$, often called the renormalization scale and $[g(\mu)] = 0$. This procedure is called dimensional transmutation where the breaking of scale invariance (present when $m_0 = 0$) introduces a scale into the theory. The requirement of independence of physical predictions on this arbitrary scale leads to the renormalization group equations and is characteristic of all regularizations. Finally, after Wick rotating to Euclidean space and using a

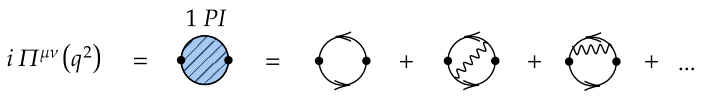


FIGURE 2.2: The vacuum polarization tensor is the sum of all one-particle irreducible (1PI) diagrams to all orders. We depict some of the diagrams which contribute to the photon vacuum polarization. In this section we calculate the first of these diagrams, the lowest order contribution, which in QED with only electrons is simply given by an electron-positron loop.

plethora of computational tricks, the integral in d-dimensions becomes, with $\Delta = m^2 - x(1-x)q^2$

$$\begin{aligned}\Pi^{\mu\nu}_{\text{1-loop}}(q) &= -f(d)\mu^{4-d}\int_0^1 dx \frac{1}{(4\pi)^{d/2}}\frac{\Gamma\left(2-\frac{d}{2}\right)}{\Delta^{2-d/2}} \\ &\quad\times\left[g^{\mu\nu}\left(-m^2+x(1-x)q^2\right)+g^{\mu\nu}\left(m^2+x(1-x)q^2\right)-2x(1-x)q^\mu q^\nu\right] \\ &= \left(q^2 g^{\mu\nu}-q^\mu q^\nu\right)\Pi_{\text{1-loop}}\left(q^2\right)\end{aligned} \tag{2.28}$$

and with $f(d) \overset{d\to 4}{\to} 4$. First we notice that the dimensionally regularized theory indeed manifestly maintains gauge invariance (at least at one-loop but this holds in fact at all orders), $q_\rho \Pi^{\rho\nu}_{\text{1-loop}} = 0$. Secondly, for $d = 4$ the integral will of course diverge. However the divergence in the limit $\eta \to 0$ in Eq. (2.28) is now an isolated singularity – a pole – in $d = 4$. Expanding in $\eta \ll 1$ giving e.g. $\Gamma(\eta) = 1/\eta - \gamma_E + \mathcal{O}(\eta)$ and $A^\eta = \exp(\eta \ln A) = 1 + \eta \ln A + \mathcal{O}(\eta^2)$ we obtain

$$\Pi_{\text{1-loop}}\left(q^2\right) = \frac{g^2(\mu)}{2\pi^2}\left\{\frac{1}{6}\left(\frac{1}{\eta}-\gamma_E\right)-\int_0^1 \mathrm{d}x\; x(1-x)\ln\left(\frac{m^2-q^2x(1-x)-i\epsilon}{4\pi\mu^2}\right)+\mathcal{O}(\eta)\right\}, \tag{2.29}$$

where γ_E is the Euler–Mascheroni constant. Now that we have isolated the singularities, we can absorb them by redefining the parameters which themselves were not measurable. The redefined, physical parameters will be finite quantities. This is done by adding counterterms to the original Lagrangian $\mathcal{L}_0 = \mathcal{L}_{\text{ren}} + \mathcal{L}_{\text{ct}}$, leading to new vertices of higher order in the coupling constants,

$$\delta\mathcal{L}_{\text{ct}} = -\frac{\delta Z_3}{2}F_{\mu\nu}F^{\mu\nu} + \delta Z_2 \bar{\psi}\mathrm{i}\,\partial\!\!\!/\,\psi - \delta m\bar{\psi}\psi + e(\mu)\mu^\eta \delta Z_1 \bar{\psi}A\psi \tag{2.30}$$

where we have defined the bare fields and parameters in terms of renormalized parameters,

$$A_0^\mu = \sqrt{Z_3}A^\mu, \quad \psi_0 = \sqrt{Z_2}\psi, \quad m_0 = \frac{Z_m}{Z_2}m, \quad e_0 = \frac{Z_1}{Z_2\sqrt{Z_3}}e(\mu)\mu^\eta \tag{2.31}$$

and we have set $g(\mu) = Z_1 Z_2^{-1} Z_3^{-1/2} e(\mu)$, giving a relation between the dimensionless bare coupling and the renormalized coupling. The coefficients $\delta Z_i = Z_i - 1$ are determined iteratively by going to higher orders in the perturbative expansion. Gauge invariance implies that $Z_1 = Z_2$, which ensures that we only need to consider the photon self energy for charge renormalization in QED, a property that does not hold in QCD and that we have tacitly ignored until now. In our case, the counterterm is therefore $\left[(Z_3 - 1)/4\right] F_{\mu\nu} F^{\mu\nu} = \left[\delta Z_3/4\right] F_{\mu\nu} F^{\mu\nu}$. The divergent part will be absorbed in δZ_3, which we calculated at $\mathcal{O}(e^2)$. The choice of how much one subtracts is somewhat arbitrary and leads to different possible subtraction (or renormalization) schemes.

Here we decide to renormalize the theory in the on-shell, or physical, renormalization scheme. Instead of fixing the parameters in the UV, we would like to fix the renormalized parameters in the IR in terms of the physical (measurable) value of the electromagnetic coupling at long distances, where α is measured to high accuracy. Since we fix physically measurable quantities, as we argued in the introduction, the bare parameters necessarily have to depend on the cutoff, $e_0 = e_0(\Lambda)$. In the on-shell scheme we require that the photon propagator should have a pole at the physical photon mass, i.e. at $q^2 = 0$, with unit residue as at tree-level[2]. The first condition, namely that the photon remains massless after including quantum corrections, is automatically satisfied as a consequence of gauge invariance. As we saw above in Eq. (2.26) we can resum the infinite series of insertions of vacuum bubbles to obtain the full photon propagator[3]

$$ie_0^2 D^{\mathrm{F}}_{\mu\nu} = -i \frac{g_{\mu\nu}}{q^2} \frac{e_0^2}{1 - e_0^2 \Pi\left(q^2\right)} \,. \tag{2.32}$$

The renormalization condition we impose is that the residue of the pole of the propagator be the measured charge in the Thompson limit, given by the fine structure constant $\alpha = 4\pi e^2 \approx 1/137.03\ldots$[4],

$$e^2 \equiv \frac{e_0^2}{1 - e_0^2 \Pi(0)} \,. \tag{2.33}$$

Inverting Eq. (2.33) allows us to express the bare coupling in terms of the renormalized coupling,

$$e_0^2 = \frac{e^2}{1 + e^2 \Pi(0)} \,. \tag{2.34}$$

Since the QED Ward identities ensure that the vacuum polarization gives the only loop corrections to the photon propagator we can plug (2.34) into Eq. (2.32) and define the full renormalized photon propagator

$$iD^{\mathrm{F}}_{\mu\nu}\left(q^2\right) = \frac{-ig_{\mu\nu}}{q^2(1 - e^2 \hat{\Pi}^{\mathrm{OS}}(q^2))} + \mathcal{O}(e^4) \,, \tag{2.35}$$

where we have defined $\hat{\Pi}^{\mathrm{OS}}(q^2) = \Pi(q^2) - \Pi(0)$. $\hat{\Pi}^{\mathrm{OS}}(q^2)$ is the finite, renormalized vacuum polarization, in which the logarithmic divergence of $\Pi(q^2)$ is subtracted on the mass shell of the photon (q^2=0), i.e. by subtracting $\Pi(0)$ and enforcing that the photon has a vanishing mass. Thus:

$$\Pi^{\mu\nu}(q) = \left(q^\mu q^\nu - g^{\mu\nu} q^2\right) \left(\Pi^{\mathrm{OS}}_{\mathrm{div}} + \hat{\Pi}^{\mathrm{OS}}\left(q^2\right)\right) \,. \tag{2.36}$$

[2]This is in fact the wave function renormalization condition for the photon field but due to $Z_1 = Z_2$ both charge and wave function renormalization can be fixed via the vacuum polarization, see Eq. (2.30).

[3]The terms proportional to q_μ will vanish in physical amplitudes since the photon propagator in physical processes connects to a conserved electron current.

[4]In momentum space, this corresponds to fixing $e^2 = p_0^2 V(p_0^2)$, i.e. measuring the Coulomb potential at large distances $r \to \infty$.

Since $\hat{\Pi}\left(q^2\right) = \Pi\left(q^2\right) + \delta Z_3$ (remember that we have pulled out the e^2 in the definition of Π) we immediately find

$$\delta Z_3^{\mathrm{OS}} = -\Pi(q^2 = 0) = -\frac{1}{12\pi^2}\left\{\frac{1}{4} - \gamma - \ln\frac{m^2}{4\pi\mu^2}\right\} + \mathcal{O}(e^2) = \frac{1}{12\pi^2}\ln\frac{m^2}{\Lambda^2} + \mathcal{O}(e^2)\ , \tag{2.37}$$

with $\Lambda^2 \equiv 4\pi\mu^2 \exp(1/\eta - \gamma_E)$ an "effective cutoff". Then

$$\hat{\Pi}^{\mathrm{OS}}\left(q^2\right) = \Pi\left(q^2\right) - \Pi(0) \overset{q^2 \ll m^2}{\longrightarrow} = \frac{1}{4\pi}\frac{1}{15\pi}\left(\frac{q^2}{m^2}\right) + \mathcal{O}\left(\frac{q^4}{m^4}\right) \tag{2.38}$$

which is independent of the regularization used. We can therefore take the limit $\Lambda \to \infty$ (infinite cutoff limit) and recover our 4-dimensional space-time. In this scheme, the renormalized mass and coupling are expressed in terms of well-measured physical quantities (the electron mass and fine-structure constant). Hence, e^2 is a physical number, so independent of Λ and, in fact, of the regularization scheme. Therefore, e_0, the bare charge must depend on both, as does $\Pi(0)$ (this can be seen by injecting Eq. (2.37) into Eq. (2.34)). The on-shell scheme is the typical renormalization scheme in QED for quantities at low energies. It cannot be used in QCD, first of all because quarks and gluons can never be on shell due to confinement. Now that we have fixed e (we can do the same with m), we can make an arbitrarily large number of non-trivial predictions.

From expression (2.35) we can define an effective charge, dependent on the momentum q^2 of the photon and which reduces to the physical charge e for $q = 0$,

$$e_{\mathrm{eff}}^2(q^2) = \frac{e^2}{1 - e^2\hat{\Pi}_{\mathrm{OS}}(q^2)}\ . \tag{2.39}$$

In terms of the fine structure constant $\alpha = 4\pi e^2$, Eq. (2.39) becomes

$$\alpha_{\mathrm{eff}}(q^2) = \frac{\alpha}{1 - \Delta\alpha(q^2)}\ , \tag{2.40}$$

where we have defined $\Delta\alpha = (\alpha/4\pi)\hat{\Pi}(q^2)$. Eq. (2.39) is the "running" of the electromagnetic coupling in the on-shell scheme. We will calculate the hadronic contribution to the Euclidean $\Delta\alpha_{\mathrm{had}}^{\mathrm{OS}}(-q^2)$ non-perturbatively (here we calculated it to one-loop order in QED) in Chap. 4.3, i.e. the energy dependence of the effective electromagnetic coupling that we just defined arising from QCD interactions.

Finally, in terms of this effective coupling, the electrical "potential" takes the form

$$\tilde{V}(p) = e_{\mathrm{eff}}^2(p)/p^2\ , \tag{2.41}$$

reducing to the Coulomb potential in the limit of static charges. The effective coupling is hence a physical observable and therefore scheme-independent (the scheme dependence arises one order higher in the perturbative expansion). The quantum corrections to the propagator give corrections to the Coulomb potential and to, for instance, the magnetic moments of the leptons. These predictions are one of the great successes of quantum field theory.

Renormalization group equations Indeed, instead of renormalizing the photon propagator in the on-shell scheme, which is often the preferred scheme for perturbative calculations at low energies in QED, we could choose a different subtraction prescription: In the so-called $\overline{\text{MS}}$-scheme, the pole in η and the finite parts γ and $\ln 4\pi$ are subtracted,

$$\hat{\Pi}_{\overline{\text{MS}}}(q^2) = \frac{1}{2\pi^2}\int_0^1 \mathrm{d}x\; x(1-x)\ln\left[\frac{m^2 + x(1-x)q^2}{\mu^2}\right] . \tag{2.42}$$

Hence, in this scheme, the addition of the counterterm amounts to a redefinition of the charge, which is finite and renormalized but no longer scale independent. However, as we argued, physical quantities O cannot (and will not) depend on the regularization or the renormalization scheme,

$$O = O_{\text{OS}}\left(\alpha, m, \ln\frac{s}{m^2}, \dots\right) = O_{\overline{\text{MS}}}\left(\alpha(\mu), m(\mu), \ln\frac{s}{\mu^2}, \dots\right) . \tag{2.43}$$

In fact, in the end all physical amplitudes can be written in terms of e and will not depend on μ[5]. From the requirement of μ-independence of the observables we obtain the renormalization group equations

$$\mu\frac{\mathrm{d}}{\mathrm{d}\mu}O = \beta(\alpha)\frac{\partial O}{\partial\alpha(\mu)} + \gamma_m(\alpha)m\frac{\partial O}{\partial m(\mu)} + \frac{\partial O}{\partial\ln\mu} = 0 , \tag{2.44}$$

where we have defined the β-function and the anomalous dimension of the electron mass,

$$\mu\frac{\mathrm{d}\alpha(\mu)}{\mathrm{d}\mu} \equiv \beta(\alpha(\mu)) , \qquad \mu\frac{\mathrm{d}m(\mu)}{\mathrm{d}\mu} \equiv \gamma_m(\alpha(\mu))m(\mu) . \tag{2.45}$$

The β-function is

$$0 = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}e_0 = \mu\frac{\mathrm{d}}{\mathrm{d}\mu}\left[\mu^\eta e(\mu)/\sqrt{Z_3}\right] \tag{2.46}$$

yielding

$$\beta(e) = -\eta e(\mu) + \frac{e(\mu)^3}{12\pi^2} + \mathcal{O}(e(\mu)^5) . \tag{2.47}$$

Solving the first equation in (2.45) we obtain at one-loop order in the $\overline{\text{MS}}$-scheme

$$\alpha(\mu) = \frac{\alpha}{1 - \frac{\alpha}{3\pi}\ln\mu^2/m^2} \tag{2.48}$$

which, at this order is equivalent to the effective coupling we defined in the on-shell scheme. Eq. (2.48) is referred to as the "running coupling", since it depends on scale. In fact, solving the RG equation is equivalent to resumming a series of leading logarithms to all orders, and the effective coupling is therefore also called leading logarithmic resummation. Hence by solving the RG equation we obtain the coupling strength $\alpha(\mu)$ appropriate for physics at energy scale μ [66]. If we choose μ very different from the energy scale of experiments, we will introduce large logarithms and the perturbative expansion breaks down.

Let us conclude this section by a last remark: From Eq. (2.48) we see that, as we go to higher energies, the running coupling increases. It diverges at a scale $\Lambda_{\text{QED}} \sim 10^{286}$ eV, QED has a Landau pole. However, we already saw in the previous section that QED should be replaced by

[5] However, at any fixed order in PT, the values of a calculated physical quantity can differ in two schemes, where this difference can only be due to the neglect of higher-order corrections. If all corrections to all orders could be computed, the two values would agree.

the electroweak theory at energies ~ 100 GeV. QED (and in fact the SM itself) only exist as a low-energy effective theory.

2.3 (Low energy) Quantum Chromodynamics

2.3.1 The QCD Lagrangian

To derive the Lagrangian of quantum chromodynamics (QCD) in the continuum in Minkowski spacetime, we follow a reasoning along the same lines as in Sec. 2.1.1. In QCD the local gauge group is $\mathrm{SU}(N_c)$ with $N_c = 3$ the number of colours.

The terms which are allowed by Lorentz symmetry let us build the Lagrangian of a free quark q, a Dirac fermion

$$\mathcal{L} = \bar{q}(x)(i\gamma_\mu \partial^\mu - m_q)q(x) \ , \tag{2.49}$$

with m_q the mass of the quark. To build a Lagrangian which is also invariant under global $\mathrm{SU}(N_c)$ transformations we can group N_c Dirac fermions in the fundamental representation of $\mathrm{SU}(N_c)$ into a N_c multiplet

$$q(x) = \begin{pmatrix} q_1(x) \\ \vdots \\ q_{N_c}(x) \end{pmatrix} . \tag{2.50}$$

$\mathrm{SU}(N_c)$ transformation will rotate the multiplets in colour space. Each component $q_a(x)$ of the multiplet is itself a Dirac fermion and comes in N_f flavours but we will restrict ourselves to one flavour in this discussion. Then Eq. (2.49) is the most general Lagrangian up to dimension 4 which is invariant under Lorentz transformations and global $\mathrm{SU}(N_c)$ transformations. There are $N_c^2 - 1$ conserved Noether charges associated with the global $\mathrm{SU}(N_c)$ symmetry. If we now require our theory to be invariant under local, space-time dependent $\mathrm{SU}(N_c)$ transformations

$$q(x) \to \Omega(x)q(x) \ , \qquad \Omega(x) \in \mathrm{SU}(N_c) \ , \tag{2.51}$$

we can follow the same arguments as in Sec. 2.1.1 - all we have to do is replace the generators of $SU(2)_L$ by the generators of $\mathrm{SU}(N_c)$ which we will denote by T_a. For $N_c = 3$ the usual choice for the fundamental representation are 8×8 dimensional matrices $T_a = \lambda_a/2$ with λ_a the Gell-Mann matrices. Hence, the full QCD Lagrangian reads

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \sum_{f=1}^{n_f} \bar{q}_f (i\not{D} - m_f) q_f \ , \tag{2.52}$$

with the covariant derivative

$$D_\mu = \partial_\mu + i g_s A^a_\mu T_a \ , \tag{2.53}$$

with g_s the strong coupling and which we have now generalized to N_f flavours.

2.3.1.1 Non-Abelian $SU(N_c)$ pure gauge theory

Actually, Eq. (2.52) is not the most general Lagrangian up to dimension 4, respecting SU(N_c) gauge symmetries and Lorentz invariance. In fact, restricting ourselves to the pure gauge part of the QCD Lagrangian, an additional term would be allowed by the gauge symmetries,

$$\mathcal{L}_{\text{QCD}} \subset -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} - \theta \frac{g_s^2}{32\pi^2} \tilde{G}^{a\mu\nu} G^a_{\mu\nu} \,, \tag{2.54}$$

where we have chosen the normalization of the second term for reasons which will be clear later. Note that we again take the sum over the colour index a. We have denoted the dual of the colour field strength tensor $\tilde{G}^{\mu\nu} \equiv \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma} G_{\rho\sigma} =^* G^{\mu\nu}$ and g_s the coupling of strong interactions.

The second term in Eq. (2.54) also violates CP [6]. One might wonder if this CP violating term has any physical effect. Indeed, this term can be written as a total derivative

$$\epsilon_{\mu\nu\rho\sigma} \operatorname{Tr}\left(G^{\rho\sigma} G^{\mu\nu}\right) = \partial^\mu K_\mu \quad \text{with} \quad K_\mu = 4\epsilon_{\mu\nu\rho\sigma} \operatorname{Tr}\left(A^\nu \partial^\rho A^\sigma - \frac{2ig_s}{3} A^\nu A^\rho A^\sigma\right) \,, \tag{2.55}$$

and K_μ is called the *Chern-Simons* current. Hence, this term does not contribute in perturbation theory, as every vertex will yield an overall factor of $\sum p_\mu = 0$ and vanish. Yet, in Sec. 2.3.3, we will argue that there exist solutions to the equations of motion with finite classical action, called *instantons*, which do contribute to physics via this term.

2.3.2 The perturbative QCD β-function: asymptotic freedom

In Sec. 2.2.1 we saw that the scale dependence of a generic coupling $\alpha_s(\mu)$ is encoded in the β-function,

$$\beta(\alpha_s) = \mu^2 \frac{\partial \alpha_s(\mu)/4\pi}{\partial \mu^2} \,. \tag{2.56}$$

For QCD, the β function is in the perturbative expansion known to five loops in the $\overline{\text{MS}}$-scheme. The first two coefficients in the perturbative β-function

$$\beta(\alpha_s) = -b_0 \left(\frac{\alpha_s}{4\pi}\right)^2 - b_1 \left(\frac{\alpha_s}{4\pi}\right)^3 + \dots \tag{2.57}$$

where

$$b_0 = \frac{11}{3} N_c - \frac{2}{3} N_f \,, \qquad b_1 = \frac{34}{3} N_c^2 - N_f \left(\frac{10}{3} N_c + \frac{N_c^2 - 1}{N_c}\right) \tag{2.58}$$

are universal whereas higher order coefficients $b_2, b_3, \dots$ depend on the RG scheme [2].

What does the β-function tell us about our system? If the fermion number is large enough, $N_f > N^{\text{IF}}$ the system becomes infrared free. For $N^* < N_f < N^{\text{IF}}$ the system is conformal. If the number of flavours N_f is small, then the β-function is negative and the system is confining at small energies: the coupling becomes larger as μ decreases. For QCD with $N_c = 3$ and $N_f = 6$ this is indeed the case, a property referred to as *asymptotic freedom*. This was recognized by

[6] To see this consider for example the term

$$\partial_0 A_1^a \partial_2 A_3^a \xrightarrow{CP} \partial_0 \left(-A_1^a\right) \left(-\partial_2\right) \left(-A_3^a\right) \to -\partial_0 A_1^a \partial_2 A_3^a .$$

where A_1 and A_3 change sign due to charge conjugation and ∂_2 due to parity.

Gross and Wilczek [70] and Politzer [71] and earned them the Nobel prize in 2004. Keeping only the first term in Eq. (2.57) and solving the β-function we obtain the running of α_s to one-loop order,

$$\alpha_s(\mu) = \frac{\alpha_s(Q)}{1 + \frac{11N_c - 2N_f}{3} \frac{\alpha_s(Q)}{2\pi} \ln(\mu^2/Q^2)} , \tag{2.59}$$

with Q some reference scale where we measure α_s precisely (e.q. $Q = M_Z$) [2]. Defining

$$\Lambda^2_{\text{QCD}} = \mu^2 e^{-\frac{4\pi}{b_0 \alpha_s(\mu)}} \tag{2.60}$$

one obtains

$$\alpha_s(Q) = \frac{4\pi}{b_0 \ln(Q^2/\Lambda^2_{\text{QCD}})} . \tag{2.61}$$

Pure Yang-Mills theory is conformal (scale invariant) classically. The running coupling arises due to quantum effects and breaks this scale invariance (this is also referred to as conformal anomaly), giving rise to a characteristic scale, Λ_{QCD}. This is an example of dimensional transmutation and is related to the fact that in order to define a QFT, we need to introduce a cutoff. Λ_{QCD} (≈ 0.2 GeV at leading order) is the scale at which the running coupling $\alpha_s(\mu)$ diverges and perturbation theory breaks down. However, it remains to be proven that below the QCD scale, quarks do indeed confine to form hadrons. Lattice QCD was introduced as a numerical tool to answer these questions quantitatively since it provides a non-perturbative tool for the calculation of physical observables at these energies [72]. The conjecture that $SU(3)$ describes indeed the strong interactions could therefore be tested experimentally in deep-inelastic scattering experiments.

2.3.3 QCD at low energies: The $U(1)_A$ problem, instantons and θ-vacua

2.3.3.1 The heavy η'

As derived in Sec. 2.3.1 the QCD Lagrangian reads (let us for the moment omit the θ-term we introduced in Sec. 2.3.1.1)

$$\mathcal{L}_{QCD} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu a} + \sum_{f=1}^{N_f} \bar{q}_f (i\not{D} - m_f) q_f , \tag{2.62}$$

where $G^{\mu\nu}$ is the gluon field strength tensor, q_f are the quark colour fields, m_f the quark masses, D_μ the covariant derivative and N_f the number of flavours. Since we are interested in phenomena at low energies, i.e. when QCD becomes strong, we typically consider just two or three light quarks ($N_f = 2$ or $N_f = 3$). We can consider that the heavy quarks have been integrated out and do not influence the dynamics [7].

[7] Let us for the moment set the number of active quark flavours to two, $N_f = 2$. If the masses of the up and the down quarks were equal, this Lagrangian would have a global $SU(2)$ symmetry, called isospin symmetry. In practice, this is a very good approximation to the order of 1%, which is why in lattice QCD simulations we often work in this isospin-symmetric limit. However, for precision observables, these small isospin-breaking effects can no longer be ignored and have to be accounted for in our simulations. We will discuss the inclusion of isospin breaking corrections in detail in Sec. 4.3.1.6.

Consider a classical symmetry with generator T^a, then the associated Noether current is

$$J^a_\mu = \sum_\Phi \frac{\delta\mathcal{L}}{\delta(\partial^\mu\Phi)} iT^a_\Phi \Phi \tag{2.63}$$

where Φ are all the fields charged under the symmetry. The Noether current is conserved, $\partial^\mu J^a_\mu = 0$. Because we are interested in typical hadronic scales $\sim \Lambda_{\rm QCD}$ and the masses of the light quarks are much smaller than $\Lambda_{\rm QCD}$, we can likewise make the approximation of massless quarks. Then, in the limit of N_f massless quarks and neglecting all interactions except QCD, $\mathcal{L}_{QCD}$ has a global symmetry, $SU(N_f)_L \times SU(N_f)_R \times U(1)_V \times U(1)_A$ [66, 73, 74]. Hence, for $N_f = 2$, according to (2.63), the associated Noether currents are

$$J^a_\mu = -\bar{q}\sigma^a/2\gamma^\mu q, \quad J^{5a}_\mu = -\bar{q}\sigma^a/2\gamma^\mu\gamma^5 q, \quad J^V_\mu = \bar{q}\gamma^\mu q, \quad J^A_\mu = \bar{q}\gamma^\mu\gamma^5 q\ , \tag{2.64}$$

with σ_a the Pauli matrices. We will show later that the axial $U(1)_A$ symmetry in fact is not a symmetry of the quantum theory since it is broken by quantum corrections. The vector symmetry corresponds to baryon number and is a symmetry of the Lagrangian, even for massive quarks [8]. As alluded to in the previous chapter, at low energies at temperatures $T_c \sim \Lambda_{\rm QCD}$ the quark bilinear obtains an expectation value,

$$\langle\bar{q}q\rangle \sim \mathcal{O}\left(\Lambda^3_{\rm QCD}\right)\ , \tag{2.65}$$

with $\Lambda_{\rm QCD} \sim 300$ MeV. This breaks J^{5a}_μ spontaneously and we are left with a global $SU(N_f)_V \times U(1)_V$ symmetry. Since we looked at QCD in the limit of massless quarks, we expect N_f^2 massless Goldstone bosons (because we have broken $U(N_f) \times U(N_f) = SU(N_f)_V \times SU(N_f)_A \times U(1)_V \times U(1)_A$ down to $U(N_f) = SU(N_f)_V \times U(1)_V$).

Let us see what happens in the case of three light quarks, i.e. $N_f = 3$. In fact, due to the small masses of the quarks, the symmetries were only approximate before the spontaneous breaking, and the Goldstone bosons are actually light pseudo-Goldstone-bosons. From the broken $\mathrm{SU}(3)_A \times \mathrm{U}(1)_A$ symmetries we appear to get nine such pseudoscalar bosons: the three pions, $\pi^\pm, \pi^0$, the four kaons, $K^\pm, K^0, \bar{K}^0$, and the two etas,

$$\eta_1 \equiv \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s}), \quad \eta_8 \equiv \frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s}),$$

where the η_1 comes from the diagonal $\mathrm{U}(1)_A$ generator and η_8 is associated with the T^8 generator of $\mathrm{SU}(3)_A$. The η_1 and the η_8 mix and can be rotated to their mass basis

$$\begin{pmatrix} \eta \\ \eta' \end{pmatrix} = \begin{pmatrix} \cos\theta_\eta & -\sin\theta_\eta \\ \sin\theta_\eta & \cos\theta_\eta \end{pmatrix} \begin{pmatrix} \eta_8 \\ \eta_1 \end{pmatrix}$$

with a mixing angle $\theta_\eta \sim -11.5°$, so $\eta' \simeq \eta_1$. The pions and the kaons are indeed very light, $\pi^{\pm,0} \sim 140$ MeV, $K^{\pm,0} \sim 500$ MeV, and one would expect the etas to have masses similar to the kaons. While the η is only slightly heavier than the kaons ($m_\eta \sim 550$ MeV), the mass of the η', however, is similar to the mass of the proton. The non-observation of a ninth Goldstone boson was dubbed the $U(1)_A$ *problem*.

[8] Note however that if we include weak interactions baryon number is also anomalous.

The $U(1)_A$ problem was solved by t'Hooft by showing that the QCD vacuum is a superposition of vacua and is called the θ-vacuum [75]. As we will see in the following paragraph, the non-triviality of the θ-vacuum effectively adds a term to the Lagrangian

$$\mathcal{L}_{QCD}^{eff} = \mathcal{L}_{QCD} - \frac{\theta g_s^2}{32\pi^2} G^a_{\mu\nu}\tilde{G}^{a\mu\nu} \ , \tag{2.66}$$

which is the same shift that a $U(1)_A$ rotation of angle θ induces. The solution of the $U(1)_A$ problem creates however another problem, the *strong CP problem*.

2.3.3.2 Instantons

Instantons are non-trivial gauge configurations that are solutions of the classical Euclidean Yang-Mills equations of motion. Indeed, in addition to the trivial stable vacuum that we usually expand around there can be other non-trivial field configurations which are classically stable due to topological obstructions. These obstructions stabilize the non-trivial configurations and prevent a smooth transformation to the trivial vacuum.

In this section we will restrict ourselves to pure Yang-Mills theory. We will later review how quarks and their interactions with gluons enter the discussion. To see how these non-trivial solutions can contribute to the QCD path integral, and since instantons are solutions of the classical Euclidean equations of motion, let us start by analytically continuing to Euclidean space, i.e.

$$x^\mu = (t, x^i) \to (-ix_E^0, x_E^i) = x_E^\mu \ .$$

As we already alluded to in section 2.3.1.1, we can add a $G^a_{\mu\nu}\tilde{G}^{a\mu\nu}$ term to the Yang-Mills action which is a total derivative, see Eq. 2.55. Since this term can be written as a total derivative we can apply Gauss' theorem

$$\int \mathrm{d}^4x \ G^a_{\mu\nu}\tilde{G}^{a\mu\nu} = \int \mathrm{d}^4x \ \partial_\mu K_\mu = \int_{\mathcal{S}_3} \mathrm{d}S_\mu K_\mu \ , \tag{2.67}$$

with $\mathcal{S}_3$ the three-sphere at infinity of our Euclidean space-time. [9]In order to obtain finite energy solutions, G has to vanish at infinity

$$G_{\mu\nu} \xrightarrow{|x|\to\infty} 0 \ . \tag{2.68}$$

As we saw in section 2.3.1.1, the gauge potentials transform under a gauge transformation as

$$A_\mu \to \Omega^{-1}A_\mu\Omega + \frac{i}{g_s}\Omega^{-1}\partial_\mu\Omega \ , \tag{2.69}$$

with $\Omega(x)$ a group element of $SU(3)$. Hence, besides $A_\mu|_{\mathcal{S}^3_\infty} = 0$ one can find additional non-trivial configurations which give a finite action, namely by gauge transforming $A_\mu = 0$, which

[9]Note that in order to obtain smooth non-trivial gauge configurations we actually need to compactify our four-dimensional spacetime to $\mathcal{S}_4$ [76].

means A_μ takes a form known as "pure gauge" [10],

$$A_\mu \to \frac{i}{g_s}\Omega^{-1}\partial_\mu\Omega\,. \tag{2.70}$$

This describes a mapping from the asymptotic $\mathcal{S}^3_\infty$ (the boundary of Euclidean spacetime) to $SU(3)$. If we first look at $SU(2)$ which is homeomorphic to $\mathcal{S}^3$ one can show that these mappings $\mathcal{S}^3_\infty \to \mathcal{S}^3$ fall into homotopy classes which are labelled by an integer – the number of times the group manifold S^3 winds around the three-sphere at spatial infinity, $\mathcal{S}^3_\infty$ [11]. The periodicity theorem by Bott [78] ensures that any continuous mapping of $\mathcal{S}^3 \to G$, where G is a simple Lie group, can be continuously deformed into a mapping $\mathcal{S}^3 \to SU(2)$, with $SU(2)$ subgroup of G. Hence, it suffices to restrict ourselves to the homotopy groups of $SU(2)$ for any simple gauge group G – so also for $G = SU(3)$ – and we have

$$\Pi_3(SU(3)) = \mathbb{Z}\,. \tag{2.71}$$

The integral can be calculated using the so-called Cartan-Maurer invariant, giving

$$\int \mathrm{d}^4x\; G^a_{\mu\nu}\tilde{G}^{a\mu\nu} = \frac{32\pi^2\nu}{g_s^2}, \qquad \nu \in \mathbb{Z} \tag{2.72}$$

Hence, the configurations fall into homotopy groups which give a contribution to the path integral proportional to an integer (the so-called *winding-number*). Configurations in one homotopy group with integer ν are related to configurations corresponding to an integer $\mu \neq \nu$ via so-called "large" gauge transformations because the maps $\mathcal{S}^3_\infty \to SU(3)$ cannot be continuously deformed into the identity map.

To see that these configurations are solutions of the Euclidean equations of motion we complete the square in the Yang-Mills action and rewrite the action as

$$S_{\mathrm{YM}} = \frac{1}{4}\int \mathrm{d}^4x\; G^a_{\mu\nu}G^{a\mu\nu} = \frac{1}{8}\int \mathrm{d}^4x\;\; \mathrm{tr}\Big[(G_{\mu\nu} \mp \tilde{G}_{\mu\nu})^2\Big] \pm \frac{1}{4}\int \mathrm{d}^4x\;\; \mathrm{tr}\Big(G_{\mu\nu}\tilde{G}_{\mu\nu}\Big) \tag{2.73}$$

and hence, making use of the positivity condition in Euclidean space, identify a lower bound for the action

$$S_{\mathrm{YM}} = \frac{1}{4}\int \mathrm{d}^4x\; G^{a\mu\nu}G_{a\mu\nu} \geq \left|\frac{1}{4}\int \mathrm{d}^4x\;\; \mathrm{tr}\Big(G_{\mu\nu}\tilde{G}_{\mu\nu}\Big)\right| = \frac{8\pi^2|\nu|}{g_s^2}\,. \tag{2.74}$$

which is the *Bogomol'nyi bound* [79]. The lower bound is reached for a dual or anti-dual curvature, $G_{\mu\nu} = \pm\tilde{G}_{\mu\nu}$. Since these configurations minimize the action, they are solutions to the classical Euclidean equations of motion $D_\mu G^{\mu\nu} = 0$. They are called *instantons* since they are localised in both space and time. Belavin, Polyakov, Schwartz and Tyupkin explicitly constructed such an instanton solution for $\nu = 1$ [80]. For nice introductions to instantons see for instance [18, 81–84].

θ-vacua To appreciate the physical implications of these instanton solutions, let us Wick-rotate back to Minkowski space-time. Here, in a specific choice of gauge, we can interpret the instanton

[10] Note that in QED $F\tilde{F} = \frac{1}{2}\partial_\mu\epsilon^\mu_{\nu\rho\sigma}A^\nu F^{\rho\sigma}$ and F falling off faster than $1/|x|^2$ implies that A falls off faster than $1/|x|$, hence surface terms can be neglected (at least if there are no magnetic monopoles).

[11] The Wikipedia page on homotopy groups of spheres has nice pictures for this (at least for the circle and the two-sphere) [77].

solution as a tunneling between vacuum states of different winding number. Indeed, in the temporal gauge, $A_0^a = 0$, the spatial components of the Chern-Simons current vanish, $K_i = 0$, and we can write for an instanton solution:

$$\nu = 1 = \frac{g_s^2}{32\pi^2} \int \mathrm{d}^4x \; \partial_0 K^0 = \frac{g_s^2}{32\pi^2} \int \mathrm{d}^3x \; K^0|_{t=-\infty}^{t=\infty} \; . \tag{2.75}$$

So the instanton is an interpolation from a configuration with winding number n to a configuration with winding number $n+1$. But the Chern-Simons current K_μ is itself not gauge invariant, hence the states at $t = \pm\infty$ with a given winding number have no real physical meaning. This can also be seen by applying a large gauge transformation Ω_ν corresponding to a winding number ν to a state with winding number n,

$$\Omega_\nu \, |n\rangle = |n + \nu\rangle \; , \tag{2.76}$$

so that all of these fixed winding number states are gauge transforms of one another. In particular, they are gauge transforms of the usual vacuum state $|0\rangle$. Gauge fixing only removes the gauge redundancy of small gauge transformations which describe physically equivalent configurations and form physical equivalence classes. Gauge fixing does not remove these so-called "large" gauge transformations in Eq. (2.76) because these describe physically distinct configurations. However, we require a gauge invariant vacuum. We can construct such a vacuum from the superposition of different vacuum states

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{in\theta} \, |n\rangle \; , \tag{2.77}$$

such that the superposition of vacua is now an eigenstate of large gauge transformations,

$$\Omega_n \, |\theta\rangle = e^{i\theta} \, |\theta\rangle \; . \tag{2.78}$$

$\theta \in [0, 2\pi)$ is often called the *θ-angle* and the new vacuum state is referred to as *θ-vacua*. For the computation of physical observables we hence have to expand around the θ-vacua which means that we have to sum over all distinct topological sectors,

$$\langle O \rangle = \frac{1}{Z} \sum_\nu e^{-\nu\theta} \int \mathrm{d}\,[A]_\nu \; O[A] \; e^{-S[A]} \; , \tag{2.79}$$

and since $\nu = g_s^2/(32\pi^2) \int \mathrm{d}^4x G\tilde{G}$, the superposition of vacua effectively adds a term

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{YM}} - \frac{\theta g_s^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \tag{2.80}$$

to the Yang-Mills Lagrangian which takes the sum over topological sectors into account [75, 85].

2.3.4 The strong CP problem

2.3.4.1 The chiral anomaly

As we discussed in Sec. 2.3.3, in the limit of $m_f = 0$, $\mathcal{L}_{QCD}$ has a global (chiral) symmetry, $SU(2)_L \times SU(2)_R \times U(1)_V \times U(1)_A$ [73]. $U(1)_V$ is an exact symmetry, however, the axial

symmetry $U(1)_A$,

$$q_j \to e^{i\alpha\gamma_5/2} q_j, \tag{2.81}$$

is anomalous, i.e. it is a symmetry at the level of the Lagrangian (so of the classical theory) which is broken by quantum corrections (so it is not a symmetry). The original way the chiral anomaly was derived is by calculating triangle graphs and imposing Ward-identities to preserve gauge invariance.

Another way anomalies can be understood is by the non-invariance of the fermionic measure in the path integral (the derivation is due to Fujikawa [86, 87]) and the Jacobian of the transformation has to be included. Since the fermionic measure is composed of Grassmann variables its Jacobian is the inverse of the determinant and we obtain for a local chiral transformation of matrix $U(x) = \exp\left(i\alpha(x)\gamma^5\right)$

$$[\mathcal{D}\psi\mathcal{D}\overline{\psi}] \to \frac{1}{\det(U)\det(\overline{U})}[\mathcal{D}\psi\mathcal{D}\overline{\psi}] = \frac{1}{\det(U)^2}[\mathcal{D}\psi\mathcal{D}\overline{\psi}] \,. \tag{2.82}$$

The Jacobian of the transformation can be written as an additional effective term in the Lagrangian. The derivation is lengthy (see original paper by Fujikawa [86, 87] or for instance [81]), so we only state the result here:

$$[\mathcal{D}\psi\mathcal{D}\overline{\psi}] \to [\mathcal{D}\psi\mathcal{D}\overline{\psi}] \exp\left(-i\alpha\frac{g_s^2}{16\pi^2}\int \mathrm{d}^4x\; G^a_{\mu\nu}\tilde{G}^{a\mu\nu}\right) . \tag{2.83}$$

Note that local (i.e. gauge) symmetries need to be anomaly free in order to ensure consistent quantization. The SM gauge group $SU(3)_C \times SU(2)_W \times U(1)_Y$ is indeed anomaly free.

2.3.4.2 Adding massive fermions: the strong CP problem

The discussion in the section 2.3.3.2 concerned pure Yang-Mills theory only. Configurations exist for which the $G\tilde{G}$ term in the Yang-Mills Lagrangian gives a finite non-trivial contribution to the action and the term should be present in the Lagrangian [12]. If we now come back to full QCD,

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{q}\left(i\not{D} - m_q \exp(i\theta_{q,\mathrm{EW}})\right) q - \frac{1}{4}G^{a\mu\nu}G^a_{\mu\nu} + \theta_{\mathrm{QCD}}\frac{g_s^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^{a\mu\nu} \,, \tag{2.84}$$

with $\theta_{q,\mathrm{EW}}$ the phases of the quark masses – we know from B- and Kaon-physics that the Yukawa couplings are complex. Now, under a $\mathrm{U}(1)_A$ chiral transformation of parameter α,

$$q \to \exp\left(i\gamma_5\alpha\right) q \,,$$

the associated fermionic Noether current $J^5_\mu = \bar{q}\gamma_\mu\gamma_5 q$ is not conserved because:

[12] The same is true for the electroweak θ-term. Hence, in pure $SU(2)_L$ gauge theory, instanton solutions can be physical. However, the fact that we add fermions and their interactions with the weak gauge bosons makes the term unphysical: The fermions transform chirally under $SU(2)_L$ which permits us to independently rotate the right handed fermions in order to cancel the θ-term. As $\mathrm{U}(1)_Y$ is abelian the corresponding θ-term integrates to zero for finite energy configurations.

1) in the quark mass term this transformation shifts the phase of the quark masses by $\theta_{q,\mathrm{EW}} \to \theta_{q,\mathrm{EW}} + 2\alpha$,

2) of the chiral anomaly, discussed in the previous subsection, which does not leave the fermionic path integral measure invariant and effectively shifts the QCD phase by $\theta_{\mathrm{QCD}} \to \theta_{\mathrm{QCD}} - 2\alpha$.

The non-invariance under a chiral rotation can be expressed in terms of the axial current Ward identity

$$\partial^\mu J^5_\mu = 2m_q \bar{q} i\gamma_5 q + \frac{g_s^2}{16\pi^2} G\tilde{G} \,. \tag{2.85}$$

The axial anomaly shifts the quark mass phases contribution to the QCD vacuum. Hence one could attempt to use the relation between the chiral anomaly and the θ-terms to remove any CP-violating effect from the QCD Lagrangian. However, once the phase is removed from the quark mass matrix a chiral rotation of the fermion fields will always reintroduce a complex phase due to the mass term. We see that only the combination of parameters

$$\bar{\theta} = \theta_{\mathrm{QCD}} + \theta_{q,\mathrm{EW}} \tag{2.86}$$

is invariant under a field redefinition and hence physical. This allows us to rewrite the Lagrangian in Eq. (2.84) where the quark mass phases have been absorbed into the QCD vacuum term

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{q}(i\not{D} - m_q)q - \frac{1}{4} G^{a\mu\nu} G^a_{\mu\nu} + \bar{\theta} \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \,. \tag{2.87}$$

Generalized to the full electroweak theory the physical parameter reads

$$\bar{\theta} = \theta_{\mathrm{QCD}} + \arg\det\left(Y_U Y_D\right) \,, \tag{2.88}$$

with Y_U and Y_D the up and down Yukawa matrices.

Observable effects of the θ-term: neutron electric dipole moment While the addition of the θ-term due to instanton effects by 't Hooft solves the $U(1)_A$ problem, a fundamentally different problem appears. As we saw in the previous sections, the last term in the Lagrangian in Eq. (2.87) is odd under the combined discrete CP transformation. We cannot get rid of this term if we assume that all quarks are massive. One consequence of such a term is a non-vanishing electric dipole moment of the neutron. However, the experimental upper bound indicates $|d_n| < 2,9 \times 10^{-26}$ e cm which corresponds to $|\bar{\theta}_{QCD}| < 10^{-10}$ (while one would expect it to be rather of $\mathcal{O}(1)$ or $\mathcal{O}(\delta)$, with $\delta \sim 10^{-3}$ the phase in the CKM matrix) [88–90]. The smallness and thus fine-tuning of $|\bar{\theta}_{QCD}|$ is known as the strong CP-problem [73, 91]. The problem really is why the combination of the QCD and EW CP violating terms is so small.

A few remarks are now in order. As we saw, instantons provide a solution to the U(1)-problem by introducing an axial-symmetry violating term. However, these non-trivial gauge configurations cause the strong CP problem which is ultimately related to the fact that all quarks in QCD are massive. A vanishing Yukawa coupling for at least one fermion would render the θ-term unphysical, see Eq. (2.88). To fully appreciate these connections, let us briefly comment on a theorem by Atiyah and Singer. Atiyah and Singer [92] showed that the index of the Dirac

operator in Euclidean space, which is defined as the difference of zero modes of left and right handed fermions, $\text{index}(i\not{D}) \equiv n_+ - n_-$, is given by

$$n_+ - n_- = \frac{g_s^2}{32\pi^2} \int \mathrm{d}^4x \; G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \; . \tag{2.89}$$

As we argued before, the right-hand side is an integer (the winding number) – this deep result also explains why the anomaly is exact and does not receive higher-order perturbative or non-perturbative corrections. Now, from the path integral, we obtain the tunnelling amplitude between two vacua,

$$\langle n \mid n+\nu \rangle \sim \int \mathcal{D}A\mathcal{D}q\mathcal{D}\bar{q} \exp\left(- \int d^4x \frac{1}{4} \operatorname{tr} G^{\mu\nu} G_{\mu\nu} + i\bar{q}\not{D}q \right) \tag{2.90}$$

$$\sim \int \mathcal{D}A \det(i\not{D}) \exp\left(- \int d^4x \frac{1}{4} \operatorname{tr} G^{\mu\nu} G_{\mu\nu} \right) , \tag{2.91}$$

where we have performed the Gaussian integral over the Grassmannian degrees of freedom analytically. In any configuration without fermions with winding number $\nu \neq 0$ we obtained a contribution to the amplitude, see Eq. (2.74). However, in the presence of fermions in configurations with $\nu \neq 0$ this now means that the massless Dirac operator has zero modes (since $\text{index}(i\not{D}) \neq 0$) and in the case of massless fermions, $m = 0$, this implies $\det(i\not{D}) = 0$. In this case, the tunnelling amplitude vanishes. This is another way of understanding why the θ-term is unphysical and can be rotated away if we have a massless quark [84].

Solutions

- **A massless quark.** As can be observed from Eq. (2.88) and as we argued in the previous paragraph, a vanishing Yukawa coupling for one of the fermions automatically resolves the strong CP problem. We could rotate the massless quark field independently by a phase to remove the θ-term without reintroducing the phase in its mass term. The θ-term would then simply correspond to a phase of a massless quark. However, lattice QCD excludes the massless quark solution by more then 24σ [93].

- **Setting $\bar{\theta} = 0$ in the ultraviolet.** If $\bar{\theta}$ is set to zero in the ultraviolet then it will stay very small since it runs very little if the effective theory contains only the SM (in the SM the lowest order contribution arises at seven loops). This contribution is much smaller than the experimental bound [94].

- **The axion.** The axion solution is probably the most prominent solution to the strong CP problem, as it requires little tuning and minimal addition of new physics. The axion solution to the strong CP problem will be discussed in detail in Chap. 6.

- ...

Part II

Testing the standard model via the electromagnetic coupling

3 Lattice Quantum Chromodynamics

It is difficult to formulate the RG approach for a particular problem and additionally, one is likely to have to carry out a complicated computer calculation, which makes most theoretical physicists cringe. Especially in the case of strong interactions of elementary particles, most theorists hope to solve the problem without turning to RG methods and computers. It will probably require several years of stagnation before theorists will accept the inevitability of the renormalization group approach despite its difficulties.
Wilson, 1975

It is no coincidence that Kenneth Wilson, who earned a Nobel prize for his contributions to the renormalization group, is also one of the founders of lattice field theories. As we mentioned in Sec. 2.3.2, Wilson introduced lattice gauge theories to address the issue of confinement of quarks [95]. In 1980, Michael Creutz performed a first promising numerical study for SU(2) gauge theories, where he made use of the heat-bath algorithm [96]. Indeed, lattice field theories rely heavily on the insights gained from renormalization group theory and the concepts could not be understood without tying the link to renormalization. We will discuss this connection in Sec. 3.6.

As we saw in Sec. 2.2.2, the naive perturbative calculation of physical observables in QED necessarily leads to ultraviolet divergences arising from loop integrals. Similarly, for QCD, perturbation theory in the bare coupling g_0 leads to divergent momentum integrals. We argued that we were too ambitious in trying to explain physics up to infinite four-momentum or zero distances: this led to a conceptual paradigm shift. By introducing a cutoff – i.e. eliminating the most energetic degrees of freedom – the integrals become convergent. Hence, the very definition of a quantum field theory requires the temporary introduction of an ultraviolet regulator. However, we required that our physical observables cannot depend on this arbitrarily chosen cutoff. This can be achieved order by order in perturbation theory by letting the coupling constants in our theory depend on this cutoff since they themselves are not measurable in experiment. Theories where this is possible for a finite fixed number of parameters are called renormalizable theories, as it is the case for QCD [97–100]. In Sec. 2.3.2 we saw that QCD is asymptotically free. Its coupling increases at low energies and perturbation theory breaks down: we need non-perturbative techniques to calculate physical observables. The regularization procedures we used in Sec. 2.2.2 are hence inappropriate and we require a non-perturbative tool to restrict the degrees of freedom in our theory. Lattice regularization, where one discretizes spacetime, is such a tool, since it naturally eliminates separations smaller than a minimal distance. In addition, the functional integral in the continuum theory is more than a high-dimensional integral and is

difficult to define. By studying QCD on a lattice both these problems are solved. Finally, for sufficiently large volumes and in the absence of massless quarks, the gap in QCD provides an IR cutoff. Otherwise, the boundary condition can be chosen so that the dimensions of the box serve as an IR cutoff. In any event, lattice QCD is the only known way to fully solve QCD in its highly non-linear non-perturbative regime.

In Minkowski spacetime the probability measure inside the path integral is complex. By analytically continuing to Euclidean spacetime, a relation to statistical physics can be established and one obtains a real normalization factor for the path integral:

$$Z = \sum_{\substack{\text{all Euclidean} \\ \text{gauge field configurations}}} \exp\left(-\frac{1}{g^2}\int d^4x\ \frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu}\right),$$

which is the partition function of a four-dimensional classical Yang-Mills system. The justification for performing this analytic continuation, or Wick rotation, is given by the Wightman axioms [101] and the Osterwalder-Schrader theorem [102, 103]. The path integral formulation in Euclidean spacetime is then suitable for efficient numerical evaluations using Monte Carlo techniques and has the advantage that numerical techniques from statistical physics can be applied. We will comment on numerical methods in Sec. 3.8.

To introduce and formalize lattice field theories and in particular QCD on the lattice, we will proceed along the same lines as in Sec. 2.3.1. We start by discretizing the free fermion action which, if done naively, gives rise to a species doubling problem. To eliminate some of these spurious degrees of freedom, we use staggered fermions in this thesis. We will then see that in order to obtain a gauge invariant action we have to introduce a parallel transporter. This parallel transporter constitutes one of the main building blocks, linking neighboring sites of the lattice. By exactly preserving gauge invariance at all stages the lattice theory inherits important properties of the continuum theory, such as renormalizability. Finally, we also discuss possibilities of defining QED in a finite volume in Sec. 3.7.

3.1 Non-perturbative regularization

Lattice regularization is a non-perturbative regularization as opposed to dimensional regularization which we used in Sec. 2.2 to regularize the photon vacuum polarization. In a theory regularized on the lattice it is possible to maintain gauge invariance at all steps of the calculation. We will see that the path integral can be defined in a natural way by discretizing space-time (in fact, this is also the way it was originally derived by Feynman [104]).

The idea is to replace continuous $d = d_s + 1$-dimensional Euclidean spacetime by a discrete hypercubic lattice [1]. On a hypercubic lattice the lattice spacing "a" is identical in all directions. Matter fields, such as quark fields, live on the sites of the lattice and will be labelled by ψ_n, where $n = (n_0, n_1, n_2, n_3)$ indicates the site on the lattice. Lattice regularization naturally provides a

[1] We will argue later that the precise way we discretize spacetime and the QCD action should not matter, due to a property called universality.

momentum cutoff; it restricts the momentum values to the first Brillouin zone

$$-\frac{\pi}{a} < p_\mu \leq \frac{\pi}{a} \,, \tag{3.1}$$

since the Fourier transform of any function on the lattice,

$$\tilde{f}(p) = a^4 \sum_n \exp(ipx) f(x) \,, \qquad \text{with} \qquad x = an \tag{3.2}$$

is $2\pi/a$ periodic. Another way of seeing it is that the minimal wavelength that can be resolved on a lattice with lattice spacing a is $\lambda_{\min} = 2a$. Hence $k_{\max} = 2\pi/\lambda_{\min} = \pi/a$.

In order to simulate QCD on a computer, we have to restrict it to a finite number of variables, which means that, in addition to discretizing spacetime, we put it in a box $T \times L^3$ with $L = N_s a$ and $T = N_t a$ and where N_t are the number of points in the temporal direction and N_s the number of points in the spatial directions. If we use periodic boundary conditions, this has the additional effect that momenta are quantized with minimal spatial quantum $p_{\min} = 2\pi/L$ and energy quantum $E_{\min}$. Finally, the hypercube is defined as

$$\Lambda_E = \left\{ x \in \mathbb{R}^4 \,|\, n_i = \frac{x^i}{a} = 0, \ldots, N_s - 1,\, i = 1,2,3\,;\, n_0 = \frac{x^0}{a} = 0, \ldots, N_t - 1 \right\} \,. \tag{3.3}$$

The lattice regularization however breaks some symmetries of the continuum theory. For instance, $SO(4)$ invariance (corresponding to Lorentz invariance in Minkowski space) is broken down to the group of hypercubic rotations. We will see in Sec. 3.3 that gauge invariance can be exactly preserved.

3.2 Lattice fermion action

Let us come back to the free fermion Lagrangian of Sec. 2.3.1. After Wick rotating to Euclidean spacetime, the Euclidean free fermion action reads

$$S_F = \int \mathrm{d}^4 x\, \bar{\psi}(x) \left(\gamma_\mu \partial_\mu + m \right) \psi(x), \tag{3.4}$$

where the Euclidean gamma matrices satisfy the Clifford algebra

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu} \,. \tag{3.5}$$

A naive discretization of the fermion action on a hypercube is

$$S_{F,\text{naive}} = a^4 \sum_{x,\mu} \bar{\psi}(x)(\gamma_\mu \nabla^\pm_\mu + m)\psi(x) = a^4 \sum_{x,y,\mu} \bar{\psi}(x) D(x,y) \psi(y) \,, \tag{3.6}$$

where we replace the derivative by a suitable difference quotient, in this case by the central difference

$$\nabla^\pm_\mu \equiv \frac{1}{2} \left(\nabla^+_\mu + \nabla^-_\mu \right) \,, \tag{3.7}$$

with ∇_μ^+ and ∇_μ^- forward and backward finite differences, respectively

$$\nabla_\mu^+\psi(x) = \frac{\psi(x+\hat{\mu}a)-\psi(x)}{a} \qquad \nabla_\mu^-\psi(x) = \frac{\psi(x)-\psi(x-\hat{\mu}a)}{a} \,. \tag{3.8}$$

In the limit $a \to 0$ $\nabla_\mu^\pm$ converges classically to the continuum derivative up to $\mathcal{O}(a^2)$ corrections. Note that, since in quantum mechanics the momentum operator of a free particle is $p \equiv -i\,\mathrm{d}/\mathrm{d}x$ and has to be hermitian in order to correspond to an observable, the derivative operator has to be anti-hermitian (a skew-symmetric matrix in a matrix representation of the operator if the elements are real). Hence, if we want to preserve unitarity, we cannot use the naive forward or backward derivative operator. As we will see in the following section, this has far-reaching consequences.

3.2.1 Fermion doubling problem

As we will see, the central derivative operator Eq. (3.7) suffers from the so-called "fermion-doubling problem" [105–107]. Indeed, the Dirac operators' Fourier transform is

$$\tilde{D}(p) = \frac{i}{a}\sum_\mu \gamma_\mu \sin(ap_\mu) + m \,, \tag{3.9}$$

where p_μ belongs to the first Brillouin zone of Eq. (3.1). Taking its inverse we obtain the fermionic propagator in momentum space,

$$G(p) = \tilde{D}^{-1}(p) = \frac{-i/a\sum_\mu \gamma_\mu \sin(ap_\mu) + m}{1/a^2\sum_\mu \sin^2(ap_\mu) + m^2} \,, \tag{3.10}$$

which has, as required, a pole at $p^2 = -m^2$ as $a \to 0$. However, for each direction μ the propagator has an additional pole (in four dimensions this gives 2^4 possibilities) in the corners of the Brillouin zone, leading to fifteen additional poles with a momentum shift

$$\pi_A^\mu \in \{\tfrac{\pi}{a}(1,0,0,0), \tfrac{\pi}{a}(0,1,0,0), \ldots, \tfrac{\pi}{a}(1,1,0,0), \ldots, \tfrac{\pi}{a}(1,1,1,1)\} \,.$$

The poles in the propagator give the particle content of the theory. Equating the denominator of Eq. (3.10) to zero, one finds finds that both ap_0 and $\pi - ap_0$ [2] have the same solutions. For the particle states the solutions are $ip_0 = E$ and $i\pi - E$ where $E > 0$ (antiparticle solutions correspond to $E < 0$) are the energies of the states and are given by the following lattice dispersion relation:

$$\sinh(Ea)^2 = \sum_i \sin^2(p_i a) + m^2 \,. \tag{3.11}$$

$E(\vec{p})$ is identical for 2^3 momenta $n_i\pi/a - p_i$, for $p_i \geq 0$ and $p_i < 0$ respectively, with $n_i \in \{0,1\}$. All of these energies survive in the continuum limit

$$\lim_{a\to 0} |E(\vec{p})|_{n_i\pi/a} = \sqrt{\vec{p}^2 + m^2} \,, \tag{3.12}$$

[2] Here we assume that $ap_0 \geq 0$ so that $\pi - ap_0$ is in the Brillouin zone defined in Eq. (3.1). If $ap_0 < 0$, it is $\pi + ap_0$ that must be considered.

so that the different poles have the same energy. Thus, the field ψ corresponds to sixteen relativistic free fermions, which is the problem of species doubling.

We will encounter the fermion doublers again in our perturbative calculation of the current-current propagator needed for the hadronic vacuum polarization on the lattice where we derive this dispersion relation. The details are given in App. B. We can understand intuitively why this problem arises: by taking the naive operator Eq. (3.7) we are skipping a site and are comparing the value of the field at a certain site with its value two sites away. However, the highest frequency that can be resolved oscillates with π/a. The naive derivative operator finds zero for the derivative of this field, instead of its maximal value.[3]

Indeed, it is possible to explicitly construct doubling symmetry transformations. This discussion closely follows Ref. [108] and it will also be very helpful for the presentation of staggered quarks below and for the construction of staggered currents in the subsequent chapter. First we define the 16 "doubling" transformations

$$\psi(x) \to \mathcal{B}_t(x)\psi(x) \,, \qquad \overline{\psi}(x) \to \overline{\psi}(x)\mathcal{B}_t^\dagger(x) \,, \tag{3.13}$$

with the $\mathcal{B}_t(x)$ defined by

$$\mathcal{B}_t(x) \equiv (-1)^{t\cdot x/a}\gamma_{\overline{t}} = e^{i\pi t\cdot x/a}\gamma_{\overline{t}} \,, \tag{3.14}$$

and labelled by the vector $t_\mu \in \mathbb{Z}_2 \equiv \{0,1\}$ and its "conjugate", given by

$$\overline{t}_\mu \equiv \sum_{\nu\neq\mu} t_\nu \bmod 2 \,. \tag{3.15}$$

The 16 matrices γ_t are given by

$$\gamma_t = \prod_{\mu=0}^{3}(\gamma_\mu)^{t_\mu} = \gamma_0^{t_0}\gamma_1^{t_1}\gamma_2^{t_2}\gamma_3^{t_3} \tag{3.16}$$

and provide a complete set of Dirac matrices, including the identity for $t = (0,0,0,0)$. Some useful properties of the γ_t are summarized in App. A of Ref. [108].

It is clear from the presence of the phase in Eq. (3.14) that the corresponding transformations map a state in one corner of the Brillouin zone into all 15 other corners as t takes on its 15 values different from $(0,0,0,0)$. With some algebra, it is also possible to show that the naive fermion action of Eq. (3.6) is invariant under the doubling transformations of Eq. (3.13). Moreover, changing basis for the doubling transformations from $\mathcal{B}_t(x)$ to the hermitian one $B_A(x) \in \{1, i(-1)^{x_\mu/a}\gamma_5\gamma_\mu, \epsilon(x)\gamma_5, (-1)^{\sum_{\nu\neq\mu} x_\mu/a}\gamma_\mu, (i/2)(-1)^{(n_\mu+n_\nu)/a}\gamma_{[\mu}\gamma_{\nu]}\}$, with $\mu < \nu$ and $\epsilon(x) = (-1)^{\sum_\mu x_\mu/a}$, one can generalize the discrete transformations of Eq. (3.13) to the continuous ones of $U(4)_V$ given by [109]:

$$\psi(x) \to e^{i\omega_A B_A(x)}\psi(x) \,, \qquad \overline{\psi}(x) \to \overline{\psi}(x)e^{-i\omega_A B_A(x)} \tag{3.17}$$

where the ω_A are real parameters. With additional work, one can further show that these $U(4)_V$ doubling transformations are a symmetry of the naive fermion action [110, 111]. In

[3] Note that this problem only arises for fermions, since in the bosonic action we need a second derivative which can be approximated by the derivative operator $\Delta_\mu^\pm\Phi(x) = 1/a^2\left[\Phi(x+a\hat{\mu}) - 2\Phi(x) + \Phi(x-a\hat{\mu})\right]$ which does not skip a site.

fact, when the fermion mass is identically zero, the symmetry of the action enlarges to a chiral $U(4)_L \otimes U(4)_R$ [112], of which the $U(4)_V$ symmetry of Eq. (3.17) is the diagonal subgroup.

In addition, in the free theory the 16 doubler states cannot transform into one another and one finds that the theory has a chiral $U(16)_L \otimes U(16)_R$ symmetry when $m = 0$, which reduces to the vector $U(16)_V$ subgroup for $m \neq 0$. In analogy with flavour, this symmetry is called a "taste" symmetry. At finite lattice spacing, and in the presence of interactions, this symmetry is broken by the exchange of gluons with some of their four-momenta components close to π/a. Clearly such gluons take a fermion from one corner of the Brillouin zone to another, thus changing its taste. However once the continuum limit is taken, π/a goes to infinity and the taste symmetry is restored even in the presence of interactions. It then corresponds to the usual flavour symmetry of 16 identical species of fermions. However, the axial anomaly receives contributions from all 16 species. Because the blocking transformations involve factors of γ_5, the doublers contribute with opposite signs in the pattern $1 - 4 + 6 - 4 + 1 = 0$ where the multiplicity and signs are associated with the corners of the Brillouin zone labelled by the following representative $t = (0,0,0,0)$, $(1,0,0,0)$, $(1,1,0,0)$, $(1,1,1,0)$, $(1,1,1,1)$, respectively [113]. Thus, the flavour singlet axial symmetry $U(1)_A$ is not broken by the anomaly and naive fermions cannot be a discretization of even 16-flavour QCD.

To eliminate doublers Wilson came up with a solution to the doubling problem by adding a Laplacian-like term to the lattice action. This term is an $\mathcal{O}(a)$ irrelevant operator and vanishes in the continuum limit,

$$S_{WF} = S_{F,naive} - \frac{a}{2} \sum_x \bar{\psi}(x) \Delta^{\pm} \psi(x) \ , \tag{3.18}$$

with $\Delta^{\pm} = \sum_\mu \nabla_\mu^- \nabla_\mu^+ = a^{-2} \sum_\mu (\nabla_\mu^+ + \nabla_\mu^- - 2)$. In this case the propagator becomes

$$\tilde{D}^{-1}(p) = \left[i \sum_\mu \gamma_\mu \frac{\sin\left(p_\mu a\right)}{a} + m + \sum_\mu \frac{2}{a} \sin^2\left(\frac{p_\mu a}{2}\right) \right]^{-1} \ . \tag{3.19}$$

Now, considering a fermion with four-momentum components much smaller than π/a, the Wilson term contributes $1/a$ corrections to the doublers' energies. Thus as the continuum limit is taken the doublers have energies that grow and therefore contribute less and less to the physics at scales much smaller than $1/a$ and fully decouple from the theory in that limit, leaving behind a single fermion with energy well below the cutoff. However, chiral symmetry is explicitly broken by this additional term that couples left and right-handed fermions.

There exists even a no-go-theorem by Nielsen and Ninomiya concerning the doubling problem [114, 115]. It states that it is impossible to construct an (Euclidean) action which is ultra local, chirally symmetric, has the correct continuum limit and does not have doublers [4] for a fermion formulation with a local, translational invariant, hermitian Hamiltonian.

Clearly, the no-go theorem implies that we have to give up one of the assumptions. We already mentioned Wilson fermions as a solution to the doubling problem which however break chiral symmetry. Other early attempts were for example the non-local SLAC operator [26, 116], for an interesting discussion see also [117] [5]. This operator can be used in systems without local gauge

[4] Ultralocal denotes an action with a coupling between sites separated by only a finite number of lattice spacings.

[5] Having been trained as an engineer, these were also my first thoughts about this problem.

invariance and its spectrum is identical to the continuum derivative operator below $p = \pi/a$. Another very interesting concept is the concept of perfect actions [118, 119]. These are lattice actions whose predictions agree with those of the continuum action. These perfect (fixed-point) actions can be constructed by applying renormalization group transformations to the continuum free fermion action. However, constructing such an action would require the tuning of the coefficients of an infinite number of irrelevant operators.

In 1982 Ginsparg and Wilson proposed that the requirement for chiral symmetry could be relaxed on the lattice to satisfy [120]

$$\{D, \gamma_5\} = aD\gamma_5 D \,. \tag{3.20}$$

Actions which satisfy the Ginsparg-Wilson relation also satisfy the Atiyah-Singer index theorem that we mentioned in Sec. 2.3.3

$$\text{index}(D) = a^5 \sum_x \frac{1}{2} \text{Tr}(\gamma_5 D) = n_- - n_+ \,. \tag{3.21}$$

It was also shown that the Ginsparg-Wilson relation implies the correct chiral anomaly in the flavour singlet case [121]. Perfect actions satisfy the Ginsparg-Wilson relation, but have not been fully constructed. Other fermion actions that do are domain-wall fermions (DWF) which bind chiral four-dimensional fermion modes to a domain wall in five dimensions [122, 123] and overlap fermions [124, 125]. These actions are therefore chiral symmetry preserving, do not suffer from a species doubling problem and are widely used.

3.2.2 Kogut–Susskind (staggered) quarks

In the staggered fermion formulation, the number of fermion doublers is reduced by a factor of four and a $U(1)$ chiral symmetry is maintained. For later convenience we will use here the notation of Ref. [108]. The first step is to perform the (unitary) spacetime-dependent field redefinition of the fermion fields [111]:

$$\psi(x) \to \Omega(x) X(x) \,, \qquad \overline{\psi}(x) \to X(x)\Omega^\dagger(x) \,, \tag{3.22}$$

where

$$\Omega(x) = \gamma_n = \prod_{\mu=0}^{3} (\gamma_\mu)^{n_\mu} \tag{3.23}$$

and $x_\mu = a n_\mu$. One can show that – using some Dirac algebra – after this field redefinition, the action in the new field variables X is diagonal in spinor space,

$$S_F(X, \bar{X}) = a^4 \sum_x \bar{X}(x) \left(\sum_\mu \eta_\mu(x) \nabla_\mu^\pm + m \right) X(x) \tag{3.24}$$

with $\eta_\mu(x) \equiv \Omega^\dagger(x)\gamma_\mu\Omega(x) = (-1)^{\sum_{\rho<\mu} n_\rho} = \pm 1$ a simple phase factor arising from the anti-commutation of the Dirac matrices, such that

$$\eta_1(x) = 1 \,, \quad \eta_2(x) = (-1)^{n_1} \,, \quad \eta_3(x) = (-1)^{n_1+n_2} \,, \quad \eta_4(x) = (-1)^{n_1+n_2+n_3} \,. \tag{3.25}$$

As will be seen in the following section, gauge interactions are introduced by modifying the symmetric derivative $\nabla_\mu^\pm$ using (3.39).

Since the action is diagonal in spinor space, three of the four components of X in Eq. (3.24) can be dropped. We will denote χ the remaining single component of the fermion field X and the staggered action is the one component version of the one given in (3.24), i.e. $S_F(\chi, \bar{\chi})$. The fact that the equations of motion given by the naive fermion action relate one component of the Dirac field ψ on even (odd) sites to a linear combination of the other components at odd (even) sites in a staggered fashion and the fact that the spin diagonalization of (3.22) reduces this structure to a single component is the reason for which the particle (antiparticle) degrees of freedom associated with χ are called a staggered fermions.

In the reduced theory the $SU(4)_V$ doubling symmetry of (3.13) is removed in a nontrivial way: all symmetries, except for the (future) gauge and vector flavour symmetries (arising when multiple one-component fields with identical masses are considered), become x-dependent [126]. Thus the properties of the physical degrees of freedom of this reduced action are not evident: we still have to show that it corresponds to a theory with 4 tastes of Dirac fermions. Since the staggered transformation mixes Dirac and space-time indices in order to recover a four-component Dirac spinor, we build linear combinations of χ-fields at different lattice sites. In particular, the fields living on the sixteen sites of a hypercube can be grouped together into a new blocked field [127, 128],

$$\psi_B^{\alpha t}(x_B) = \frac{1}{8} \sum_{\delta n_\mu \in a\mathbb{Z}_2} \Omega(\delta x)^{\alpha t} \chi(x_B + \delta x) \,, \qquad \overline{\psi}_B^{t\alpha}(x_B) = \frac{1}{8} \sum_{\delta n_\mu \in a\mathbb{Z}_2} \overline{\chi}(x_B + \delta x) \Omega^\dagger(\delta x)^{t\alpha} \,, \quad (3.26)$$

where we have divided the original lattice in hypercubes which are labelled by $x_{B\mu} = 2aN_\mu$, $x_{B\mu} \bmod 2 = 0$ and $x_B + \delta x$, $\delta x_\mu \in \{0, a\}$ correspond to the sixteen sites of the hypercube and the new lattice has a doubled lattice spacing $2a$. We have made explicit the spin α and "taste" t indices which we label by Greek and Latin letters, respectively. They each run from $1, \ldots, 4$ and denote the rows and columns of the Dirac matrix $\Omega(\delta x)$. Then the staggered action Eq. (3.24) can be re-expressed in terms of the blocked fields as

$$S_F\left[\psi_B, \overline{\psi}_B\right] = (2a)^4 \sum_{x_B} \overline{\psi}(x_B) \left\{ m(\mathbb{I} \otimes \mathbb{I}) + \sum_\mu \left[(\gamma_\mu \otimes \mathbb{I}) \nabla_\mu^\pm - a(\gamma_5 \otimes \xi_\mu \xi_5) \Delta_\mu^\pm \right] \right\} \psi_B(x_B) \,, \quad (3.27)$$

with $2a$ the lattice spacing of the blocked lattice. The notation $\gamma_\Gamma \otimes \xi_{\Gamma'}$ indicates that the Dirac matrix γ_Γ acts on the Dirac indices and $\xi_{\Gamma'}$, with $\xi_\mu = \gamma_\mu^*$, on the taste ones. While the first two terms are diagonal in taste space, the last one, which looks like a "Wilson-type" term, explicitly breaks taste symmetry. This term vanishes in the naive continuum limit and we recover an action of four flavours of free massive Dirac spinors. This justifies *a posteriori* the statement made above that t denotes the four tastes associated with staggered fermions.

The above argument is useful because it gives a sense of why a staggered fermion field generates four tastes. However it is misleading [128]:

1. the taste symmetry of free staggered fermions is not broken in the absence of interactions;
2. there are no $\mathcal{O}(a)$ discretization errors with free (also interacting) staggered fermions: these start at $\mathcal{O}(a^2)$.

The problem is that the blocking transformation of (3.26) only projects the staggered field on the given taste t approximately, up to discretizations errors of $\mathcal{O}(ap)$ for a fermion of momentum $p \in (-\pi/2, \pi/2]^4$. This can be seen by going back to the original naive fermion field ψ of (3.6). In terms of that field, the four-component generalization of the blocking transformation of (3.26) can be written [108]:

$$\psi_B(x_B) = \frac{1}{8} \sum_{\delta x_\mu \in a\mathbb{Z}_2} \mathcal{B}_t(x_B + \delta x)\psi(x_B + \delta x) = \frac{e^{i\pi t \cdot x_B/a}}{8} \sum_{\delta x_\mu \in a\mathbb{Z}_2} e^{i\pi t \cdot \delta x/a} \gamma_{\bar{t}} \psi(x_B + \delta x) \ , \tag{3.28}$$

Clearly the sum over the corners of the hypercube is not sufficient to Fourier transform a fermion of momentum $p \in (-\pi/2, \pi/2]^4$, annihilated by ψ, to a momentum $t\pi/a + p$ required to give it taste t. It does so only up to $\mathcal{O}(ap)$ corrections.

One of the benefits of staggered fermions is that they possess an unbroken $U(1)$, non-anomalous axial symmetry that is softly broken by the fermion mass m. In the blocked field notation of (3.26), it is (e.g. [129]):

$$\psi_B(x_B) \to \exp[i\theta(\gamma_5 \otimes \xi_5)]\psi_B(x_B) \qquad \bar{\psi}_B(x_B) \to \bar{\psi}_B(x_B)\exp[i\theta(\gamma_5 \otimes \xi_5)] \ , \tag{3.29}$$

with θ a real parameter. This symmetry is non-anomalous because it is not a taste singlet. In the single-component staggered field theory, the symmetry becomes:

$$\chi(x) \to e^{i\theta\epsilon(x)}\chi(x), \qquad \bar{\chi}(x) \to \bar{\chi}(x)e^{i\theta\epsilon(x)} \ . \tag{3.30}$$

Because of the presence of the sign-oscillating function, $\epsilon(x)$, defined in the paragraph following (3.16), this chiral symmetry is known as staggered fermion's $U(1)_\epsilon$ symmetry. These fermions have other symmetries that give rise to the spacetime symmetries of Euclidean fermions (e.g. translations and $SO(4)$ rotations) as well as the chiral flavour symmetries $SU(4)_L \times SU(4)_R \times U(1)_V \times U(1)_A$ in the massless and continuum limits. [6]

In the continuum, the complete set of fermion bilinears are constructed from the usual combinations of Dirac matrices, yielding operators that transform as $SO(4)$ scalars, pseudoscalars, vectors, etc. and possibly under a flavour group, if there is more than one Dirac fermion field. Ignoring flavour, these continuum bilinears can be written, e.g. $J_s = \bar{\psi}\gamma_s\psi$ with $s \in \mathbb{Z}_2^{\otimes 4}$. With staggered fermions there is an additional taste index, $t \in \mathbb{Z}_2^{\otimes 4}$, and in the free theory bilinears are given by,

$$J_{s,t}(x_B) = \overline{\psi}_B(x_B)(\gamma_s \otimes \xi_t)\psi_B(x_B) \ . \tag{3.31}$$

In the interacting theory one must include suitable link variables of the kind discussed in the next section, so as to obtain gauge invariant operators.

As done in (3.27), instead of writing the bilinears in terms of Dirac matrices labelled by s and t, we can give them in terms of the usual set of 16 Dirac and 16 taste matrices. In that case, for instance, the operator corresponding to $t = 0$ with $s_0 = 0$ and only one of the three spatial $s_i = 1$, is written

$$J_{\gamma_i \otimes I}(x_B) = \overline{\psi}_B(x_B)(\gamma_i \otimes I)\psi_B(x_B) \tag{3.32}$$

[6] Note that if the theory includes N_f degenerate staggered field, the continuum symmetries become those of $4N_f$ flavours instead of 4.

and would be called a "Lorentz" [7] vector and taste singlet. However, as discussed around (3.28), the blocked fields have a taste that is only well defined up to $\mathcal{O}(a)$ corrections. The operator that will create/annihilate a meson with vector Lorentz and taste quantum numbers at time $x_{B,0}$ with zero three-momentum is actually $\sum_{\vec{x}_B} J_{\gamma_i \otimes I}(x_B)$. And even then, this bilinear will create/annihilate two mesons: the one of interest, here the ρ and its time-doubler of opposite parity. As will be seen in two-point functions of staggered mesons below, this time-doubler contribution gives these functions an oscillatory behavior in time.

We argued above that a staggered fermion field gives rise to 4 tastes of quarks. Not surprisingly, we find that the bilinear operators of (3.31) lead to 16 tastes, t, of mesons for each of the 16 possible Lorentz components, s, of $J_{s,t}$ [130].

In the continuum limit in which quark tastes transform in the fundamental representation of $SU(4)_V$ for $m \neq 0$ and $SU(4)_L \times SU(4)_R$ when $m = 0$, these mesons are mass degenerate. At finite lattice spacing and in the presence of interactions, the doubling symmetry of staggered fermions reduces to a shift symmetry which forms a discrete Clifford group, Γ_4, generated by the four taste matrices ξ_μ [131]. In addition, the staggered action is symmetric under improper hypercubic transformations, under the usual $U(1)_V$ baryon number transformations and under the axial $U(1)_\epsilon$ transformations of (3.30). Particle operators and states, however, are classified according to the irreducible representations (irreps) of the rest-frame subgroup formed by the spacetime symmetries of the action which commute with the transfer matrix [131]. This group is the geometric time-slice group [132], composed of the three spatial shifts, spatial rotations and spatial inversions. It has 8 irreps in the case of pseudoscalar mesons. Within each of these multiplets, the mesons have the same mass, but these masses differ by $\mathcal{O}(a^2)$ discretization errors from multiplet to multiplet. Of course in the continuum limit these multiplets coalesce into the adjoint representation of a flavour $SU(4)_V$.

The taste violations that lead to the lifting of the continuum flavour $SU(4)_V$ multiplets are particularly visible in the pion spectrum. This is because the QCD contribution to pion masses is particularly small, so that taste-breaking effects stand out. However, Lee and Sharpe [133] have shown that corrections to staggered would-be Goldstone bosons are symmetric under an $SO(4)$ subgroup of the continuum flavour group, up to discretization errors that are NNLO in staggered chiral perturbation theory, i.e. of $\mathcal{O}(a^4, a^2m, a^2p)$. The quarks reside in its fundamental representation. This representation is the taste equivalent of the one of Dirac spinors in the Euclidean Lorentz group. Thus up to those NNLO discretization errors the 16 tastes of light pseudoscalar mesons are expected to roughly organize into 5 irreducible representations (irreps), which we denote by their Lorentz names: S for scalar or singlet, V for vector, T for tensor, A for axial and P for pseudoscalar. As in the Lorentz case, these 5 representations have multiplicities 1, 4, 6, 4, 1, respectively, and members of a given representation are mass-degenerate (up to the discretization errors discussed above). Thus to NLO in the staggered chiral expansion one has

$$m^2_{\pi,T} \simeq 2Bm + a^2\delta_T \ , \tag{3.33}$$

with taste $T = P, A, T, V, S$ and where δ_T is independent of a. In (3.33), B is the usual, continuum low-energy constant proportional to the chiral condensate [134]. Importantly, $a^2\delta_T$ and all

[7] In the lattice part of this thesis, in which the Lorentz group becomes $SO(4)$ because of the Euclidean metric, we will still frequently call this group the Lorentz group.

higher-order discretization errors do not vanish in the limit of vanishing quark mass because of the remnant, non-anomalous $U(1)$ axial symmetry of staggered fermions. This guarantees that the pseudoscalar-taste, light pseudoscalar meson is a pseudo-Goldstone boson. Moreover the other tastes become the fifteen remaining pseudo-Goldstone bosons of the spontaneously broken chiral $SU(4)_L \times SU(4)_R$ in the continuum limit. Especially for long distance quantities which are dominated by two-pion states these distortions lead to significant cutoff effects and become visible in the continuum extrapolation. We will discuss in Sec. 4.3.2.2 how we deal with taste-breaking effects.

Why use staggered fermions? Since in the staggered formulation, three of the four components are dropped and we have only one component per lattice site, they are very cheap to simulate. In particular, this allows us to perform simulations at several lattice spacings with quark masses close to the physical point, see Sec. 4.3. As we already mentioned, the staggered action preserves a remnant chiral symmetry, which often simplifies renormalization compared to e.g. Wilson fermions. Moreover, in the colour-gauged theory, this symmetry is spontaneously broken, yielding a pseudo-Goldstone boson whose mass vanishes when the quark mass is taken to zero. Finally, discretization errors show up only at $\mathcal{O}(a^2)$.

Why not to use staggered fermions? In order to remove three of the four tastes, the fourth root of the staggered determinant is taken (called "rooting"). This results in an action which has non-localities of $\mathcal{O}(a^2)$. Thus these non-localities disappear in the naive continuum limit, but it has not been shown rigorously that they do in the presence of interactions. The issue is that the theory to which a given lattice action leads in the continuum limit is determined, in a large part, by its universality class. In turn, universality arguments not only depend on the number of dimensions and the symmetries of the order parameter, but they also rely on the fact that the theories considered are *local* (see Sec. 3.6). The question is then whether the rooting procedure changes the universality class of staggered fermions. However, there is abundant theoretical and numerical evidence that this is not the case – see e.g. [109, 129] and references within. In particular no results obtained with rooted staggered fermions have been shown to deviate from those obtained with other fermion formulations in the continuum limit.

There is another way to understand the issue of rooting. As discussed above, the staggered action has an exact $SU(4)_V$ doubling symmetry. This symmetry ensures that a staggered field gives rise to four tastes which become degenerate flavours in the continuum limit. If the symmetry between those tastes were exact also at finite lattice spacing, then the rooting procedure would be an exact way to reduce the four tastes associated with a staggered field to a single flavour in the continuum limit. However, at finite lattice lattice spacing the symmetry between tastes is broken by highly virtual gluons, either by emission or absorption of a gluon with $p \approx \pi/a$, which changes the quark's taste instead of going of-shell [108]. Hence, one would have to show that the restoration of the flavour symmetry in the continuum limit survives rooting. As discussed above, however, all evidence strongly suggests that this is the case.

3.3 Lattice gauge action

So far we have restricted our discussion to the definition of free fermions on the lattice which bore a serious problem, the fermion doubling problem. We also discussed a number of possible solutions. In order to simulate full interacting QCD, we have to step back to our discussion of the beginning of this thesis, see Sec. 2.1.1, and require our theory to be invariant under not only global but space-time dependent $\mathrm{SU}(N_c)$ transformations. We aim at constructing a lattice action which exactly satisfies this gauge invariance. As the derivative in the continuum theory, the symmetric difference quotient in Eq. (2.49) is no longer well defined under a $\mathrm{SU}(N_c)$ gauge transformation. The idea is to parallel transport the field $\psi(x)$ to $x + a\hat{\mu}$. We define a so-called parallel transporter, which transports ψ from x to y as $U(x,y)\psi(x)$. This transporter depends on the path, i.e. it matters how we get from x to y. The derivative in Eq. (2.49) will then, as in the continuum, be replaced by a covariant derivative. $\bar{\psi}(x)U(x, x + a\hat{\mu})\psi(x + a\hat{\mu})$ is gauge invariant provided $U(x,y)$ transforms in a specific way under local $\mathrm{SU}(N_c)$ transformations, namely

$$U(x,y) \to \Omega(x)U(x,y)\Omega^\dagger(y) \,, \tag{3.34}$$

and we see that $U(y,x)$ is an element of the local symmetry group. Hence we can write

$$U(x,y) = \mathcal{P}\exp\left(ig\int_x^y A_\mu^a(z)T^a \,\mathrm{d}z_\mu\right) \tag{3.35}$$

where $\mathcal{P}$ is a path-ordering operator (this is needed since the generators at different points do not commute). This object is called a *Wilson line*. $U(x,y)$ is a non-local object but we will see that in the continuum limit it will be possible to express it in terms of purely local objects[8]. Hence, in order to make Eq. (2.49) gauge invariant, we have to use Wilson lines between adjacent points, so called "link"-variables, i.e.

$$\nabla_\mu^+\psi(x) = \frac{U_\mu(x)\psi(x+\hat{\mu}a) - \psi(x)}{a} \qquad \nabla_\mu^-\psi(x) = \frac{\psi(x) - U_{-\mu}(x - a\hat{\mu})\psi(x - \hat{\mu}a)}{a} \,. \tag{3.39}$$

where we write $U_{\pm\mu}(x) = U(x, x \pm a\hat{\mu})$ and we notice that $U_{-\mu}(x) = U(x, x - a\hat{\mu}) = U_\mu^\dagger(x - a\hat{\mu}, x)$ [9].

We have defined a parallel transporter which provides us with a transformation rule to go from one reference frame to another, allowing us to compare field values at different points and reduces to the continuum expression for infinitesimally close points. The Wilson line constitutes one of the fundamental objects (the basic gauge variables) on the lattice. Again, as in the continuum

[8] To make this explicit, let us expand the parallel transporter around x, since we require a local theory and we hence only need $U(x,y)$ infinitesimally close to x

$$U(x, x + \varepsilon n^\mu) = U(x,x) + \varepsilon n^\mu \partial_\mu U(x,x) + \ldots = 1 - ig\varepsilon n^\mu A_\mu(x) + \ldots \,, \tag{3.36}$$

where we have naturally set $U(x,x) = 1$ and have defined $\partial_\mu U(x,x) \equiv gA_\mu(x)$. Then, the covariant derivative can be written as

$$D_\mu \equiv \partial_\mu - igA_\mu(x) \,, \tag{3.37}$$

expressed only in terms of local data. Performing an infinitesimal group transformation, i.e. $\Omega \approx 1$, writing $\Omega = \exp(ig\alpha_a T^a)$ with $|\alpha_a| \ll 1$ we find

$$A_\mu^a(x) \to A_\mu^a(x) - \partial_\mu\alpha_a(x) + gf^{abc}\alpha_b(x)A_\mu^c(x) \,, \tag{3.38}$$

which is indeed the transformation rule from Eq. (2.10).

[9] We restrict ourselves to the naive fermion action in this discussion, other fermionic actions can be obtained from analogous reasoning – in particular, the staggered fermionic action in Eq. (3.24).

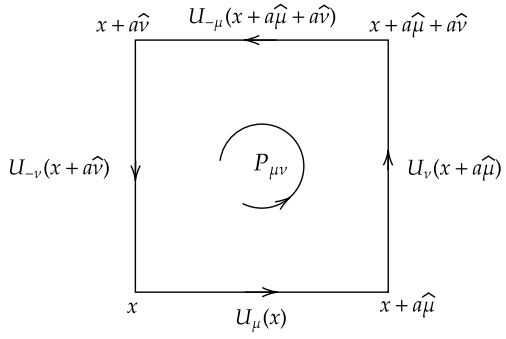


FIGURE 3.1: The plaquette $P_{\mu\nu}$ can be constructed by multiplying links in a closed loop.

theory we have to give the gluons dynamics. A gauge invariant operator can be built from the parallel transporter by constructing an ordered loop of parallel transporters. Since we require a local action we go around the smallest possible loop,

$$P_{\mu\nu}(x) \equiv U_\mu(x) U_\nu(x + a\hat{\mu}) U_\mu^\dagger(x + a\hat{\nu}) U_\nu^\dagger(x) \; . \tag{3.40}$$

This object is called the *plaquette* and is depicted in Fig. 3.1 [10]. However, for non-abelian gauge theories as $\mathrm{SU}(N_c)$ the plaquette in Eq. (3.40) is not yet gauge invariant – in fact, it transforms as $\Omega(x) P_{\mu\nu} \Omega^\dagger(x)$, see Eq. (3.34). By taking the trace, making use of the trace's cyclic property, we find a gauge invariant object and we will be able to construct an action for the gauge fields that reduces to the continuum gauge action in the naive continuum limit. To see this, let us Taylor-expand the operator in Eq. (3.40) for $a \to 0$, where we have to make use of the Baker-Hausdorff formula

$$e^A e^B = e^{A + B + \frac{1}{2}[A,B] + \ldots} \; , \tag{3.42}$$

yielding

$$P_{\mu\nu} \simeq \exp(i a^2 g (\partial_\mu A_\nu - \partial_\nu A_\mu) - a^2 g^2 [A_\mu, A_\nu]) \; . \tag{3.43}$$

Finally, upon expanding the exponential we find

$$P_{\mu\nu} = 1 + i a^2 g G_{\mu\nu}(x) - \frac{a^4}{2} g^2 G_{\mu\nu}^2(x) + \mathcal{O}(a^5) \; . \tag{3.44}$$

Taking the trace of $P_{\mu\nu}$, the second term term vanishes since the generators are traceless. Hence, we define the Wilson action to be [95, 135]

$$S_W = \frac{\beta}{N_c} \sum_{x,\mu,\nu} \mathrm{Re}\,\mathrm{Tr}\{1 - P_{\mu\nu}(x)\} \tag{3.45}$$

which reduces to the continuum Yang-Mills action provided the constant in front is chosen appropriately, namely

$$\beta = \frac{2N_c}{g^2} \; . \tag{3.46}$$

[10] In analogy to the Wilson line the continuum version of the plaquette is called a *Wilson loop*, i.e. a Wilson-line with same initial and end point

$$U_\gamma = P \exp\left(i g \oint_\gamma A_\mu^a(z) T^a \, \mathrm{d}z_\mu \right) \; . \tag{3.41}$$

The Wilson action is of course again by no means unique as we will in particular see in Sec. 3.6.3. The Wilson action (3.45) respects the symmetries of the lattice (hypercubic rotations, lattice translations) and respects parity and charge conjugation.

3.4 Quantization and Euclidean path integral of lattice QCD

3.4.1 Gauge measure

As stated in the introduction, the measure becomes well defined after regularization. A natural measure for compact Lie groups such as $\mathrm{SU}(N_c)$ that is gauge invariant is the Haar measure,

$$\mathrm{d}U = \mathrm{d}\,(UV) = \mathrm{d}\,(VU) = \mathrm{d}U^{-1}\,, \qquad \text{for any } V \in \mathrm{SU}(N_c)\,. \tag{3.47}$$

The measure is normalized to unity,

$$\int \mathrm{d}U = 1\,. \tag{3.48}$$

The advantage is now that we are integrating over the (compact) gauge group instead of integrating over the Lie algebra. Integrating over the Lie algebra requires gauge fixing in order to quotient out physically equivalent configurations. Here we do not require gauge fixing; the overcounting of physical degrees of freedom will simply be normalized by the partition function.

3.4.2 (Fermionic) path integral

Finally, the QCD expectation value of an observable on the lattice can be obtained by evaluating the Euclidean path integral

$$\langle O\rangle = \frac{1}{Z}\int \mathcal{D}U\mathcal{D}\bar{\psi}\mathcal{D}\psi\;\mathcal{O}(\psi,\bar{\psi},U)\;e^{-S_U-S_F(\bar{\psi},\psi,U)}\,, \tag{3.49}$$

with Z the partition function of the system. Usually, the Lagrangian is bilinear in the Dirac fields, $\mathcal{L} = \bar{\psi}(x)D\psi(x)$ with D the Dirac operator. Hence, the integration over the Grassmann valued Dirac fields amounts to solving a Gaussian integral and the integration over $\bar{\psi}$ and ψ can be carried out analytically, yielding

$$\langle O\rangle = \frac{1}{Z}\int \mathcal{D}U\,\det D[U]e^{-S_U}\,\tilde{O}(U)\,, \tag{3.50}$$

where $\tilde{O}$ is an effective operator with the effect of the integrated Grassmann fields taken into account, i.e. where fermion fields are Wick contracted. Calculating the fermion determinant numerically is computationally demanding and was historically first treated in mean-field theory in lattice simulations (so-called *quenched* approximation). For the determination of light hadron masses, the error made by this approximation is $\mathcal{O}(10-15\%)$. We will briefly comment on numerical methods in Sec. 3.8.

3.4.3 Relation to statistical mechanics

We have already mentioned in the introduction that the partition function after Wick rotation can be interpreted as the partition function of a four-dimensional classical statistical system,

$$Z = \int \mathcal{D}\Phi e^{(-1/\hbar)\int \mathrm{d}^d x \mathcal{L}_E(\Phi)} , \tag{3.51}$$

where we have reintroduced the factor $\hbar$. By identifying $\hbar$ as a temperature $1/\beta$ we obtain the formal relation between an Euclidean QFT in d-dimensional spacetime and a classical statistical system in d-dimensional space. We will make use of this formal relation when discussing the conceptual links between renormalization in statistical physics and in quantum field theory. In particular, for our discussion concerning the continuum limit it can also be helpful to identify the coupling $\beta^{-1} = g^2/2N$ (in the infinite time limit), whose symbol we picked in hindsight, as the role of temperature in this analogy. Most important for practical purposes is that due to this formal relation to statistical physics, we can use algorithms and methods which were originally developed for that field.

Since we work in a finite box with temporal extent T and periodic boundary conditions, we can even establish a further formal relation, namely to a quantum statistical system in equilibrium at temperature $1/T$, as we now see. The time evolution of a Heisenberg operator is $\mathcal{O}(t) = e^{iHt}\mathcal{O}(0)e^{-iHt}$, which after Wick rotating to $t \to -ix_0$ becomes $\mathcal{O}(x_0) = e^{x_0 H}\mathcal{O}(0)e^{-x_0 H}$. Then it is possible to show that the path integral representation for the Euclidean correlation functions of two operators $\mathcal{O}_{1,2}$ with Euclidean time separation x_0 with periodic/anti-periodic boundary conditions imposed on bosonic/fermionic fields in the temporal direction [106],

$$\langle \mathcal{O}_1(x_0)\mathcal{O}_2(0)\rangle_T = \frac{1}{Z_T}\int_{\mathrm{PBC}} \mathcal{D}\Phi e^{-\int_0^T \mathrm{d}x_0 \int \mathrm{d}^D x \mathcal{L}(\Phi)} \mathcal{O}_1[\Phi(x_0)]\mathcal{O}_2[\Phi(0)] \tag{3.52}$$

is

$$\langle \mathcal{O}_1(x_0)\mathcal{O}_2(0)\rangle_T = \frac{\mathrm{tr}\left[e^{-(T-x_0)H}\mathcal{O}_1 e^{-x_0 H}\mathcal{O}_2\right]}{\mathrm{tr}\left[e^{-TH}\right]} = \frac{\mathrm{tr}\left[e^{-TH}\mathcal{O}_1(x_0)\mathcal{O}_2(0)\right]}{\mathrm{tr}\left[e^{-TH}\right]} . \tag{3.53}$$

which is the expectation value of a quantum statistical system at temperature $1/\beta = 1/T$. This formal relation between temperature and cyclic time[11] will become more important in the second part of this thesis, where we use finite temperature field theory to calculate expectation values of observables in the early universe, see Sec. 7.4. Eq. (3.53) are the Euclidean correlation functions that we calculate on the lattice. Hence, we obtain a relation between an Euclidean quantum field theory in $(D+1)$ dimensional space-time at finite β and a quantum statistical system in D-dimensional space. In the limit $\beta \to \infty$ and after analytic continuation to Minkowski spacetime we recover our vacuum QFT. In practice, in order to simulate a QFT in vacuum, one chooses $T \gg L$ and takes the infinite time and infinite volume limit.

From the Euclidean correlation function Eq. (3.53) we can extract the energy gap between the ground state and the first excited state: By inserting a complete set of eigenstates $|\alpha\rangle$ of the Hamiltonian in Eq. (3.53) and by taking the limit $T \to \infty$ we obtain

$$\langle \mathcal{O}_1(x_0)\mathcal{O}_2(0)\rangle = \lim_{T\to\infty}\langle \mathcal{O}_1(x_0)\mathcal{O}_2(0)\rangle_T = \sum_\alpha e^{-x_0(E_\alpha - E_0)}\langle 0|\mathcal{O}_1(x_0)|\alpha\rangle\langle\alpha|\mathcal{O}_2(0)|0\rangle , \tag{3.54}$$

[11] I agree with A.Zee: "Some physicists, myself included, feel that there may be something profound here that we have not quite understood." [68]

with $E_\alpha - E_0$ the so-called mass gap, which is the energy of the state $|\alpha\rangle$ relative to the vacuum energy. For large Euclidean time separations x_0 the ground state dominates while excited states become exponentially suppressed. As we will see in the following section, this allows us to extract physical quantities, as for instance hadron masses, from the exponential fall-off of two-point correlation functions.

3.5 Example: Hadron Spectroscopy

How can we make physical predictions with our simulations? As we saw in the previous section, in order to recover physical results in principle one has to analytically continue the results of the simulations back to Minkowski spacetime, which can be highly non trivial. However, there are many observables which can be extracted directly in the spacelike region [12]. The easiest example is the way hadron masses can be extracted from Euclidean correlation functions. Here we will sketch a simple example for the charged pion, whose mass can be obtained by extracting the exponential fall-off of the time evolution of a Euclidean two-point correlation function. Other hadron masses can be obtained in a similar fashion.

In general, we construct meson correlation functions[13]

$$M(x) \equiv \bar{\psi}\Gamma\psi(x) \ , \tag{3.55}$$

where $\Gamma = \left\{1, \gamma_5, \gamma_\mu, \gamma_\mu\gamma_5, \gamma_\mu\gamma_\nu\right\}$ is a combination of γ-matrices (the typical scalar, pseudoscalar, vector, axial-vector and tensor) and we have suppressed Dirac and colour indices. To improve the signal of the correlator, one projects to zero spatial momentum

$$C_M(t) = \sum_{\vec{x}} \langle M(t,\vec{x})M(0,\vec{0})\rangle \ , \tag{3.56}$$

where one of the currents can be placed at $x = 0$ due to translational invariance. The integration over fermionic degrees of freedom can again be performed analytically, yielding

$$C_M(t) = \frac{1}{Z}\int \mathcal{D}U e^{-S_U[U]} \det(D) \int_{\vec{x}} \left\{ -\operatorname{Tr}\left[D^{-1}_{0,x}\Gamma D^{-1}_{x,0}\Gamma\right] + \operatorname{Tr}\left[D^{-1}_{0,0}\Gamma\right]\operatorname{Tr}\left[D^{-1}_{x,x}\Gamma\right]\right\} \tag{3.57}$$

where D is the massive Dirac operator. The first term corresponds to the quark connected contribution, the second term is the quark disconnected contribution which is numerically much more challenging. This contribution arises for equal quark flavours in the operator (3.55).

To be more specific, let us construct the correlation function for the extraction of the mass of a charged pion. We have to construct an operator which has the same quantum numbers as the π^- so that the operator will have non-vanishing overlap,

$$M_\pi(x) = (\bar{u}\gamma_5 d)(x) \, . \tag{3.58}$$

[12] Similarly, in Chap. 4 we obtain the hadronic contribution to the running of the electromagnetic coupling in the spacelike region.

[13] In this section and the following sections we denote Euclidean time again by t.

Due to periodic boundary conditions, the correlator is symmetric under the exchange $t \leftrightarrow T - t$. Inserting a complete set of eigenstates the spectral decomposition of the operator is ($t \geq 0$)

$$C_\pi(t, \vec{p}) = \sum_\alpha \frac{|\langle 0|M_\pi(0)|\alpha\rangle|^2}{2E_\alpha(\vec{p})} \left(e^{-E_\alpha(\vec{p})t} + e^{-E_\alpha(\vec{p})(T-t)} + \mathcal{O}\left(e^{-(T+t)}\right)\right) , \tag{3.59}$$

where the sum runs over all states which have non-zero overlap with the operator in Eq. (3.58) and we have used the usual relativistic normalization $\langle\alpha|\beta\rangle = 2E_\alpha\delta_{\alpha\beta}$. $E_\alpha(\vec{p})$ is the mass gap relative to the vacuum energy. By projecting to zero spatial momentum the groundstate is the mass of the π^-, $m_{\pi^-} = E_0(\vec{0})$ and assuming non-degenerate states $E_0 < E_1 < \ldots$ we find

$$\lim_{t\to\infty} C_\pi(t, \vec{0}) \simeq \frac{|\langle 0|M_\pi(0)|\pi\rangle|^2}{m_{\pi^-}} e^{-m_{\pi^-}T/2} \cosh(m_\pi(T/2 - t)) , \tag{3.60}$$

since in the limit $t \to \infty$ the contributions of excited states get exponentially suppressed. In terms of the Wick contraction we find for this specific example

$$C_\pi(t) = \frac{1}{Z} \int \mathcal{D}U e^{-S_U[U]} \det(D^u) \det(D^d) \int_{\vec{x}} \left\{ -\mathrm{Tr}\left[(D^u_{0,x})^{-1} \gamma_5 (D^d_{x,0})^{-1} \gamma_5 \right] \right\} . \tag{3.61}$$

3.6 Continuum limit

The lattice regularization is only a regularization. In particular, it breaks certain symmetries. For instance, it breaks the O(4) symmetry of a relativistic field theory in Euclidean spacetime down to the subgroup of hypercubic rotations. In addition, due to cutoff effects arising at finite lattice spacing, we do not recover exactly the continuum results, even for long distance quantities. This regularization should therefore be removed by taking the limit $a \to 0$. The existence of this limit and whether/how this limit can be taken is a very subtle issue. In particular, the naive (classical) continuum physics is recovered by sending $a \to 0$ without changing the value of the coupling. In the quantum theory this is of course no longer possible and we have to adjust the parameters of our theory as we change the cutoff (see also Sec. 2.2).

Wilson established the relation between renormalizability in QFTs and universality in critical statistical systems. To understand this connection we first revisit renormalization group (RG) transformations in configuration space, which is the appropriate way of performing RG transformations in a lattice-regularized theory. Performing RG transformations corresponds to a flow in theory space. This flow can exhibit fixed points and we will see how the universal behaviour near fixed points helps us understand why our "simple" quantum field theories (and our algorithm for doing physics in general, see 2.2) work so well.

Many very good references have helped me in (trying to) understand(ing) the basic concepts, see for instance [105, 136–141] and this section is based on these references. Below we make explicit which quantities are dimensionful (df) and which are lattice quantities (lat).

3.6.1 Renormalization group transformation, fixed point and behaviour near the fixed point

On the lattice, an RG transformation averages out fluctuations of $\mathcal{O}(a)$ in such a way that their effect on long-distance observables is exactly taken into account and the physical predictions at distances longer than the new cutoff remain unchanged. An RG transformation therefore relates the predictions at cutoff $\Lambda = 1/a$ to the prediction at cutoff $\Lambda' = 1/a'$ with $a' > a$. To preserve the physical prediction the coupling constants have to change with the cutoff in a specific way. The way this is implemented in configuration space is by performing a blocking procedure. The fields φ_n which live on the original lattice with spacing a are blocked together and one defines a new blocked field χ_{n_B} on a lattice with spacing $a' = 2a$ as an average over the original fields in the block. The averaging is performed in such a way that the partition function expressed in terms of the blocked fields does not change,

$$Z = \prod_n \int d\varphi_n \mathrm{e}^{-S(\varphi)} = \prod_{n_B} \int d\chi_{n_B} \mathrm{e}^{-S'(\chi)} = Z' \,.$$

$\exp(-S'(\chi)) = \prod_n \int d\varphi_n \delta\left(\chi_{n_B} - b\sum_{n\epsilon n_B} \varphi_n\right) \exp(-S(\varphi))$ is called the effective action. In particular, S and S' are equally well suited for describing the long distance physics at scales $L \gg a$. However, RG transformations should respect the symmetries of the original action.

In general, an RG transformation will generate all kinds of interactions allowed by the symmetries of the original action. We therefore write our action in most general form with dimensionless couplings K_α – this is possible by using appropriate powers of $\Lambda = 1/a$ – such that the functional form remains the same. RG transformations then correspond to a flow in theory space, $\{K_\alpha\} \to \{K'_\alpha\} \to \{K''_\alpha\} \to \ldots$. At each RG step, the lattice spacing gets doubled, $a \to 2a = a' \to 2^2 a = a'' \to \ldots$ and the dimensionless correlation lengths are therefore halved, $\xi_{\mathrm{lat}} \to \xi_{\mathrm{lat}}/2 \to \xi_{\mathrm{lat}}/2^2 \to \ldots$, while the physical correlation lengths ξ_{df} do not change. This flow can exhibit fixed points (FP). At a fixed point, $\{K^*_\alpha\} \to \{K^*_\alpha\}$ the correlation length hence has to diverge, $\xi_{\mathrm{lat}} \to \infty$, or else vanish. The set of points with $\xi_{\mathrm{lat}} \to \infty$ forms the so-called critical surface.

Let us study the behaviour near fixed points. By linearizing and expanding in $\Delta K'_\alpha = K'_\alpha - K^*_\alpha$ one finds that one can classify the eigenvalues of the eigenoperators in relevant ($|\lambda_\alpha| > 1$), marginal ($|\lambda_\alpha| = 1$) and irrelevant ($|\lambda_\alpha| < 1$) eigenvalues, hence determining whether the corresponding operator will be enhanced (relevant) or suppressed (irrelevant) by repeated RG transformations:

$$S \to S^* + \sum_i \lambda^i c_i h^i \to S^* + \sum_i (\lambda^i)^2 c_i h^i \to \ldots \,, \tag{3.62}$$

with h^i the eigenvectors which define the eigenoperators and c_i the corresponding couplings, which are small close to the FP. To determine the fate of marginal operators one would have to go to higher orders in the expansion around the fixed point, the linear approximation is not sufficient. The ensemble of only irrelevant operators forms the critical surface since all RG trajectories starting on this hypersurface will flow into the fixed point. If we do not start exactly on the critical surface but close to it, the correlation length will get halved at each step and we move away.

In the case of pure Yang-Mills interactions in $d = 4$ dimensions, the following picture emerges: The theory has one marginally relevant operator

$$\frac{g^2}{6} F^a_{\mu\nu} F^{\mu\nu a} \, .$$

All other operators are irrelevant. The coupling g increases as we perform RG transformations from the UV to the IR. This is the property of asymptotic freedom that we discussed in Sec. 2.3.2. We will restrict our discussion here for simplicity to pure Yang-Mills theory.

3.6.2 Universality, asymptotic scaling, taking the continuum limit

As we discussed above, in order to take the continuum limit we need to determine the RG flow, i.e. we need to know how to change the value of the coupling as a function of the cutoff to preserve the predictions at low energies (to make the connection with Sec. 2.2, the lattice parameters are the bare couplings at $\Lambda \sim 1/a$).

All quantities on the computer are dimensionless. We therefore have to determine the lattice spacing for a particular value of g_0 (our input) through dimensionless properties of the system. As we saw in Sec. 3.5, we can for instance determine the masses of hadrons from the exponential fall-off of Euclidean correlation functions. Schematically,

$$\langle \Phi(x_{\rm df}) \Phi(0) \rangle \sim \exp\left(-\frac{x_{\rm df}/a}{\xi_{\rm df}/a} \right) \, ,$$

where both $x_{\rm lat} = x_{\rm df}/a = n_x$ and $\xi_{\rm lat} = \xi_{\rm df}/a$ are dimensionless. The correlation length $\xi_{\rm df}$ is the length scale over which a massive particle can propagate with a significant amplitude,

$$M_{\rm df} = \frac{1}{\xi_{\rm df}} = \frac{1}{a \xi_{\rm lat}} = \frac{f(g_0)}{a} \, , \tag{3.63}$$

where $f(g_0) \sim \mathcal{O}(1)$ for some generic g_0 and $M_{\rm df}$ would therefore be of the order of the UV cutoff. However, in order to keep the physical mass $M_{\rm df}$ finite as we decrease a, we have to tune the bare coupling g_0 to some g_0^*. Specifically, $f(g_0) \xrightarrow[g \to g_0^*]{a \to 0} 0$ and consequently,

$$\xi_{\rm lat} = \frac{1}{a M_{\rm df}} = \frac{1}{f(g_0)} \xrightarrow[g \to g_0^*]{a \to 0} \infty \, . \tag{3.64}$$

For $a \to 0$ the dimensionless correlation length $\xi_{\rm lat}$ diverges. Therefore, taking the continuum limit corresponds to a second order phase transition in the corresponding statistical system, and the QFT in the continuum limit corresponds to a critical statistical system. It is empirically known that many physical systems near criticality behave in a similar way and fall into a small number of equivalence classes, whose members exhibit the same critical behaviour, independent of their microscopic properties, a characteristic called *universality*. The typical example is of course the liquid-gas and ferromagnetic-paramagnetic phase transitions which are microscopically very different but were found to exhibit the same critical exponents. In fact, these universality classes are distinguished only by the number of space dimensions, the number of components of the order parameter and the symmetry of the system [106].

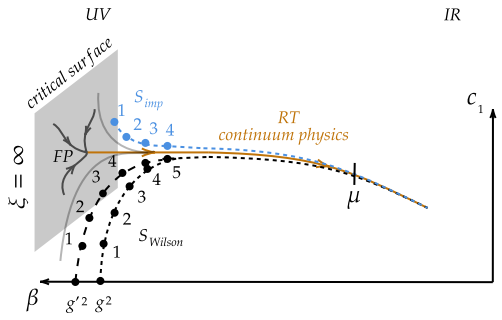


FIGURE 3.2: Renormalization group flow and the renormalized trajectory in pure Yang-Mills theory. The gray surface forms the critical surface. The orange line is the renormalized trajectory.

There is a specific trajectory, called the *renormalized trajectory* (RT), which describes the flow coming out of the fixed point (more precisely, infinitely close to the fixed point) away from the critical surface. This is the orange line in Fig. 3.2. Flows starting close to the critical surface but away from the RT will first evolve close to the critical surface towards the FP. Once they approach the FP they will deviate away from the critical surface and gradually approach the renormalized trajectory. This general picture is depicted in Fig. 3.2. Luckily, fixed points are in general universal and contain only a small finite number of relevant operators. It is the essence of renormalizability in perturbation theory, namely that one only has to adjust a few relevant parameters (in fact, as we argued above, for pure Yang-Mills theory only one parameter). In continuum perturbation theory the FP is at $g^2 = 0$ and we keep only relevant operators. In the Wilsonian RG the renormalized running coupling is the flow along the RT, determined by the continuum β function. In particular, due to universality we expect the physical predictions to be independent of the specific lattice action chosen. This property is therefore crucial for having the freedom to choose different discretizations of the QCD action.

To take the continuum limit we need to determine the RG flow, i.e. the relation $a \leftrightarrow g_0$. The RG equation follows from the simple requirement that physical observables do not depend on the cutoff,

$$a\frac{\mathrm{d}}{\mathrm{d}a}O(g_0, a) = 0 \tag{3.65}$$

for instance $O(g_0, a) = M_{\mathrm{df}}(g_0, a) = f(g_0)/a$, see Sec. 3.5, from which we obtain the Callan-Symanzik equation

$$-f(g_0) + \left(\frac{\partial}{\partial g_0} f(g_0)\right)\beta_{\mathrm{lat}} = 0\,. \tag{3.66}$$

We have recovered an alternative definition of the β-function introduced in Sec. 2.2,

$$\beta_{\mathrm{lat}} \equiv a\frac{\partial g_0}{\partial a}\,. \tag{3.67}$$

One way to determine the β-function would again be perturbatively, i.e. by an expansion in small bare coupling g_0

$$\beta_{\mathrm{lat}}(g_0) = 4\pi b_0\left(\frac{g_0}{4\pi}\right)^3 + 4\pi b_1\left(\frac{g_0}{4\pi}\right)^5 + \dots\,, \tag{3.68}$$

where, as discussed in Sec. 2.3.2, one can show that the first two coefficients in the expansion are universal (see Eq. (2.58) with $N_f = 0$ for pure Yang-Mills theory). Higher-order terms are scheme

dependent. With the knowledge of b_0 and b_1 Eq. (3.68) can be integrated and one obtains

$$M_{\mathrm{df}} = \frac{c}{a} \exp\left(-\frac{(4\pi)^2}{2b_0 g_0^2}\right)\left(b_0\left(\frac{g_0}{4\pi}\right)^2\right)^{-\frac{b_1}{2b_0^2}} \left[1 + \mathcal{O}(g_0^2)\right] \equiv c\Lambda_{\mathrm{lat}} \tag{3.69}$$

with Λ_{lat} the integration constant at $g_0 = \infty$ which becomes the characteristic scale. It is a cutoff independent external parameter and carries a mass dimension of 1. Finally, we see from Eq. (3.69) that in order to take the continuum limit $a \to 0$, one has to tune the bare coupling $g_0^2(a) \xrightarrow{a\to 0} (4\pi^2)/(2b_0 \ln(1/a\Lambda_{\mathrm{lat}})) \xrightarrow{a\to 0} 0$ (or, to make the analogy with Sec. 3.4.3 $\beta = 6/g_0^2 \xrightarrow{a\to 0} \beta_c = \infty$).

In the case of the standard Wilson action, see Eq. (3.45), we have one bare parameter, $g_0^2 \in (0, \infty)$ and all other parameters vanish $c_1 = c_2 = \ldots = 0$. This is depicted in Fig. 3.2: The two black dashed lines correspond to the RG flows of two Wilson actions S and S', with two different values for the bare parameters $g' < g$. If we are close enough to the critical surface, we can tune g' in such a way that after five (S) and six (S') RG transformations, respectively, both trajectories join. However, for the primed action, we require one additional RG step. Hence the correlation length is doubled, which means that the lattice spacing is halved $a(g') = a(g)/2$. Both actions describe the same long distance physics $a \ll \xi_{\mathrm{phys}}$, which is the requirement of renormalizability; only their scale is different.

If we could rely on perturbation theory we could tune our bare parameters as a function of the cutoff according to Eq. (3.69). Then, all simulations would be related by RG transformations, i.e. they would all lie on a line of constant physics. The regime where g_0 is small enough such that the terms beyond the two universal coefficients are negligible is called *asymptotic scaling* regime (the regime where perturbation theory is applicable but higher order corrections are needed is called *scaling* regime). However, in practice the bare couplings are too large (we are not close enough to the critical surface) and are therefore not in this asymptotic scaling regime. Hence, simulations with bare parameters tuned according to Eq. (3.69) do not lie on the same RG trajectory, the right-hand side of the Callan-Symanzik equation (3.66) with β_{lat} does not vanish but is equal to terms $\mathcal{O}\left(a^P\right)$ [137, 142], where the exponent P depends on the specific lattice action. These terms are called scaling violations.

As we discussed, we cannot rely on perturbation theory to determine the relation between the bare coupling g_0 and the cutoff, since in practice we are too far away from criticality and the asymptotic scaling regime. Hence, we require a non-perturbative way of determining the lattice spacing for a given value of the bare coupling. One can for instance determine $(aM_{\mathrm{had}})_{g_0}$ as discussed in the previous section from the exponential fall-off of a Euclidean time correlator and then obtain the lattice spacing from

$$a = \frac{(aM_{\mathrm{had}})_{g_0}}{M_{\mathrm{had,exp}}} \,.$$

However, one does not have to restrict oneself to experimentally measurable quantities. A different way is to measure quantities related to the Wilson-flow of gauge fields [143], such as w_0 [144], or to the force between a static quark and a static antiquark separated by a distance r, such as the Sommer scale [145].

Finally, to take the continuum limit in practice, one performs several simulations at different bare couplings and determines the relation $a \leftrightarrow g_0$ non-perturbatively as described in the previous paragraph. Once we have fixed this relation we can make predictions for other physical quantities, at some long-distance scale μ (the position where we would arrive after a few RG steps, see Fig. 3.2). At long distances $\mu \ll 1/a$, we are already close to the continuum trajectory and we should not observe significant cutoff effects. We perform several such simulations and extrapolate the predictions for these long-distance quantities to $a \to 0$. To perform these extrapolations one relies on assumptions on the functional form of the discretization effects. This becomes particularly important for high precision observables, as the continuum extrapolation necessarily introduces a systematic uncertainty in the final value. It is therefore extremely important to have an analytic understanding of the asymptotic dependence on the lattice spacing. This understanding might also offer ways of finding different formulations where cutoff effects are greatly reduced, as we discuss in the next section.

A few final remarks are now in order. As we argued, we heavily make use of the universality/renormalizability property of physical theories when building our lattice actions and it is one of the key properties for lattice field theories to work. It allows us to choose actions with a lattice cutoff which are very different from the continuum action – in particular, we will see in the following section that they contain irrelevant operators when expanded as an effective continuum action – but to still obtain the same IR physics. Just as in the case of the ferromagnetic-paramagnetic and the liquid-gas phase transitions the microscopic properties of our lattice actions might be very different but their long-distance behaviour looks the same. They lie in the basin of attraction of the fixed point of our theory and the continuum limit is the fixed point of the corresponding statistical system.

More generally, we make use of universality in all physical theories – in fact, this is the reason we can do physics, and why the algorithm we presented at the beginning of Chap. 2 works so well. One might of course argue that there is no reason to take the continuum limit (in fact, the UV physics might very well not be described by a field theory at all). In PT we only keep relevant couplings or marginally ones, such as e in QED or λ in $\lambda\Phi^4$ theory, which is technically convenient, because it allows us to take the infinite cutoff limit. For IR physics, choosing the RT is a very good approximation because it is impossible to tell from which initial condition the flow emerged. It would require high precision experiments to detect a deviation from the RT and is in some sense the basic idea behind the search for new physics at the intensity frontier (e.g. via EW precision tests).

3.6.3 Symanzik effective theory

One can make use of the concepts introduced in the previous section to construct lattice actions which provide a faster convergence to the continuum limit as we send the lattice spacing to zero. This can be achieved within the framework of Symanzik effective theories. Formally, one can write the lattice action as an effective (continuum) theory of the continuum action (in a purely mathematical sense), where the inverse lattice spacing plays the role of a new physics scale, with Wilson coefficients in front of the higher dimensional operators:

$$S_{\text{eff}} = \int \mathrm{d}^4x \, \{\mathcal{L}_{\text{cont}}(x) + a\mathcal{L}_1(x) + a^2\mathcal{L}_2(x) + \ldots\} \, . \tag{3.70}$$

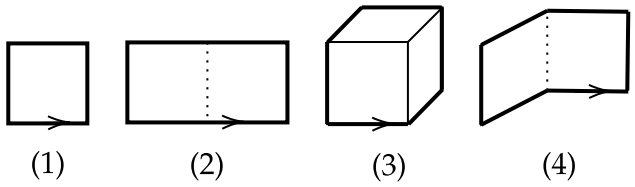


FIGURE 3.3: Closed paths that can be chosen as a lattice discretization for the dimension-4 continuum Yang-Mills operator and the dimension-6 operators in Eq. (3.72).

The $\mathcal{L}_i$ are higher dimensional local operators of dimension $4+i$,

$$\mathcal{L}_i = \sum_k c_k(g(\mu), a\mu)\mathcal{O}_k(\mu) \ , \tag{3.71}$$

where μ is the renormalization scale and $\mathcal{O}_k(\mu)$ are (renormalized) local operators. One can deduce the general form of the continuum operators $\mathcal{O}_k$ by constructing all possible operators of dimension $4+i$ allowed by the lattice symmetries. For on-shell quantities one can further establish relations between the possible operators and eliminate redundancies by using the field equations.

Symanzik improvement and its relation to renormalized trajectories The idea behind Symanzik's improvement programme is to add irrelevant operators (in the RG sense) to the lattice action which cancel the effective terms in Eq. (3.70). This removes the leading term a^P on the right-hand-side of the Callan-Symanzik equation (3.66), therefore reducing discretization effects and accelerating the approach to the continuum limit. These operators vanish as we take the continuum limit $a \to 0$. Conceptually, this improvement procedure is similar to using higher order derivative stencils in finite-difference numerical solutions to differential equations. Determining the Wilson coefficients in front of these operators is however a non-trivial task due to the quantum nature requiring renormalization and the non-perturbative nature of QCD at low energies. The standard textbook example is the $\mathcal{O}(a)$-improvement of Wilson's fermion action. Since we do not employ Wilson fermions in our simulations, let us briefly sketch the main steps for on-shell improvement of the pure Yang-Mills action. For the Yang-Mills action, effective operators start at dimension six. There are three independent operators respecting the symmetries [146, 147]

$$O_1 = \sum_{\mu,\nu} \mathrm{tr}\, D_\mu F_{\mu\nu} D_\mu F_{\mu\nu}, \quad O_2 = \sum_{\mu,\nu,\rho} \mathrm{tr}\, D_\mu F_{\nu\rho} D_\mu F_{\nu\rho}, \quad O_3 = \sum_{\mu,\nu,\rho} \mathrm{tr}\, D_\mu F_{\mu\rho} D_\nu F_{\nu\rho} \ . \tag{3.72}$$

To represent the four continuum operators (one of dimension four and three of dimension six) we need four lattice operators of the general form

$$\mathrm{Re}\ \mathrm{Tr}\left[1 - U(\mathcal{C})\right] \ ,$$

where $U(\mathcal{C})$ is an ordered product of link variables along a closed curve $\mathcal{C}$ (such as the plaquette). The choice for the lattice discretization of the terms in Eq. (3.72) is again of course not unique but differs by higher order $\mathcal{O}(a^4)$ effects. A common choice is to use the smallest loops possible. This leaves us with the lattice operators depicted in Fig. 3.3.

$$S_{\text{gauge}} = -\beta \left\{ c_0 \sum P_{\text{plaq}} + c_1 \sum P_{\text{rect}} + c_2 \sum P_{\text{chair}} + c_3 \sum P_{\text{para}} \right\} . \tag{3.73}$$

To recover the correct continuum normalization for the dimension four operator we require for the corresponding Wilson coefficients of the operators in Fig. 3.3

$$c_0(g^2) + 8c_1(g^2) + 8c_2(g^2) + 16c_3(g^2) = 1 . \tag{3.74}$$

To obtain improvement at tree-level one finds [105]

$$c_0(g) = \frac{5}{3} , \qquad c_1(g) = -\frac{1}{12} , \qquad c_2(g) = c_3(g) = 0 . \tag{3.75}$$

One can go to higher orders in perturbation theory or attempt to determine the coefficients non-perturbatively [140, 146–150].

On a more conceptual basis, this is precisely what we have been alluding to in the previous section, however here we see it explicitly: the lattice action is equal to the continuum theory up to irrelevant operators in the RG picture. From what we learned in the previous section, the effect of these operators becomes irrelevant as we perform RG transformations to the infrared. Low energy physics cannot distinguish between a theory emerging from the fixed point (the renormalized trajectory) and a theory with irrelevant operators, since all RG trajectories focus onto this renormalized trajectory at low energies. This is the principle of universality. We took advantage of this and constructed a lattice action where we removed the dimension-6 operators (however here only at tree-level). This will approach us further to the (continuum) renormalized trajectory and the lattice action will have reduced discretization errors – in fact, they begin at $\mathcal{O}(a^4)$. In our simulations we use an $\mathcal{O}(a^2)$-improved Lüscher-Weisz action [151], see Sec. 4.3.1.

This conceptual picture is also depicted in Fig. 3.2. We already discussed the RG trajectories of the Wilson action ($c_1 = c_2 = \ldots = 0$). By turning on the coupling c_1 it is now possible to tune c_1 in such a way that for the same distance from the critical surface, the renormalized trajectory is reached after fewer block transformations than for the Wilson action, indicated by the blue dotted line. In fact, if we could tune an infinite number of parameters we would be able to sit exactly on the renormalized trajectory. Actions on this trajectory are therefore called "perfect actions" [138], see also Sec. 3.2.1. These are actions with the same predictions as the continuum action, even at finite lattice spacing and independently whether the lattice is fine or coarse. There have been approaches to approximate this fixed point action with a finite number of parameters [152].

3.6.4 Recovering full QCD

So far we have restricted our discussion to pure gauge theory, where one relevant operator g_0 has to be tuned. Including massive fermions, also the bare masses of the quarks have to be renormalized. Full QCD is then parametrized by the set of bare couplings

$$\{g_0, m_u, m_d, m_s, m_c, m_b, m_t\} .$$

Again, instead of using perturbation theory, one usually fixes these bare parameters in so-called hadronic schemes. One measures dimensionless combinations of hadronic masses and/or decay

constants and uses those to fix the bare quantities. Once this finite number of parameters is determined, one can make physical predictions. In our simulations we use a mass-dependent scheme, which is a scheme where two simulations with the same bare g_0 can correspond to a different lattice spacing, if the bare parameters for the quark masses differ. In mass-independent schemes the value of the lattice spacing is determined at the physical value of the quark masses (hence independent of the bare quark masses), resulting in a single lattice spacing per g_0. Both schemes yield the same results in the continuum limit. We will come back in Sec. 4.3.2.4 on how exactly we fix the "physical point", i.e. results for observables that we can compare to experiment.

3.7 Electromagnetism on the lattice

As we saw in section 2.3, in the limit of equal up and down quark masses, QCD possesses an exact $SU(2)$-isospin symmetry,

$$\begin{pmatrix} u \\ d \end{pmatrix} \longrightarrow \exp\left[i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}\right]\begin{pmatrix} u \\ d \end{pmatrix} . \tag{3.76}$$

The small breaking of this isospin symmetry has important consequences for the universe as we know it today. For example, as we will discuss in Sec. 5.2 the small difference in the masses of the neutron and proton plays an important role in big bang nucleosynthesis during which the lightest elements were produced. It also guarantees the mere existence and stability of atoms [153, 154].

This symmetry is not only broken by the difference in the quark masses but also by the interactions of quarks with photons, due to their different electric charges. The difference in mass is very small compared to the typical hadronic scale Λ_{QCD} and so is the electromagnetic interaction strength compared to the strong interactions at low energies,

$$\frac{\hat{m}_d - \hat{m}_u}{\Lambda_{\mathrm{QCD}}} \ll 1, \qquad (e_u - e_d)\, e_f \hat{\alpha} \ll 1 .$$

Hence, isospin symmetry is only broken very mildly and is therefore an excellent approximation at the level of percent precision. However, if one aims at subpercent precision these effects have to be included. Here we want to discuss possibilities of simulating QED on the lattice, which holds some intrinsic and fundamental difficulties.

3.7.1 QED in finite volume

In our lattice simulations we impose (mostly) periodic boundary conditions. Hence, we compactify our Euclidean manifold and look at the torus:

$$\mathbb{T}^4 = \prod_{\mu=0}^{4} \mathbb{R}/\mathbb{Z}L_\mu . \tag{3.77}$$

However, in order to satisfy Gauss' law, the charge in the interior of our finite volume necessarily has to vanish,

$$Q = \int_{\mathbb{T}_3} \mathrm{d}^3x \; j_0(t,\mathbf{x}) = \int_{\mathbb{T}_3} \mathrm{d}^3x \; \partial_k E_k(t,\mathbf{x}) = \int d^3x \; \rho(t,\vec{x}) = \int_{\partial\mathbb{T}_3} d\vec{S}\cdot\vec{E}(t,\vec{x}) = 0 \; , \tag{3.78}$$

since we require $\vec{E}(x_k = 0) = \vec{E}(x_k = L_k)$ in order to satisfy our periodic boundary conditions. This constitutes a fundamental obstacle for simulating QED on a lattice.

Let us sketch how this obstacle arises formally. The field strength tensor (containing electric and magnetic fields) has to satisfy periodic boundary conditions. The electric and magnetic fields are determined by the electromagnetic potential A_μ, albeit only up to a gauge transformation,

$$A_\mu \to A_\mu - \partial_\mu \lambda \; . \tag{3.79}$$

As we will see, the gauge freedom has two consequences in a finite volume with periodic boundary conditions: 1) a remaining aperiodic part in the gauge potential A_μ (without this freedom A_μ would be completely periodic); 2) an additional (periodic) shift symmetry, which does not get fixed by local gauge fixing and which can be identified with the zero mode of A_μ.

1) From the requirement of periodicity it is possible to construct certain relations between gauge potentials

$$\Delta_\nu A_\mu(x) \equiv A_\mu(x + L_\nu \hat{\nu}) - A_\mu(x) = \partial_\mu \omega_\nu \; , \tag{3.80}$$

where the so-called "twist"-functions ω_ν have to satisfy [155]

$$\Delta_\nu \omega_\rho - \Delta_\rho \omega_\nu \equiv \Phi_{\nu\rho} = \text{const.} \tag{3.81}$$

$\Phi_{\nu\rho}$ is antisymmetric and constant and it can be shown that it is a gauge invariant quantity and plays the role of a constant electromagnetic flux going through the boundaries of the hypercube. The aperiodic part $A_\mu^{(0)}$ of A_μ is then determined by this antisymmetric tensor (again up to a gauge transformation) [155]. It is a topological term since $F_{\mu\nu}$ depends on the choice of A_μ. By requiring A_μ to be periodic (i.e. eliminating the aperiodic part), the flux is set to zero.

2) Since we have compactified our spacetime, A_μ is invariant under so-called large gauge transformations [156, 157], with c_ν a real constant vector

$$\lambda = \frac{c_\nu}{\prod_\rho L_\rho} x_\nu \; ,$$

which are not continuously connected to the identity and do not get fixed by local gauge fixing procedures [157], cf. Sec. 2.3.3.2.[14].

[14] In infinite volume physical quantities are invariant under a gauge transformation $A_\mu \to A_\mu + \partial_\mu \lambda$. After fixing the gauge, for instance in the Lorenz gauge

$$\partial_\mu A^\mu = 0 \; ,$$

one is left with an additional freedom, namely all functions verifying $\Box\lambda = 0$ (or $\Delta\lambda = 0$ in the Euclidean). By choosing appropriate boundary conditions for the gauge potential – those which leave the action finite – λ is then uniquely determined. See for instance [66].

Decomposing the gauge field A_μ using Hodge's theorem [155] and integrating over the hypercube $C = \prod_\mu [0, L_\mu]$ one finds

$$\int_C A_\mu \, \mathrm{d}^4x = \underbrace{\int_C (\partial_\mu f + \partial_\nu h_{\mu\nu}) \, \mathrm{d}^4x}_{=0 \text{ using Stokes' theorem}} + \int_C \left(\frac{c_\mu}{\prod_\nu L_\nu} + A_\mu^{(0)} \right) \mathrm{d}^4x \, , \tag{3.82}$$

with f a scalar field on $\mathbb{T}^4$ and $h_{\mu\nu}$ an antisymmetric tensor field on $\mathbb{T}^4$. Since we have eliminated the aperiodic part $A_\mu^{(0)}$, the periodic shift symmetry c_μ can be identified with the zero mode of the photon field:

$$\int_C \sum_{k \in \mathbb{T}^4} \frac{1}{\prod_\nu L_\nu} \tilde{A}_\mu(k) e^{ik\cdot x} \, \mathrm{d}^4x = \tilde{A}_\mu(k=0) \, . \tag{3.83}$$

The zero mode, being simply a shift of the electromagnetic potential, does not contribute to $F_{\mu\nu}$.

This zero-mode, however, causes severe problems in finite volume, where momenta are quantized. Consider for instance the perturbative photon propagator in finite volume,

$$D_{\mu\nu}(x) = \frac{1}{\prod_\nu L_\nu} \sum_{k \in \mathbb{T}^4} \frac{\delta_{\mu\nu}}{k^2} e^{ik\cdot x} \, . \tag{3.84}$$

It clearly diverges because of the $k = 0$ term. Hence, QED in a finite volume with periodic boundary conditions introduces plain infrared divergences. Clearly, in order to obtain finite results, these divergences have to be removed. Since this shift symmetry cannot be removed by local gauge fixing (not relating different sites) we have to impose yet another, global constraint. In fact, as we just showed, the zero mode problem and the obstruction of putting a charge in a periodic box from Gauss' law are closely intertwined. To circumvent this obstruction we have to change the equations of motion (which of course converge to Maxwell's equations in the infinite-volume limit) which will be equivalent to fixing the zero mode or changing the boundary conditions altogether. This will also regularize the IR divergences. Several prescriptions have been proposed and tested in the literature:

- **QED**$_{TL}$ One way to change the equations of motion is to add a uniform, time independent background current to the action (a Lagrange multiplier)

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu} F_{\mu\nu} + \left(j_\mu - \frac{c_\mu L_\mu}{\prod_\rho L_\rho} \right) A_\mu \, , \qquad \int_{\mathbb{T}^4} A_\mu \, \mathrm{d}^4x = 0 \tag{3.85}$$

  where c_μ is a constant Lagrange multiplier and A_μ hence has to satisfy the constraint in the second equation. This action leads to the modified equations of motion

$$\partial_\mu F_{\mu\nu} = j_\nu - \frac{L_\nu}{\prod_\rho L_\rho} c_\nu \, . \tag{3.86}$$

  The constraint on A_μ is equivalent to removing the global zero-mode of the gauge field, $\tilde{A}_\mu(k=0) = 0$ on $T \times L^3$ four-torus [158]. c_μ is chosen such that

$$c_0 = \int \mathrm{d}^3x j_0(x) \quad \text{which implies} \quad \int \mathrm{d}^3x \; \partial_i F_{0i}(x) = 0 \, .$$

  This prescription is non-local (it couples gauge fields at arbitrary times/distances) and violates reflection positivity [159]. As we can see from Eq. (3.86), it is physically equivalent

to adding a canceling uniform charge distribution to ensure the periodicity condition at the boundary and still allow for a non-vanishing charge in the interior.

- **QED$_L$** This prescription is similar to QED$_{TL}$ but with a Lagrange-multiplier $c_\mu = c_\mu(t)$ and the corresponding constraint

$$\int_{\mathbb{T}^3} A_\mu(t,\vec{x})\,\mathrm{d}^3x = 0 \qquad \forall t \Rightarrow \quad \tilde{A}_\mu\left(k_0, \vec{k} = \vec{0}\right) = 0, \quad \forall k_0 \tag{3.87}$$

It hence amounts to removing the spatial zero mode at every timeslice [153, 160]. Clearly, this prescription breaks Lorentz invariance and is non-local [157]. However, it is reflection-positive which is not the case for QED$_{TL}$ and therefore has a transfer matrix [153]. We will use this prescription to remove the photon zero-mode in this thesis.

- **Massive photon** Giving the photon a small mass $m_\gamma \neq 0$ screens the long-range interactions and regularizes the IR divergences. However, this prescription (mildly) breaks gauge invariance. In practice, one performs simulations at various values of m_γ and extrapolates the results to $m_\gamma = 0$ (one however has to be careful since the $L \to \infty$ and $m_\gamma \to 0$ limits do not commute).

- **C* boundary conditions** Another possibility proposed in the literature is to impose so-called C^* boundary conditions on the gauge and fermion fields in at least one direction [161],

$$A_\mu(x + L\mathbf{k}) = -A_\mu(x) \quad \psi(x + L\mathbf{k}) = C^{-1}\bar{\psi}^T(x) \quad \bar{\psi}(x + L\mathbf{k}) = -\psi^T(x)C\,, \tag{3.88}$$

with C the charge-conjugation matrix for spinors, which also removes the zero-mode. QED with C^* boundary conditions yields a local QFT in finite volume. It is however expensive to simulate since the size of the lattice is at least doubled and breaks cubic invariance if the boundary conditions are not applied in all directions. Charge and flavour conservation are broken but the remaining symmetries ensure the stability of all stable mesons and most of the stable baryons [157].

All prescriptions mentioned above should be equivalent if the infinite volume limit is taken before any other limit. Since QED is a long-range interaction, it introduces large finite-volume effects. In the QED$_L$ scheme, which we use in this thesis to simulate QED on the lattice, finite-volume corrections of observables with QED interactions are proportional to powers of $1/L$ instead of exponentially suppressed, as is the case for pure QCD [159]. We will discuss finite-volume effects in Sec. 4.3.2.3.

3.8 Numerical methods

As we saw, lattice QCD relies on the numerical evaluation of the QCD path integral which makes lattice QCD a numerically challenging field. The precision of calculated observables therefore depends foremost on the available computer power and algorithmic advances. Since the very first simulations in the 80's numerical methods have advanced enormously, with nowadays a whole plethora of numerical recipes and novel developments, such that also computer science profits from the advancements in lattice QCD.

Although I am very interested in numerical mathematics and algorithms, I unfortunately did not have the chance to work in this sub-field of lattice QCD during my thesis and will hence only briefly outline the most important ideas. More details can for instance be found in [162], for overviews of recent advances see for instance Refs. [163, 164].

Importance sampling: the Metropolis algorithm A direct evaluation of the path integral using a numerical quadrature rule is technically unfeasible. Instead, one evaluates the path integral using Markov chain Monte Carlo importance sampling techniques. The idea is to generate a series of N random configurations and approximate the expectation value of an observable O as

$$\langle O \rangle = \frac{\sum_i O(U_i) e^{-S(U_i)}}{\sum_i e^{-S(U_i)}} \ . \tag{3.89}$$

We already mentioned in Sec. 3.4.3 the formal relation between an Euclidean quantum field theory in d-dimensional spacetime and a classical statistical system in d-dimensional space. Therefore, the probability measure can be interpreted as the Boltzmann factor of a canonical ensemble. $e^{-S(U_i)}$ is typically strongly peaked around the minimal action configurations and a huge number of random configurations would be necessary to accurately converge to the value of the integral. Rather, one generates a sequence of gluonic field configurations distributed according to the Boltzmann factor, i.e. according to the canonical equilibrium distribution

$$\mathrm{d}P(U) = \frac{1}{Z} \mathcal{D}[U] \, e^{-S_G[U]} \det[U] \tag{3.90}$$

and approximates the value of the integral as

$$\langle O \rangle \approx \frac{1}{N} \sum_{i=1}^{N} O[U_i] \ . \tag{3.91}$$

The exact value of the integral is obtained for $N \to \infty$ and for a finite number of samplings the statistical error is in general $\mathcal{O}(1/\sqrt{N})$.

Markov chains are a way of generating a set of configurations consistent with the canonical distribution. One starts from an arbitrary configuration and each subsequent configuration is obtained from the previous one

$$U_0 \to U_1 \to U_2 \to \dots$$

Markov chains possess the ergodicity property, which means that after a certain time all configurations can be reached (ensemble average equals time average).

In the quenched approximation or for pure gauge simulations a local update algorithm can be used. The classic example of such an importance sampling algorithm is the Metropolis algorithm [165]: the rule $T(U|U')$ is split into a generating $g(U \to U')$ and an accept/reject $a(U \to U')$ step, such that the total probability is the product of the generating and acceptance probabilities. The acceptance probability

$$a(U \to U') = \min\left\{1, P(U)/P(U')\right\} \ , \tag{3.92}$$

with $P(U) = e^{-S(U)}$ satisfies the detailed balance condition. For further details about the implementation of the Metropolis algorithm to Wilson's gauge action see for instance Refs. [162, 166, 167].

Hybrid Monte Carlo algorithm To include dynamical fermions in our simulation we perform the integral over the Grassmannian fields analytically. This results in a non-local effective action $S_{\text{eff}} = S_U - \text{Tr}(\ln D(U))$, where the second term takes into account effect of sea quarks. Local update algorithms such as the Metropolis algorithm become impractical and we therefore require a suitable updating algorithm for these effective non-local actions. The hybrid Monte Carlo algorithm is nowadays the standard algorithm for simulations with sea quarks. The idea of the hybrid Monte Carlo algorithm is to find the minimal energy solution of the system by constructing an (artificial) Hamiltonian and evolving the system according to Hamilton's equations. This is in practice again followed by an accept/reject step – i.e. accept the new configuration with probability $\exp(-\Delta E)$ where ΔE is the energy difference with the previous configuration. This accept/reject step is necessary, since the evolution in "time" of Hamilton's equations is performed numerically and therefore not exact [166, 168–170].

As we saw in Sec. 3.4.2, the integral over the Grassmanian quark degrees of freedom can be calculated analytically. This is advantageous, as one does not have to deal with representing Grassmann numbers on a computer. The calculation of the fermion determinant can however become computationally demanding. Historically, simulations were performed in the so-called quenched approximation, where the determinant is set to a constant and dynamical quark effects are thus treated as a mean field. With the HMC algorithm it is possible to efficiently add dynamical quark effects by introducing so-called pseudo-fermion fields. The net effect of this formulation is that, in the case of two quark flavours, one only has to evaluate $(DD^\dagger)^{-1}$ instead of a determinant. This renders the inclusion of the dynamical fermions much cheaper [171]. However, this reformulation is only possible if an even number of quark flavours is included. A variant of this algorithm, called the rational hybrid Monte Carlo algorithm allows for a generalization to an arbitrary number of flavours [172–174].

In our work configurations are generated with a Rational Hybrid Monte Carlo algorithm [174] including force gradient [175], multiple timescales [176] and preconditioning of the fermion determinant by multiple pseudofermion fields [177].

Calculation of quark propagators As we saw in Sec. 3.5, hadron propagators are constructed from quark propagators G, requiring the inversion of the Dirac operator D,

$$\sum_{y,\beta,b} D^{ab}_{\alpha\beta}(x,y) G^{bc}_{\beta\gamma}(y,z) = \delta_{x,y}\delta_{ac}\delta_{\alpha\gamma} \ , \tag{3.93}$$

where (x, α, a) correspond to spacetime, Dirac and colour indices, respectively. The backward propagator can then easily be obtained by using γ_5-hermiticity. Calculating the full inverse of the Dirac operator would provide us the all-to-all propagator from all sources y to all sinks z. Since the Dirac matrix has a size $12N \times 12N$ with N the number of lattice sites, this is however very costly and in addition requires much storage. Hence, one uses sources to approximate the

inverse. This amounts to solving a linear system of equations,

$$D\psi(x) = \eta(x) \ , \tag{3.94}$$

where $\eta(x)$ is the source field and $\psi(x)$ the solution. For instance, by choosing point sources $\eta(x) = \delta(x)$ one calculates the one-to-all propagator which corresponds to one row of the full propagator $G(y,z)$. To solve this linear system of equations Krylov-space solvers are most efficient. In scientific simulations, much computational effort is often needed to solve linear systems of equations. These systems can be quite large and additionally result in dense matrices, which makes it very costly to solve such systems of equations with the usual elimination methods – such as the Gaussian. The Gaussian elimination method (a direct one) requires a computational effort of $\mathcal{O}(N^3)$ for dense matrices ($N^3/3$ for LR decomposition, $N^2/2$ for forward and backward substitution [178]), thus grows cubically with the dimension of the system and leads relatively soon to a high computational effort[15] We won't go into the details of Krylov subspace solvers, a nice overview can be found in [179]. The first of these Krylov subspace solvers is the conjugate gradient method proposed by Hestenes and Stiefel in 1952 [180]. It relies on an iterative minimization of the functional

$$J[\psi] \equiv \frac{1}{2}\psi^T D\psi - \psi^T \eta$$

whose global minimum ψ^* solves $D\psi = \eta$. One takes advantage of the fact that the level sets of J are circles with respect to the algebraic norm, and hence takes the steepest descent direction orthogonal to those level sets which point directly to the minimum (while in the steepest descent method one takes the level sets in Euclidean coordinates which are ellipses). The original CG-algorithm works for positive-definite, symmetric matrices only. There exist many variants of the original conjugate gradient method, such as the bicgstab algorithm [181] which also works for non-symmetric general matrices. To improve the condition number $\kappa(D) = \lambda_{\max}(D)/\lambda_{\min}(D)$, with $\lambda(D)_{\max/\min}$ the maximal/minimal eigenvalue of the Dirac operator, one can use preconditioning (such as for instance even-odd preconditioning for Wilson fermions) to accelerate the convergence of the solver [162]

$$LDR\Phi = L\eta \tag{3.95}$$

and the solution is obtained by calculating $\psi = R\Phi$. From the condition number we also see that lowering the mass of the lightest quark becomes numerically more challenging. In our work we use a technique which is called low-mode-averaging [182–184]. The basic idea is that the low energy spectrum of the Dirac operator is solved exactly,

$$G(x,y) = \sum_i^L \frac{1}{\lambda_i} v^{(i)}(x) \prod v^{(i)}(y)^\dagger + \tilde{G}(x,y) \ , \tag{3.96}$$

with $v^{(i)}$ and λ_i the L low-lying eigenvectors and corresponding eigenvalues. $\tilde{G}(x,y)$ is then easier to invert numerically since the low modes have been removed. This significantly reduces the noise in the propagator, also because it allows averaging not only over the source position, but also over that of the sink, augmenting the statistics by a number proportional to the lattice volume.

[15] This was already recognized by Gauss himself and so he proposed in 1823 an iteration method for the solution of a system of equations with four unknowns. *I recommend this mode to you for imitation. You will hardly ever again eliminate directly, at least not when you have more than 2 unknowns. The indirect method can be carried out while half asleep, or while thinking about other things.*. (Gauss to Gerling, letter of December 26, 1823)

In recent years, progress has also been made in the application of domain decomposition techniques to lattice QCD, see for instance [185–188].

Smearing In order to cope with strong UV fluctuations of the gauge fields, one additionally smooths them out over a certain finite range. This is done by taking averages over products of some neighbouring links, followed by a projection to an $SU(3)$ group element. This smearing procedure should not affect long-distance quantities. In our simulations we use *stout* [189] and *HEX* [190–192] smearing techniques, cf. Sec. 4.3.1. Stout smearing has the advantage that it is differentiable with respect to the link variables, rendering Monte Carlo sampling based on molecular dynamics evolution (such as the Hybrid Monte Carlo algorithm) highly efficient. The HEX smearing technique uses an exponential smearing of an average of plaquettes together with the hypercubic nesting trick (HYP). Smearing the link variables before coupling them to staggered fermions is particularly advantageous because it helps suppressing taste-breaking effects which arise from interactions with highly virtual gluons. The net effect is that the spectrum of the Dirac operator resembles one with perfect 4-fold degeneracy [193].

3.9 Analysis

3.9.1 Statistical error

The expectation value of an observable O can be obtained by calculating the sample average of this observable measured on N configurations,

$$\langle O \rangle = \frac{1}{N} \sum_{i=1}^{N} O_i \left(1 + \mathcal{O}(1/\sqrt{N})\right) , \tag{3.97}$$

with O_i the value of the observable on configuration i. Our gauge configurations are generated using MCMC methods whose mere definition implies that subsequent configurations are not independent. Hence, in order to obtain an estimate for the statistical uncertainty, one cannot simply use the sample standard deviation which would scale as $\mathcal{O}(1/\sqrt{N})$. The number of update steps separating two configurations which can be regarded as independent is called the autocorrelation time. As we discussed in Sec. 3.6, taking the continuum limit corresponds to a second order phase transition in the classical statistical system. For $a \to 0$ the correlation length diverges, $\xi \to \infty$. This implies that the autocorrelation time blows up when going to smaller and smaller lattice spacings. Hence, this represents a fundamental obstacle in simulating QCD on lattices with extremely small lattice spacings. This effect is called *critical slowing down*. A possible solution to this problem might be the use of machine learning techniques, see for instance [194, 195].

Imagine we have determined the autocorrelation time for a specific ensemble (see Refs. [162, 166, 167]). Then, a way to estimate the statistical uncertainty and to eliminate correlations as best as possible, is the following: First, the data is binned into N_J blocks with length $L_B \gg \tau$ where

τ is the (dimensionless) autocorrelation time and averages over the values in each bin are taken,

$$O_b = \frac{1}{L_B} \sum_{i=1}^{L_B} O_{b+i} \,, \qquad b = 1, \ldots, N_J \,. \tag{3.98}$$

This leaves us with a set of n_B blocked measurements O_b, $b = 1, \ldots, N_J$. Now, since we are interested in calculating derived quantities $f(O_b)$ of these O_b measurements, taking the naive average would introduce a bias of $\mathcal{O}(1/N)$. The bias of the estimator can be reduced by performing a so-called resampling of the data. Here we use the jackknife method to resample our data. This reduces the bias to $\mathcal{O}(1/N^2)$. The method works as follows. From the original (already blocked) dataset of size N_J one element at a time is removed, resulting in N_J new samples, where each new jkn-th jackknife sample vector contains $N_J - 1$ elements

$$\vec{O}_{jkn} = \{O_1, O_1, \ldots, O_{jkn-1}, O_{jkn+1}, \ldots, O_{N_J-1}, O_{N_J}\} \,. \tag{3.99}$$

The new list of data O_{jkn} with $jkn = 1, \ldots N_J$ is then obtained by taking the average of each new jackknife sample vector. Finally, the estimator for $f(O)$ can be obtained

$$\bar{f} = \frac{1}{N_J} \sum_{jkn=1}^{N_J} f(O_{jkn}) \tag{3.100}$$

and its variance is [12, 162, 166, 167, 196, 197]

$$\sigma_{\bar{f}}^2 = \frac{N_J - 1}{N_J} \sum_{jkn=1}^{N_J} \left(f(O_{jkn}) - \bar{f} \right)^2 . \tag{3.101}$$

In this work we use the jackknife method to estimate the statistical uncertainty of our quantities. For each ensemble we block the data in $N_J = 48$ blocks, resulting in blocks with lengths much longer than the typical autocorrelation time. We choose an equal number of blocks on all ensembles which is very convenient for the analysis.

3.9.2 Systematic error

When analysing the simulation data, multiple "choices" have to be made, such as fit ranges for hadron correlators to extract masses, forms of fit functions for the continuum extrapolation, and many others. To estimate the uncertainty originating from these choices, we perform a variety of analyses – in practice several hundred thousand analyses – and average over these analyses to obtain a final value. Each of these analyses gets a weight, determining how important the particular analysis is in the final estimate of the physical quantity. In specific cases, we consider the choices we make to be equally likely, hence these analyses are weighted uniformly. In other cases, we wish to choose the analysis which is preferred by the data. For this we use the Akaike Information Criterion (AIC) [198], which chooses the "best" model among the various choices of models, where "best" – i.e. the distance of the model from the true distribution – is measured in the sense of the Kullback-Leibler divergence. For a more detailed discussion see also [153]. Let us assume we have a total number of M different models. Each of these M models i has a certain number of fit parameters, p_i. By fitting model i to the data, we obtain an associated value for the chi-square, χ_i^2. Then, for Gaussian distributed errors, the AIC weight w_i for model

i is given by

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{m=1}^{M} \exp(-\Delta_m/2)} \tag{3.102}$$

with $\Delta_i = \chi_i^2 + 2p_i$. This procedure gives weights to all available procedures according to the fit quality of the model. Models with lower χ^2 and fewer fit parameters are hence preferred. If we now wish to consider only a subset of datapoints (e.g. we discard some of the coarsest lattice spacings), the AIC-weight has to be slightly modified. Assuming normally distributed errors one finds $\Delta_i = \chi_i^2 + 2p_i - d_i$, where d_i are the number of data points. This modification ensures that models where datapoints have been discarded get punished [7, 199]. We detail the procedure to estimate the systematic uncertainties in our final analysis in Sec. 4.3.2.4.

4 The running of the electromagnetic coupling

4.1 Introduction

As we saw in Sec. 2.2.2, for pure QED, the full photon propagator is given by infinite insertions of the vacuum polarization tensor. By resumming these bubbles we were able to define an effective coupling, whose "running" we are interested in in this section. In the on-shell scheme, the scheme which we will use in Sec. 4.3, we fixed the renormalized coupling via the Coulomb potential at large distances.

At low energies, in the Thomson limit, α is given by the fine structure constant and is known extremely precisely from low-energy measurements. The current most precise value is determined by measuring the recoil velocity, using matter-wave interferometry, of a rubidium atom after absorption of a photon [200]. The value is $\alpha^{-1} = 137.035999206(11)$, i.e. a remarkable relative precision of 81 parts per trillion. This measurement is in tension with an earlier experiment using cesium-133 atoms in matter-wave interferometer measurements that obtained $\alpha^{-1} = 137.035999046(27)$ [201]. They disagree by over 5 standard deviations. Alternatively, the fine structure constant can be determined by measuring the anomalous magnetic moment of the electron, a_e $(= 1159652180.73(28) \times 10^{-12})$ [202], resulting in $\alpha^{-1} = 137.035999150(33)$ [32], in agreement with the combined result from atomic interferometry measurements. Finally, the current world average $\alpha^{-1} = 137.035999084(21)$ still has a relative precision below a part per billion [2]. As we discussed in Sec. 2.2, the connection between different energy scales in different experiments is given by the running of the coupling constants. Of particular interest is hence the value of the electromagnetic coupling at the Z-scale since it enters electroweak precision tests of the SM. In calculating the running up to the Z-scale, five orders of magnitude in precision are lost [203]. This makes $\alpha(M_Z)$ one of the least well determined input parameters of the SM.

Including the entire SM, the photon couples to the quarks and the leptons, i.e. the fermionic current in Eq. (2.23) includes a sum over both leptons and quarks. Hence, the vacuum polarization bubble that we calculated in Sec. 2.2 can contain photons, leptons, quarks, gluons, etc. Due to the very different nature of these interactions and consequently very different computational tools, it is convenient to split the vacuum polarization into a leptonic part, the hadronic contribution from the 5 lightest quarks and the contribution from the top. In principle there are also weak gauge-boson contributions, such as that of W^+W^-. However, when QED is embedded in the non-abelian electroweak theory, the presence of three and four gauge-boson interactions do

not allow a unique definition of a gauge invariant effective coupling via the resummation of one-particle irreducible (1PI) vacuum polarization bubbles [1]. These issues are not relevant in this thesis, because we are interested here in helping to determine the contributions to electroweak precision processes that can be obtained using resummations of the 1PI photon self-energy defined in the absence of weak interactions. These can be calculated using the following running electromagnetic coupling defined in the on-shell scheme [42, 205–207]

$$\alpha(s) = \frac{\alpha}{1 - \Delta\alpha(s)} , \tag{4.1}$$

where

$$\Delta\alpha(s) = \Delta\alpha_{\text{lep}}(s) + \Delta\alpha^{(5)}_{\text{had}}(s) + \Delta\alpha_{\text{top}}(s) . \tag{4.2}$$

Other possible contributions to these electroweak processes must be included separately.

In (4.2), the leptonic contribution is defined to include only leptons and photons, i.e. only QED, and can therefore be obtained in perturbation theory. We have calculated this contribution at one-loop order in the on-shell scheme in perturbation theory in Sec. 2.2. It has been calculated up to four loops and is therefore known very precisely (see [208] and references therein). Similarly, the very small contribution from short distances from the top can be calculated in perturbation theory and is known to $\mathcal{O}(\alpha_s^2)$ [209]. The current values and precision of these contributions is $\Delta\alpha_{\text{lep}}(M_Z^2) = 314.979(2) \times 10^{-4}$ [208, 210] and $\Delta\alpha^{\text{top}}(M_Z^2) = -0.7201(37) \times 10^{-4}$ [211–214]. However, elow the scale of a few GeV, the contribution from the five lightest quarks becomes very difficult to calculate due to the infrared slavery property of QCD. This contribution receives large non-perturbative contributions and is hence not accessible via known analytic methods. Due to the complications in computing it, this contribution dominates the uncertainty in the running of α, by more than 90% at M_Z. Its current average in the on-shell scheme is $\Delta\alpha^{(5)}_{\text{had}}(M_Z) = 0.02768(7)$ [44].

However, future colliders will significantly reduce the uncertainties on $\alpha(M_Z)$ obtained indirectly via fits to electroweak precision observables. We won't go into the details of electroweak precision fits, for recent overviews see for instance [2, 42]. Instead, we will only briefly outline the general idea and review the current situation. The parameters in the electroweak theory are intricately connected and can ultimately be expressed in terms of the five parameters g, g', g_s, λ and μ (see Sec. 2.1.2). Alternatively, one uses a different set of parameters (cf. also Sec. 2.1.2),

$$\alpha \equiv \frac{e^2}{4\pi} = \frac{g^2 g'^2}{4\pi(g^2 + g'^2)} , \ \alpha_s \equiv \frac{g_s^2}{4\pi} , \ \sqrt{2}G_F \equiv v^{-2} = \frac{\lambda^2}{2|\mu^2|} , \ M_Z = \frac{1}{2}\sqrt{g^2 + g'^2}v , \ M_H = \lambda v .$$

Other quantities, such as the mass of the W-bosons M_W or the weak mixing angle $\sin^2\theta_W$ can be expressed in terms of these parameters. Since the discovery of the Higgs boson in 2012 [31] and the determination of its mass, all free SM parameters are fixed. Global fits therefore allow for important consistency tests of the SM. One experimentally measures quantities which are sensitive to these parameters (e.g. the masses M_Z, M_W and M_H themselves or their decay widths Γ_Z and Γ_W or so called forward-backward asymmetries which are sensitive to the effective mixing angle) in hadron or e^+e^- colliders and compares these measurements to the predictions from performing global fits. This is the idea behind electroweak precision physics. A discrepancy between

[1] See e.g. [204] for a discussion of the problems encountered in defining a gauge-invariant effective coupling in a non-abelian gauge theory and of possible, partial solutions.

measurement and theoretical prediction might reveal contributions of new physics which were not taken into account. It is hence crucial that both experimental and theoretical uncertainties are of the same size to be most sensitive to signs of new physics. Finally, electroweak fits play an important role in answering the question of the value of λ at higher energies and consequently the stability of the Higgs vacuum [215, 216].

The GAPP package [43] and the Gfitter package [217, 218] can be used to perform global SM fits to electroweak precision data. Both programs differ in computer language (Fortran and C^{++}), renormalization scheme and details in the data sets employed, which makes them complementary and allows for important consistency checks. Results of these fits, including and excluding the corresponding measurement value and likewise including and excluding the experimentally measured value of the Higgs mass are summarized in Refs. [2, 42]. Here we only mention the most important points relevant for the following discussion. Firstly, the global fits are in general in very good agreement with the measurements and have good chi-square[2]. However, the Higgs mass determined indirectly in the global fit, $M_H = 91.0^{+20}_{-17}$ GeV, is in slight tension (1.7 σ) with the directly measured value, see also Fig. 4.1 of Ref. [42] [3]. For our discussion most important is the fact that the running of the effective electromagnetic coupling is for many observables the dominant radiative correction and the precision is limited by the theoretical precision in its hadronic contribution.

As already mentioned, the precision in the running of α could become the bottleneck in precision tests with measurements from future colliders. The FCC-ee Tera-Z is a planned e^+e^- collider at CERN at the Z resonance region. The FCC-ee "Electroweak Factory" plans to produce all massive bosons of the SM and the top quark (and hence should also significantly improve the determination of the top mass). FCC-ee plans to run a FCC-ee-Z phase for four years with centre-of-mass energies of 88-95 GeV and an integrated luminosity of 150 ab^{-1}. They expect to see $3 \cdot 10^{12}$ visible Z decays [221]. The goal is to achieve an at least 20-100 times higher precision in electroweak precision measurements. In particular, by measuring the forward-backward asymmetry around the Z pole, $\alpha(M_Z)$, could be directly determined [222]. The ILC/GigaZ is a linear collider about 30-50 km long planned in Japan, with a center-of-mass collision energy of 500 GeV and a possibility to increase it to 1 TeV. All standard model particles would be within reach. For M_H for the ILC/GigaZ the uncertainties on M_Z and $\Delta\alpha^{(5)}_{\text{had}}(M_Z)$ become equally important and a total precision of less than 10 GeV on the Higgs mass could be achieved. In a nutshell, the situation can therefore be summarized as follows [223, 224]:

	$\delta\alpha(M_Z)/\alpha(M_Z)$
current precision	$\sim 10^{-4}$
precision needed for future colliders	$\sim 5 \times 10^{-5}$

In fact, the uncertainties on the direct calculation of $\alpha(M_Z)$ will have to be reduced by a factor of two to fully leverage these future measurements in the search for new fundamental physics [3]. Fig. 4.1 sketches a summary of the current situation. I show a recent determination using

[2]Note that the recent measurement of the Z-boson mass from CDF at Fermilab is of the same level of precision with the EW fits and disagrees with the SM prediction by more than 7σ deviations. Using the new measurement in the gobal fits leads to bad χ^2 and significant tensions. For discussions on consequences for electroweak fits and new physics scenarios see for instance [219, 220].

[3]The value obtained by the Mainz collaboration for $\Delta\alpha^{(3)}_{\text{had}}(2\ \text{GeV}^2)$ [8] results in a decrease of the indirectly determined Higgs mass by 13 GeV [2]. We will comment more on this result and the relation between recent lattice determinations of the muon anomalous magnetic moment and the running of α in Sec. 4.4

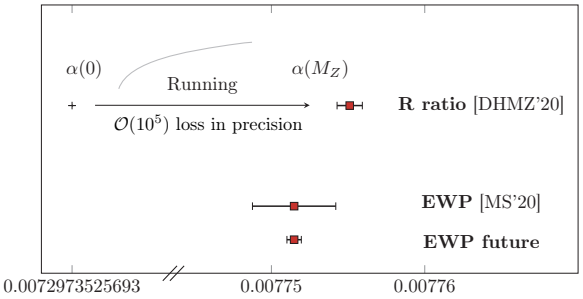


FIGURE 4.1: The electromagnetic coupling in the Thomson limit, given by the fine-structure constant, compared to its value at the Z-scale. By running $\alpha(0)$ up to the mass of the Z-boson, five orders in precision are lost. We show the value of $\alpha(M_Z)$ obtained using a data driven approach for determining the hadronic contribution to the running of α in the low energy regime [4] versus the value obtained indirectly from global fits to electroweak precision observables [49]. The forecast for estimated uncertainties at future colliders such as ILC or FCC-ee is depicted at the bottom and would require the uncertainty on the theoretical prediction to be halved in order to match the predicted precision at future colliders.

a data driven approach for determining the QCD running of α in the low energy regime [4]. While this determination has smaller errors than those of the value indirectly obtained via fits to electroweak precision observables [49], future colliders will significantly reduce the uncertainties of the latter.

The traditional way of obtaining the HVP in this non-perturbative regime is to use measurements of the $e^+e^- \to$ hadrons cross section as a function of the centre-of-mass energy and a dispersion relation [4, 5]. However, for space-like momenta, the HVP is accessible to ab-initio calculations in lattice QCD and thus provides a complementary approach which does not depend on cross-section data [7–9, 225, 226]. In App. A we show an (incomplete) compilation of the current precision on $\Delta\alpha^{(5)}_{\text{had}}(M_Z)$, obtained either via the dispersive approach or on the lattice. The phenomenological determinations differ in many ways, e.g. by how much and in which energy regions perturbative QCD is used, by how correlations and tensions between or within the various data sets are accounted for, by how to interpolate between different measurement bins to perform the necessary integrals over center-of-mass energy, etc. We will briefly outline the dispersive approach in Sec. 4.2.1. The hadronic contribution to the running of α in the Euclidean region, obtained in lattice QCD simulations, will be the main focus of this part of the thesis.

4.2 The dispersive approach: relating the HVP to $e^+e^- \to$ hadrons cross-section data

4.2.1 Analyticity: dispersion relations and vacuum polarization

The contribution to hadronic vacuum polarization (HVP) from the five lightest quarks is defined as [4]

$$\Pi_{\mu\nu}(q^2) = i \int d^4x e^{iq\cdot x} \langle 0|T\{j_\mu(x)j_\nu(0)\}|0\rangle = \left(q_\mu q_\nu - g_{\mu\nu}q^2\right)\Pi\left(q^2\right) \tag{4.3}$$

[4] Here as often in this chapter of the thesis we drop the superscripts and subscripts on $\Pi^{(5)}_{\text{had}}(q^2)$ and denote this five-quark-flavour contribution simply by Π except in cases where this simplified notation is ambiguous.

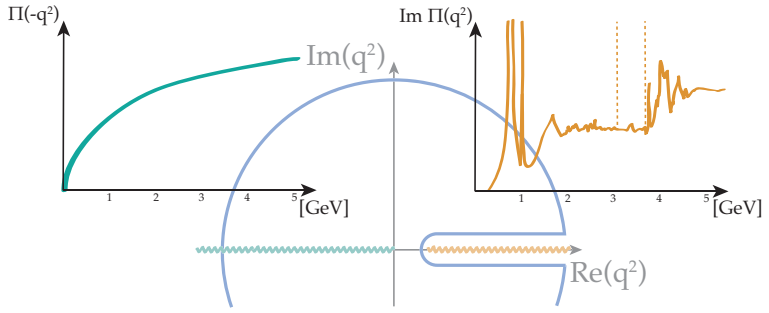


FIGURE 4.2: Analyticity properties of the hadronic vacuum polarization function $\Pi(q^2)$ in the q^2 complex play and Cauchy contour $\mathcal{C}$ used to relate its real and imaginary parts via a dispersion relation. The circle in $\mathcal{C}$ has infinite radius. The imaginary part of the HVP arises from the discontinuity along the cut on the positive real axis shown in brown. In the spacelike region with negative q^2, the Euclidean HVP function $\Pi(Q^2 = -q^2)$ is smooth (green curve) while the imaginary part, which corresponds the spectral function of hadrons in the $J^{PC} = 1^{11}$ channel (see Sec. 4.2.2), has a very rich structure (brown curve). The Euclidean vacuum polarization function can, for values of Q^2 that are sufficiently below the lattice cutoffs accessible in simulations, be computed numerically on the lattice.

with

$$j^\mu(x) = \sum_{f=u,d,s,c,b} q_f \bar{f}(x)\gamma^\mu f(x) = \frac{2}{3}\bar{u}\gamma_\mu u - \frac{1}{3}\bar{d}\gamma_\mu d - \frac{1}{3}\bar{s}\gamma_\mu s + \frac{2}{3}\bar{c}\gamma_\mu c - \frac{1}{3}\bar{b}\gamma_\mu b \ . \tag{4.4}$$

One can rewrite the time ordered current correlator in Eq. (4.3) in the Kaellen-Lehmann representation in order to understand its analytic properties and to relate its imaginary part to physical states going on-shell. By doing this, one can show, making use of the causal and unitary structure, that Π is analytic and real except for a cut along the positive real axis and that its imaginary part is given by on-shell intermediate physical states. We will not show this here but the derivation can be found in e.g. [227]. The idea is to insert a completeness relation and to write the correlator as an integral over a spectral function of physical intermediate states in the Hilbert space. The cut along the positive real axis starts at the threshold for pair creation, $s_{\rm th} = 4m^2$, i.e. where particles can go on-shell. Here m is the mass of the lightest pair-producible particle. This property is, as we mentioned above, a consequence of causality and unitarity. Let us now make use of the fact that we know the analytic structure of the vacuum polarization function and let us choose the integration contour as depicted in Fig. 4.2, such that it does not contain any singularities.

Using Cauchy's theorem for the contour $\mathcal{C}$ depicted in Fig. 4.2, i.e. excluding the cut, one finds

$$\Pi\left(q^2\right) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathrm{d}s\ \frac{\Pi(s)}{s-q^2} = \frac{1}{2\pi i}\int_{s_{\rm th}}^{\infty} \mathrm{d}s\ \frac{\Pi(s+i\epsilon)-\Pi(s-i\epsilon)}{s-q^2} + \mathcal{C}_\infty \tag{4.5}$$

where $\mathcal{C}_\infty$ is the contribution from the circle at infinity. The minus sign in front of $\Pi(s-i\epsilon)$ appears because of the orientation of the curve. Then, using the Schwarz reflection principle, $\Pi(s-i\epsilon) = \Pi^*(s+i\epsilon)$ and letting $R \to \infty$ one obtains

$$\Pi(q^2) = \lim_{\epsilon\to 0^+}\Pi(s+i\epsilon) = \frac{1}{\pi}\lim_{\epsilon\to 0^+}\int_{4m^2}^{\infty} \mathrm{d}s \frac{\mathrm{Im}\Pi(s)}{s-q^2-i\epsilon} + \mathcal{C}_\infty \ . \tag{4.6}$$

The function is determined by $\mathrm{Im}\Pi(s)$, up to a possibly non-zero $\mathcal{C}_\infty$. If $\Pi(s)$ were to fall off sufficiently rapidly for $|s| \to \infty$, that contribution would vanish and VP would be completely determined by its imaginary part along the cut. However this fall-off condition is not satisfied for VP. As we saw in the previous section, $\Pi(q^2)$ is logarithmically divergent. Renormalizing in

the on-shell scheme we defined the subtracted VP $\hat{\Pi}(Q^2)$ which satisfies the fall-off condition which can be checked order by order in pQCD. This gives rise to the subtracted dispersion relation

$$\hat{\Pi}(q^2) = \Pi(q^2) - \Pi(0) = \lim_{\epsilon\to 0^+} \frac{q^2}{\pi} \int_{s_{\rm th}}^{\infty} \mathrm{d}s \frac{\mathrm{Im}\Pi(s)}{s(s-q^2-i\epsilon)} \,. \tag{4.7}$$

In fact, $\Pi(q^2)$ is real everywhere except along the cut, $q^2 \geq 4m^2$, where it also develops an imaginary part. Nevertheless the real part of $\Pi(q^2)$ along the cut can also be obtained from its imaginary part, via the principal value integral

$$\mathrm{Re}\ \Pi(q^2) = \frac{q^2}{\pi} \mathcal{P} \int_{s_{\rm th}}^{\infty} \mathrm{d}s \frac{\mathrm{Im}\Pi(s)}{s(s-q^2)} \,. \tag{4.8}$$

4.2.2 Unitarity: the optical theorem and the R-ratio

Unitarity leads to an important relation between scattering amplitudes and cross sections, called the optical theorem. From unitarity of the S-matrix, $SS^\dagger = 1$, and writing $\mathcal{S} = 1 - i\mathcal{T}$, with $\mathcal{T}$ the transfer matrix,

$$\langle f|\mathcal{T}|i\rangle = (2\pi)^4 \delta^4(p_i - p_f)\mathcal{M}(i \to f) \,, \tag{4.9}$$

we find that $2\mathrm{Im}(\mathcal{T}) = i(\mathcal{T}^\dagger - \mathcal{T}) = \mathcal{T}^\dagger\mathcal{T}$ and hence (sandwiching between $\langle f|$ and $|i\rangle$) and using the completeness relation $1 = \sum_X \int \mathrm{d}\Pi_X\, |X\rangle\langle X|$ and $\mathrm{d}\Pi_{\rm LIPS} = (2\pi)^4\delta^{(4)}(\sum p)\,\mathrm{d}\Pi_X$

$$\mathcal{M}(i\to f) - \mathcal{M}^*(f\to i) = i\sum_X \int \mathrm{d}\ \Pi_X (2\pi)^4 \delta^4\left(p_i - p_X\right) \mathcal{M}(i\to X)\mathcal{M}^*(f\to X) \tag{4.10}$$

which is the *optical theorem* and must hold order by order in perturbation theory. Hence the imaginary part of the photon propagator is proportional to the total cross section $\sigma_{\rm tot}(e^+e^- \to \gamma^* \to \text{anything})$. In order to extract only the hadronic vacuum polarization function, we have to get rid of the kinematics of the initial e^+e^- pair and of the bubble resummation in the photon propagator. To do so, we realize that in the limit $s \gg 4m_e^2$

$$\sigma(e^+e^- \to \text{hadrons}) = \frac{1}{2s}\mathrm{Im}\mathcal{M}(e^+e^- \overset{\text{via } \Pi_{\rm had}(q^2)}{\to} e^+e^-) \tag{4.11}$$

with the insertion of the hadronic part of the vacuum polarization bubble $\Pi^{\mu\nu}_{\rm had}(q^2)$. For the matrix element one obtains

$$i\mathcal{M} = (-ie)^2 G_\mu \frac{-i}{s}\left(i(q^2 g^{\mu\nu} - q^\mu q^\nu)\Pi_{\rm had}(q^2)\right)\frac{-i}{s} G_\nu \,, \tag{4.12}$$

where $s = q^2$ and $\Pi_{\rm had}(q^2)$ is the hadronic part of the vacuum polarization function and G_μ and G_ν encode the spinor structure of the initial and final states, respectively. Contracting and averaging/summing over the initial/final spin configurations one obtains [227, 228]

$$\sigma(e^+e^- \to \text{ hadrons}) = \frac{e(s)^2}{s}\mathrm{Im}\Pi_{\rm had}(s) = \frac{4\pi\alpha(s)}{s}\mathrm{Im}\Pi_{\rm had}(s) \,, \tag{4.13}$$

where we have used the effective electromagnetic coupling $e(s)$ to account for the resummation of vacuum bubbles in the photon propagators, see Sec. 2.2.2. One then defines the R-ratio as

the normalized $e^+e^- \to$ hadrons cross section

$$R_{\rm had}(s) \equiv \frac{\sigma\,(e^+e^- \to \text{ hadrons })}{4\pi\alpha(s)^2/(3s)} \,. \tag{4.14}$$

The term in the denominator is equal to the tree-level cross-section $e^+e^- \to \mu^+\mu^-$ in the limit $s \gg 4m_\mu^2$ containing the effective photon propagator. At high enough energies, due to the asymptotic freedom property of QCD, the R-ratio can be calculated perturbatively. However, for center-of-mass energies $\sqrt{s} \lesssim 2$ GeV the quarks confine and hadronize. The process becomes non-perturbative. The R-ratio can then be related to the experimentally measured ratio of $\sigma(e^+e^- \to \text{hadrons})$ over $\sigma(e^+e^- \to \mu^+\mu^-)$, see Fig. 4.3. [5] Then the precise relationship becomes [6]

$$R(s) = 12\pi \mathrm{Im}\Pi(s) \,, \tag{4.16}$$

and finally, inserting $R(s)$ in the dispersion relation (4.7) we find

$$\hat{\Pi}(q^2) \equiv \Pi(q^2) - \Pi(0) = \frac{q^2}{12\pi^2} \int_{s_{\rm th}}^{\infty} \mathrm{d}s \, \frac{R(s)}{s(s-q^2-i\varepsilon)} \,. \tag{4.17}$$

The reference approach of calculating the low energy contribution to the hadronic vacuum polarisation consists therefore in evaluating the dispersion integral Eq. (4.17) from a compilation of precise experimental data for the R-ratio as a function of the squared centre-of-mass energy. In regions where there is no data available one uses perturbative QCD to interpolate. $\sigma_{\rm had}$ is measured at e^+e^- colliders. At low $\sqrt{s}$ there are two different exclusive approaches, i.e. where the final hadronic states are reconstructed. One either scans the centre-of-mass energy of the collider itself (energy scan) as can be done at symmetric e^+e^- colliders such as VEPP-2000. In the second technique, called *radiative return*, one considers events with a hard photon emitted in the initial state, so-called initial state radiation (ISR). This permits colliders as KLOE or BABAR to scan low energy regions (as the initial lepton has lost energy due to the ISR) although the collider is operating at a fixed centre-of-mass energy. The data analysis associated with this technique requires that one has full control over initial-state radiation, but it allows the use of very large data sets from recent fixed-energy high-luminosity experiments. One usually combines these data in fits with data from CMD and BES which scan the energy [229]. Hence, to sum up, the hadronic contribution to the running of α can be obtained via a dispersive approach

$$\Delta\alpha_{\rm had}^{(5)}(s) = -\frac{\alpha(0)\,s}{3\,\pi}\left(\int_{4m_\Pi^2}^{E_{\rm cut}^2} \frac{R^{\rm data}(s')}{s'(s'-s)}ds' + \int_{E_{\rm cut}^2}^{\infty} \frac{R^{\rm pQCD}(s')}{s'(s'-s)}ds'\right) \,. \tag{4.18}$$

[5] Besides its importance for calculating the hadronic contribution to the photon vacuum polarization function, the R-ratio has other interesting phenomenological applications. At leading order in perturbative QCD for massless quarks the R-ratio becomes a step-function

$$R_{\rm had}(s) = N_c\left(\sum_f Q_f^2\right)\Theta(s) + \mathcal{O}(\alpha_s) \tag{4.15}$$

which effectively counts the number of quark degrees of freedom. By comparing this prediction to experimental data, see Fig. 4.3, we deduce that there are $N_c = 3$ colours and we find the charges of the quarks in units of e, $R_{\rm had}(s) \simeq N_c\left((2/3)^2 + (1/3)^2 + (1/3)^2\right) = 2/3N_c$ for $4m_s^2 \ll s < 4m_c^2$, etc.

[6] In the following, we suppress the supscript "had".

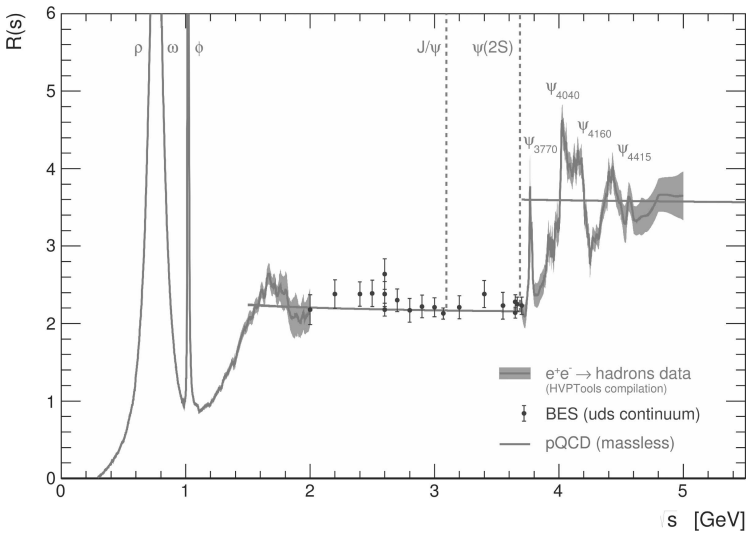


FIGURE 4.3: Experimental cross section for $e^+e^- \to$ hadrons as a function of the centre-of-mass energy. The bands correspond to the combined experimental measurements with uncertainties. The plot also contains the prediction from perturbation theory. The data points come from inclusive measurements from BES. Figure taken from [229].

HVP via the Adler function and the Euclidean split method As can be observed in Fig. 4.3, due to resonances and thresholds, it is challenging to determine at what value of $\sqrt{s}$ perturbation theory and data start to match. As has been suggested by Adler [230] and implemented and advertised by Jegerlehner [3, 203, 231], this comparison is more straightforward in the Euclidean region via the Adler function [3, 203, 231]. The Adler function is defined as the derivative of the HVP

$$D(Q^2) \equiv 12\pi^2 Q^2 \frac{\mathrm{d}\Pi(q^2)}{\mathrm{d}q^2} \,, \tag{4.19}$$

with $Q^2 = -q^2$ the Euclidean momentum transfer squared. Since $D(Q^2)$ is the derivative of the HVP it is finite and becomes constant at high energies. The Adler function is obtained from a dispersive integral over the narrow regions of the R-ratio in the time-like region,

$$D\left(Q^2\right) = Q^2 \int_{4m_\pi^2}^{\infty} \frac{R(s)}{\left(s+Q^2\right)^2} \,\mathrm{d}s \,, \tag{4.20}$$

is therefore smooth and easier to compare to perturbation theory than the R-ratio. One then obtains the Euclidean running from the integral of the Adler function,

$$\Delta\alpha_{\mathrm{had}}(Q_0^2) = \frac{\alpha}{3\pi} \int_0^{Q_0} \mathrm{d}Q'^2 \frac{D(Q'^2)}{Q'^2} \,. \tag{4.21}$$

A further advantage of this approach is that the experimental data used in the dispersive integral are reshuffled in such a way that the relevance of the different energy ranges is weighted differently and the total error thus reduced [3, 203]. Finally, as we are interested in the shift of the electromagnetic coupling constant at M_Z in the time-like region, the general strategy to obtain its time-like counterpart lies in using a telescopic sum, the so-called Euclidean split method

[3, 203][7],

$$\Delta\alpha_{\text{had}}^{(5)}\left(M_Z^2\right) = \Delta\alpha_{\text{had}}^{(5)}\left(-Q_0^2\right)_{\text{LQCD/R-ratio}} + \left[\Delta\alpha_{\text{had}}^{(5)}\left(-M_Z^2\right) - \Delta\alpha_{\text{had}}^{(5)}\left(-Q_0^2\right)\right]_{\text{pQCD/ R-ratio}} + \left[\Delta\alpha_{\text{had}}^{(5)}\left(M_Z^2\right) - \Delta\alpha_{\text{had}}^{(5)}\left(-M_Z^2\right)\right]_{\text{pQCD}} . \tag{4.22}$$

The smooth running in the Euclidean region from $-Q_0^2$ to $-M_Z^2$ can reliably be obtained in perturbation theory – for example from the perturbative Adler function in the Euclidean region or from the R-ratio –, as well as the last tiny contribution from M_Z^2 to $-M_Z^2$, since in both terms the non-perturbative contribution $\Pi(0)$ cancels. As we will see in the following section, the first term in Eq. (4.22) can also be calculated on the lattice for values of Q_0^2 well below the lattice cutoff. Here, the result from the lattice and from perturbation theory and/or the R-ratio should match onto each other. The general strategy of obtaining the first term in Eq. (4.22) – which is the main observable of this study – will be discussed in the next section.

4.3 Lattice QCD calculation of the hadronic running of α

In this chapter we detail the calculation of the main observable of the first part of this thesis, namely the determination of the hadronic contribution to the running of α based on ab-initio lattice QCD calculations. The paper on this part of the thesis is not yet finished but we present here most of the results that will go into it. As the running of α is closely related to the determination of the leading hadronic contribution to the anomalous magnetic moment of the muon, many techniques developed for a recent calculation of our collaboration could be adapted to the present work [7]. In particular, many of the lattice QCD simulations and of the computations of the necessary correlation functions have been completed on national and European supercomputers. Details on the way the simulations were performed can be found in [7] and are briefly summarized in Sec. 4.3.1. However, the analysis of the running of α holds its own difficulties that had to be understood and controlled. I started by performing a thorough analysis to yield the running of α up to ~ 10 GeV^2 in the isospin limit. While promising, this analysis pointed to difficulties, in particular at high energies, where discretization errors become important. I hence started an extensive study of discretization errors which will be detailed in Sec. 4.3.2.1. In particular, I was surprised to discover that the hadronic vacuum polarisation receives logarithmically-enhanced cutoff effects which render the continuum extrapolation more difficult [232]. I calculated the coefficient of this logarithmic term analytically at leading order in lattice perturbation theory and designed a number of improvement procedures to better control the continuum extrapolation. Concerning the continuum extrapolation at high energies, I tried many different approaches and am now confident that a conservative procedure for the estimation of systematic uncertainties has been achieved.

I then finally implemented a full systematic analysis, which is detailed in Sec. 4.3.2.

The original motivation is, as stated in the introduction, to significantly improve the precision of $\alpha(M_Z)$ to allow for more stringent SM electroweak precision tests. While the dispersive

[7] Recall from Sec. 2.2 that the shift in the running of α is related to the vacuum polarization function via $\Delta\alpha(q^2) = 4\pi\hat{\Pi}(q^2)$.

calculations to obtain the running of α have been performed since many years, see App. A, there is only one full lattice calculation by the Mainz group so far [8]. Burger et al. have performed the first lattice study of the Euclidean running using a fit ansatz [226]. Francis et al. [233] performed a first exploration of the Adler function on the lattice, based on the time-momentum-representation method that we detail below. In earlier works of our collaboration, the Euclidean running up to 10 GeV^2 was explored [6, 7]. In particular, we are the first ones to include the leading order QED and strong isospin breaking effects.

4.3.1 Lattice details

4.3.1.1 Ensembles

The main part of the simulations was performed using our so-called 4stout lattice action. It is a tree-level improved Symanzik gauge action [151] and a one link staggered fermion action. The gauge links are smeared with four steps of stout smearing [189], and smearing parameter $\rho = 0.125$. $2+1+1$ dynamical flavours were used for the generation of ensembles with equal up and down quark masses, $m_u = m_d = m_l$. The masses of the light and strange quarks were set to bracket the physical point and the charm mass was set by its ratio to the strange mass $m_c/m_s = 11.85$. This value was determined in Ref. [234] and is within 1% of the most recent lattice FLAG average [235]. We use five different values for the gauge coupling parameter $\beta = 6/g^2$, all of which were also used for the determination of the anomalous magnetic moment of the muon [7]. Here we omit however a sixth coarser ensemble because we are studying processes up to relatively large momenta. All lattices have a size of $L \times T \approx 6$ fm $\times$ 9 fm. The details of our 4stout ensembles are listed in Tab. 4.1. The physical point is determined using the mass of the Ω^- baryon, the pseudoscalar mass M_π and the mass of the connected pseudoscalar meson M_{ss}. The physical value of M_{ss} was determined in Ref. [7], where in addition the mass of the π^+, K^+ and K^0 were used for the definition of the physical point. The scattering around the physical point is shown in Fig. 4.4 where we plot the deviation of light and strange quark masses from their physical values for each of our 26 ensembles. We discuss the scale setting and definition of the physical point further in Sec. 4.3.2.4. We also calculate $\mathcal{O}(m_d - m_u)$ strong isospin-breaking effects and $\mathcal{O}(e_v^2)$ QED isospin-breaking effects on a subset of these ensembles.

QED isospin-breaking effects $\mathcal{O}(e_s^2)$ and $\mathcal{O}(e_v e_s)$ originating from the sea are calculated on a set of smaller 4stout ensembles with spatial extent $L = 3$ fm, since FV effects in the dynamical QED effects on those smaller lattices were found to be within the statistical uncertainties. We collect details on these ensembles in Tab. 4.2.

In order to ensure a reliable extrapolation to infinite volume, four dedicated $N_f = 2+1$ simulations were performed using an action specifically designed for this purpose: Since for large boxes only coarse lattice spacings are achievable, taste violations become significant and need to be suppressed. We therefore use a DBW2 gauge action [236] and one-link staggered quarks with four steps of HEX smearing applied to the gauge links [191]. In the following we will refer to this action as the "4HEX" action. Two of these ensembles have spatial extent $L_{\text{ref}} = 6.272$ fm, the other two were simulated in a very large box with spatial extent $L_{\text{big}} = 10.572$ fm. To further cope with taste-violations, the mass of the Goldstone pion is set to the harmonic mean square of all the pion taste partners, resulting in a Goldstone pion mass of ~ 110 MeV which is smaller

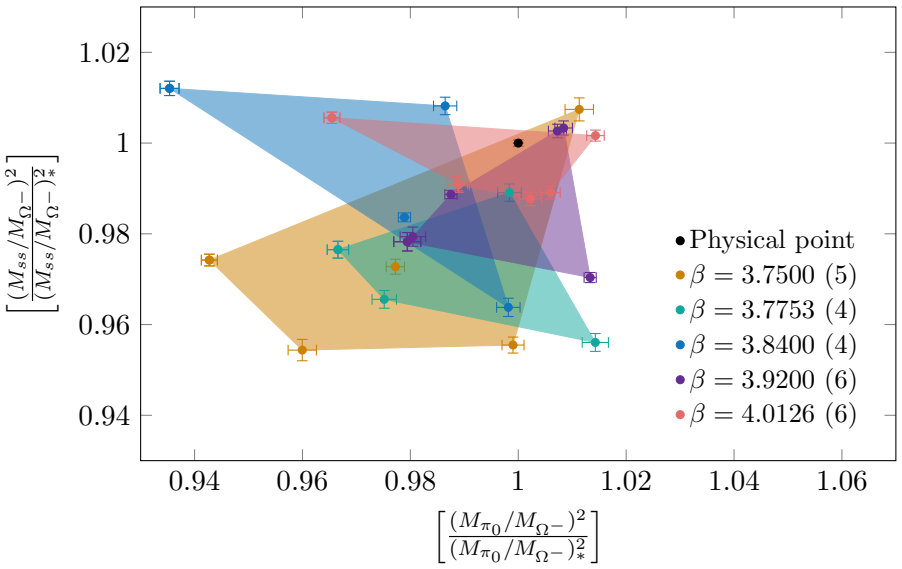


FIGURE 4.4: Landscape of our 4stout ensembles. Each colour corresponds to a different value of β, as listed in Tab. 4.1. We plot the deviations of the light and strange quark masses from their physical values which were chosen such that they scatter around the physical point, indicated by a black dot. M_{Π_0} is the mass of the Goldstone pion, M_{Ω^-} the mass of the Ω^- baryon, M_{ss} the mass of the meson built from a strange and an anti-strange quark and $*$ subscripts indicate the physical value. For more details see the main text and Sec. 4.3.2.4.

than the experimental value. This is supported by arguments from NLO chiral perturbation theory [7]. We list the details of these ensembles in Tab. 4.3.

4.3.1.2 Staggered electromagnetic currents of quarks

As discussed below in Sec. 4.3.1.3, to compute the hadronic contribution to the running of α we need the Euclidean quark electromagnetic current-current correlator in the time-momentum representation. Thus, we need the spatial components of the quark electromagnetic current. Here we focus on a single generic flavour contribution to this current with all charge factors removed, i.e. the current associated with flavour number. In the continuum this conserved vector current is uniquely defined:

$$J_i(x) = \bar{\psi}(x)\gamma_i\psi(x) \ , \tag{4.23}$$

with $i = 1, 2, 3$.

On the lattice, there are many possible discretizations of this current which all reduce to the one above in the naive continuum limit. In the presence of interactions, all but the conserved one will require a finite renormalization.

To display these currents we begin with the naive fermion bilinears with spin index s and taste index t which become diagonal after the staggering transformation of (3.22). These are the only ones that survive when three of the four Dirac components of the staggered field X are discarded [108]. Moreover, we are interested in currents with vanishing three-momentum. They can be written [108]:

$$J^{\mathrm{TL}}_{s,t}(x_0) = \frac{1}{N_{\delta\mathbf{x}_{st}}} \sum_{\mathbf{x}} \sum_{\delta\mathbf{x}_{st}} \bar{\psi}(x) B^\dagger_{\bar{t}}(x) \gamma_s U(x, x + \delta\mathbf{x}_{st}) \psi(x + \delta\mathbf{x}_{st}) \tag{4.24}$$

β	a[fm]	$L \times T$	am_s	m_s/m_l	# config	M_π[MeV]
3.7500	0.118	56×96	0.049593	28.038	315	129.613
			0.049593	26.939	516	132.218
			0.051617	29.183	504	130.010
			0.051617	28.038	522	132.605
			0.055666	28.038	215	138.370
3.7753	0.1116	56×84	0.047615	27.843	510	131.780
			0.048567	28.400	505	131.899
			0.046186	26.479	507	132.687
			0.049520	27.852	385	134.556
3.8400	0.0952	64×96	0.043194	28.500	510	135.029
			0.043194	30.205	190	131.207
			0.043194	30.205	436	131.286
			0.040750	28.007	1503	131.734
			0.039130	26.893	500	131.441
3.9200	0.0787	80×128	0.032440	27.679	506	132.011
			0.034240	27.502	512	136.304
			0.032000	26.512	1001	133.713
			0.032440	27.679	327	131.896
			0.033286	27.738	1450	133.697
			0.034240	27.502	500	136.225
4.0126	0.0640	96×144	0.026500	27.634	446	132.926
			0.026500	27.124	551	133.962
			0.026500	27.634	2248	132.997
			0.026500	27.124	1000	134.243
			0.027318	27.263	985	136.104
			0.027318	28.695	1750	132.529

TABLE 4.1: List of the 4stout ensembles used in this work. We list the gauge coupling, lattice spacing, lattice size, strange quark mass and ratio of the mass of the strange and light quarks, number of configurations and the approximate mass of the Goldstone pion. The charm mass is fixed via $m_c/m_s = 11.85$ [234].

β	a[fm]	$L \times T$	m_s	m_s/m_l	#conf	M_π[MeV]
3.7000	0.1315	24×48	0.057291	27.899	716	133.880
		48×64	0.057291	27.899	300	130.636
3.7753	0.1116	28×56	0.047615	27.843	887	135.879
3.8400	0.0952	32×64	0.043194	28.500	1110	138.496
			0.043194	30.205	1072	134.800
			0.040750	28.007	1036	135.771
			0.039130	26.893	1035	134.646

TABLE 4.2: List of the 4stout ensembles used for the calculation of dynamical QED effects. We list the gauge coupling, lattice spacing, lattice size, strange quark mass and ratio of the mass of the strange and light quarks, number of configurations and the approximate mass of the Goldstone pion.

where the superscript TL is explained below, s and t are the spin and taste indices, each belonging to $\mathbb{Z}_2^{\otimes 4}$, where $U(x,y)$ is the average of the links over the shortest paths between x and y and where the sum over $\delta\mathbf{x}_{st}$ is over the three components $\delta x_{st,i}$, $i = 1,2,3$. These sums include $\delta x_{st,i} = 0$ if $(s+t)_i \mod 2 = 0$ and $\delta x_{st,i} = \pm a$ for $(s+t)_i \mod 2 = 1$, i.e. an average over spatial reflections. $N_{\delta\mathbf{x}_{st}}$ is the total number of values of $\delta\mathbf{x}_{st}$ is summed over. In (4.24), $B_t(x)$ is the doubling tranformation operator given in (3.14) and γ_s in (3.16).

Coming back to the spatial component of the vector current of interest, in the case of staggered fermions there are (3 spatial indices) × (16 tastes) = 48 distinct components. Thanks to the geometric time-slice symmetry (GTS) of the transfer matrix discussed in Sec. 3.2.2, these organize

β	a[fm]	$L \times T$	m_s	m_s/m_l	# conf	M_π[MeV]
0.7300	0.112	56×84	0.060610	44.971	7709	104
		96×96			962	
		56×84		33.728	8173	121
		96×96			813	

TABLE 4.3: List of the 4HEX ensembles used for the calculation of finite-volume effects. We list the gauge coupling, lattice spacing, lattice size, strange quark mass and ratio of the mass of the strange and light quarks, number of configurations and approximately the mass of the Goldstone pion. Note that the masses of the Goldstone pions are supposed to bracket $M_\pi = 110$ MeV, where the harmonic mean square of the pion taste partners corresponds to the experimental pion mass.

Irrep spin⊗taste	$[s,t]$	mult.	# hops	# paths	ID
$\gamma_i \otimes \xi_i$ (local)	$[(0,\hat{\mathbf{i}}),(0,\hat{\mathbf{i}})]$	3	0	0×0	IV
$\gamma_i \otimes I$ (conserved)	$[(0,\hat{\mathbf{i}}),(0,\mathbf{0})]$	3	1	2×1	V
$\gamma_i \otimes \xi_i\xi_j$ $(j \neq i)$	$[(0,\hat{\mathbf{i}}),(0,\hat{\mathbf{i}}+\hat{\mathbf{j}})]$	6	1	2×1	X
$\gamma_i \otimes \xi_i\xi_j\xi_k$ $(j,k \neq i,\ k \neq j)$	$[(0,\hat{\mathbf{i}}),(0,\hat{\mathbf{i}}+\hat{\mathbf{j}}+\hat{\mathbf{k}})]$	3	2	4×2	XI
$\gamma_i \otimes \xi_j$ $(j \neq i)$	$[(0,\hat{\mathbf{i}}),(0,\hat{\mathbf{i}}+\hat{\mathbf{j}})]$	6	2	4×2	XV
$\gamma_i \otimes \xi_j\xi_k$ $(j,k \neq i,\ k \neq j)$	$[(0,\hat{\mathbf{i}}),(0,\hat{\mathbf{j}}+\hat{\mathbf{k}})]$	3	3	8×6	XX
# irreps = 6		# total multiplicity = 24			

TABLE 4.4: Irreps of time-local (TL) staggered discretizations of the continuum spatial-vector current of (4.23). The first two columns identify the irrep, first in spin⊗taste notation and second in terms of the spin s and taste t indices of (4.24). The third column indicates the multiplicity of the irrep and the fourth, the number of spacetime dimensions along which the quark field in $J_{s,t}(x)$ is shifted by a away from x. The fifth column provides the number of paths over which connecting links must be included in the notation (number of n-cubes with common vertex x) × (number of paths connecting x with the vertex along the longest diagonal), where n is the dimension the space delimited by the n hop directions. The last column is the number of the operator to which the irrep corresponds in Table. 1 of [237] (also that of [132]). In this table, the three sets of unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ with $i,j,k = 1,2,3$ each correspond to the same orthonormal basis of three-space.

into 12 irreps, each one containing the three spatial component of the vector current [130, 132]. Moreover these 12 irreps can be separated into two sets of 6 irreps each.

The first set of 6 irreps correspond to time-local (TL) currents for which $(s+t)_0 \mod 2 = 0$ and that are obviously local in time. Because we are considering spatial coordinates of vectors currents for which $s_0 = 1$, these currents must have $t_0 = 0$. Their properties are summarized in Tab. 4.4. It is interesting to note that the conserved current (i.e. the Noether current that results from the $U(1)_V$ quark-flavour symmetry) is a taste singlet while the local one is a taste vector.

The second set of irreps are non-local in time (TNL) and correspond to $(s+t)_0 \mod 2 = 1$. The same argument as above implies that the corresponding currents have $s_0 = 0$ and $t_0 = 1$ These are not given directly by the expression of the TL operators of (4.24) but can be constructed from them. We want operators that have the same quantum numbers as the TL ones but that include shifts in time. To obtain these operators we follow [132] and define them as:

$$J^{\mathrm{TNL}}_{s,t}(x_0) = \frac{1}{2}\left[J^{\mathrm{TL}}_{s,t}(x_0+a) - (-1)^{\sigma_s} J^{\mathrm{TL}}_{s,t}(x_0-a)\right] , \tag{4.25}$$

where σ_s is the number of hops associated with the corresponding TL operator in Tab. 4.4. The properties of these TNL operators are summarized in Tab. 4.5.

From these naive fermion operators it is straightforward to obtain the corresponding staggered ones in terms of the single componenent field $\xi(x)$ defined in Sec. 3.2.2. First we use the staggering transformations of (3.22) to rewrite the currents $J^{\mathrm{TL}}_{s,t}(x_0)$ in terms of the four-component

spin⊗taste	Irrep $[s,t]$	mult.	# hops	# paths	ID
$\gamma_i \otimes \xi_0\xi_i$	$[(0,\hat{\mathbf{i}}),(1,\hat{\mathbf{i}})]$	3	1	0×0	IVtnl
$\gamma_i \otimes \xi_0$	$[(0,\hat{\mathbf{i}}),(1,\mathbf{0})]$	3	2	2×2	Vtnl
$\gamma_i \otimes \xi_0\xi_i\xi_j\ (j\neq i)$	$[(0,\hat{\mathbf{i}}),(1,\hat{\mathbf{i}}+\hat{\mathbf{j}})]$	6	2	2×2	Xtnl
$\gamma_i \otimes \xi_0\xi_i\xi_j\xi_k\ (j,k\neq i,\ k\neq j)$	$[(0,\hat{\mathbf{i}}),(1,\hat{\mathbf{i}}+\hat{\mathbf{j}}+\hat{\mathbf{k}})]$	3	3	4×6	XItnl
$\gamma_i \otimes \xi_0\xi_j\ (j\neq i)$	$[(0,\hat{\mathbf{i}}),(1,\hat{\mathbf{i}}+\hat{\mathbf{j}})]$	6	3	4×6	XVtnl
$\gamma_i \otimes \xi_0\xi_j\xi_k\ (j,k\neq i,\ k\neq j)$	$[(0,\hat{\mathbf{i}}),(1,\hat{\mathbf{j}}+\hat{\mathbf{k}})]$	3	4	8×24	XXtnl
# irreps = 6		# total multiplicity = 24			

TABLE 4.5: Irreps of time-non-local (TNL) staggered discretizations of the continuum spatial-vector current of (4.23). The columns have the same meaning as in Tab. 4.4. Here the ID of the operator is the one taken from the corresponding TL operator in Tab. 4.4 (see also Table 2 of [132]

staggered field $X(x)$:

$$J^{\mathrm{TL}}_{s,t}(x_0) = \frac{1}{N_{\delta\mathbf{x}_{st}}}\sum_{\mathbf{x}}\sum_{\delta\mathbf{x}_{st}} \bar{X}(x)\Omega(x)^\dagger B^\dagger_{\bar{t}}(x)\gamma_s U(x,x+\delta\mathbf{x}_{st})\Omega(x+\delta\mathbf{x}_{st})X(x+\delta\mathbf{x}_{st}) \tag{4.26}$$

Then using the definition of $B_t(x)$ in (3.14), of $\Omega(x)$ in (3.23) and reducing $X(x)$ to $\chi(x)$, the currents can be given in the one-component staggered formalism [108],

$$J^{\mathrm{TL}}_{s,t}(x_0) \to \frac{1}{N_{\delta\mathbf{x}_{st}}}\sum_{\mathbf{x}}\sum_{\delta\mathbf{x}_{st}} \beta_{st}(x)\bar{\chi}(x)\chi(x+\delta\mathbf{x}_{st})\ , \tag{4.27}$$

where the phase $\beta_{st}(x)$ is given by

$$\beta_{st}(x) = \frac{1}{4}\mathrm{Tr}\left[\gamma_t^\dagger\gamma^\dagger_{x/a}\gamma_s\gamma_{x/a+s+t}\right] = (-1)^{(t^<+s^>)\cdot x/a}(-1)^{s\cdot(s+t)^<}\ , \tag{4.28}$$

with

$$m^<_\mu = (\sum_{\nu<\mu} m_\nu) \mod 2 \qquad \text{and} \qquad m^>_\mu = (\sum_{\nu>\mu} m_\nu) \mod 2\ , \tag{4.29}$$

for any $m \in \mathbb{Z}_2^{\otimes 4}$.

The point of considering all of these discretizations of the quark-flavour vector current is the following. If taste-symmetry were exact, the various discretizations would all give the same result at non-zero lattice spacing. However, due to taste violations, their approach to the continuum is different. In particular, as we will see below, the leading-order discretization error in the corresponding correlation functions is $\mathcal{O}(a^2\ln a^2)$ and its coefficient changes for different current discretizations. These different discretizations provide therefore an ideal way of assessing the systematic uncertainty of our results, since in the continuum, due to an exact taste-symmetry at zero lattice spacing, they should extrapolate to the same point.

In this thesis, the 12 irreps of currents defined above have only been implemented for the charm contribution to the HVP in the isospin limit which has particularly large discretizations due to the fact that the charm quark mass is large in lattice units. For all other contributions, only the conserved current has been considered.

4.3.1.3 Definition of the hadronic vacuum polarization on the lattice

We argued in Sec. 4.2 that the vacuum polarization function is analytic in the whole complex q^2 plane, except for a cut along the real axis. On the lattice, we have access to the HVP in the space-like region, where the vacuum polarization function is smooth, see Fig. 4.2. The HVP function in the Euclidean $\Pi(Q^2 = -q^2)$ is again related to $\Pi(q^2)$ in the time-like region via a dispersion relation (cf. Sec. 4.2.1)

$$\Pi(Q^2) = \frac{1}{\pi}\int_{s_{\rm th}}^{\infty} \mathrm{d}s\ \frac{Q^2}{s(s+Q^2)}\mathrm{Im}\Pi(s)\ . \tag{4.30}$$

Time-momentum-representation method. On the lattice, we study the time evolution of Euclidean correlation functions ($Q^2 = -q^2$). From the two point correlation function

$$C_{\mu\nu}(x) = \langle J_\mu(x) J_\nu(0)\rangle \tag{4.31}$$

of the quark electromagnetic current

$$J_\mu = \sum_{f=u,d,s,c} q_f \bar{\psi}\gamma_\mu\psi_f \quad \left(q_u = q_c = \frac{2}{3}, q_d = q_s = -\frac{1}{3}\right) \tag{4.32}$$

we obtain the Euclidean vacuum polarization tensor via a Fourier transform

$$\Pi_{\mu\nu}(Q) = \int d^4x\ e^{iQ\cdot x}\left\langle J_\mu(x)J_\nu(0)\right\rangle = \left(Q_\mu Q_\nu - \delta_{\mu\nu}Q^2\right)\Pi\left(Q^2\right)\ . \tag{4.33}$$

The last equality follows, in analogy to Sec. 2.2, from $\mathcal{O}(4)$ invariance and current conservation. Note however that the vacuum polarization is fully transverse only in infinite volume and infinite time. As we will see in the following, in finite volume, care has to be taken. In this study we include the four lightest quarks u, d, s and c. The very small contribution from the b quark is obtained from the results of [13]. We project the correlation function to zero spatial momentum and additionally flavour decompose it. This has the advantage that the various contributions have very different statistical and systematic uncertainties which can be addressed separately (note that we do not yet sum over i),

$$C_{ii}(t) = \int \mathrm{d}^3\vec{x}\left\langle J_i(x)J_i(0)\right\rangle = C_{ii}^l(t) + C_{ii}^s(t) + C_{ii}^c(t) + C_{ii}^{\rm disc}(t)\ . \tag{4.34}$$

Explicitly, the split of the connected part $\sum_f q_f^2 C_{ii}^{\rm conn}(m_f, e_f)$ is

$$C_{ii}^l \equiv \tfrac{4}{9}C_{ii}^{\rm conn}(m_u, \tfrac{2}{3}e) + \tfrac{1}{9}C_{ii}^{\rm conn}(m_d, -\tfrac{1}{3}e), \tag{4.35}$$

$$C_{ii}^s \equiv \tfrac{1}{9}C_{ii}^{\rm conn}(m_s, -\tfrac{1}{3}e), \tag{4.36}$$

$$C_{ii}^c \equiv \tfrac{4}{9}C_{ii}^{\rm conn}(m_c, \tfrac{2}{3}e)\ . \tag{4.37}$$

Further, in the isospin symmetric limit, the separation in IB singlet and triplet is

$$\begin{aligned}
\langle J_i J_i \rangle_{I=1} &\equiv \langle \frac{9}{10} C_{ii}^{l} \rangle_0 = \left[\frac{1}{2} C_{ii}^{\text{conn}} \right]_0 \\
\langle J_i J_i \rangle_{I=0} &\equiv \left[\frac{1}{10} C_{ii}^{l} + C_{ii}^{s} + C_{ii}^{c} + C_{ii}^{\text{disc}} \right]_0 \\
&= \left[\frac{1}{18} C_{ii}^{\text{conn}}(m_l, 0) + \frac{1}{9} C_{ii}^{\text{conn}}(m_s, 0) + \frac{4}{9} C_{ii}^{\text{conn}}(m_c, 0) + C_{ii,0}^{\text{disc}} \right]_0
\end{aligned} \tag{4.38}$$

For now, we work in the isospin-symmetric limit $m_u = m_d$. We will discuss isospin-breaking effects in Sec. 4.3.1.6. Analogously to what we saw in Sec. 3.5, the first three terms correspond to quark-connected diagrams, whereas the last term is a quark-disconnected contribution which mixes different flavours. From Eq. (4.34) we obtain the spatial component of the vacuum polarization tensor projected to zero spatial momentum from a modified Fourier transform,

$$\Pi_{ii}(Q_0, \vec{Q} = \vec{0}) = \int \mathrm{d}^4 x \; e^{iQ_0 t} \langle J_i(t, \vec{x}) J_i(0) \rangle \; , \tag{4.39}$$

which via Eq. (4.33) is related to the vacuum polarization function,

$$\Pi_{ii}(Q_0, \vec{Q} = \vec{0}) = -Q_0^2 \; \Pi(Q_0^2) \; . \tag{4.40}$$

Hence we can obtain the vacuum polarization in the Euclidean region, the quantity of interest here, via

$$\Pi(Q_0^2) = \frac{1}{Q_0^2} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{-iQ_0 t} C_{ii}(t) \; . \tag{4.41}$$

For later discussions let us briefly mention that the zero-momentum two-point electromagnetic current correlator is related to the R-ratio, which we introduced in the previous section, via a Laplace transform. This can be seen by inverting the Fourier transform in Eq. (4.39)

$$C_{ii}(t) = -\int \frac{\mathrm{d}Q_0}{2\pi} \; \Pi_{ii}(Q_0, \vec{Q} = \vec{0}) \; e^{iQ_0 t} = \int \frac{\mathrm{d}Q_0}{2\pi} \; Q_0^2 \; \Pi(Q_0^2) \; e^{iQ_0 t} \tag{4.42}$$

and subsequently substituting the dispersion relation (4.30) for $\Pi(Q_0^2)$

$$C(t) = \frac{1}{24\pi^2} \int_0^{\infty} \mathrm{d}s \; \sqrt{s} \; R(s) \; e^{-\sqrt{s}|t|} \tag{4.43}$$

for $t \neq 0$. In finite volume (on a 4-dimensional torus of size L^4) $O(4)$ invariance is reduced to $H(4)$, the hypercubic group. Furthermore, the Ward identities still hold: $q_\mu \Pi_{\mu\nu}(q) = 0$[8]. However, since in FV the vacuum polarization tensor is not fully transverse (this is of course analogous to the photon obtaining a mass at finite temperature),

$$\Pi_{\mu\nu}(0) \sim L^4 e^{-2\pi m_\pi L} \neq 0 \; , \tag{4.44}$$

$\Pi_{\mu\nu}(Q = 0)$ in finite volume contaminates $\Pi(Q^2) \sim \Pi_{\mu\nu}(Q)/Q^2$ for $Q^2 \to 0$ with very large finite volume effects. To remove these large finite volume effects we subtract $\Pi_{ii}(0)$,

$$\Pi(Q_0^2) = -\frac{1}{Q_0^2} \left[\Pi_{ii}(Q^2) - \Pi_{ii}(0) \right] \; . \tag{4.45}$$

[8] However, once space-time is discretized and depending on the choice of action, q in the Ward identities is replaced by $\hat{q}_\mu \Pi_{\mu\nu}(q) = 0$ with $\hat{q} = 2/a \sin(Qa)/2$ [9]. This observation will become important later.

To deal with ultraviolet divergences we also need to subtract $\Pi(0)$. We Taylor expand Eq. (4.41) for small Q^2,

$$\Pi(Q_0^2) = \frac{1}{Q_0^2}\int_{-\infty}^{\infty} \mathrm{d}t\; C_{ii}(t) - \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}t\; t^2 C_{ii}(t) + \dots\,. \tag{4.46}$$

The first term in Eq. (4.46) corresponds to the large FV effect, whereas we remove the second term to renormalize in the on-shell scheme. In our lattice calculation, to improve statistics we further impose cubic symmetry by averaging over the spatial components of the two-point correlator, $C(t) \equiv 1/3\sum_{i=1}^{3} C_{ii}(t)$ and we obtain

$$\hat{\Pi}(Q^2) = \frac{1}{3}\sum_{i=1}^{3}\int_{-\infty}^{\infty} \mathrm{d}t\; \left[\frac{e^{iQt}-1}{Q^2} + \frac{t^2}{2}\right] C_{ii}(t)\,. \tag{4.47}$$

$C(t)$ is an even function in t and the expression is real. In discrete space-time it finally takes the form [9]

$$\hat{\Pi}(Q^2) \equiv \Pi(Q^2) - \Pi(0) = 2a\sum_t \left[\frac{\cos(Qt)-1}{Q^2} + \frac{t^2}{2}\right] \mathrm{Re}\, C(t) \tag{4.48}$$

which formally is defined $\forall Q \in \mathbb{R}$. Hence, the HVP can be obtained from a zero-momentum-projected correlator multiplied by a Q^2-dependent function. In the following we will refer to

$$k_0^{\hat{\Pi}}(Q,t) = \left[\frac{\cos(Qt)-1}{Q^2} + \frac{t^2}{2}\right] \tag{4.49}$$

as the "kernel function" (the subscript "0" will become clear in the following section). In principle, since we have analytically continued $\hat{\Pi}(Q^2)$, we can inject any Q^2 we wish, but on the lattice we are of course limited at high Q^2, i.e. for short distance quantities, by the lattice spacing a which provides a momentum cutoff π/a – in particular, we have access only to $Q^2 \ll (\pi/a)^2$ to avoid large cutoff effects. At low Q^2, or large distances, finite-volume and taste-breaking effects are large. We will discuss these issues in the following sections.

4.3.1.4 Definition of observables

In order to avoid mixing between cutoff effects at different scales and to address these short- and long-distance challenges systematically, we split the Euclidean running of α into various contributions. In this section we define the different quantities we compute.

Split in observables Due to the subtraction of $\Pi(0)$ the HVP mixes for large Q^2 short- and long-distance scales. Hence, during a long time while working on this project we decided to focus on the Adler function, see Eq. (4.19). While the Adler function will not enter our final results, we studied it in depth and used it to develop our analysis strategy. In particular, we performed a detailed study of discretization effects based on which we designed an improvement procedure for our final analysis. We will present this study in Sec. 4.3.2.1. For later convenience, let us hence also introduce the definition of the Adler function, which is essentially the derivative of the vacuum polarization function [238],

$$D(Q^2) \equiv 12\pi^2 Q^2 \frac{\mathrm{d}\hat{\Pi}(Q^2)}{\mathrm{d}Q^2} = 24\pi^2 a Q^2 \sum_t \frac{\mathrm{d}k^{\hat{\Pi}}(t,Q^2)}{\mathrm{d}Q^2} C(t) = 2a\sum_t k^D(t,Q^2) C(t)\,. \tag{4.50}$$

The advantage is that, for massless quarks, Q^2 alone determines whether the Adler function is a short- or long-distance quantity (it is a single-scale quantity). This again allows a separation into contributions which have very different systematic uncertainties. However, since our final aim is to calculate the shift of the electromagnetic coupling up to some reasonable choice of Q^2, this would require integrating the Adler function, since it is the derivative of the HVP. This would introduce a few difficulties, namely

- Performing the analysis for various (many) different values of Q^2, in particular in the low-energy region, where the slope of the Adler function is large.
- Finding an appropriate parametrization to obtain a smooth function (e.g. a Padé approximation).
- Finding a way to assess the systematic uncertainty related to this parametrization and the fact that we have only access to a finite number of Q^2 values.

Hence, for our final analysis we decided to calculate finite differences of the HVP, which we call the discrete Adler function

$$\Delta\Pi(Q^2) \equiv \Pi(Q^2) - \Pi(Q^2/2)\ . \tag{4.51}$$

The HVP is obtained from a telescopic sum

$$\hat{\Pi}(Q_n^2) = \Pi(Q_n^2) - \Pi(0) = \sum_{i=1}^{n-1} \Delta\Pi(2^i Q_0^2) + \hat{\Pi}(Q_0^2) \tag{4.52}$$

and we choose $Q_0^2 = 1\ \mathrm{GeV}^2$. For the light and the disconnected contribution we further split the low-energy contribution $\hat{\Pi}(Q_0^2 = 1\ \mathrm{GeV}^2)$ into a short-distance and a long distance window,

$$\begin{aligned}\hat{\Pi}(Q_0^2) = \hat{\Pi}_{00-04}(Q_0^2) + \hat{\Pi}_{04-\infty}(Q_0^2) = &\int \mathrm{d}t\ w(t; 0, t_2 = 0.4\ \mathrm{fm})\ C(t)\ k(t, Q^2) + \\ &\int \mathrm{d}t\ w(t; t_1 = 0.4\ \mathrm{fm}, \infty)\ C(t)\ k(t, Q^2)\end{aligned} \tag{4.53}$$

where we convoluted the integrand with a smooth, so-called window function [10]

$$w(t; t_1, t_2) \equiv \Theta(t; t_1, \Delta) - \Theta(t; t_2, \Delta) \quad \text{with} \quad \Theta(t; t', \Delta) \equiv \tfrac{1}{2} + \tfrac{1}{2}\tanh[(t - t')/\Delta] \tag{4.54}$$

Δ is the width of the smoothed step function and we employ $\Delta = 0.15$ fm, which is appropriate for the range of lattice spacings used in this thesis. This smoothing is necessary to avoid jumps in the different contributions that might arise if one were to separate them at precise values of t. In that case the values of a given contribution can depend quite strongly on the exact separation time because the given interval may include, or not, a particular correlator value.

As we will show in the following sections, at short distances the HVP receives logarithmically-enhanced cutoff effects whereas at long distances, the spectrum of the two-point correlation function gets distorted by taste-breaking effects. A mixing of these effects would jeopardize a controlled continuum extrapolation and the split in short- and long-distance quantities is hence necessary. This split into long- and short-distance quantities can be observed in Fig. 4.5 where we plot the integrand of the time-momentum-representation (TMR) integral as a function of Euclidean time t for our various observables, normalized by the total value of the integral.

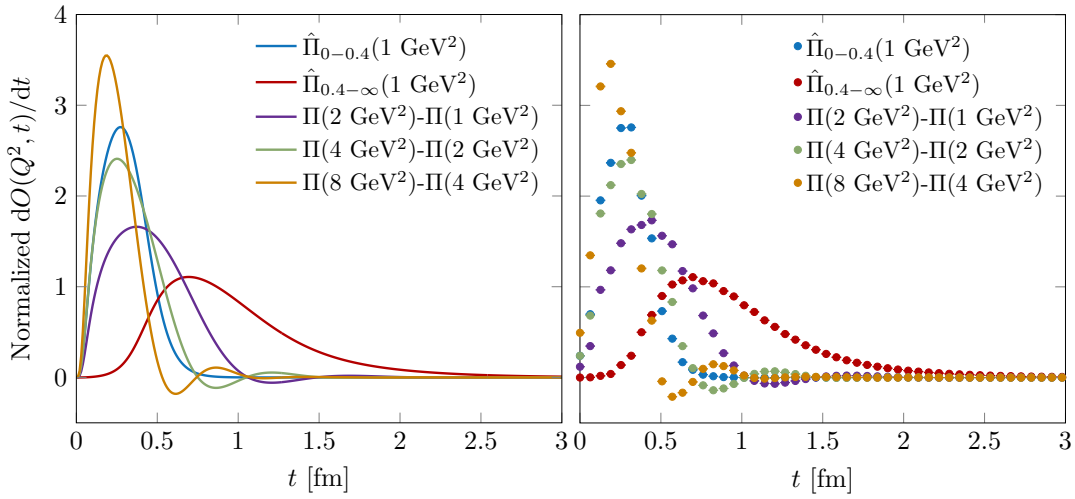


FIGURE 4.5: Integrands of the time-momentum-representation (TMR) integral (4.59) for the different observables that we calculate as a function of Euclidean time t. All integrands are normalized by the total value of the integral. The lines in the left plot correspond to the integrand using a correlator $C(t)$ that was obtained using R-ratio data from Ref. [5]. The data points with error bars in the right plot are obtained using lattice data at the physical pion mass for the light connected contribution to the conserved current correlator, at $a \approx 0.064$ fm.

Clearly, $\Delta\Pi(4\ \mathrm{GeV}^2)$, $\Delta\Pi(8\ \mathrm{GeV}^2)$ and $\hat{\Pi}_{00-04}(1\ \mathrm{GeV}^2)$ are strongly peaked at short distances, while the window function cuts off the short distance part in $\hat{\Pi}_{0.4-\infty}(1\ \mathrm{GeV}^2)$. This quantity extends to large values of t. $\Delta\Pi(2\ \mathrm{GeV}^2)$ is also peaked at small Euclidean times, but extends a bit further to $t \approx 1$ fm. The upshot is that the behaviour of our observables is enhanced at specific scales which allows for an optimal treatment of the continuum and infinite-volume limits.

Kernels As we saw in Sec. 3.2.1, the momentum that appears in a bosonic propagator on the lattice is related to the continuum-limit one via $\hat{Q} = 2/a \sin\left(aQ/2\right)$. It is hence equally well justified to replace the momentum in Eq.(4.49) by this lattice momentum. Note however that the Q inside the cosine comes from the Fourier transform and should therefore not be replaced by $\hat{Q}$. Notice also that the Ward identities read $\hat{Q}_\mu \Pi_{\mu\nu} = 0$ on the lattice [9]. Hence, a different definition of the HVP on the lattice is given by

$$\hat{\Pi}(Q^2) = 2\int_0^\infty \mathrm{d}t\ k_1(t,Q)\ C(t)\ , \qquad k_1(t,Q) = \frac{\cos(Qt)-1}{\hat{Q}^2} + \frac{t^2}{2}\ . \tag{4.55}$$

There are, in fact, several well-motivated choices for the kernel which reduce to $k(t,Q)$ of (4.49) in the continuum limit. We investigated two additional choices for the kernel functions:

$$k_1(t,Q) = \frac{\cos(Qt)-1}{\hat{Q}^2} + \frac{t^2}{2}, \qquad k_2(t,Q) = \frac{\cos(Qt)-1}{\hat{Q}^2} + \frac{(Qt)^2}{2\hat{Q}^2} \tag{4.56}$$

These kernel functions change the asymptotic behaviour of the observables. Firstly, at small Q, the two kernel functions behave as

$$k_1(t,\hat{Q}) \xrightarrow{Q\to 0} \frac{1}{\hat{Q}^2}\left(1 - \frac{(Qt)^2}{2} + \mathcal{O}\left(t^4\right) - 1\right) + \frac{t^2}{2} \tag{4.57}$$

$$k_2(t,\hat{Q}) \xrightarrow{Q\to 0} \frac{1}{\hat{Q}^2}\left(1 - \frac{(Qt)^2}{2} + \mathcal{O}\left(t^4\right) - 1\right) + \frac{(Qt)^2}{2\hat{Q}^2}\ , \tag{4.58}$$

i.e. when employing $k_1(t,Q)$, the $t^2/2$ term vanishes only up to discretization errors. This was our original motivation for also employing k_2. However, as t^2 grows in Eq. (4.58), the discretization errors grow and this kernel function introduces large additional discretization errors at long distances. In the following we will therefore only consider the kernel function k_0 (see Eq. (4.49)) and k_1. A crucial observation to make is that, as we discuss in the following, this kernel function changes the leading logarithmic $\mathcal{O}(a^2)$ coefficient in the asymptotic behaviour of our short-distance quantities, allowing for an additional assessment of systematic effects. Finally, any one of our observables can be denoted by

$$O_i^J(Q^2,Q_R^2;t_1,t_2) = 2\int_0^\infty \mathrm{d}t\; C^J(t)\; k_i(t,Q^2,Q_R^2)\; w(t;t_1,t_2)\;, \tag{4.59}$$

where J labels the staggered current, i labels the kernel function, w the time window and O is the observable we consider, i.e. either the HVP $\hat{\Pi}$, the discrete Adler function $\Delta\Pi$ or the Adler function D. In the following, if we do not specify the values for t_1 and t_2 explicitly, this indicates that we are considering the full time interval $t_1 = 0$, $t_2 = \infty$, without any window function.

4.3.1.5 Renormalization of non-conserved currents

β	a[fm]	Z^2_{IV}	Z^2_{IVtnl}	Z^2_{XX}	Z^2_{Xtnl}
3.7500	0.1191	0.9075	0.9901	1.3315	1.1426
		0.9039	0.9771	1.3319	1.1315
		0.9039	0.9817	1.3231	1.1308
		0.9013	0.9840	1.3410	1.1410
3.7753	0.1116	0.9133	0.9900	1.2957	1.1286
		0.9092	0.9844	1.2968	1.1228
		0.9107	0.9851	1.2936	1.1227
		0.9151	0.9927	1.3000	1.1313
3.8400	0.0952	0.9304	0.9995	1.2396	1.1119
		0.9300	0.9984	1.2441	1.1143
		0.9300	0.9984	1.2441	1.1143
		0.9329	1.0014	1.2384	1.1130
		0.9318	0.9971	1.2411	1.1113
3.9200	0.0787	0.9467	1.0056	1.1942	1.0964
		0.9450	1.0071	1.1900	1.0977
		0.9489	1.0087	1.1874	1.0972
		0.9467	1.0056	1.1942	1.0964
		0.9466	1.0070	1.1895	1.0961
		0.9450	1.0071	1.1900	1.0977
4.0126	0.0640	0.9553	1.0082	1.1516	1.0812
		0.9580	1.0104	1.1494	1.0808
		0.9553	1.0082	1.1516	1.0812
		0.9580	1.0104	1.1494	1.0808
		0.9571	1.0140	1.1521	1.0861
		0.9591	1.0095	1.1522	1.0824

Table 4.6: Approximate renormalization factors of the four staggered currents IV, IVtnl, XX and Xtnl (see Tabs. 4.4 and 4.5 for the definition of the currents) for the list of our 4stout ensembles used in this work, used in the analysis of the charm correlator.

Non-conserved staggered, vector currents require an additional finite renormalization:

$$J_\mu^{J,\mathrm{ren}} = Z_J J_\mu^{J,\mathrm{bare}}\;, \tag{4.60}$$

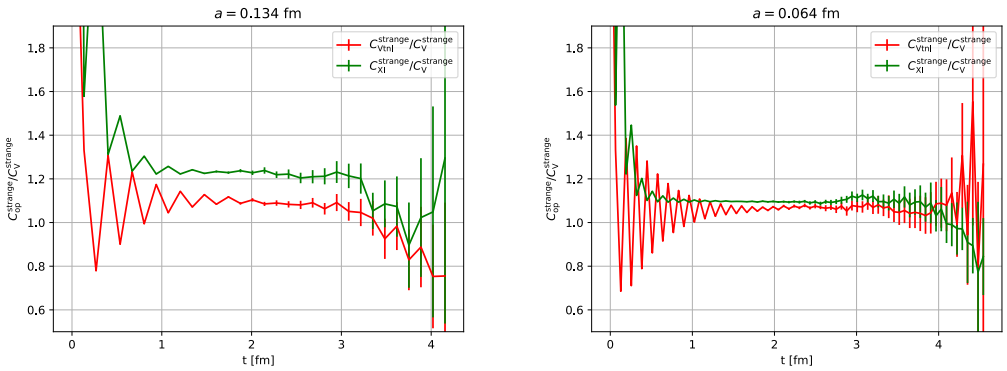


FIGURE 4.6: Ratio of strange quark current-current correlators determined with non-conserved currents and the conserved current V. We present the results for the coarsest and the finest lattice spacings and two non-convserved currents XI and Vtnl. For other currents and lattice spacings the plots look similar.

where J_μ^J denotes one of the vector-current discretizations listed in Tabs. 4.4 and 4.5. To determine the Z_J factors we use the strange current-current correlator which has smaller taste-breaking effects than the light one. In Fig. 4.6 we present the results for the ratio of non-conserved to conserved strange quark current-current correlators for the coarsest and the finest lattice spacings used in the current analysis and for two typical operators. The ratio of the two correlators should be constant up to cutoff effects and give the renormalization factor: $C_J^{\text{strange}}/C_{\text{V}}^{\text{strange}} = 1/Z_J^2$. Note that the t-dependence of this ratio indeed exhibits a plateau, at times $t \gtrsim 1$ fm, while at very large times it becomes very noisy. In practice we define the renormalization constant via the strange connected HVP contribution to the anomalous magnetic moment of the muon in the window $[1.0, 2.0]$ fm (see [10] for definition). Thus we obtain:

$$Z_J^2 = a_{\mu,\text{strange}}^{\text{V},\ [1:2]\,\text{fm}} / a_{\mu,\text{strange}}^{J,\ [1:2]\,\text{fm}} \ . \tag{4.61}$$

4.3.1.6 QED and strong isospin breaking corrections

As discussed in Sec. 3.7, working in the isospin-symmetric limit is no longer sufficient if one aims at a precision < 1%. In this section we consider how QED and strong-isospin breaking (SIB) effects are included in our calculations. We collectively call these two corrections isospin-breaking (IB) effect. The various strategies to compute them which have been explored can be summarized as follows [156, 239]

- **Operator insertion method.** [240, 241] $\mathcal{O}(m_u - m_d)$ and $\mathcal{O}(\alpha_{\text{em}})$ operators are inserted and evaluated on isospin-symmetric QCD configurations. No new configurations need to be generated and one directly obtains the desired order in α_{em}. However, the observables are difficult to compute, in particular for quark-disconnected diagrams.
- **Direct method.** [242–247] masses and QED are included directly in the simulation. The observables are much simpler than in the operator insertion method and include quark-disconnected diagrams. The disadvantage is that new configurations including QED need to be generated.
- **Any combination of the first two approaches.**

The strategy we follow in this study is the third approach, i.e. a combination of simulations including QED and $m_u \neq m_d$-effects and an expansion in leading IB breaking effects (see below). We use the same formalism as in [7] and briefly summarize the procedure. Details can be found in the supplementary material of [7] and in [12, 248]. We include IB effects to first order in the IB parameters $\delta m = m_u - m_d$ and $\alpha_{\rm em}$ by expanding the QCD+QED path integral in these parameters and measuring these observables on isospin symmetric configurations, where we also distinguish between sea and valence charges. In such a way, one obtains a fixed order in the expansion. Further, one can factorize the coefficients and obtain a large numerical signal. Our QED+QCD action is given by

$$S[U, A, \bar{\psi}, \psi] = S_g[U; g] + S_\gamma[A] + \sum_f \bar{\psi}_f D\left[U, A; e, q_f, m_f\right] \psi_f \ . \tag{4.62}$$

Performing the integral over the fermionic degrees of freedom in the path integral analytically, we can write the QED+QCD partition function for 1+1+1+1 staggered fermions as

$$Z = \int [dU] e^{-S_g[U]} \int [dA] e^{-S_\gamma[A]} \prod_f \det M_f[V_U e^{ieq_f A}, m_f]^{1/4} \ , \tag{4.63}$$

with quark flavours $f = \{u, d, s, c\}$ and their corresponding electric charges $e\, q_f = e\left\{\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{2}{3}\right\}$. In the following, the product of the determinants of the fermion matrices will be denoted by

$$\text{dets}[U, A; \{m_f\}, \{q_f\}, e] \equiv \prod_f \det M_f^{1/4}, \tag{4.64}$$

where the explicit form of the fermionic matrix M_f reads

$$M_f[V_U e^{ieq_f A}, m_f] = \sum_\mu D_\mu[V_U e^{ieq_f A}] + m_f \ , \tag{4.65}$$

i.e. QED is implemented in a non-compact form. Here V_U and A denote the smeared gluon gauge field and the unsmeared photon field, respectively. QED in this calculation is defined in the QED$_L$ scheme [160].

We now define the various derivatives of an observable $O(e, \delta m)$ with respect to the IB parameters e and $\delta m \equiv m_d - m_u$

$$O_0 \equiv O(0,0), \quad O'_m \equiv m_l \frac{\partial O}{\partial \delta m}(0,0), \quad O'_1 \equiv \frac{\partial O}{\partial e}(0,0), \quad O''_2 \equiv \frac{1}{2}\frac{\partial^2 O}{\partial e^2}(0,0) \ . \tag{4.66}$$

We further make the distinction between valence, e_v, and sea, e_s, electric charges, $O\left(e_v, e_s\right)$ which we denote by

$$O''_{20} \equiv \frac{1}{2}\frac{\partial^2 O}{\partial e_v^2}(0,0), \quad O''_{11} \equiv \frac{\partial^2 O}{\partial e_v \partial e_s}(0,0), \quad O''_{02} \equiv \frac{1}{2}\frac{\partial^2 O}{\partial e_s^2}(0,0) \ . \tag{4.67}$$

Finally, the IB expansion of any observable to $\mathcal{O}(e_v^2, e_s^2, \delta m)$ can be written as

$$\langle O \rangle \simeq \langle O_0 \rangle_0 + \tfrac{\delta m}{m_l} \langle O \rangle'_m + e_v^2 \langle O \rangle''_{20} + e_v e_s \langle O \rangle''_{11} + e_s^2 \langle O \rangle''_{02} \ . \tag{4.68}$$

We will detail below which expressions the various terms take. Note that the derivative of the fermion determinant evaluated on isospin-symmetric configurations vanishes

$$\left\langle \frac{\mathrm{dets}_1'}{\mathrm{dets}_0} \right\rangle_0 = 0 \,.$$

Similarly,

$$\frac{\partial \mathrm{dets}}{\partial \delta m} = 0$$

because the derivative for M_u differs from the one for M_d only by a minus sign (and we take the sum because of the product rule). In order to obtain the explicit expressions for the terms in Eq. (4.68), let us expand the expectation value of an observable O to $\mathcal{O}(\alpha_{\mathrm{em}}, \delta m)$. Explicitly,

$$\begin{aligned}\langle O \rangle &= \frac{\int [dU] \int [dA] e^{-S_g} e^{-S_\gamma}\, \mathrm{dets}[U, A; \{m_f\}, \{q_f\}, e]\; O}{\int [dU] \int [dA] e^{-S_g} e^{-S_\gamma}\, \mathrm{dets}[U, A; \{m_f\}, \{q_f\}, e]} \\ &\simeq \frac{\int [dU] \int [dA] e^{-S_g} e^{-S_\gamma}\, \mathrm{dets}_0 \left(1 + e_s \frac{\mathrm{dets}_1'}{\mathrm{dets}_0} + e_s^2 \frac{\mathrm{dets}_2''}{\mathrm{dets}_0}\right) \left(O_0 + \frac{\delta m}{m_l} O_m' + e_v O_1' + e_v^2 O_2''\right)}{\int [dU] \int [dA] e^{-S_g} e^{-S_\gamma}\, \mathrm{dets}_0 \left(1 + e_s \frac{\mathrm{dets}_1'}{\mathrm{dets}_0} + e_s^2 \frac{\mathrm{dets}_2''}{\mathrm{dets}_0}\right)}\end{aligned} \tag{4.69}$$

dets_0 is the product of the fermionic determinant in the IB symmetric limit. Using

$$\frac{\mathrm{ddet}(A)}{\mathrm{d}\alpha} = \mathrm{Tr}\left(\det(A) A^{-1} \frac{\mathrm{d}A}{\mathrm{d}\alpha}\right) \tag{4.70}$$

we find for the derivatives of the fermion determinants

$$\frac{\mathrm{dets}_1'}{\mathrm{dets}_0} = \sum_f \frac{q_f}{4} \mathrm{Tr}\left\{M_f^{-1} D[iAV_U]\right\} \tag{4.71}$$

$$\begin{aligned}\frac{\mathrm{dets}_2''}{\mathrm{dets}_0} = \frac{1}{2}\Bigg[&\left(\frac{\mathrm{dets}_1'}{\mathrm{dets}_0}\right)^2 - \sum_f \frac{q_f^2}{4} \mathrm{Tr}\left\{M_f^{-1} D[iAV_U] M_f^{-1} D[iAV_U]\right\} \\ &- \sum_f \frac{q_f^2}{4} \mathrm{Tr}\left\{M_f^{-1} D[A^2 V_U]\right\}\Bigg]\end{aligned} \tag{4.72}$$

Expanding the denominator in expression (4.69) as $1/(1+\varepsilon) = 1 - \varepsilon + \mathcal{O}(\varepsilon^2)$ we find after some algebra[9]

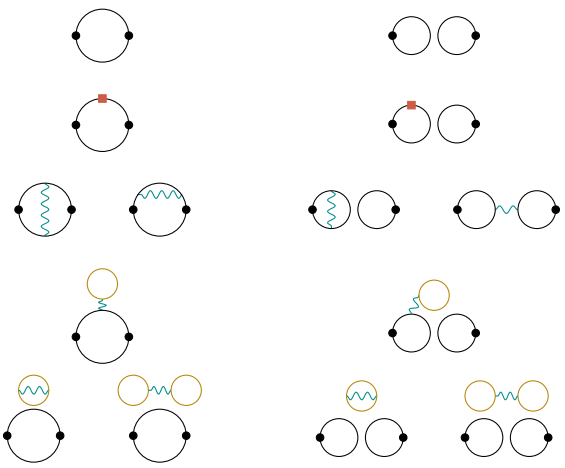
$\langle O \rangle = \langle O_0 \rangle_0$

$+\frac{\delta_m}{m_l} \langle O_m' \rangle_0$

$+e_v^2 \langle O_2'' \rangle_0$

$+e_v e_s \left\langle O_1' \frac{\mathrm{dets}_1'}{\mathrm{dets}_0} \right\rangle_0$

$+e_s^2 \left\langle \left[O_0 - \langle O_0 \rangle_0\right] \frac{\mathrm{dets}_2''}{\mathrm{dets}_0} \right\rangle_0$



[9] A very special thanks to Letizia for this great illustration which I adapted from her proceedings [248]

Note that $\langle \ldots \rangle_0$ denotes evaluation of the operator on isospin-symmetric configurations. In the following we will often further make explicit whether the path integral is evaluated on gluon $\langle \ldots \rangle_U$ or photon $\langle \ldots \rangle_A$ configurations. Diagrammatically, the red square denotes the insertion of the strong IB operator, while blue lines are photons and yellow circles sea-quark loops. As usual, black lines and dots are valence quarks and current insertions, respectively.

We include isospin corrections from the strange and light quarks, and neglect IB corrections coming from the charm because the charm contribution has no leading δm corrections and the electromagnetic ones are small corrections to an already small contribution to the observables studied in this thesis.

4.3.1.7 Calculation of various isospin-breaking components

Generally speaking, derivatives with respect to the valence charge, e_v, are evaluated using finite differences of the observables computed in the presence of a quenched quantized photon field. Our approach has the advantage that we do not have to construct and code the diagrams that would be present in the operator insertion method. Indeed, that approach would give rise to 4-point functions, i.e. 2-point functions with two quark electromagnetic currents. In order to facilitate the computation of the isospin breaking derivatives of hadron masses, the masses are extracted using an effective mass functional of the hadron propagator $\mathcal{M}[H]$ which is given in a closed analytic form [249]. This permits us to calculate derivatives $\partial\mathcal{M}/\partial H$ analytically. For more details on the specific procedure for the various IB derivatives see [7, 12].

The list of ensembles used to evaluate SIB and QED effects was presented in Sec. 4.3.1. We will discuss finite volume effects on QED derivatives in Sec. 4.3.2.3.

Strong IB corrections As we saw in the expansion above, the strong isospin derivative of an observable can be calculated by inserting a mass derivative operator into the path integral. To reduce the noise in the SIB contribution to the light connected propagator, it is evaluated at larger than physical m_l, i.e. at $[C^{\text{conn}}(\kappa m_l, 0)]'_m$ with $\kappa = 3, 5, 7, 9, 11$ which subsequently lets us extrapolate to $\kappa = 1$. Both the SIB derivative of the disconnected current propagator and of a hadron mass M are evaluated using a forward finite difference quotient:

$$M'_m \approx \frac{m_l}{\delta m}\left(\mathcal{M}[\langle H_{\delta m}\rangle_0] - \mathcal{M}[\langle H_0\rangle_0]\right) \tag{4.73}$$

with $\delta m = 2m_l \frac{1-r}{1+r}$ and $r = \frac{m_u}{m_d} = 0.485$ [250] and we use $I(m_l, 0)$ and $I(0.9\, m_l, 0)$ for $\partial C_0^{\text{disc}}/\partial m_l$.

Valence-valence contribution Similarly, the second valence-valence derivative O''_2 of an observable $O(e, \delta m)$ can be computed using a central finite difference quotient

$$O''_2 \approx \frac{1}{2e_v^2}\left[(O_+ + O_-) - 2O_0\right] ,$$

where $O_\pm \equiv O(\pm e_*, 0)$ and $e_* = \sqrt{4\pi\alpha_{\text{em}}}$ is the physical value of the electric coupling. Since the light contribution to C''_{20} is noisy, to estimate $C''_{l,20}$ we proceed along the same lines as for the calculation of the strong IB derivative of the connected light current propagator, i.e. we evaluate $C^{\text{conn}}(\kappa m_l, 0)$ and $C^{\text{conn}}(\kappa m_l, \pm\frac{1}{3}e_*)$ at larger m_l for $\kappa = 3, 5, 7, 9, 11$ and chirally extrapolate to

$\kappa = 1$. This is not necessary for the strange contribution which is estimated using finite differences in $\pm\frac{1}{3}e_*$ at m_s.

To evaluate $[C^{\text{disc}}]''_{20}$ to $\mathcal{O}(e_v^2)$ it actually suffices to calculate a linear combination of $I(m_l, 0)$, $I(m_l, \pm\frac{1}{3}e_*)$, $I(m_s, 0)$, and $I(m_s, \pm\frac{1}{3}e_*)$ [7].

M''_{20} is estimated using finite differences for the hadron propagators measured at $\delta m = 0$ and $e_v = \{+e_*, -e_*, 0\}$, respectively. In addition, we remove finite volume effects in M''_{20} by evaluating them in the QED$_L$ scheme (see Sec. 4.3.2.3).

Sea-valence contribution The sea-valence contribution is given by

$$\langle O \rangle''_{11} = \left\langle \left\langle O'_1 \frac{\text{dets}'_1}{\text{dets}_0} \right\rangle_A \right\rangle_U . \tag{4.74}$$

For the evaluation of $\text{dets}'_1/\text{dets}_0$ one photon field per gluon field is generated and on each of those $\sim 10^4$ random sources. The valence part is, as we summarized above, again estimated using finite differences, $O'_1 \approx \frac{1}{2e_v}(O_+ - O_-)$.

Sea-sea contribution The sea-sea contribution is given by

$$\langle O \rangle''_{02} = \left\langle [O_0 - \langle O_0 \rangle_U] \left\langle \frac{\text{dets}''_2}{\text{dets}_0} \right\rangle_A \right\rangle_U . \tag{4.75}$$

For the estimation of $\text{dets}''_2/\text{dets}_0$ entering Eq. (4.75), approximately 2000 photon fields A are generated for each gluon field configuration with 12 random sources on each photon field. The generation of photon fields from the (free) photon action e^{-S_γ} is particularly cheap, since it simply requires sampling from a Gaussian distribution.

4.3.1.8 Blinding

The lattice results for the HVP function in this thesis were blinded in a straightforward fashion. My thesis advisor rescaled all correlation functions, including those associated with QED and SIB corrections, by an overall scaling factor between −5 and 5 percent. In late September this rescaling factor (1.037375878694366) was revealed to me so that I could include a comparison of my results with those obtained by another lattice group or using R-ratio results.

As these results have not yet been published, they will have to be re-blinded before being re-analyzed after at least one additional lattice spacing is added to the computation. How this will be done has not yet been fully decided.

4.3.2 Analysis strategy

While formally Eq. (4.48) can be used to define $\hat{\Pi}(Q^2)$ for any $Q \in \mathbb{R}$, we are limited at large Q by the momentum cutoff on the lattice, and at small Q the observable will feel the finite size of the lattice. At those large distances, finite-volume and, since we are using staggered quarks,

taste-breaking effects will play a role. In this section we discuss the improvements we perform on our data to control the extrapolation to the continuum. We also discuss our global fit procedure with the inclusion of leading-order isospin breaking corrections and our procedure to estimate statistical and systematic uncertainties on the quantities we calculate. Finally, we discuss how we combine the various observables in a consistent manner, preserving correlations.

4.3.2.1 Challenges at short distances: discretization errors

At large values of Q^2, discretization effects proportional to $(aQ)^n$, $n \geq 2$, become important, eventually spoiling the continuum extrapolation of the HVP function. In addition, all of our short-distance observables receive logarithmically-enhanced cutoff effects which render the continuum extrapolation more difficult. In order to perform a reliable continuum extrapolation, we propose a number of improvement procedures based on lattice perturbation theory. To illustrate the improvement programme, we present an exemplary study based on the Adler function at $Q^2 = 5\ \mathrm{GeV}^2$.

Logarithmic cutoff effects Naively, from Symanzik effective theory (cf. Sec. 3.6.3) for on-shell staggered quantities as for instance masses, one expects power corrections in the lattice spacing, modified by powers of α_s,

$$M(a) = M(0)\left\{1 + a^2 \sum_{n=0}^{\infty} c_n \alpha_s^n(1/a) + \mathcal{O}\left(a^4\right)\right\} . \tag{4.76}$$

For a detailed study see also [251]. However, as recently shown in [11, 232], the contribution of the light valence quarks of mass m_l to the Adler function receives logarithmically-enhanced $\mathcal{O}(a^2)$ lattice artefacts, even at leading order (LO) in lattice perturbation theory. For a generic logarithmically UV divergent quantity O, such as the Adler function, one obtains

$$O\left(Q^2, a\right) = O\left(Q^2\right)\left\{1 + \Gamma(aQ)^2 \ln(aQ)^2 + \mathcal{O}\left[(aQ)^2, (am_l)^2\right]\right\} , \tag{4.77}$$

where $O\left(Q^2\right)$ is the value of O in the continuum and Γ is a constant. As we will show, this is also the case for the short-distance window of the HVP and discrete Adler functions. These logarithmically-enhanced cutoff effects arise from small separations between the two electromagnetic currents: $D(Q^2)$, $\hat{\Pi}(Q^2)$ and $\Delta\Pi(Q^2, Q_R^2)$ are not on-shell quantities. Note also that perturbative corrections to the logarithmically-enhanced term, of the form $\alpha_s^n(1/a)(aQ)^2 \ln(aQ)^2$, are of order $\alpha_s^{n-1}(1/a)(aQ)^2$ because $\alpha_s \sim -1/\ln(a\Lambda_{\mathrm{QCD}})$ and are therefore no longer logarithmically enhanced. For a more detailed discussion see App. D.

To understand heuristically why these logarithmically enhanced corrections arise, we expand the correlator and the kernel function in powers of a^2. By dimensional analysis, $C(t)$ scales like $1/t^3$ while all our kernel functions at short distances scale like t^4,

$$C(t) \underset{a \ll t}{\overset{t \to 0}{\sim}} \frac{1}{t^3}\left(\alpha_0 + \alpha_1 \frac{a^2}{t^2} + \alpha_2 \frac{a^4}{t^4} + \cdots\right) \tag{4.78}$$

$$O\left(Q^2, a\right)\Big|_{a^2} \sim \int_a \mathrm{d}t\ Q^2 t^4 C(t) \sim a^2 Q^2 \int_a \frac{\mathrm{d}t}{t} \sim \frac{1}{2} a^2 Q^2 \ln a^2 . \tag{4.79}$$

Current ID	IV	V	X	XI	XV	XX
$\Gamma^J_{0,LO}$	$-1/48$	$-1/120$	$1/240$	$7/240$	$5/192$	$1/24$
$\Gamma^J_{1,LO}$	$1/80$	0	$1/40$	$1/20$	$3/64$	$1/16$

TABLE 4.7: Coefficient of the logarithmically-enhanced discretization error from leading order lattice perturbation theory in the massless limit for all different staggered currents for kernel k_0 (first row) and kernel k_1 (second row). The calculations are detailed in App. B.3. The coefficients obtained for the time non-local operators are identical to the ones for the time local operators which is why we do not make them explicit here. Note however that for the definition of the Adler function in Eq. (4.50) all coefficients have to be multiplied by an overall factor of $12\pi^2$.

Hence the coefficient of the a^2 term is clearly logarithmically divergent; the discretization errors retain some sensitivity to the ultraviolet.

As we will show below, it is crucial to control the behaviour of these logarithmic cutoff effects to ensure a reliable continuum extrapolation. For this purpose, we explicitly calculate the coefficient of the leading $\ln(aQ)^2$ coefficient analytically in lattice perturbation theory in the massless limit for the different discretizations of the staggered vector currents. The calculations are outlined in App. B.3. There are various sources of logarithmically-enhanced cutoff effects. In particular, an expansion in powers of a in LO lattice perturbation theory yields at $\mathcal{O}(a^2)$

$$O_i^J\left(Q^2,t,a\right)\Big|_{a^2} = k_i^O(t,Q,a)\Big|_{a^0} C^J(t)\Big|_{a^2} + k_i^O(t,Q,a)\Big|_{a^2} C^J(t)\Big|_{a^0} \ , \tag{4.80}$$

consequently the coefficient of the logarithmically-enhanced term changes for the different staggered currents (J labels the ID of the current, as given in Tab. 4.4) and the two different kernel functions k_i, $i = 0,1$ that we employ. This allows for a reliable assessment of systematic effects related to the continuum limit. To summarize our findings, the coefficient is, up to a factor $12\pi^2$ coming from the definition of the Adler function, identical for the observables we consider[10],

$$O^J_{i,\mathrm{LO}}\left(Q^2,a\right) \xrightarrow[a\to 0]{m\to 0} O^J_{i,\mathrm{LO}}\left(Q^2,0\right) + \frac{q_f^2}{\pi^2}\Gamma^J_{i,\mathrm{LO}}(aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \mathcal{O}\left((aQ)^2\right) \ . \tag{4.82}$$

The coefficient for the Adler function receives an additional factor of $12\pi^2$, due to our definition in Eq. (4.50). Note that in the above expression we have set the number of colours to $N_c = 3$. In addition, we find the same coefficients in the massive case but with an additional term that can however be absorbed in the $(aQ)^2$ coefficient. Finally, we find that the coefficients are identical for the time local and time non-local currents. In Tab. 4.7 we collect the logarithmic coefficients for the various staggered currents and kernel functions employed in our analysis. Let us also remark that logarithmic cutoff effects are not limited to $\mathcal{O}(a^2)$. In fact, both the the kernel function k_1 and – from a similar heuristic argument as above – the correlator generate a whole tower of logarithmic cutoff effects. For a more detailed discussion see App. B.3. We neglect these higher-order terms here.

[10] Only the expansion parameter changes for the discrete Adler function. Specifically,

$$\Delta\Pi_i^J\left(Q^2,Q_R^2,a\right) \xrightarrow[a\to 0]{m_l\to 0} \Delta\Pi_i^J\left(Q^2,Q_R^2,0\right) + \frac{q_f^2}{\pi^2}\Gamma^J_{i,\mathrm{LO}} a^2\left[Q^2\ln\left(\frac{aQ}{2\pi}\right)^2 - Q_R^2\ln\left(\frac{aQ_R}{2\pi}\right)^2\right] + \mathcal{O}\left(a^2(Q^2-Q_R^2)\right) \ . \tag{4.81}$$

Note also that in order to take the massless limit for the HVP, we need to cut off the TMR integral at some time t since it becomes infrared divergent for $m \to 0$. In this section, we therefore always assume implicitly that we consider only the short-distance window of the HVP.

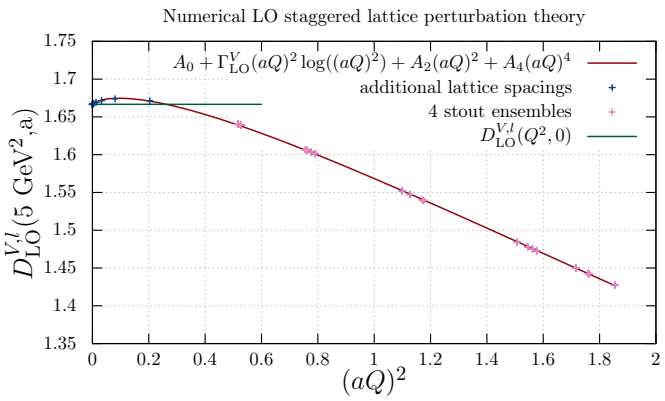


FIGURE 4.7: Light contribution to $D(5\ \text{GeV}^2)$ as a function of $(aQ)^2$ at LO in staggered, lattice perturbation theory. The pink points correspond to the lattice spacings available in our 4-stout ensembles [7], blue points are additional, smaller lattice spacings. The green line is the continuum value known analytically at LO.

Parenthesis: exploratory study of discretization errors of the Adler function When one aims at calculating the running of α up to $\sim 8\ \text{GeV}^2$, challenges at short distances due to large cutoff effects necessarily arise. In order to better understand these cutoff effects, we started with an exploratory study of the Adler function. The reason for focusing on this observable was that, as mentioned earlier, the Adler function is an IR safe single-scale quantity, for $Q^2 \gg \Lambda^2_{\rm QCD}$. Thus, all discretization errors involve only the dimensionless combination aQ which can be dialed to any desired value by tuning Q. Further, we focused on the conserved current V which is the one we mostly use in our final analysis. Based on this understanding, we designed our final analysis. In what follows we will hence focus on the challenges which show up in the connected light quark contribution to $D(Q^2)$ for Euclidean $Q^2 \gg \Lambda^2_{\rm QCD}, m_l^2$.

In Figure 4.7 we plot the light contribution of the Adler function at $Q^2 = 5\ \text{GeV}^2$ at LO in staggered, lattice perturbation theory as a function of the lattice spacing squared. Clearly, a naive extrapolation using a simple linear function in $(aQ)^2$ would completely miss the continuum limit. Even worse, the logarithmic term becomes important for small lattice spacings and the function turns around. Note also that this turnover is shifted towards smaller a for larger values of Q^2. Hence, to ensure a reliable continuum extrapolation even at large Q^2, it is crucial to have an analytic understanding of the asymptotic dependence on the lattice spacing. To simplify the discussion below, let us absorb the additional factors in this section and let us define

$$D\left(Q^2, a\right) = D\left(Q^2\right)\left\{1 + \tilde{\Gamma}(aQ)^2 \ln(aQ)^2 + \mathcal{O}\left[(aQ)^2, (am_l)^2\right]\right\}\ , \tag{4.83}$$

i.e. we have set $\tilde{\Gamma} \equiv 12\ \Gamma/N_c$. For the leading-order coefficient of the conserved current V of the logarithmically-enhanced discretization error in staggered, lattice perturbation theory we hence have (cf. Sec. 4.3.2.1 and App. B.3)

$$\tilde{\Gamma}^V_{\rm LO} = -1/30\ . \tag{4.84}$$

Clearly, these logarithmically enhanced cutoff effects have to be well understood for a reliable continuum extrapolation. In order to tackle these logarithmic cutoff effects we have composed a number of improvement procedures based on lattice perturbation theory:

I. **Removal of discretization effects at $\mathcal{O}(\alpha_s^0)$** One way to tackle large cutoff effects

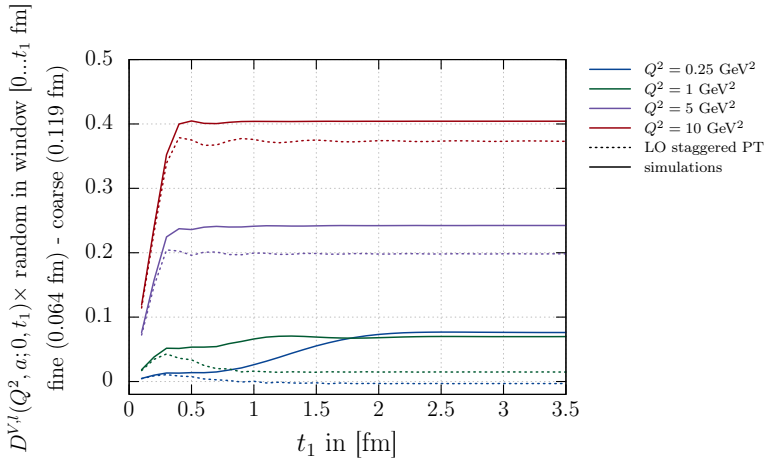


FIGURE 4.8: Connected light quark contribution to $D(Q^2)$ in a time window $[0 \dots t_1]$ fm for different values of Q^2. We plot the difference between a fine ($a = 0.064$ fm) and a coarse ($a = 0.118$ fm) lattice, in comparable volumes. The solid lines are computed using simulation results. The dotted curves are the predictions from leading-order lattice perturbation theory. The kernel used here is $k_0^D(t, Q)$.

consists in removing some of the discretization errors by improving the data using lattice perturbation theory. Hence, we define our LO-improved observables,

$$\tilde{O}(Q^2, a) \equiv O(Q^2, a) + O_{\rm LO}(Q^2, 0) - O_{\rm LO}(Q^2, a) \ , \tag{4.85}$$

where $O_{\rm LO}(Q^2, 0)$ and $O_{\rm LO}(Q^2, a)$ are the observable in the continuum and in staggered, lattice perturbation theory at LO, respectively. This should cure the data from the leading discretization errors at large momenta, where perturbation theory works well. More importantly, it removes the leading, logarithmically-enhanced discretization errors, up to small $\alpha_s(Q^2)$ suppressed terms, as we further discuss in App. D. The LO-corrections in infinite volume and infinite time for the observables we consider can be obtained from the modified Fourier transform of the current correlator (see Eq. (B.24)),

$$O^J_{i,\rm LO}(Q^2, a) = \sum_{t=0}^{\infty} k_i^O(t, Q) C^J_{\rm LO}(t, a) \ , \tag{4.86}$$

where the sum in t can be calculated analytically. We outline the computation and collect the results for the various observables in App. B.2. The results for their continuum counterparts are presented in App. C.

Needless to say, this improvement is only viable at short distances, where perturbation theory is expected to work well. To investigate whether Eq. (B.41) reproduces the discretization errors in our simulations for large Q^2, we depict in Figure 4.8 the cutoff effects to the connected light quark contribution to the Adler function $D_l(Q^2, a)$ as a function of Euclidean time, i.e. the integrand is convoluted by the smooth window function Eq. (4.54)

$$D_i^{V,\rm win}(Q^2, a; 0, t_1) \equiv 2a \sum_{t=0}^{\infty} {\rm d}t \ w(t; 0, t_1) \ C^V(t, a) \ k_i^D(t, Q) \ , \tag{4.87}$$

where $k_i^D(t, Q)$, $i = 0, 1$, determined through (4.50). We plot this window quantity for $k_0^D(Q, t)$ from our 4-stout simulations (solid line), taking the difference of $D^{\rm win}$ obtained on our finest and coarsest lattice. These discretization errors are compared to those obtained

using lattice perturbation theory for the same lattice parameters (dotted curves). For $Q^2 \lesssim 1$ GeV^2, the discretization errors are less than 10% in general but in the lattice simulations they increase at large distances due to taste-breaking effects. Clearly, for times longer than 0.3 fm Eq. (B.41) fails to describe the discretization errors properly. However, as Q^2 increases between 1 and 10 GeV^2 discretization errors become larger and larger but saturate quickly. In fact, here, LO lattice perturbation theory describes the discretization errors to better than 10% for $Q^2 = 10$ GeV^2.

II. **Removal of an additional discretization effect** Let us factorize the expansion of the Adler function as in Eq. (4.77)

$$D\left(Q^2, a\right) = D\left(Q^2\right)\left\{1 + \tilde{\Gamma}_{\mathrm{LO}}(aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \mathcal{O}\left((aQ)^2\right)\right\} , \tag{4.88}$$

with

$$D\left(Q^2\right) = D_{\mathrm{LO}}\left(Q^2\right) + D_{\mathrm{NLO}}\left(Q^2\right)\alpha_s\left(Q^2\right) + \mathcal{O}\left(\alpha_s^2\right) , \tag{4.89}$$

where $D_{\mathrm{LO}}(Q^2)$ and $D_{\mathrm{NLO}}(Q^2)$ are the LO and one-loop Adler function in the continuum, respectively [214]. Since $\tilde{\Gamma}_{\mathrm{LO}}$ and $D_{\mathrm{NLO}}(Q^2)$ are known, by expanding equation (4.88), we find that we can define an additionally subtracted $\bar{D}(Q^2)$ which will have yet formally smaller logarithmically-enhanced discretization errors,

$$\bar{D}\left(Q^2, a\right) = \tilde{D}\left(Q^2, a\right) - (aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 \tilde{\Gamma}_{\mathrm{LO}} D_{\mathrm{NLO}}\left(Q^2\right)\alpha_S\left(Q^2\right) , \tag{4.90}$$

where $\tilde{D}\left(Q^2, a\right)$ is the leading-order improved Adler function defined in Eq. (4.85). After removal of this additional discretization error, logarithmically-enhanced discretization errors begin at $\mathcal{O}\left(\alpha_s^2(Q^2)(aQ)^2\ln(aQ)^2\right)$ and $\mathcal{O}\left((aQ)^4\ln(aQ)^2\right)$ and should be small. All the other terms are regular and begin at $\mathcal{O}\left((aQ)^2\right)$. In Figure 4.9 we plot the continuum extrapolation of $D_l(5\text{ GeV}^2)$. Two types of improvements are shown: one where the leading logarithmically-enhanced cutoff effect is removed and one where we also subtract the additional mixed term in eq. (4.90). As a fit to the unimproved lattice results shows (dark red line), the logarithmic coefficient is close to the one expected from lattice perturbation theory, see Eq. (4.84). Removing the logarithmic term at leading order divides this coefficient by a factor of ~five (violet line), removing the additional discretization effect reduces the logarithmic cutoff effect further: it vanishes within errors (orange line).

III. **Taking the continuum limit using the lattice momentum $\hat{Q}$** As can be observed from Figure 4.7, since the logarithmically-enhanced cutoff effect becomes important for small values of $(aQ)^2$, the asymptotic form of the function changes direction and approaches the continuum value from above. In order to reliably capture this turnover behaviour, results at small enough values of $(aQ)^2$ are needed. Reassuringly, as discussed in the previous paragraph, for $Q^2 = 5$ GeV^2, the logarithmic term that we obtain by a fit to the data is close to the value expected from lattice perturbation theory. We will come back to this observation in our final analysis in Sec. 4.3.3. One can further test that the leading logarithmically-enhanced discretization error is correctly picked up in fits to the lattice

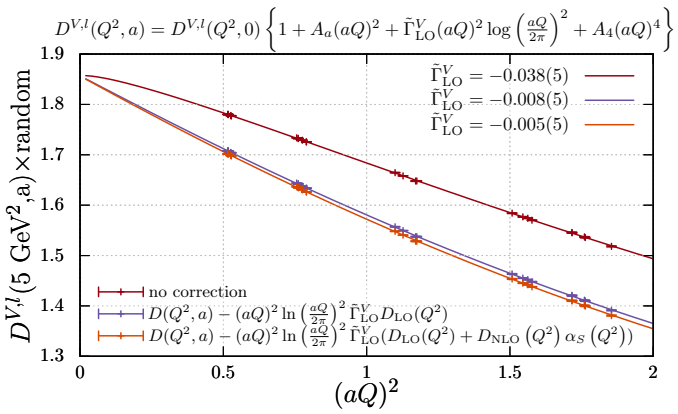


FIGURE 4.9: Continuum extrapolation of $D^{V,l}(5\ \mathrm{GeV}^2)$ with the results from our 4-stout ensembles [7]. The lattice results have been blinded by a random factor between 1.01 and 0.99. The dark red points correspond to unimproved data, for the purple line we have removed the leading logarithmically-enhanced cutoff effect and from the orange datapoints we have removed the additional mixed term. We also depict the logarithmic coefficient $\tilde{\Gamma}_{\rm LO}$, cf. (4.88), obtained from a fit to the data, using the fit function in the title.

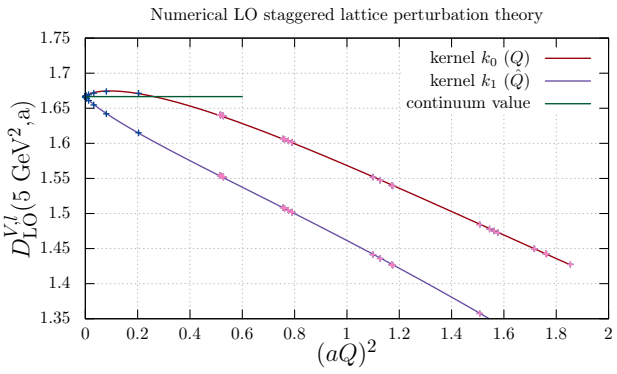


FIGURE 4.10: Light contribution to $D_{\rm LO}(5\ \mathrm{GeV}^2)$ as a function of $(aQ)^2$ in leading-order lattice perturbation theory, employing $k_0(t,Q)$ (pink) and $k_1(t,Q,a)$ (violet) in eq. (4.48). The pink points correspond to the lattice spacings available from our 4-stout ensembles [7], blue points are additional smaller lattice spacings. The green line is the continuum value. By employing $k_1(t,Q,a)$ the continuum value is approached from below.

results by modifying the kernel function, which we already alluded to and further discuss now.

Interestingly, as we show in App. B.3, when employing $k_1(t,Q,a)$ as a kernel function, the leading-order coefficient in front of the $\mathcal{O}(a^2)$ logarithmically-enhanced cutoff effect changes its sign, as now also $k_1(t,Q,a)$ receives $\mathcal{O}(a^2)$ corrections, see also Eq. (4.80). This is what can be observed in Figure 4.10, where the light contribution to $D(5\ \mathrm{GeV}^2)$ in lattice perturbation theory employing $k_0(t,Q)$ (pink) and $k_1(Q,t)$ (violet) are depicted. Since the logarithmically-enhanced cutoff effect for $k_1(Q,t)$ has a different sign, it approaches the continuum limit from below and can therefore serve as an additional systematic check.

For further details and a preliminary continuum extrapolation of $D(5\ \mathrm{GeV}^2)$ see Ref. [252] from which I also adapted this section.

Corrections for the final analysis As discussed above, these improvements significantly reduce discretization errors and allow a controlled continuum extrapolation. However, it remains

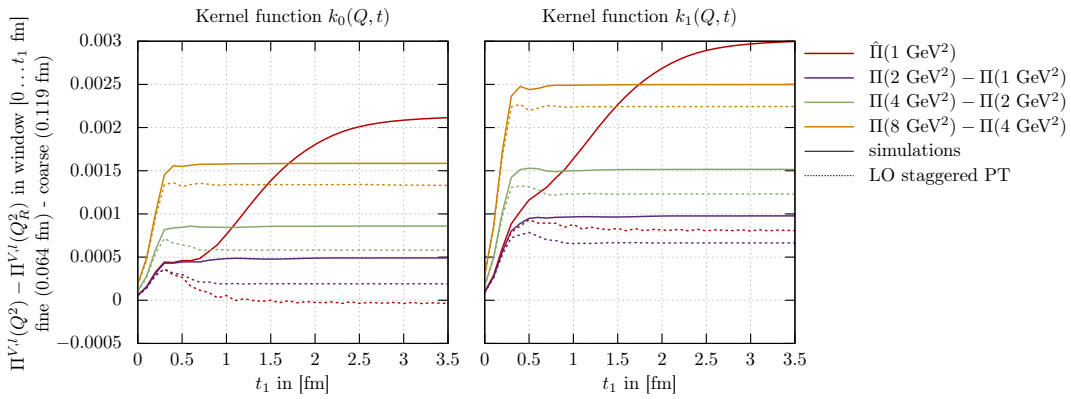


FIGURE 4.11: Light contribution to our final choice of observables $\hat{\Pi}^V(Q^2)$ and $\Delta\Pi^V(Q^2)$ in a time window $[0\ldots t_1]$ fm employing the kernel function $k_0(t,Q)$ (*left*) and $k_1(t,Q,a)$ (*right*), respectively. We plot the difference between a fine (a = 0.064 fm) and a coarse (a = 0.118 fm) lattice, in comparable volumes. The solid lines are obtained from the simulation. The dotted curves are the predictions from leading-order lattice perturbation theory.

to be examined where lattice perturbation theory can be applied for our final choice of quantities. Concerning the removal of leading order discretization effects in the discrete Adler function and the HVP, a similar picture emerges. In Fig. 4.11 we plot our observables in a cumulative time window, $O^V(Q^2,a;0,t_1)$, again taking the difference between the finest lattice and the coarsest lattice for two of our 4-stout simulations (solid line). This difference is again compared to the discretization errors obtained using lattice perturbation theory for the same lattice parameters (dotted curves). In the left-hand-side plot we employ the kernel function k_0. On the right we plot the same quantities but employing our second kernel function k_1. As in the case of the Adler function, for large momenta, LO lattice perturbation theory describes the discretization errors quite well. The discretization errors in the same set of observables but using the lattice momentum $k_1(t,Q,a)$ inside the kernel function, are larger since the lattice momentum introduces additional discretization errors. For all observables but $\hat{\Pi}(1\ \mathrm{GeV}^2)$, LO staggered perturbation theory captures the discretization errors increasingly well in the full t-range as Q^2 is increased. For our largest momentum, $\Delta\Pi(8\ \mathrm{GeV}^2)$, it reproduces ~ 90% of the discretization errors.

These insights can now be applied to the short distance contributions to the running of α. We incorporate some of the improvements mentioned in the previous section in our analysis:

- While we investigated in removing an additional cutoff effect, we do not include this variation in the final analysis. Its effect is very small compared to other systematic variations. However, as we argued above, the removal of discretization effects at leading order in lattice perturbation theory may not remove the logarithmic term at $\mathcal{O}(a^4)$ and above. To be conservative, we perform extrapolations with a logarithmic and without a logarithmic term in the fit to the improved data.

- We improve the short-distance connected quantities, namely $\Delta\Pi(8\ \mathrm{GeV}^2)$, $\Delta\Pi(4\ \mathrm{GeV}^2)$, $\Delta\Pi(2\ \mathrm{GeV}^2)$ and $\hat{\Pi}(1\ \mathrm{GeV}^2, 0.0\ \mathrm{fm}, 0.4\ \mathrm{fm})$, using lattice perturbation theory, Eq. (4.85). In the case of the light and the strange connected components, we also perform continuum extrapolations with the unimproved data which is a very conservative approach. Remarkably, as we will see below, they largely extrapolate to the same point. This again suggests that we pick up the logarithmic term correctly in the unimproved case.

- For the discrete Adler function we always apply the improvement in the full t-range. Firstly, the χ^2/dof are much better and secondly, only fits to the corrections in the full t-range vanish within errors in the continuum limit. We checked for all quantities explicitly that the results agree. Hence, mainly in light of the much better χ^2, we decided to apply perturbation theory for all t. However, we always also include the fits to the uncorrected data in our final analysis.

- For $\hat{\Pi}(1\ \text{GeV}^2)$ we employ the prescription in Eq. (4.85) only in the short distance window from $0\ldots0.4$ fm. Clearly, for distances beyond $t \sim 0.4$ fm, PT cannot be applied for this quantity, which also receives significant contributions from longer distances, see Fig. 4.5. In this case the sum in Eq. (4.86) has to be evaluated numerically.

4.3.2.2 Long-distance taste-breaking corrections

For low values of Q^2, the light and disconnected contribution are dominated by two-pion states including the ρ-resonance (cf. Sec. 4.2) and are hence affected by additional discretization effects originating from taste-symmetry breaking in the staggered formulation. At low energies, these taste violations become in fact the most important cutoff effects of staggered fermions, since they distort the pion spectrum. Here we discuss the improvements we apply to our data at long distances. They improve the extrapolation to the continuum limit, based on a physically motivated model. Of course, in the continuum limit these corrections vanish since taste-breaking effects also vanish in the model. As we saw in Sec. 3.2.2, due to these taste-breaking effects which break the taste symmetry at finite lattice spacing, the pion spectrum splits into 16 tastes. We extract the mass splitting by measuring the difference between the mass of the non-Goldstone pions and the Goldstone pion for each of our ensembles, see also Fig. 2 of Ref. [7][11]. The 16 tastes can be grouped into taste-singlet I, vector taste V, tensor taste T, axial vector taste A and pseudoscalar taste P multiplets. The remnant $SO(4)$ taste symmetry is only very mildly broken and the masses inside a multiplet are to a good degree of accuracy degenerate (the pion mass splittings inside a multiplet are less than a factor ~ 50 smaller than the splittings between different multiplets on our coarsest lattice). We use the measurements of the mass splitting to model taste-violations and reduce discretization errors at long distances, as we will now briefly review.

Based on the extensive study that was performed in Ref. [7] regarding taste-breaking effects, we model these discretization errors using the ρ-model in its staggered version (we call it SRHO). This model was already used by the HPQCD collaboration to model taste-breaking effects in a_μ [253]. The ρ-model is an effective theory, including the ρ, π and photons. In the continuum, it was proposed by Sakurai [254] and was later used by Jegerlehner and Szafron to model $\rho-\gamma$ mixing in $e^+e^- \to \pi^+\pi^-$ [255]. Its Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F^2_{\mu\nu}(A) - \frac{1}{4}F^2_{\mu\nu}(\rho) + (D_\mu\pi)^\dagger(D_\mu\pi) + \frac{1}{2}m^2_\rho\rho_\mu\rho^\mu - \frac{e}{2g_\gamma}F_{\mu\nu}(\rho)F^{\mu\nu}(A) - m^2_\pi\pi^\dagger\pi\ , \quad (4.91)$$

with ρ_μ, π and A_μ the rho, pion and photon fields, $F_{\mu\nu}$ the abelian field strength tensor, the covariant derivative $D_\mu = \partial_\mu - ieA_\mu - ig\rho_\mu$ and five effective parameters $e, g, g_\gamma, m_\pi, m_\rho$. The

[11] Theoretically the taste violations should vanish with $\alpha_S a^2$ as $a \to 0$ since they arise from the exchange of highly virtual gluons. However, the differences in the mass splitting in our simulations approach the continuum limit much faster, suggesting that they are dominated by higher-order terms of the type $\alpha_S^n a^2$, $n > 1$ [251].

renormalized photon propagator to $\mathcal{O}(g^2, g_\gamma^2, gg_\gamma)$ can be obtained by resumming a Dyson series,

$$iG_{\gamma\gamma}^{\mathrm{RHO}}(q) = \frac{1}{q^2}\frac{1}{1-e^2\Sigma(q^2)} + \frac{e^2}{g_\gamma^2}\frac{\left[1-gg_\gamma\Sigma(q^2)\right]^2}{q^2\left[1-g^2\Sigma(q^2)\right]-m_\rho^2} \ . \tag{4.92}$$

The resulting propagator has two poles, one at $q^2 = 0$ and a second one at $q^2 = m_\rho^2$, the renormalized ρ-mass. By matching this expression at the rho-pole to a Breit-Wigner propagator,

$$iG_{\gamma\gamma}^{\mathrm{RHO}}(q) \to \frac{F_\rho^2/\left(2M_\rho^2\right)}{q^2 - M_\rho^2 + iM_\rho\Gamma_\rho} \tag{4.93}$$

the parameters g, g_γ, m_ρ can be fixed. Here we use $M_\rho = 775\,\mathrm{MeV}$, $\Gamma_\rho = 150\,\mathrm{MeV}$ and $F_\rho = 210\,\mathrm{MeV}$.

To isolate staggered taste-violations we include the mass splitting of the pions, keeping the remaining parameters fixed. The mass splitting enters via the average of the masses of the pion taste partners, hence we replace $\Sigma(q^2; M_\pi^2)$ by the taste-averaged pion vacuum polarization function $\frac{1}{16}\sum_\alpha \Sigma(q^2; M_{\pi,\alpha}^2)$, where we sum over all sixteen pion taste partners. Finally, one finds that the HVP, including finite volume and taste-breaking effects, takes the expression

$$\begin{aligned}\hat{\Pi}^{\mathrm{SRHO}}(q^2) &= \hat{\Sigma}_{\mathrm{latt}}(q^2) + \\ &+ \frac{F_\rho^2}{2M_\rho^2}\cdot\left[1 + gg_\gamma\hat{\Sigma}_{\mathrm{rho}} - \tfrac{1}{2}g^2\hat{\Sigma}_{\mathrm{rho}} - \tfrac{1}{2}g^2M_\rho^2\hat{\Sigma}'_{\mathrm{rho}}\right]^2 \cdot \frac{q^2[1-gg_\gamma\hat{\Sigma}_{\mathrm{latt}}(q^2)]^2}{q^2[1-g^2\hat{\Sigma}_{\mathrm{latt}}(q^2)] - M_\rho^2[1-g^2\hat{\Sigma}_{\mathrm{rho}}]}\end{aligned} \tag{4.94}$$

where $\hat{\Sigma}_{\mathrm{latt}}(q^2)$ denotes the pion vacuum polarization function, averaged over taste, in finite volume and $\hat{\Sigma}_{\mathrm{rho}} = \mathrm{Re}\hat{\Sigma}_\rho + \Sigma(0) - \Sigma_{\mathrm{latt}}(0)$.

This effective theory is expected to work well at long distances, where taste breaking effects become important. We would therefore like to isolate the long-distance contribution and convolute the HVP by our window function,

$$\hat{\Pi}_{\mathrm{win}}\left(Q^2; t_1, t_2\right) = 2\int_0^\infty \mathrm{d}t\frac{1}{Q^2}\left[\cos(Qt) - 1 + \frac{1}{2}(Qt)^2\right]w(t; t_1, t_2)C(t) \ . \tag{4.95}$$

The window function Eq. (4.54) that we employ is smooth and symmetric in t. Hence, we can write

$$\hat{\Pi}_{\mathrm{win}}\left(Q^2; t_1, t_2\right) = \int_{-\infty}^{\infty}\frac{\mathrm{d}P}{2\pi}\frac{1}{Q^2}\left[\tilde{w}(P-Q) - \tilde{w}(P) - \frac{1}{2}Q^2\tilde{w}''(P)\right]P^2\hat{\Pi}\left(P^2\right), \tag{4.96}$$

where we have denoted the Fourier transform of the window function by

$$\tilde{w}(Q; t_1, t_2) = \int_{-\infty}^{\infty}\mathrm{d}t\cos(Qt)w(t; t_1, t_2) \ .$$

As in the previous section, we would like to investigate whether and where the SRHO model provides a sensible description of these long-distance cutoff effects. In Fig. 4.12 we plot the discretization errors of the light connected contribution to $\hat{\Pi}(1\ \mathrm{GeV}^2, t_1, t_1 + 0.5\ \mathrm{fm})$ in a sliding window of width 0.5 fm. The solid lines correspond to the difference between a fine and a coarse lattice of two of our 4-stout ensembles, using the kernel function k_0 (green) and k_1 (yellow) in

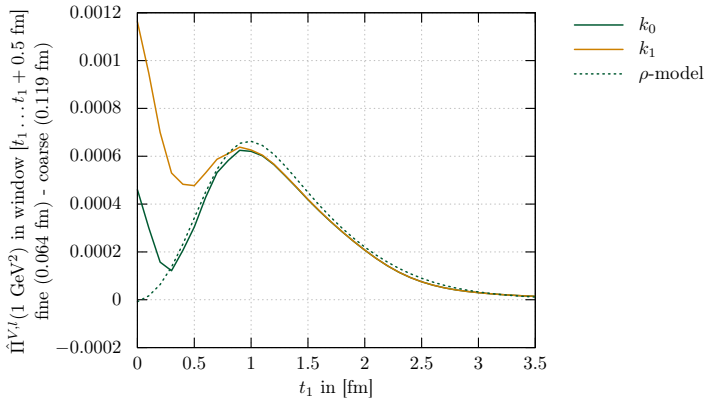


FIGURE 4.12: Light contribution of $\hat{\Pi}(1\ \mathrm{GeV}^2)$ in a sliding window $[t_1 \ldots t_1 + 0.5]$ fm. We plot the difference between a fine and a coarse lattice, in volumes $L = 6.14$ fm and $L = 6.67$ fm. The solid lines are obtained from the simulation. The dotted curves are the predictions of the SRHO model.

the TMR integrand. They should give us a description of the discretization errors arising in our simulations. We compare these discretization errors to those obtained using the SRHO-model for the same lattice parameters (dotted curves). Remarkably, the SRHO taste improvement describes our simulation data well all the way down to $t_1 \simeq 0.4$ fm for $k_0(t, Q)$[12]. For $\Pi_1^{V,l}$, i.e. with the kernel function k_1, the lattice momentum $\hat{Q}$ introduces additional discretization errors at short distances. To be conservative, we decided to consider only uncorrected lattice results for $0 \leq t_1 \leq 1$ fm. Finally, to cure our data from taste breaking effects at long distances, we apply corrections

$$\begin{aligned}\hat{\Pi}(Q^2)^{\mathrm{light}}(L,T,a) \to \hat{\Pi}(Q^2)^{\mathrm{light}}(L,T,a) \\ + \tfrac{10}{9}\left[\hat{\Pi}(Q^2)^{\mathrm{RHO}}_{t\geq t_{\mathrm{sep}}}(L_{\mathrm{ref}},T_{\mathrm{ref}}) - \hat{\Pi}(Q^2)^{\mathrm{SRHO}}_{t\geq t_{\mathrm{sep}}}(L,T,a)\right],\end{aligned} \tag{4.97}$$

with $L_{\mathrm{ref}} = 6.272$ fm to the light and similarly to the disconnected contribution. We include in total three different additional variations for a systematic error estimate. We apply the taste-improvement in time windows $t_{\mathrm{sep}} = 1.0, 1.5$ fm. To be conservative, we also perform fits without taste-improvement for the light connected contribution and include this variation in our error budget. For the light connected contribution we have included a factor 10/9 which changes to $-1/9$ for the disconnected contribution, since the correction in Eq. (4.94) corresponds to the $I = 1$ contribution. We also interpolate the simulation results to the same reference volume $L_{\mathrm{ref}}^3 \times T_{\mathrm{ref}}$.

4.3.2.3 Finite-volume corrections

While finite-volume effects become small for high Q^2, at long distances, the observable feels the finite size of the lattice. Hence, in particular for the $I = 1$ contribution to $\hat{\Pi}(1\ \mathrm{GeV}^2)$, finite-size effects become non-negligible. For the infinite volume limit of the $I = 1$ isospin-symmetric part of the HVP and discrete Adler functions, we performed a dedicated finite size study, similar to the procedure in Ref. [7]. The general idea is also depicted in Fig. 4.13. For details see the supplementary material in Ref. [7]. Let us briefly summarize the most important steps:

[12] This is, in fact, in agreement with what was found earlier for taste-violations in a_μ^{light}, see Fig. 15 in Ref. [7].

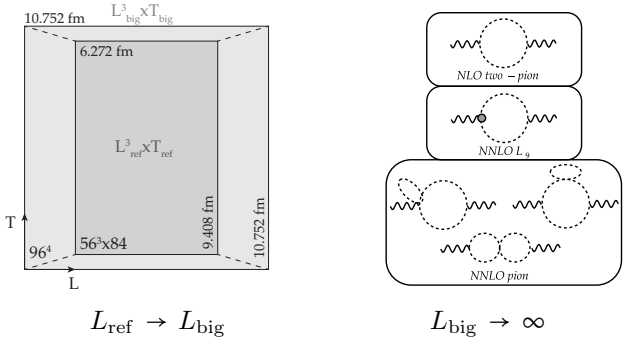


FIGURE 4.13: Schematic representation of our strategy for the calculation of finite volume effects. For the extrapolation to a big reference volume $L_{\rm big} = T_{\rm big} = 10.752$ fm we perform a dedicated lattice study with a special action designed for that purpose (4HEX). For the final residual extrapolation to infinite volume we use NNLO chiral perturbation theory [7].

- All ensembles with the 4stout action are performed at a volume of (6 fm)$^3\times$(9 fm) ($L_{\rm ref}$ = 6.272 fm, $T_{\rm ref} = 3/2L_{\rm ref}$). The results are continuum extrapolated, yielding the result $\hat{\Pi}(Q^2)^{\rm 4stout}_{a\to 0}(L^3_{\rm ref}, T_{\rm ref})$. These results are collected in Sec. 4.3.3.

- To calculate the largest FV effect, a dedicated lattice study, using a special 4HEX action with $n_f = 2+1$ and 4 steps of HEX smearing (see Sec. 3.8) to reduce taste violations is performed, see Tab. 4.3. They are performed in two different volumes, $L_{\rm ref}$ = 6.272 fm ($56^3\times 84$) and in $L_{\rm big} = T_{\rm big}$ = 10.752 fm ($96^3\times 96$). In addition, the mass of the Goldstone pion is lowered so that the FV corrections in these staggered simulations corresponds to those one would get in the absence of taste corrections at the physical value of the pion mass.

- The physical pion mass is defined as the harmonic mean square of the sixteen taste partners, $M^{-2}_{\pi,HMS} = \frac{1}{16}\sum_\alpha M^{-2}_{\pi,\alpha} = M^*_{\pi,0}$. This corresponds to a mass of the Goldstone-pion of ~ 110 MeV. This choice is motivated by the observation that the slope of the HVP function (the first moment) is proportional to $\sum_\alpha M^{-2}_{\pi,\alpha}$. In the extensive finite volume study in Ref. [7] it was found that using this definition of the physical pion mass very good agreement with NNLO SXPT is obtained, see Sec. 18 of Ref. [7] for an extensive discussion. Two simulations with two different Goldstone pion masses, namely M_π = 104 MeV and M_π = 121 MeV were performed. The physical point is then set by interpolating the results with these two pion masses to M_π = 110 MeV.

- We have only one lattice spacing for the 4HEX simulations, therefore no continuum extrapolation can be performed. Instead we compare the continuum extrapolated observables from the 4-stout action in the reference volume to the value of our observable in the 4HEX simulation in the reference volume. The final FV correction from the 4HEX simulations is rescaled by this factor with a 100% uncertainty. We give the various factors for our observables in Tab. 4.8. We then add up linearly the statistical error from the 4HEX simulations and this estimated systematic error to give the total error of the difference of our observables between the big and the reference volume.

- For the small extrapolation from $L_{\rm big}$ to infinite volume we use NNLO XPT, which is found to agree very well with the lattice result for the difference $[\hat{\Pi}(Q^2)_{L,\rm big} - \hat{\Pi}(Q^2)_{L,\rm ref}]$. See Ref. [7] for a more detailed discussion about the calculation of finite volume effects using NNLO XPT. We also calculated finite time effects which were found to be negligible.

$O(Q^2, Q_R^2) \cdot 10^4$, Q^2 in [GeV^2]	$\hat{\Pi}(1)$	$\Delta\Pi(2,1)$	$\Delta\Pi^*(4,2)$	$\Delta\Pi^*(8,4)$
$O^{\text{4HEX}}(56^3 \times 84)$	356.17(18)	103.6642(39)	97.6837(21)	86.2593(10)
$O^{\text{4HEX}}(96^3 \times 96)$	359.45(13)	103.8030(20)	97.75805(92)	86.29846(40)
$O^{\text{4HEX}}(96^3 \times 96)$ -$O^{\text{4HEX}}(56^3 \times 84)$	2.96(21)	0.1249(41)	0.0669(21)	0.0353(10)
$O^{\text{NNLO–XPT}}(\infty - 10.752\text{ fm})$	0.15	0.0025	0.0013	0.00063
$O^{\text{4stout}}_{a\to 0}(56^3 \times 84)/O^{\text{4HEX}}(56^3 \times 84)$	0.97	1.06	1.12	1.24
final result	3.10[22]	0.135[12]	0.076[10]	0.0445[99]

TABLE 4.8: Finite-volume corrections for our set of observables. We list the results from the 4HEX simulations in boxes $L_{\text{ref}} = 6.272$ fm ($56^3 \times 84$) and in $L_{\text{big}} = T_{\text{big}} = 10.752$ fm ($96^3 \times 96$), interpolated to the physical point with $M_\pi = 110$ MeV, the differences between the big and the reference box (calculated under the jackknife), the result from NNLO XPT for the extrapolation from L_{big} to ∞, the factor between the continuum extrapolated observables from the 4-stout action in the reference volume and the value of our observable in the 4HEX simulation in the reference volume and the final FV correction. The errors inside the round brackets are statistical. The error on the final correction corresponds to the total error. *The results for these two short-distance quantities are not used to correct our results for reasons explained in the main text.

Finally, the infinite-volume limit for any of our observables will be obtained via a telescopic sum

$$\hat{\Pi}(Q^2)_\infty - \hat{\Pi}(Q^2)_{L,\text{ref}} = \underbrace{[\hat{\Pi}(Q^2)_\infty - \hat{\Pi}(Q^2)_{L,\text{big}}]}_{\text{NNLO XPT}} + \underbrace{[\hat{\Pi}(Q^2)_{L,\text{big}} - \hat{\Pi}(Q^2)_{L,\text{ref}}]}_{\text{4HEX}} \tag{4.98}$$

Finite volume effects are added as a number to the final result. We collect the final numbers for this procedure in Tab. 4.8. For $\hat{\Pi}(1)$, which receives contributions from long distances, in particular due to the subtraction of $Q^2 = 0$, finite size effects coming from the extrapolation to the big volume are important. They give a $\sim 1\%$ correction to the continuum extrapolated result in the reference volume, see Sec. 4.3.3. However, the small residual correction from the extrapolation $L_{\text{big}} \to \infty$ is already well below the uncertainty on this quantity and even below the statistical uncertainty on the correction coming from the extrapolation to the big box. For both $\hat{\Pi}(1\,\text{GeV}^2)$ and $\Delta\Pi(2\text{ GeV}^2)$, the correction due to the extrapolation to the big box already lies below the systematic uncertainty on the continuum extrapolated value in the reference volume. For the short distance quantities $\Delta\Pi(4\,\text{GeV}^2)$ and $\Delta\Pi(8\,\text{GeV}^2)$, finite-volume effects are expected to be negligible, as confirmed. In this case, the fact that we only have one lattice spacing in the big box and therefore cannot perform a continuum extrapolation is an issue because discretization errors on these short-distance quantities become large. Furthermore, we do not expect NNLO-XPT to work well for these short distance quantities. Thus, we do not correct our lattice results for $\Delta\Pi(4\,\text{GeV}^2)$ and $\Delta\Pi(8\,\text{GeV}^2)$ with the results and rather neglect these tiny effects altogether. The numbers in Tab. 4.8 are given here for completeness. We also calculated finite-time effects which we found to be entirely negligible for our set of quantities.

Note that the finite-volume effects described up to now are those associated with the $I = 1$ channel. These are dominated by two-pion contributions that appear at NLO in XPT. In principle quantities in the $I = 0$ channel are also affected by the finite size of the lattice. However these are dominated by three-pion contributions that only appear at N^{4}LO in XPT [7]. Thus, they are highly suppressed compared to those arising in the $I = 1$ channel. Since the latter are already smaller than the uncertainties on the values of the HVP function in finite volume, the $I = 0$ effects are completely negligible.

We also considered finite-volume effects in the masses that we use to determine the physical point. For the Ω-baryon mass, these are negligible as the leading ones require popping a light quark-anti-quark pair in the vacuum. Finite-volume effects were calculated for the pion mass in chiral perturbation theory, yielding a 0.022% relative correction for $M_\pi(L) - M_\pi(\infty)$ [7, 256]. For M_{ss} these are even smaller. Thus, we can neglect finite-volume corrections to all of those masses.

Since QED is a long-range interaction, the QED corrections are expected to be particularly sensitive to finite-volume effects. Indeed, QED effects decrease only as powers of $1/L$ instead of exponentially. For hadron masses, finite volume effects in the QED_L scheme have been calculated analytically to $\mathcal{O}(1/L^2)$:

$$M(L) - M = -\frac{(Qe)^2 c}{8\pi}\left[\frac{1}{L} + \frac{2}{ML^2} + O(L^{-3})\right], \quad \text{with} \quad c = 2.837297... \tag{4.99}$$

where $c = 2.837297...$ and Q is the electric charge of the hadron [257, 258]. These finite-volume effects were first investigated by comparing QED contributions which were both calculated in a $L = 6$ fm and a smaller $L = 3$ fm box on our coarsest lattice, $\beta = 3.700$. It was found that the sea-sea and the sea-valence contribution did not show any relevant volume dependence compared to the statistical precision. They are therefore evaluated on a dedicated subset of smaller $L = 3$ fm lattices, see Sec. 4.3.1. The valence-valence contribution however exhibits non-negligible finite volume effects. These contributions are calculated in $L = 6$ fm boxes and are corrected using Eq. (4.99) for the mesons. SIB contributions which suffer from exponentially small FV effects are also calculated in $L = 6$ fm boxes.

FV effects in the QED corrections to the current-current correlator were studied extensively in Ref. [259] and in Ref. [7]. These were shown to behave like $\alpha/(LM_\pi)^3$ for our QCD+QED to QCD matching conditions [7] and are thus negligible compared to the other uncertainties on the QED corrections.

4.3.2.4 Global fit procedure and definition of the physical point

The simulations contain the lattice spacing and four bare parameters,

$$\{a, m_u, m_d, m_s, e\},$$

which need to be fixed to their physical values. In this study, they are fixed by using the set of observables

$$\{M_{\Omega^-}, M^2_{\pi_\chi}, M^2_{ss}, \Delta M^2, e^2_*\}, \tag{4.100}$$

where $*$ sub- and superscripts indicate that it is the physical value that is considered and where we have defined

$$\begin{aligned} M^2_{\pi_\chi} &\equiv \tfrac{1}{2}\left(M^2_{uu} + M^2_{dd}\right) \simeq M^2_{\pi_0} \\ \Delta M^2 &\equiv \left(M^2_{dd} - M^2_{uu}\right), \end{aligned} \tag{4.101}$$

and $e^2_* = \sqrt{4\pi\alpha}$, with α the fine structure constant, see Sec. 4.1. In partially-quenched chiral perturbation theory coupled to photons it has been shown that $M_{\pi_\chi} \simeq M_{\pi_0}$ [260], up to terms

that are beyond leading order in isospin breaking. The lattice spacing is set via

$$a = (aM_{\Omega^-})/M^*_{\Omega^-} \,, \tag{4.102}$$

where $M^*_{\Omega^-} = 1672.45(29)$ MeV is the experimentally measured mass of the Ω^- baryon [2] and (aM_{Ω^-}) the extracted mass in lattice units. This is a so-called mass-dependent scheme, cf. Sec. 3.6.4.

The mass of the ground-state Ω^- baryon is determined using the staggered operator $\Omega_{\mathrm{VI}}(t)$ first introduced in Ref. [261] and using either a four-state fit with two positive and two negative parity states and with a prior term in the correlated χ^2, or a generalized eigenvalue problem (GEVP). The variation between the three determinations of M_{Ω^-} is included as a systematic error, see Sec. 4.3.3. Similarly, for the pseudoscalar meson $M^2_{\pi_\chi}$ an operator corresponding to the pseudo-Goldstone taste was used. For the extraction of the mass from the correlator a correlated $\cosh\left[M(t-T/2)\right]$ fit at sufficiently late times was performed. A variation in the fit range is again included as a systematic error, cf. Sec. 4.3.3. We use here the results from Ref. [7], where details can be found in the supplementary material in Sec. 4. For details on the construction of the staggered operators see also Ref. [12].

M_{ss} and ΔM^2 are not experimentally measurable but they are "physical observables" in the sense that they have a well-defined continuum limit[13]. Their physical values have been determined in Ref. [7],

$$\begin{aligned}[M_{ss}]_* &= 689.89(28)(40)[49]\,\mathrm{MeV} \\ [\Delta M^2]_* &= 13170(320)(270)[420]\,\mathrm{MeV}^2 \,,\end{aligned}$$

where the errors are statistical, systematic, and total, respectively. This set of observables is also well suited for an IB decomposition, as we will argue in Sec. 4.3.2.4.

Our observables take the form (4.68), with the definition of the various components given in (4.3.1.6). Observables, such as the HVP function, are non-linear functions of the lattice spacing and the correlator, see Eq. (4.59), which themselves are affected by isospin breaking effects. In order to calculate the isospin breaking corrections to the HVP function, these corrections have to be expanded to $\mathcal{O}(e_v^2, e_s^2, \delta m)$. The expanded terms will then give $\langle\hat{\Pi}\left(Q^2\right)\rangle_0$, $\langle\hat{\Pi}\left(Q^2\right)\rangle''_{20}$ etc., i.e. the isospin breaking corrections to the HVP function. The same is true for any other observable, see Eq. (4.108). In my code, this is done via an automatic derivation which only keeps the first order terms in the expansion.

In order to obtain the physical result for an observable O, we parametrize its behaviour and perform a global fit which includes a continuum extrapolation, an interpolation to the physical

[13] Alternatively, the bare parameters could be fixed using a set of only experimentally measurable quantities $\{M_{\Omega^-}, M^2_{\pi_\chi}, M^2_{K_\chi}, \Delta M^2_{K_\chi}, e^2_*\}$, as in [7], where

$$\begin{aligned} M^2_{K_\chi} &\equiv \tfrac{1}{2}\left(M^2_{K_0} + M^2_{K_+} - M^2_{\pi_+}\right) \\ \Delta M^2_K &\equiv M^2_{K_0} - M^2_{K_+} \,. \end{aligned} \tag{4.103}$$

We have checked explicitly that, using this set of physical observables, we obtain entirely compatible results.

point and the determination of strong-isospin breaking (SIB) and QED corrections,

$$O(a,Q^2) = O(0,Q^2) + \underbrace{A(a)}_{\substack{\text{cont.}\\ \text{extrap.}}} + \underbrace{BX_l + CX_s}_{\substack{\text{interpolation to}\\ \text{physical point}}} + \underbrace{DX_{\delta m} + Ee_v^2 + Fe_v e_s + Ge_s^2}_{\substack{\text{determination of}\\ \mathcal{O}(\delta m, e^2)\text{ corrections}}}\,, \tag{4.104}$$

In the following, we will denote $O(0,Q^2) = A_0$. Simulations are performed close to the physical point, see Fig 4.4. Hence, X_l, X_s are proxies to parametrize the small difference in the light and strange quark masses from their physical values. They are built from the set of observables (4.100) and are defined as

$$X_l = \left[\frac{(M_{\pi_0}/M_{\Omega^-})^2}{(M_{\pi_0}/M_{\Omega^-})^2_*}\right] - 1, \quad X_s = \left[\frac{(M_{ss}/M_{\Omega^-})^2}{(M_{ss}/M_{\Omega^-})^2_*}\right] - 1\,. \tag{4.105}$$

Since our ensembles closely bracket the physical point, see Fig. 4.4, this parametrization suffices for an interpolation to the physical point. Similarly, we parametrize the dependence in isospin-breaking parameters via

$$X_{\delta m} = \frac{\Delta M^2}{M^2_{\Omega^-}}\,, \qquad X_{vv} = e_v^2\,, \qquad X_{vs} = e_v e_s\,, \qquad X_{ss} = e_s^2\,. \tag{4.106}$$

Note that we also distinguish valence, e_v, and sea, e_s, electric charges as in [7, 12]. In our ensembles, $X_{\delta m}$ is scattered around $[\Delta M^2/M^2_{\Omega^-}]_*$.

Close to the critical point, in the scaling regime (cf. Chap. 3.6) the lattice spacing dependence of our observable can be parametrized as

$$A(a) = A_2[a^2\alpha_s^n(1/a)] + A_{2l}a^2\log(a^2) + A_4[a^2\alpha_s^n(1/a)]^2\,, \tag{4.107}$$

where A_2, A_{2l}, A_4 are fit parameters. Note that $\mathcal{O}(a)$ and $\mathcal{O}(a^3)$ lattice artefacts are absent for staggered fermions, see also Sec. 3.2.2. As we discussed in great detail in Sec. 4.3.2.1, the HVP receives logartihmically-enhanced cutoff effects $a^2\log(a^2)$ that have to be controlled. There is also a source of logarithmically-suppressed effects that originate from higher order quantum corrections in powers of $\alpha_s(1/a) \sim -1/\log(a\Lambda_{\rm QCD})$. If large and if the anomalous dimensions of the operators which they result from are known, these discretization errors can be resummed using renormalization group equations. Here we make the simplifying assumption that only one higher-dimensional operator contributes to this scaling behavior with an anomalous dimension n. These anomalous dimensions were studied for hadron masses in Yang-Mills theory and for hadron masses and flavour currents in [251, 262, 263], see also [264]. These are found to be small but the coefficient for the HVP is unknown and the authors argue that such anomalous dimensions could be as large as $n = 3$ [265]. The value of $\alpha_s(1/a)$ is obtained by using five-loop perturbation theory to run $\alpha(M_Z)$ [2] down to $1/a$. To estimate the uncertainty originating from our ignorance of the precise value of the anomalous dimension, we perform each analysis using two extreme values: $n = \{0,3\}$. The choice for $n = 3$ is additionally motivated by the fact that it reproduces well the behaviour of taste-breaking effects, see Fig. 2 of the supplementary material in [7].

Note that the fit functions B, C, D, E, F, G can also depend on lattice spacing and/or the light and strange quark masses:

$$\begin{aligned}
B &= B_0 + B_2 a^2 \\
C &= C_0 + C_2 a^2 \\
D &= D_0 + D_2 a^2 + D_l X_l + D_s X_s, \\
E &= E_0 + E_2 a^2 + E_l X_l + E_s X_s, \\
F &= F_0 + F_2 a^2, \\
G &= G_0 + G_2 a^2.
\end{aligned}$$

In practice we find that, except in the charm contribution, we never need a lattice spacing dependence in B and C to describe our data well.

Considering each isospin derivative separately, (4.104) can be split to a system of five equations:

$$[O(Q^2)]_0 = [A + BX_l + CX_s]_0 \tag{4.108}$$

$$[O(Q^2)]'_m = [DX_{\delta m}]'_m \tag{4.109}$$

$$[O(Q^2)]''_{20} = [A + BX_l + CX_s + DX_{\delta m}]''_{20} + [E]_0 \tag{4.110}$$

$$[O(Q^2)]''_{11} = [A + BX_l + CX_s + DX_{\delta m}]''_{11} + [F]_0 \tag{4.111}$$

$$[O(Q^2)]''_{02} = [A + BX_l + CX_s + DX_{\delta m}]''_{02} + [G]_0 \tag{4.112}$$

Then, the fit parameters $A_0, A_2, \ldots, G_2$ are obtained by fits to the data. The final physical value for each analysis is obtained by taking the fit function (4.104) with the parameters obtained by fitting to the data and by plugging in the expansion parameters corresponding to the physical, continuum extrapolated value, i.e.

$$a = 0, \quad X_s = 0, \quad X_l = 0, \quad X_{\delta m} = [\Delta M^2 / M^2_{\Omega^-}]_*, \quad X_{vv} = X_{vs} = X_{ss} = e^2_* \,.$$

Finally, the physical result is given by

$$O(Q^2) = O(Q^2, 0) + D_0 [X_{\delta m}]_* + (E_0 + F_0 + G_0) e^2_* \,. \tag{4.113}$$

Isospin breaking decomposition We would like to decompose the observables into an isospin symmetric part, i.e. the value of the observable in a world where isospin symmetry is exact, and isospin breaking parts. To do so, we need to define observables which are equal in both worlds and which allow us to match the two theories. Different observables will of course give different decompositions (scheme dependence). For further discussions see Ref. [266]. Different IB components (like e.g. strong IB part) are not physical observables and therefore scheme dependent. Only the sum of all components is physical and scheme independent. Hence, the decomposition in Eq. (4.68) is not an isospin breaking decomposition but an expansion in bare parameters.

While in Ref. [7] the Wilson flow w_0 was used for a decomposition into isospin-symmetric and isospin breaking components – the "BMW-scheme" – in this study we decided to use another prescription for the separation. The scheme relies on M_Ω, and the meson masses M_{ss}, M_{uu} and M_{dd}. The meson masses are not physically observable (they cannot be measured in experiments)

but have a well-defined continuum limit,

$$\langle O\rangle = \langle O\rangle(M_{\pi_\chi}/M_\Omega,\, M_{ss}/M_\Omega,\, LM_\Omega,\, \Delta M/M_\Omega,\, e)\,. \tag{4.114}$$

This choice of observables is well suited for a decomposition in IB components: The masses of the quark-connected mesons $s\bar{s}$, $u\bar{u}$ and $d\bar{d}$ are absent of QED effects. $\Delta M^2 = \left(M_{dd}^2 - M_{uu}^2\right)$ is also a suitable quantity to measure the strong isospin breaking effect coming from the difference in mass of the up and the down quark because it is directly proportional to $m_d - m_u$ and its QED corrections are even more suppressed than those of the individual M_{uu} and M_{dd}. In chiral perturbation theory with photons it can be shown that up to terms that are second order in isospin breaking $\Delta M^2 \simeq 2B_2\delta m$. The same is true for $M_{\pi_\chi}^2 \simeq M_{\pi^0}^2$, as was mentioned earlier [260]. QED effects in both quantities are therefore very small.

The main reasons for this choice of scheme are that

- The fit qualities using M_Ω are in general better and the final results in general slightly more precise.
- In using M_{ss} and M_Ω to obtain both the final result and to perform an isospin separation, we have to perform the several hundred thousand fits for each observable only once and obtain the isospin breaking decomposition "for free". This is in contrast to the work in Ref. [7], where the final result was determined in so-called "type-I fits" using the Ω and the K mass for the scale setting and the determination of the physical strange mass, respectively, and M_{ss} and w_0 were used in "type-II" fits to perform the separation of isospin-symmetric and isospin breaking contributions.
- A new and more precise determination of w_0 is under way but has not yet been finalized.

Nevertheless, we explicitly cross-checked that – as expected – both the new and old prescriptions yield highly compatible results. Finally, we decompose a physical observable into an isospin symmetric contribution, a contribution from QED effects and a strong isospin breaking part,

$$\langle O\rangle = \langle O\rangle_{\mathrm{ISO}} + \langle O\rangle_{\mathrm{QED}} + \langle O\rangle_{\mathrm{SIB}}\,. \tag{4.115}$$

Isospin breaking effects originating from QED interactions are defined by the expansion term in e_*^2,

$$\langle O\rangle_{\mathrm{QED}} \equiv e_*^2 \cdot \tfrac{\partial\langle O\rangle}{\partial e^2}(M_{\pi_\chi}/M_\Omega,\, M_{ss}/M_\Omega,\, LM_\Omega,\, \Delta M/M_\Omega,\, e=0)\,. \tag{4.116}$$

Similarly, the strong isospin breaking part is defined as

$$\langle O\rangle_{\mathrm{SIB}} \equiv [\Delta M^2/M_\Omega^2]_* \cdot \tfrac{\partial\langle O\rangle}{\partial(\Delta M/M_\Omega)^2}(M_{\pi_\chi}/M_\Omega,\, M_{ss}/M_\Omega,\, LM_\Omega,\, (\Delta M^2/M_\Omega^2)=0,\, e=0). \tag{4.117}$$

Finally, the remaining part, obtained by switching off the electric charge and the strong isospin breaking parameter $\Delta M/M_\Omega = 0$, yields the isospin-symmetric part

$$\langle O\rangle_{\mathrm{ISO}} \equiv \langle O\rangle(M_{\pi_\chi}/M_\Omega, M_{ss}/M_\Omega,\, LM_\Omega,\, (\Delta M^2/M_\Omega^2)=0,\, e=0). \tag{4.118}$$

Hence, in each analysis, the IB components are obtained by

$$[O]_{\text{iso}} = O(Q^2, 0) = A_0 \,, \tag{4.119a}$$

$$[O]_{\text{sib}} = D_0[\Delta M^2/M^2_{\Omega^-}]_* \,, \tag{4.119b}$$

$$[O]_{\text{qed}} = [O]_{\text{qed-vv}} + [O]_{\text{qed-sv}} + [O]_{\text{qed-ss}} = (E_0 + F_0 + G_0)e^2_* \,. \tag{4.119c}$$

Note that this is the first lattice determination of $\Delta\alpha^{\text{had}}$ where leading-order IB breaking corrections are fully calculated and a decomposition is performed. We therefore cannot compare to other groups.

Correlated fits If the fit function is linear in the fit parameters, the "best-fit parameters" from minimizing the squared difference can be found analytically. Here we use however a modified definition of the chi-squared

$$\chi^2 = \sum_{i,j}^{N_{\text{ens}}} (Y_i - f(x_i))C^{-1}_{ij}(Y_j - f(x_j)) \tag{4.120}$$

with N_{ens} the number of ensembles, Y_i the lattice results of interest (e.g. $\hat{\Pi}^l(Q^2)$), x_i the lattice quantities on which they depend ($M^2_{\pi_\chi}$, $M^2_{ss}, \ldots$) on ensemble i and f the fit function which also depends on fit parameters. The chi-squared definition is modified since we replace the usual covariance matrix (that including only statistical fluctuations on Y_i) by the covariance matrix of the residuals $Y_i - f(x_i)$,

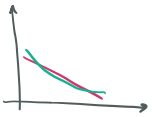
$$C_{ij} = \frac{N_J - 1}{N_J} \sum_{J=1}^{N_J} \left[\left(Y_i^{(J)} - f(x_i^{(J)})\right) - \left(Y_i^{(0)} - f(x_i^{(0)})\right)\right]\left[\left(Y_j^{(J)} - f(x_j^{(J)})\right) - \left(Y_j^{(0)} - f(x_j^{(0)})\right)\right] \,, \tag{4.121}$$

where N_J are the number of jackknife samples and sample $J = 0$ corresponds to the values obtained on the full set of ensembles. In this approach the covariance matrix takes into account the statistical fluctuations of both the x_i and Y_i as well as the correlations between them, see also Ref. [197]. With this covariance matrix, in general, one cannot find an analytical solution for the minimization of the χ^2 since it becomes non-linear. In practice, I first approximate the solution by using the usual definition of the covariance matrix. Then, the covariance matrix is updated and the procedure iterated, with the parameters obtained in the previous step as initial guess. Finally, a minimization algorithm (CG, etc.) can be applied.

Estimation of statistical uncertainties The statistical uncertainty is estimated using the jackknife method. We use $N_J = 48$ jackknife samples on all ensembles, which results in a blocksize much larger than the typical autocorrelation time.

Estimation of systematic uncertainties Having gone through the various challenges in the previous subsections, we are finally able to give a summary of the various possible sources of systematic uncertainties. Here we use the histogram method for an estimation of the systematic uncertainties, see also Refs. [7, 153, 267]. The philosophy we follow is to include all possible variations (such as higher order terms in the expansion in Eq. (4.107), different kernel functions

etc.) which give a reasonable description of our data. We summarize these variations below. In our analysis, we consider six main categories of sources for systematic uncertainties, namely:

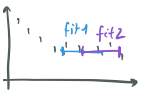


Fit function. Using Symanzik effective theory we can determine the behaviour of our observables as a function of a. Here, we usually take into account the leading cutoff effects. In order to estimate the uncertainty originating from subdominant terms, we vary the function in Eq. (4.107) by including possible higher-order corrections. The same is true for the interpolation to the physical point and the parametrization that we use in order to determine the IB corrections. In practice, we include all terms necessary for a reasonable χ^2 (i.e. a p-value above 0.05) and that should be there from our theoretical understanding. We then switch on/off subdominant terms which do not vanish within errors and which improve the χ^2 slightly. Of course, the choice of fit parameters varies for the different observables. Details about fit parameters included for each observable are listed in Sec. 4.3.3.

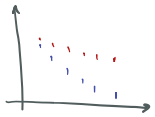
Cuts. For the iso-symmetric part of our observables we have at our disposal twenty-six large scale simulations varying over five different lattice spacings ranging from 0.118 fm to 0.064 fm and quark masses bracketing the physical point. Hence, in order to further estimate the systematic error associated with the continuum extrapolation, we can remove some of our coarser lattice spacings and perform the extrapolation without these data points. In practice, we eliminate between zero and three of the coarsest lattices, depending on the number of parameters in the fit function (4.107) (we always keep at least $n+1$ lattice spacings if we have n fit parameters). Cuts in the fits to the IB corrections are also applied and varied independently from those in the iso-symmetric part.

PDG

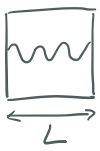
Physical masses. We take into account the errors on the experimentally measured hadron masses and on the continuum extrapolated values for the masses of the connected $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ mesons by performing the analysis with both $m_{\text{phys}} + \Delta m$ and $m_{\text{phys}} - \Delta m$.



Mass extraction ranges. Hadron masses are extracted by fitting to the effective mass curve of the corresponding correlation functions, for which a certain fit range has to be chosen. They are fitted once a plateau is reached in order to extract the mass of the groundstate. Excited states still contribute but are exponentially suppressed at large times. These effects should be small compared to the statistical error and the final result should not depend on the specific choice of the fit range. To estimate the systematic uncertainty originating from the mass extraction range, we choose 2-3 fit ranges per hadron (likewise for the IB components of the hadrons) and perform each analysis with the extracted mass from each of the fit ranges.



Choice of kernel functions and improvements. As we discussed in Sec. 4.3.2.1, there are various viable ways of defining the kernel function in the TMR-integrand and improvements vanishing in the continuum limit that we can apply to our data. All these variations should extrapolate to the same result in the continuum limit. For example, we do and do not improve the data employing the prescription in Eq. (4.85), or we employ k_0 and k_1 as a kernel in Eq. (4.59). These variations represent, together with the cuts in the lattice spacing, usually the most important sources of systematic errors.



Finite volume. Since calculations are necessarily performed in a finite box, the limit to infinite volume has to be taken. As we discussed in Sec. 4.3.2.3, here we performed a dedicated lattice study to reliably take this limit to a big volume of $L \sim 11$ fm and use NNLO chiral perturbation theory to extrapolate to $L \to \infty$. The details of the way we estimate the systematic error originating from this procedure can also be found in that section.

A variation of all different choices results in several hundred thousand fits for each observable, see Sec. 4.3.3, where we provide a more detailed list. We call each collection of choices an analysis. To keep track of the statistical uncertainty, we perform the global fit for each choice "under the jackknife", i.e. the global fit is performed on each jackknife sample, so $N_J + 1$ times. As mentioned above, due to our modified definition of the covariance matrix, these fits are non-linear. To efficiently minimize the χ^2 for each of the several hundred thousand fits I use the Fortran MINUIT-minimizer [268].

In our approach to determine the systematic uncertainty we include all possible variations that describe the lattice data reasonably well. Each possible analysis gets a weight. We consider that all choices except the fit functions and cuts in the lattice spacing are equally likely and are therefore weighted with a flat weight. Other choices related to the parametrization of the fit function and cuts in the lattice spacing are weighted according to the Akaike information criterion (AIC), see Eq. (3.102) and Ref. [7]. Then, from the various fits a histogram can be constructed, where the mean or the median is taken as the final estimate and the spread as an estimate of the systematic uncertainty of the observable.

For later convenience, let us denote by a Latin letter the analysis variations that correlate two quantities, $a = 1, \ldots, n_c$, such as for instance the mass extraction range for the pseudoscalars. Note that the results obtained from these variations are all weighted uniformly. Further, those analysis variations which are independent for each observable are labelled by a Greek letter, $\alpha = 1, \ldots, n_i$. These include the variation of fit functions, the β-cuts, the variation of the kernel functions and the improvements we apply to our data. Some of them are uniformly, some AIC-weighted. Note also that the weights do not depend on the jackknife sample, but they are constructed from the fit to the original sample. The statistical uncertainties are assumed to be Gaussian distributed, $\mathcal{N}(O; \bar{O}_i, \sigma_i)$, where O is the observable of interest, $\bar{O}_i$ the central value (the jackknife average) and σ_i the jackknife standard deviation for a particular analysis. Then, for each observable O, we calculate a probability distribution function (PDF), which is

the weighted normal distribution

$$PDF(O) = \frac{1}{n_c} \sum_{a,\alpha} \frac{w_{a\alpha}}{\sum_\beta w_{a\beta}} \mathcal{N}\left(O, \bar{O}_{a\alpha}, \sigma_{a\alpha}\right) , \tag{4.122}$$

where we have denoted by $w_{a,\alpha}$ the weight of a particular fit and by $\mathcal{N}(x, \mu, \sigma)$ the normal distribution with mean μ and variance σ. To obtain the final result and the error on the final value we build the cumulative distribution function,

$$CDF(O, \lambda) = \int_{-\infty}^{O} \mathrm{d}R \; PDF(R, \sigma_i \sqrt{\lambda}) , \tag{4.123}$$

where we have rescaled the statistical error by a factor λ for later convenience. This rescaling will allow us to disentangle the statistical and systematic error: From the above defined CDF, a reasonable definition for the total error would be the difference between the 16% and the 84% percentiles of the CDF, i.e.

$$\sigma_{\text{tot}}^2 \equiv \left[0.5\left(O_{84\%} - O_{16\%}\right)\right]^2 , \tag{4.124}$$

where we have defined O_x as the x'th percentile such that $CDF(O_x, \lambda = 1) = x$. However, in our distributions we sometimes encounter "outlier" peaks which render the distribution asymmetric and which would not necessarily be taken into account in the full error, see Sec. 4.3.3. To be conservative, we therefore define our total error as the maximum of the distance between the 16% and the 84% percentiles to the median of the CDF,

$$\sigma_{\text{tot}}^2 \equiv \max\left\{(O_{50\%} - O_{84\%})^2, (O_{50\%} - O_{16\%})^2\right\} . \tag{4.125}$$

The rescaling of the statistical error of each individual analysis should propagate into the same rescaling for the total statistical error. Hence we have

$$\sigma_{\text{tot}}^2(\lambda) = \lambda \sigma_{\text{stat}}^2 + \sigma_{\text{sys}}^2 , \tag{4.126}$$

where we require that the statistical and systematic error add up in quadrature to the total error. Finally, the central value for a specific quantity is the median of the CDF without any rescaling, i.e. the median of $CDF(O, 1)$. The statistical and systematic errors are disentangled by solving a system of equations, namely by constructing the CDF for $\lambda = 1$ and for $\lambda = 2$ and by solving for σ_{sys}^2 and σ_{stat}^2. Choosing $\lambda = 2$ instead of $\lambda = 0$, which would be the natural choice for isolating the systematic uncertainty, has the advantage that the CDF is smooth.

Error budget In order to resolve the relative contributions of the various sources of systematic uncertainties to the final error, we follow the same procedure as outlined in detail in Ref. [7]: First we build a CDF corresponding to each variation of a specific source of error, with M possible variations. If we consider for instance the choice of kernel function, we build $M = 2$ CDFs, one from all analyses with k_0 as a kernel function and one from all analyses performed with k_1. We then determine M total errors from these M CDFs. Here, we use the definition of Eq. (4.124) for the total error. Subsequently, from these M (Gaussian) CDFs we again build a total CDF, with σ the total error according to Eq. (4.124) of this choice and w_i the sum of the weights of all analyses entering this choice. To build the final CDF, we use the mean instead of the median of the individual CDFs. The systematic error obtained from this total CDF corresponds

to the error budget of this particular source of uncertainty. Since the systematic errors can be (anti)correlated, the quadratic sum of the error budgets does not necessarily sum up to the full systematic error.

Uncertainties on sum of contributions We perform a "final" analysis for each of the observables we listed above, namely the connected light, strange, charm and the disconnected contribution for which we calculate the step scaling function in powers of two in the Euclidean momentum squared, see Sec. 4.3.1.4. This split allows for an optimal distribution of computer time and in particular for an optimal assessment of systematic choices in the analysis of a specific observable. For each of these quantities we calculate the statistical and systematic error as outlined above. To obtain the Euclidean running of α up to 8 GeV^2, the sum of all our observables has to be taken. For simplicity here we discuss our strategy for calculating the statistical and systematic uncertainties on the sum of two observables. In particular, to do so, the distributions' shapes and possible correlations have to be preserved. For the statistical uncertainty, this can be achieved straightforwardly by keeping track of all jackknife samples. In principle, one could compute the systematic uncertainty by taking all possible combinations of analyses for each quantity. Since we perform several hundred thousand of analyses for each quantity and we have a total of eighteen quantities to sum, taking all possible combinations would result in $(\sim 100.000)^{18}$ possibilities and would clearly not be feasible. Instead, our strategy relies on reconstructing the distributions via random sampling.

- We randomly sample S_c samples from the correlated systematic choices in a correlated way, i.e. we make sure we select the *same* combinations for all quantities to be summed. Since they are flat-weighted this amounts to a uniform distribution.
- For each of the two quantities, we perform an importance sampling with S_i samples of the remaining combinations, according to the weight

$$\frac{1}{n_c}\frac{w_{a\alpha}}{\sum_\beta w_{a\beta}}\,,$$

  where we have already selected the correlated ingredients a, which are identical for both quantities. In this step, we build a CDF for each combination of correlated systematics. Then we pick a random number r and select the particular analysis which lies right above r.
- For those systematic correlated ingredients which are not present in some of our quantities (such as, for instance, the mass extraction range of the IB pseudoscalar masses in the charm contribution) we produce identical copies and divide the weight by the number of systematic variations.

This leaves us with two importance sampled distributions of a single systematic ingredient and where each sample has a weight $1/(S_c S_i)$. To obtain the sum of several quantities we reconstruct the distributions for all quantities, keeping track of the correlated choices and of all the jackknife samples. The sum will then be given by the sum of the distributions, where each distribution is now simply uniform. In practice, in order to correctly reconstruct the distributions from the original ones, we need a sample size of approximately one million samples.

4.3.3 Lattice results

We are now finally in the position to present the lattice results for the strange, light, charm and disconnected components, including electromagnetic and strong-isospin breaking effects, of the hadronic contribution to the Euclidean running of the electromagnetic coupling α in the on-shell scheme, in the continuum and infinite volume limits. As discussed in the previous sections, the two limits are taken separately. We present in the following sections the various procedures used for the set of observables given in Sec. 4.3.1.4. In Tab. 4.9 we give a summary of the systematic variations employed. They depend of course on the specific observable and flavour. We also provide a short descriptive explanation for each variation, for more details see Sec. 4.3.2. We further list the number of variations performed for each systematic effect and specify whether the variation is uniformly weighted or by using the Akaike-Information-criterion (AIC). We ascribe equal weights to those fits obtained by varying subtractions of discretization errors using the prescription (4.85) or (4.97), choices of fit ranges for the determination of hadronic masses, the error in the physical value of the hadron masses and the variation of the kernel function in the TMR integrand in Eq. (4.59) because we consider these equally probable descriptions of the lattice data. The choices purely related to the continuum extrapolation, such as the parameters in the fit function, the variation of the power of the strong coupling constant in the lattice spacing dependence of the A coefficient from the commonly used $n = 0$ to $n = 3$ in (4.104) and the removal of the coarsest lattice spacings are weighted according to the Akaike-information criterion. Finally, for each variation we list in Tab. 4.9 the corresponding labels used in the error budget tables which we provide for each individual quantity in the following sections.

Let us make some general remarks. As explained in detail in Chap. 4.3.1.4, we perform the running in discrete step scaling steps of powers of two in the Euclidean momentum squared to optimally disentangle possible systematic effects. Since the long-distance part of the light and disconnected contribution suffers from taste-breaking effects, we further split these two contributions to $\hat{\Pi}(1\ \mathrm{GeV}^2)$ into a long- and a short-distance window. As explained in detail in Sec. 4.3.2, for some quantities we also perform analyses where we apply an improvement on the data on each ensemble before performing the continuum extrapolation. This efficiently removes large cutoff effects for short-distance quantities and those related to taste violations in long-distance $I = 1$ quantities. We will make explicit for each quantity below if and what kind of improvements have been applied. Note that we also always include fits to unimproved data in the continuum extrapolations.

4.3.3.1 Connected light contribution

We list in Tab. 4.10 the continuum extrapolated results obtained from the global fits for the light connected contribution to our set of observables, in units of 10^4, together with the total error and the separation into statistical and full systematic error, as detailed in Sec. 4.3.2.4. The results are obtained in a finite box and the finite-size corrections calculated in Sec. 4.3.2.3 will be added subsequently. We also list the number of fits and the error budget for all variations performed. For the determination of the isospin-symmetric component, we include 25 4-stout ensembles with lattice spacings ranging from 0.118 fm to 0.064 fm and quark masses bracketing the physical point, as detailed in Sec. 4.3.1. In total, we performed between ~ 55000 and ~ 100000

Variation label	Description	#	Weight
Pseudoscalar meson fits M_π fit	mass extraction fit range for isospin-symmetric part of π: $1.8\ldots3.0$ fm and $2.0\ldots3.8$ fm	2	flat
M_Ω fits M_Ω fit	mass extraction fit range for isospin-symmetric part of Ω^- mass; see Suppl. information Table 6 in Ref. [7]	3	flat
IB M_Ω fits $M_{\Omega,\mathrm{IB}}$	mass extraction fit range for IB parts of Ω^- mass	2	flat
IB meson fits for meson masses $M_{\pi,\mathrm{IB}}$	mass extraction fit range for IB parts of M_0, M''_{20}, M''_{11}, M''_{02}, M'_m	2	flat
Physical M_Ω mass $M_{\Omega,\mathrm{phys}}$	$M_\Omega - \Delta M_\Omega$, $M_\Omega + \Delta M_\Omega$ taken from Ref. [2]	2	flat
"Physical" M_{ss} mass $M_{\Omega,\mathrm{phys}}$	$M_{ss} - \Delta M_{ss}$, $M_{ss} + \Delta M_{ss}$ taken from (168) of Ref. [7]	2	flat
Finite volume (FV)	discretization error on lattice determination of FV correction, see Sec. 4.3.2.3	/	flat
Taste improvement taste imp.	different time windows where SRHO is applied, no improvement, $t = 1.0\ldots\infty$ fm and $t = 1.5\ldots\infty$ fm, see Sec. 4.3.2.2	2-3	flat
LO improvement tree-level	subtraction of discretization errors at LO in staggered perturbative QCD	2	flat
Kernel $k_{0,1}$ kernel	vary kernel function in TMR integrand (4.59)	2	flat
Fit variations A_4 on/off etc.	switch on/off parameters in the fit function (4.104); see individual sections for more details	1-6	AIC
β-cuts cuts	remove coarsest lattice spacings; depends on number of parameters in $A(a)$	1-4	AIC
β-cuts in IB parts cuts	vary independently from β-cuts in isospin-symmetric contribution	3	AIC
Variation $n = 0, 3$ anomalous dim.	anomalous dimension for $\alpha_s(1/a)$ in expansion (4.107)	2	AIC

TABLE 4.9: Summary of systematic variations employed in the lattice determination of the set of observables given in Sec. 4.3.1.4. The variations depend on the specific observable and flavour contribution. We list the type of variation, a short descriptive explanation, the number of variations and whether the variation is flat or AIC weighted. In dark grey we list the label we use in the error budget tables in the sections below. For more details see also Sec. 4.3.2.4.

fits for each individual quantity. For all short-distance quantities we performed fits to data points that have undergone a subtraction of discretization errors based on LO staggered PT, as outlined in Sec. 4.3.2.1. For the fits to the unimproved data we always keep the logarithmic term, which can be well determined in the fit to the data. For all short-distance quantities, except for $\Delta\Pi(2\ \mathrm{GeV}^2)$, we find that we always need both an A_{2l} and an A_4 coefficient in the fit function to obtain a reasonable χ^2. After the LO improvement we are still left with a residual discretization error that we model with an $a^2\alpha_s(1/a)^n$ and $(a^2\alpha_s(1/a)^n)^2$ dependence, with $n = \{0, 3\}$. The first term is always kept while the latter term is also varied. Although the LO improvement should remove most of the logarithmic $a^2 \ln a^2$ cutoff effects (see App. D), to be conservative, we include both fits where the coefficient is constrained to zero and fits where the logarithmic term is kept free in the fit function to the improved data.

To resolve the composition of the various systematic errors we calculate the error budget of the individual quantities as detailed in Sec. 4.3.2.4. Clearly, the dominant source of error comes

from the choices made related to the continuum extrapolation: the variation in the anomalous dimension, the LO improvement, the removal of the coarsest lattice spacings and the variation of the coefficients A_{2l} and A_4 usually introduce the largest uncertainty. For the highest momentum, $\Delta\Pi(8\ \mathrm{GeV}^2)$, the variation of the kernel function becomes the dominant source of error. Note also that several of these errors are (anti-)correlated and a clear separation into error budgets is therefore not possible. As a consequence, the sum of the individual errors does not necessarily correspond to the total systematic error. The systematic uncertainty related to the scale setting and the setting of the physical point is very small for all quantities.

In Fig. 4.14 we show the continuum extrapolation of the short-distance window of the light contribution to $\hat{\Pi}(1\ \mathrm{GeV}^2, 0.0\ \mathrm{fm}, 0.4\ \mathrm{fm})$. We show here a number of illustrative fits, for more details see the caption of Fig. 4.14. They were obtained by performing global fits as detailed in Sec. 4.3.2.4. The continuum extrapolated value includes the leading order isospin breaking corrections, as discussed in Sec. 4.3.1.6. We perform fits to data where we have employed the kernel function k_0 (red) and the kernel function k_1 (blue) in the definition of the TMR integrand. As explained in Sec. 4.3.2, this has the particular advantage that the coefficient in front of the $a^2 \ln(a^2)$ cutoff effect in leading order staggered perturbation theory changes sign. This leads to a turnover of the curve for k_0 at small a and both curves approach the continuum limit from different directions. The dashed lines correspond to the fits to the data on which we applied a LO improvement before performing the global fit. The above improvement procedure is only applied to the isospin-symmetric component. The improvement removes up to 25 percentage points from the discretization effects on our coarsest lattice and allows a controlled continuum limit. The continuum limit, plotted in orange, has two errors, the statistical error and the combined statistical and systematic error, as outlined in Sec. 4.3.2.4.

In Fig. 4.16 we show the continuum extrapolation of the long-distance window of the light contribution to $\hat{\Pi}(1\ \mathrm{GeV}^2, 0.4\ \mathrm{fm}, \infty)$. The long-distance window gives the largest contribution to $\hat{\Pi}(1\,\mathrm{GeV}^2)$. The continuum extrapolation of this quantity is particularly challenging due to taste-violations in the pion spectrum. While we performed a thorough study of the removal of taste-breaking effects in the light contribution using the SRHO model, which seems to be working extremely well in modeling the discretization effects even down to 0.4 fm, to be very conservative, we applied here the taste-corrections only in time windows above 1.0 fm and even included fits to the uncorrected datapoints in the final result. As can be observed in the fits to the uncorrected data, taste-breaking effects introduce a large curvature and significantly increase the final value and the final error and thus have an important impact on the final result. We would therefore like to perform a more in-depth investigation of taste-breaking effects, for instance in NNLO staggered chiral perturbation theory, as was done in Ref. [7]. To improve the approach to the continuum limit, we further aim at including two additional ensembles at a smaller lattice spacing of $a \approx 0.048$ fm in our analysis which should exhibit reduced taste violations and should reduce the final error significantly.

In Fig. 4.18 we depict the continuum extrapolation for the light connected contribution to $\Pi(8\,\mathrm{GeV}^2) - \Pi(4\,\mathrm{GeV}^2)$. At these large momenta, discretization errors become significant: Our coarsest lattice spacing suffers from discretization effects of $\sim 20\%$ for k_0 and almost 40% for k_1. LO staggered perturbation theory removes practically all discretization effects, resulting in datapoints which almost lie on a horizontal line. Despite these sizeable discretization effects in the unimproved data, remarkably, the fits to the improved and the unimproved data extrapolate

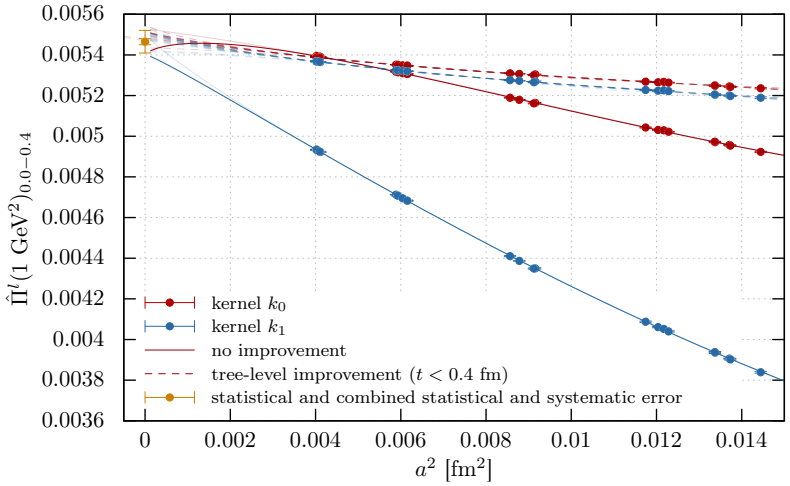


FIGURE 4.14: Example continuum extrapolations of $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$, including QED and SIB effects. These are obtained from the global fit procedure described in Sec. 4.3.2.4. The four sets of 25 points with errors are results for $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$ computed using the 25 ensembles listed in Tab. 4.1. The red points are obtained with kernel $k_0(Q,t)$ and the blue ones with $k_1(Q,t)$. Moreover the points that lie along dashed curves have undergone a subtraction of discretization errors using LO staggered perturbation theory as described in (4.85). Those that lie along solid curves are uncorrected. Each of these 2×2 sets of points is fitted independently. A few typical fits for each set are plotted as a function of a^2 as curves whose color and dashing match the corresponding set of points. The different curves correspond to variations in the continuum extrapolation dependence of the fit function and in the number of coarser lattice spacings included. In plotting these curves only the dependence of the corresponding fit function on a^2 is kept: the dependence on the other observables, used to define the physical point of 1+1+1+1 flavour QCD+QED (see (4.100)), is fixed by setting those observables to their physical value. Similarly, one of the fit functions is used to interpolate the corresponding lattice points to the physical value of each of the observables of (4.100). The transparency of the curves is determined by the AIC weight of the corresponding fit: darker curves have larger weight. Finally, the orange point at $a^2 = 0$ corresponds to our final value for $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$ with IB corrections: it includes two errors, the statistical and the combination of the statistical and all systematic errors, associated with variations beyond those illustrated here.

to the same point – note again that these fits were all performed independently. In addition, the advantage of including both kernel functions k_0 and k_1 becomes again apparent: these approach the continuum limit from above and below. They join in the continuum limit, suggesting that we correctly pick up the logarithmic coefficient, which becomes important at small lattice spacings, and this already with our current set of ensembles. The addition of a smaller lattice spacing will allow us to further constrain this coefficient. Again, after tree-level improvement, the difference between the two kernel functions basically vanishes. There are however two significant outliers which are only shown for illustrative purposes. These particular fits are clearly disfavoured by the data with bad χ^2/dof ~3-4 and therefore do not contribute to the final value. One of these outliers corresponds to a fit to uncorrected data with kernel function k_0 and anomalous dimension $n = 3$. Firstly, as already mentioned, this fit is clearly disfavoured by our lattice-data, resulting in a tiny AIC weight. This indicates that the $n = 3$ scaling which was found to fit well the cutoff effects at low energies [7] and which is well-tuned for taste-breaking effects associated with the pion mass does not necessarily work well at short distances where those taste-breaking effects are less important. Secondly, in that case, we find that the the $a^2 \ln a^2$ fit coefficient has a sign opposite to the one expected from LO staggered perturbation theory and shoots up, probably stemming from the combination of $a^2\alpha_s(1/a)^3$ and the $a^2 \ln a$ logarithmically-enhanced terms.

We also plot in Figs. 4.15, 4.17 and 4.19 the corresponding probability (red) and cumulative distribution functions (blue) of these particular quantities, together with the median and the combined statistical and systematic error as green bands, as explained in Sec. 4.3.2.4. The PDFs clearly display a multi-peak structure (less pronounced for $\Delta\Pi(8\,\mathrm{GeV}^2)$), which mainly originates

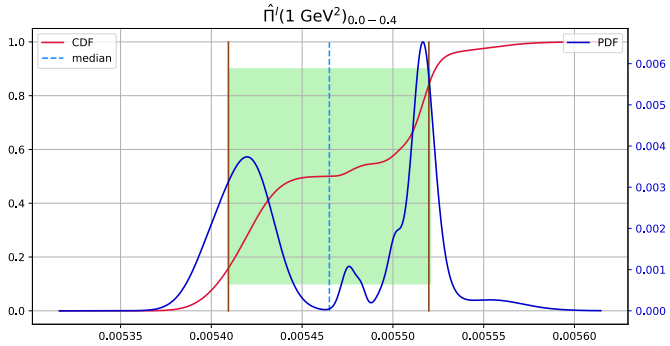


FIGURE 4.15: Probability distribution function (PDF) (blue curve) and cumulative distribution function (CDF) (red curve) from which the total uncertainty on $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$ is derived. These are plotted as a function of the value of $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$. They are obtained from all of the statistical fluctuations and all of the systematic variations considered in the determination of $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$, as described in (4.122), (4.123). Our final central value for $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$ is defined as the median of the distribution, shown as a vertical blue dashed line. The brown vertical lines correspond to the 15.9 and the 84.1 percentiles. Their distance from the median provide a lower and upper uncertainty, respectively. Our final total error is given by taking the largest of these two uncertainties and treating it as a symmetric error around the median. The range of values of $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.0-0.4}$ covered by this total uncertainty is shown as the green band..

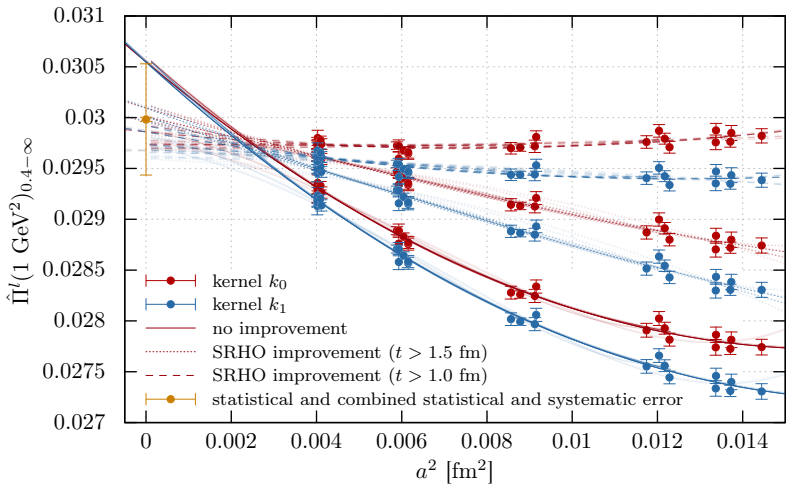


FIGURE 4.16: Same as Fig. 4.14, but for $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.4-\infty}$, and with the subtraction of discretization errors using LO staggered perturbation theory replaced by two levels of taste improvement, using the SRHO model for Euclidean times in the TMR integration above 1 fm (dashed curves) or above 1.5 fm (dotted curves), as described in Sec. 4.3.2.2. Also shown are the unimproved data and a corresponding set of fits.

from the fact that we also include fits to the unimproved data in the final error estimate. In order to cover these peaks in the final error, the latter is obtained by taking the largest of the distances from the median to the 15.9 and 84.1 percentiles, as explained in Sec. 4.3.2.4.

By performing the isospin decomposition as detailed in Sec. 4.3.2.4 we obtain the individual isospin contributions in units of 10^4, which are given in Tab. 4.11. The errors correspond to the statistical, full systematic and total error, respectively. We apply again the PDG conventions for the rounding and the significant error digits [2]. The rounding is determined based on the total error. All remaining plots of the continuum extrapolations and the corresponding probability and cumulative distribution functions of the quantities in Tab. 4.10 can be found in App. E.

Finally, in Tab. 4.12 we list the results for the light contribution to the Euclidean HVP function, $\hat{\Pi}^l(Q^2)$, for $Q^2 = 1, 2, 4, 8$ GeV^2, together with the total, statistical and full systematic error. These were obtained by summing the contributions in Tab. 4.10 based on the sampling procedure

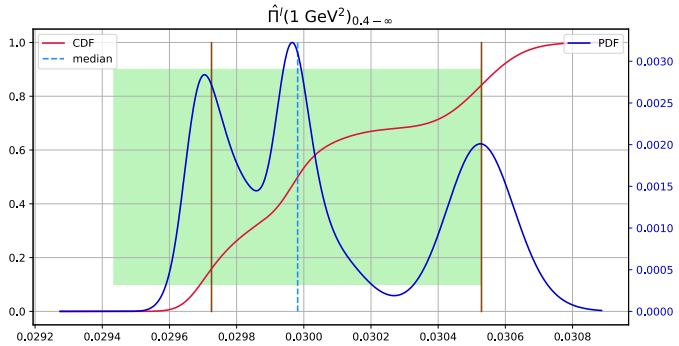


FIGURE 4.17: Same as Fig. 4.15, but for $\hat{\Pi}^l(1\,\mathrm{GeV}^2)_{0.4-\infty}$

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^l_{00-04}(1\ \mathrm{GeV}^2)$	$\hat{\Pi}^l_{04-\infty}(1\ \mathrm{GeV}^2)$	$\Delta\Pi^l(2\ \mathrm{GeV}^2)$	$\Delta\Pi^l(4\ \mathrm{GeV}^2)$	$\Delta\Pi^l(8\ \mathrm{GeV}^2)$
Number fits	55296	55296	110592	55296	110592
median	54.7	300	110.12	109.06	107.0
total error	6 (1.02%)	5 (1.83%)	32 (0.29%)	34 (0.31%)	4 (0.39%)
statistical error	1 (0.27%)	0 (0.07%)	15 (0.14%)	13 (0.12%)	2 (0.16%)
systematic error	5 (0.98%)	5 (1.82%)	28 (0.25%)	31 (0.29%)	4 (0.36%)
M_π fit	<1	<1	<1	<1	<1
$M_{\pi,\mathrm{IB}}$ fit	<1	<1	<1	<1	<1
M_Ω fit	<1	<1	<1	<1	<1
$M_{\Omega,\mathrm{IB}}$ fit	<1	<1	<1	<1	<1
$M_{\Omega,\mathrm{phys}}$	<1	<1	<1	<1	<1
$M_{ss,\mathrm{phys}}$	<1	<1	<1	<1	<1
kernel	<1	<1	<1	<1	1
anomalous dim.	3	2	23	20	<1
tree-level	5	off	11	15	<1
taste imp.	off	4	off	off	off
cuts	3	2	4	4	<1
A_0 on/off	on	on	on	on	on
A_2 on/off	on	on	on	on	on
A_{2l} on/off	<1	off	<1	2	<1
A_4 on/off	2	2	18	22	<1
B_0 on/off	<1	on	4	5	<1
C_0 on/off	on	on	on	on	on
D_0 on/off	on	on	on	on	on
D_a on/off	off	<1	on	<1	<1
D_l on/off	<1	<1	<1	on	<1
D_s on/off	off	off	off	off	off
E_0 on/off	on	on	on	on	on
E_a on/off	on	off	on	on	on
E_l on/off	on	off	<1	off	on
E_s on/off	<1	off	off	off	off
F_0 on/off	on	on	on	on	on
F_a on/off	on	off	on	on	on
G_0 on/off	on	on	on	on	on
G_a on/off	on	off	3	6	<1

TABLE 4.10: Continuum extrapolated results and error budget for the connected light contribution to our observables, including leading-order isospin-breaking corrections. *The errors are to be understood on the last digits of the central value* and we apply the PDG conventions for the rounding, which is determined based on the total error [2]. We also list the number of analyses which were carried out for each individual quantity. The results correspond to a box size $L_{\mathrm{ref}} = 6.272$ fm and $T_{\mathrm{ref}} = 3/2L_{\mathrm{ref}}$.

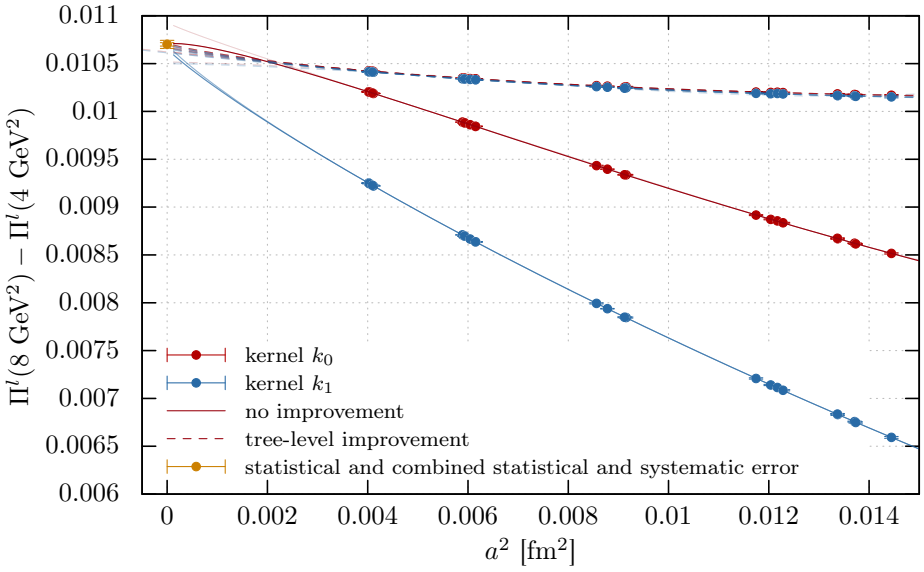


FIGURE 4.18: Same as Fig. 4.14, but for $\Pi^l(8\,\mathrm{GeV}^2) - \Pi^l(4\,\mathrm{GeV}^2)$.

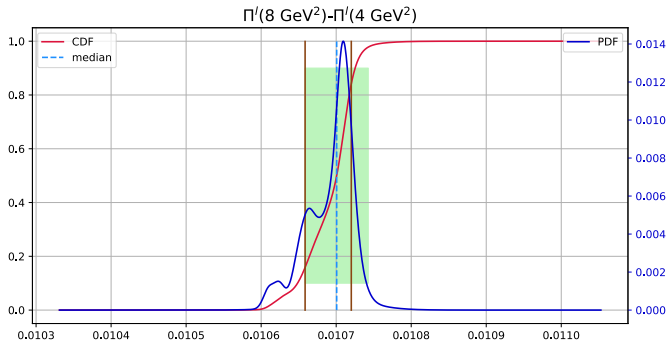


FIGURE 4.19: Same as Fig. 4.15, but for $\Pi^l(8\,\mathrm{GeV}^2) - \Pi^l(4\,\mathrm{GeV}^2)$.

outlined in Sec. 4.3.2.4, thereby preserving possible correlations. The total errors on the results for $\hat{\Pi}^l(Q^2)$, for $Q^2 = 1, 2, 4, 8$ GeV^2 are completely dominated by the taste-breaking systematic on $\hat{\Pi}^l(1\,\mathrm{GeV}^2)$. Thus, once those effects are better controlled the uncertainties on our results for $\hat{\Pi}^l(Q^2)$ for all Q^2 will be greatly reduced.

4.3.3.2 Strange contribution

In Tab. 4.13 we collect the continuum extrapolated results obtained from the global fits for the strange connected contribution to our set of observables, in units of 10^4, together with the total error and the separation into statistical and full systematic error, as detailed in Sec. 4.3.2.4. The variations listed in Tab. 4.13 are part of the error budget that we also give in that table. Strong-isospin breaking does not enter in the strange contribution, hence the D coefficients in Eq. (4.104) are always switched off in the fit function. Since taste-breaking effects are negligible in the strange contribution, we do not have to split $\hat{\Pi}(1\ \mathrm{GeV}^2)$ into a long- and a short-distance window. However, we only apply tree-level improvement in the short-distance window $t < 0.4$ fm. The fits to the uncorrected data in this quantity do not always allow us to sufficiently constrain

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^l_{00-04}(1\ \mathrm{GeV}^2)$	$\hat{\Pi}^l_{04-\infty}(1\ \mathrm{GeV}^2)$
iso	54.6(2)(5)[6]	298(0)(5)[5]
qed	0.34(3)(6)[7]	−2.5(3)(10)[10]
qed-vv	0.21(0)(7)[7]	−0.6(5)(2)[6]
qed-vs	−0.022(5)(1)[5]	−0.5(5)(0)[5]
qed-ss	0.150(32)(13)[34]	−1.3(7)(3)[8]
sib	0.01133(28)(8)[29]	2.01(13)(23)[27]

$O(Q^2)\cdot 10^4$	$\Delta\Pi^l(2\ \mathrm{GeV}^2)$	$\Delta\Pi^l(4\ \mathrm{GeV}^2)$	$\Delta\Pi^l(8\ \mathrm{GeV}^2)$
iso	109.94(15)(29)[33]	108.97(13)(31)[34]	107.0(2)(4)[4]
qed	0.61(11)(15)[19]	0.58(08)(08)[11]	0.47(5)(6)[8]
qed-vv	0.47(5)(4)[6]	0.44(2)(6)[6]	0.40(0)(12)[12]
qed-vs	−0.086(22)(1)[22]	−0.051(13)(1)[13]	−0.026(7)(1)[7]
qed-ss	0.23(9)(17)[19]	0.20(8)(3)[9]	0.11(6)(8)[9]
sib	0.124(7)(2)[7]	0.0397(13)(23)[26]	0.0114(5)(3)[6]

TABLE 4.11: Continuum extrapolated results for the different isospin-breaking components of the light connected contribution to our set of observables. The errors are to be understood on the last digits of the central value and we apply the PDG conventions for the rounding, which is determined based on the total error [2]. The decomposition was performed as detailed in Sec. 4.3.2.4. The results correspond to a box size $L_{\mathrm{ref}} = 6.272$ fm and $T_{\mathrm{ref}} = 3/2L_{\mathrm{ref}}$.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^l(1\,\mathrm{GeV}^2)$	$\hat{\Pi}^l(2\,\mathrm{GeV}^2)$	$\hat{\Pi}(4\,\mathrm{GeV}^2)$	$\hat{\Pi}^l(8\,\mathrm{GeV}^2)$	$\Delta\Pi^l((8,1)\,\mathrm{GeV}^2)$
median	355	465	574	681	326.2
total error	5 (1.53%)	5 (1.17%)	5 (0.94%)	5 (0.79%)	6 (0.18%)
statistical error	0 (0.07%)	0 (0.06%)	0 (0.04%)	0 (0.03%)	3 (0.10%)
systematic error	5 (1.53%)	5 (1.16%)	5 (0.94%)	5 (0.79%)	5 (0.15%)

TABLE 4.12: Connected light contribution to the Euclidean $\hat{\Pi}^l(Q^2)$ vs. Q^2. These were obtained by summing the individual contributions to $\hat{\Pi}(1\,\mathrm{GeV}^2)$ and to the discrete Adler function in Tab 4.10, following the sampling procedure outlined in Sec. 4.3.2.4 in order to preserve possible correlations. The errors are to be understood on the last digits of the central value and we apply the PDG conventions for the rounding, which is determined based on the total error [2]

the fit coefficient in front of the logarithmic $a^2 \ln a$ term, and we hence include both fits where the coefficient is left free and where it is constrained to zero. Once tree-level improvement is applied to the data, the logarithmic coefficient vanishes within errors and we do not include it in the fit function. For the steps in the discrete Adler function, we apply the same procedure as in the case of the connected light contribution. However, we fix the anomalous dimension in the fit function to $n = 0$ because, with $n = 3$, the χ^2/dof are bad. In total, we perform between ~5.000 and ~83.000 fits, depending on the quantity.

We depict in Fig. 4.20 the continuum extrapolation of the connected strange contribution to $\Delta\Pi^s(4\,\mathrm{GeV}^2)$, and in Fig. 4.21 the corresponding probability and cumulative distribution functions. Similarly to the continuum extrapolations of the light connected contribution at large momenta, discretization errors become large. However, LO staggered perturbation theory works very well in removing lattice artefacts. At 8 GeV^2, the improvement removes up to 32 percentage points of the discretization effects on our coarsest lattice and allows a controlled continuum limit. Remarkably, the LO correction also corrects for almost all discretization errors differing between the kernel k_0 and k_1, such that the two curves almost lie on top of each other. A decomposition into IB components as detailed in Sec. 4.3.2.4 results in the individual isospin contributions in units of 10^4 given in Tab. 4.14, together with the statistical, full systematic and total error.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^s(1\ \mathrm{GeV}^2)$	$\Delta\Pi^s(2\ \mathrm{GeV}^2)$	$\Delta\Pi^s(4\ \mathrm{GeV}^2)$	$\Delta\Pi^s(8\ \mathrm{GeV}^2)$
Number fits	5760	13824	27648	82944
median	41.6	18.81	20.45	20.84
total error	4 (0.90%)	8 (0.41%)	8 (0.41%)	14 (0.68%)
statistical error	2 (0.38%)	5 (0.28%)	5 (0.24%)	2 (0.10%)
systematic error	3 (0.81%)	6 (0.30%)	7 (0.33%)	14 (0.67%)
M_π fit	<1	<1	<1	<1
$M_{\pi,\mathrm{IB}}$ fit	<1	<1	<1	<1
M_Ω fit	1	1	<1	<1
$M_{\Omega,\mathrm{IB}}$ fit	<1	1	1	2
$M_{\Omega,\mathrm{phys}}$	<1	<1	<1	<1
$M_{ss,\mathrm{phys}}$	<1	<1	<1	<1
kernel	<1	<1	2	2
anomalous dim.	off	off	off	off
tree-level	2	3	3	5
cuts	1	1	1	2
A_0 on/off	on	on	on	on
A_2 on/off	on	on	on	on
A_4 on/off	off	3	4	2
A_{2l} on/off	2	2	3	10
B_0 on/off	off	off	<1	<1
C_0 on/off	on	on	on	on
D_0 on/off	off	off	off	off
D_a on/off	off	off	off	off
D_l on/off	off	off	off	off
D_s on/off	off	off	off	off
E_0 on/off	on	on	on	on
E_a on/off	<1	on	on	on
E_l on/off	off	off	off	<1
E_s on/off	off	1	1	2
F_0 on/off	on	on	on	on
F_a on/off	on	<1	<1	on
G_0 on/off	on	on	on	on
G_a on/off	off	off	off	<1

TABLE 4.13: Same as Tab. 4.10 but for the connected strange contribution. Remember that the errors are to be understood on the last digits of the central value.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^s(1\ \mathrm{GeV}^2)$	$\Delta\Pi^s(2\ \mathrm{GeV}^2)$	$\Delta\Pi^s(4\ \mathrm{GeV}^2)$	$\Delta\Pi^s(8\ \mathrm{GeV}^2)$
iso	41.5(2)(3)[4]	18.81(5)(6)[8]	20.45(5)(7)[9]	20.84(2)(14)[14]
qed	0.14(4)(2)[4]	0.013(3)(24)[24]	0.00(0)(5)[5]	0.006(15)(9)[18]
qed-vv	0.195(10)(12)[16]	0.027(0)(9)[9]	0.015(2)(12)[12]	0.012(0)(22)[22]
qed-vs	−0.03(1)(0)[1]	−0.0050(23)(0)[23]	−0.0028(9)(17)[20]	−0.0012(7)(1)[7]
qed-ss	−0.03(3)(1)[4]	−0.013(8)(15)[17]	−0.020(0)(31)[31]	−0.002(15)(15)[21]

TABLE 4.14: Same as Tab. 4.11 but for the IB decomposition of the strange connected contribution. Note that strong isospin-breaking corrections are absent in the strange contribution. Remember that the errors are to be understood on the last digits of the central value.

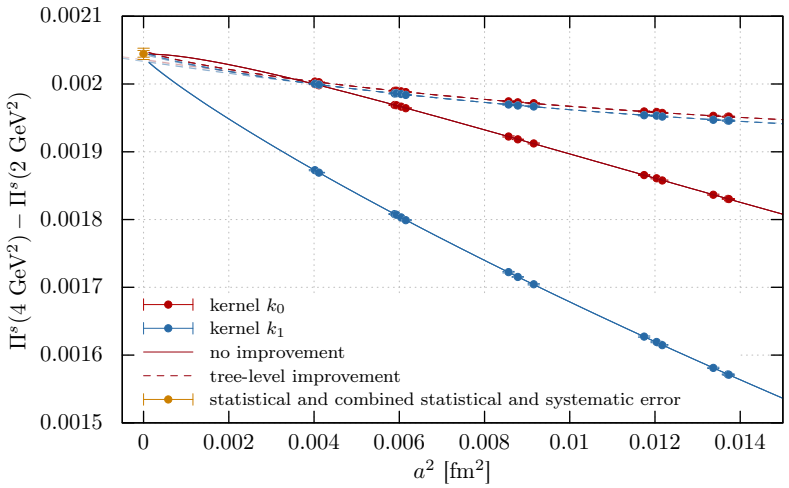


FIGURE 4.20: Same as Fig. 4.14, but for $\Delta\Pi^s(4\,\mathrm{GeV}^2)$, and with the subtraction of discretization errors using LO staggered perturbation in the full time range.

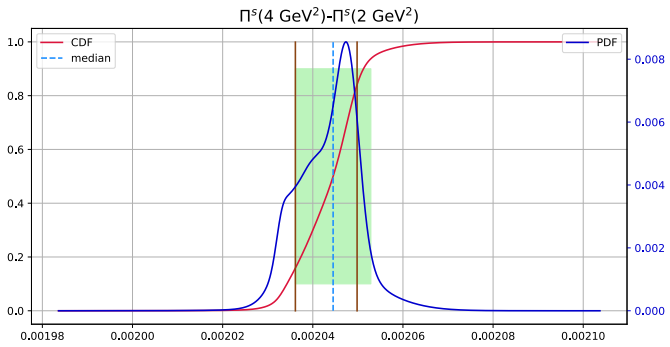


FIGURE 4.21: Same as Fig. 4.17, but for $\Delta\Pi^s(4\,\mathrm{GeV}^2)$.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}(2\,\mathrm{GeV}^2)$	$\hat{\Pi}(4\,\mathrm{GeV}^2)$	$\hat{\Pi}(8\,\mathrm{GeV}^2)$	$\Delta\Pi((8,1)\,\mathrm{GeV}^2)$
median	60.4	80.8	101.70	60.08
total error	4 (0.60%)	4 (0.44%)	35 (0.35%)	17 (0.28%)
statistical error	2 (0.28%)	2 (0.23%)	19 (0.19%)	11 (0.18%)
systematic error	3 (0.53%)	3 (0.38%)	30 (0.29%)	12 (0.21%)

TABLE 4.15: Same as Tab. 4.12, but for the strange connected contribution to $\hat{\Pi}(Q^2)$. Note that the errors are to be understood on the last digits of the central value.

By summing the various quantities in Tab. 4.13 via the sampling procedure outlined in Sec. 4.3.2.4 we obtain the strange contribution to the Euclidean HVP function, $\hat{\Pi}^s(Q^2)$ for $Q^2 = 1, 2, 4, 8\,\mathrm{GeV}^2$. The results, together with the total, statistical and full systematic error are given in Tab. 4.15.

4.3.3.3 Connected charm contribution

The charm current-current correlator is significantly more precise than the light and strange contributions. If we use the full statistics for the charm contribution to the discrete Adler function $\Delta\Pi(Q^2)$ it becomes very difficult to perform the global fit and get reasonable values of χ^2/d.o.f. For this reason in the evaluation of the charm contribution we restrict our analysis to 48 configurations per each ensemble, uniformly distributed over the whole set of configurations.

We also use 24 random sources. In general, the analysis for the charm contributions is analogous to the light and strange part. As for the strange, we can safely neglect finite volume and taste-breaking corrections which are even smaller here. We also do not study isospin-breaking corrections for the same reason.

Apart from the conserved current we use other current discretizations, as detailed in Sec. 4.3.1.2, which helps constrain the continuum extrapolation of the charm contribution which is expected to be difficult due to discretization effects proportional to powers of am_c that can be as large as 0.66 on our coarsest lattices. In the global fit procedure we then simultaneously continuum extrapolate the observables O^J determined using the different current operators, labelled by "J" via:

$$O^J(a,Q^2) = O(0,Q^2) + \underbrace{A^J(a)}_{\substack{\text{cont.}\\ \text{extrap.}}} + \underbrace{B^J X_l + C^J X_s}_{\substack{\text{interpolation to}\\ \text{physical point}}} . \tag{4.127}$$

Note that observables corresponding to different currents have different discretization errors, but should give the same result in the continuum limit. Thus the continuum extrapolated values $O(0,Q^2)$ are constrained to be the same for all current discretizations. The coefficients of A^J, B^J and C^J are allowed to differ.

We have found for the charm contribution that the data before the LO perturbative lattice correction cannot be described reasonably by our functions, so in the analysis we used only the LO corrected correlator. The LO correction is performed separately for each staggered current. The data for different currents computed on a given ensemble are correlated, so we take these correlations into account in our global fit. We find that if we use all 12 current correlators available the covariance matrix is close to singular and the fits become ill behaved. For this reason we use only 5 different current correlators in the analysis. These include the one constructed from the conserved current (V) and four currents which have the largest and the smallest values, so that we can constrain the continuum extrapolated values for our observables from above and below.

In Tab. 4.16 we summarize information about the different fit functions employed. Since we do not include IB corrections in the charm contribution and since we do not include fits to unimproved data, the number of fits that we perform for each quantity is reduced to 2304. All other parameters are always set to zero. In Tab. 4.16 we also summarize our final results for the corresponding discrete Adler function and for the various contributions to systematic uncertainties. Here, the uncertainty originating from the scale/physical point setting, in particular from the various fit ranges used to extract the Ω mass and from the continuum extrapolated value of the M_{ss} meson mass, become comparable to the uncertainties originating from the continuum extrapolations.

In Fig. 4.22 we present the dependence, on the lattice spacing, of the (tree-level corrected) HVP and discrete Adler function $\hat{\Pi}^c(1\ \text{GeV}^2)$, $\Delta\Pi^c(2\ \text{GeV}^2)$, $\Delta\Pi^c(4\ \text{GeV}^2)$, $\Delta\Pi^c(8\ \text{GeV}^2)$ on the lattice spacing.

In Tab. 4.17 we summarize the results for the charm connected contribution to the Euclidean HVP function, $\hat{\Pi}^c(Q^2)$ for $Q^2 = 1,2,4,8\,\text{GeV}^2$. These are obtained by summing the HVP and the discrete Adler functions.

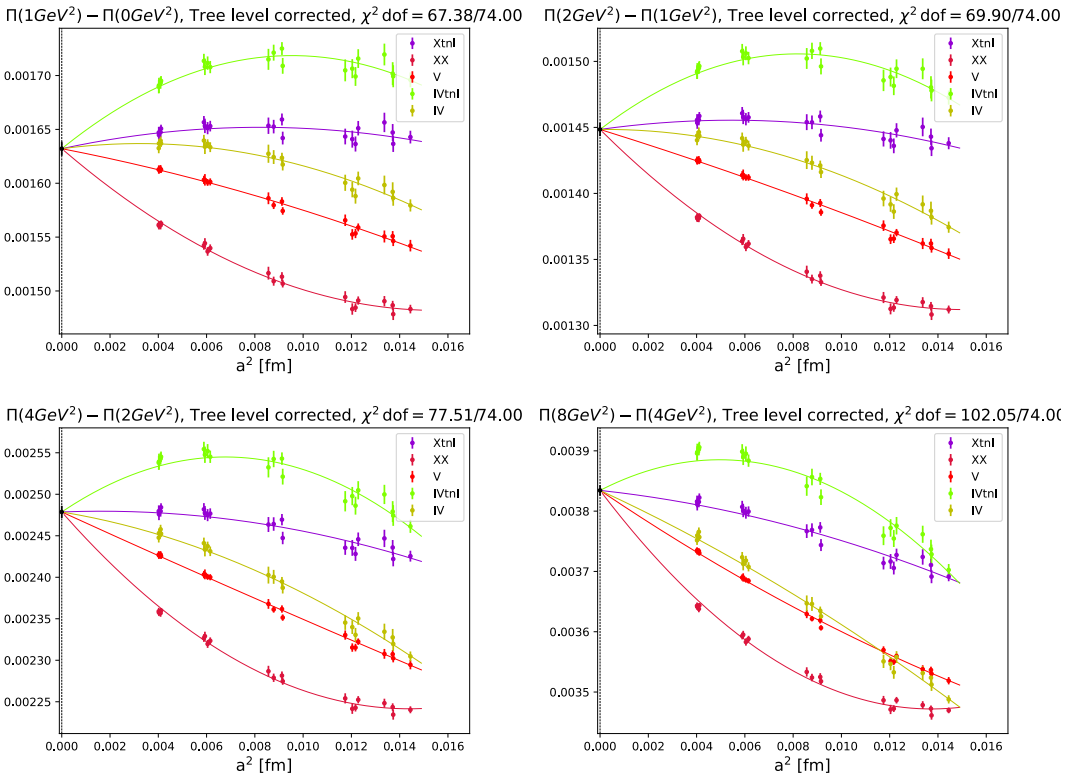


FIGURE 4.22: Illustrative examples for continuum extrapolations of the connected charm contribution to $\hat{\Pi}(1\,\mathrm{GeV}^2)$ (upper left), $\Delta\Pi(2\,\mathrm{GeV}^2)$ (upper right), $\Delta\Pi(4\,\mathrm{GeV}^2)$ (lower left), $\Delta\Pi(8\,\mathrm{GeV}^2)$ (lower right). For each quantity we depict a typical fit, together with the associated χ^2/dof. The datapoints have undergone a subtraction of discretization effects based on LO staggered perturbation theory (see App. B.2 for explicit expressions) prior to performing the fit. The different colours correspond to the five different staggered current discretizations of the quark electromagnetic current correlator which we fit simultaneously and constrain to extrapolate to the same point in the continuum limit. In plotting these curves only the dependence of the corresponding fit function on a^2 is kept: the dependence on the other observables, used to define the physical point of 2+1+1 flavour QCD (see (4.100)), is fixed by setting those observables to their physical value. Similarly, the fit function is used to interpolate the corresponding lattice points to the physical value of each of the observables of (4.100). Finally, the black point at $a^2 = 0$ corresponds to our final continuum extrapolated value for this particular fit, including its statistical uncertainty.

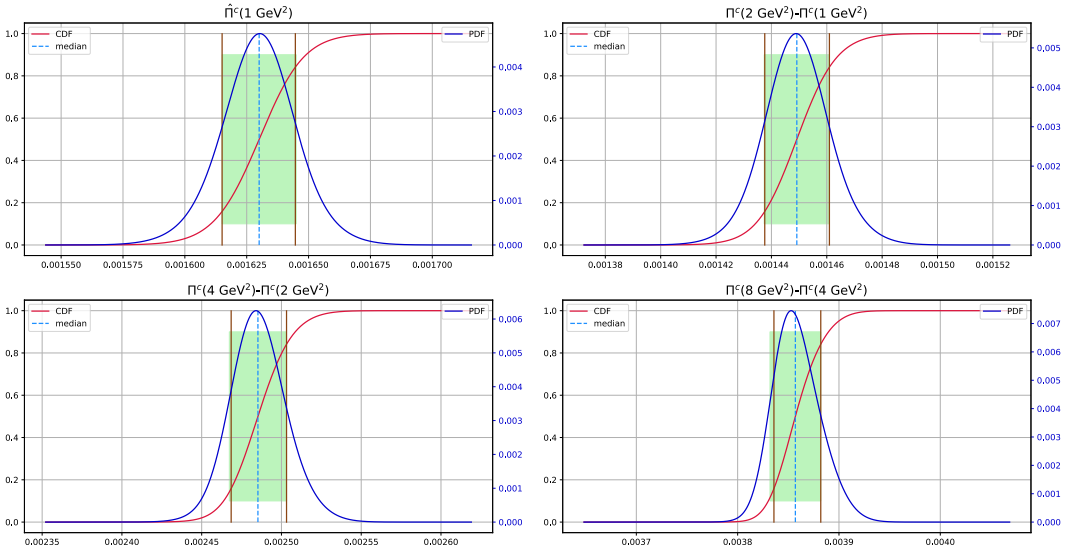


FIGURE 4.23: Same as Fig. 4.17, but for $\hat{\Pi}^c(1\,\mathrm{GeV}^2)$ (upper left), $\Delta\Pi^c(2\,\mathrm{GeV}^2)$ (upper right), $\Delta\Pi^c(4\,\mathrm{GeV}^2)$ (lower left), $\Delta\Pi^c(8\,\mathrm{GeV}^2)$ (lower right), respectively.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^c(1\ \mathrm{GeV}^2)$	$\Delta\Pi^c(2\ \mathrm{GeV}^2)$	$\Delta\Pi^c(4\ \mathrm{GeV}^2)$	$\Delta\Pi^c(8\ \mathrm{GeV}^2)$
Number fits	2304	2304	2304	2304
median	16.30	14.49	24.85	38.57
total error	15 (0.92%)	12 (0.81%)	18 (0.72%)	25 (0.65%)
statistical error	14 (0.84%)	11 (0.73%)	15 (0.62%)	18 (0.47%)
systematic error	6 (0.38%)	5 (0.35%)	9 (0.37%)	17 (0.45%)
M_π fit	<1	<1	<1	<1
$M_{\pi,\mathrm{IB}}$ fit	<1	<1	<1	<1
M_Ω fit	2	2	3	4
$M_{\Omega,\mathrm{IB}}$ fit	<1	<1	<1	<1
$M_{\Omega,\mathrm{phys}}$	1	1	1	1
$M_{ss,\mathrm{phys}}$	4	4	5	7
kernel	1	1	1	2
anomalous dim.	<1	1	4	11
tree-level	on	on	on	on
cuts	2	1	3	4
A_0 on/off	on	on	on	on
A_2 on/off	on	on	on	on
A_{2l}, A_4 on/off	<1	<1	<1	<1
B_0 on/off	on	on	on	on
B_2 on/off	on	on	on	on
C_0 on/off	on	on	on	on
C_2 on/off	on	on	on	on

TABLE 4.16: Same as Tab. 4.10 but for the connected charm contribution. Remember that the errors are to be understood on the last digits of the central value.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}(2\,\mathrm{GeV}^2)$	$\hat{\Pi}(4\,\mathrm{GeV}^2)$	$\hat{\Pi}(8\,\mathrm{GeV}^2)$	$\Delta\Pi((8,1)\,\mathrm{GeV}^2)$
median	30.79	55.6	94.2	77.9
total error	26 (0.86%)	4 (0.77%)	6 (0.68%)	5 (0.64%)
statistical error	24 (0.79%)	4 (0.71%)	6 (0.61%)	4 (0.56%)
systematic error	10 (0.33%)	2 (0.31%)	3 (0.30%)	2 (0.31%)

TABLE 4.17: Same as Tab. 4.12, but for the charm connected contribution to $\hat{\Pi}(Q^2)$. Note that the errors are to be understood on the last digits of the central value.

4.3.3.4 Disconnected contribution

In Tab. 4.18 we list the results and the error budgets for the disconnected contribution to our set of observables. As in the case of the light connected contribution, we split $\hat{\Pi}(1\,\mathrm{GeV}^2)$ into a short-distance window from $t = 0 - 0.4$ fm and into a long-distance one from $t = 0.4 - \infty$ fm. For the long-distance window we apply the same analysis procedure as in the case of the light connected contribution. We use two values for the anomalous dimension $n = 0, 3$ and we improve the continuum limit by applying the taste-improvement procedure based on the SRHO model. However, here we do not include the fits to the unimproved data in the final analysis, as it does not significantly alter the result. We do not apply tree-level improvement since the disconnected component of the correlator vanishes at leading order, see Sec. 4.3.2.1. For the same reason we do not include a logarithmic term in the short-distance components. For all quantities (and for the long-distance window of $\hat{\Pi}^d(1\,\mathrm{GeV}^2)$ after applying the taste-improvement) we find an A_4 term that is consistent with zero and therefore use only linear continuum extrapolations. For

all quantities except for $\Delta\Pi^d(2\,\mathrm{GeV}^2)$ and $\Delta\Pi^d(4\,\mathrm{GeV}^2)$ we even find that a constant describes the a^2-dependence of the isospin-symmetric contribution reasonably well and we include the variation $A_2 = 0$ and A_2 free as a systematic uncertainty. The number of fits we have to perform in the case of the disconnected component gets therefore significantly reduced for the discrete Adler function. In particular, we find that all fits (and in the case of $\hat{\Pi}(1\,\mathrm{GeV}^2, 0.0\,\mathrm{fm}, 0.4\,\mathrm{fm})$ more than 95%) have a p-value larger than 5%.

The plots of the continuum extrapolations and the corresponding plots of the probability and cumulative distribution functions for the disconnected component are collected in App. E.

In Tab. 4.19 we summarize the results for the isospin-breaking components for the disconnected contribution. As for the previous quantities, we also provide in Tab. 4.20 the results for the disconnected contribution to the Euclidean $\hat{\Pi}^d(Q^2)$ in discrete Q^2-steps.

4.3.3.5 Hadronic running of α up to 8 GeV^2: lattice results

Based on the sampling procedure outlined in Sec. 4.3.2.4, we collect below the results for the sum, preserving correlations, of the connected light, strange, charm and disconnected contributions to the Euclidean running of the electromagnetic coupling up to $Q^2 = 8$ GeV^2. Tab. 4.21 contains the median, total, statistical and full systematic error for the HVP function $\hat{\Pi}(Q^2)$ and for our step scaling function. The contributions of the different steps in the discrete Adler function are more or less of the same size since the running is logarithmic (see Sec. 2.2) and we perform the step scaling in powers of two in the Euclidean momentum transfer squared. While for the steps above $1\,\mathrm{GeV}^2$ we have reached a very small relative error of $0.2 - 0.3\%$, the error on the low energy contribution is still above 1%. The total error is in general dominated by systematic effects, except for the running from $1\,\mathrm{GeV}^2$ to $8\,\mathrm{GeV}^2$, where the charm contribution, which is statistics dominated, becomes important. In Fig. 4.24 we show the probability and cumulative distribution functions for the running of the sum of connected light, strange, charm and disconnected contributions to $\hat{\Pi}(Q^2)$. They clearly display a double-peak structure, mainly originating from the variation between improvement and no improvement of the data prior to the continuum extrapolation.

In Tab. 4.22 we provide the IB decomposition for the four-flavour hadronic vacuum polarization function up to $Q^2 = 8\,\mathrm{GeV}^2$, based on the decomposition scheme introduced in Sec. 4.3.2.4. Isospin-breaking effects arising from electromagnetic interactions play a sizable role and contribute $\sim 1\%$ to $\hat{\Pi}(1\,\mathrm{GeV}^2)$ (they are negative). At higher momenta, these contribute only at the sub-percent level. IB effects originating from the difference in the light quark masses m_u and m_d contribute much less, $\sim 0.2\%$ at $1\,\mathrm{GeV}^2$ and become negligible at higher energies. It is instructive to compare the various flavour contributions, as well as FV corrections, that we have collected in Tab. 4.23. We have also included the contributions from the bottom quark, which were obtained using the results from Ref. [13]. The sum of the connected light, strange, charm, disconnected, bottom and finite-size contributions yields the final five-flavour hadronic $\hat{\Pi}(Q^2)$, given in the last coloumn of Tab. 4.23. Note that the sum of the connected light, strange, charm and disconnected contributions was obtained by performing the sampling procedure, and the results were given in Tab. 4.21. The finite-size and bottom contributions were simply added, and the errors were added in quadrature since these contributions are completely independent

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^d_{00-04}(1\ \mathrm{GeV}^2)$	$\hat{\Pi}^d_{04-\infty}(1\ \mathrm{GeV}^2)$	$\Delta\Pi^d(2\ \mathrm{GeV}^2)$	$\Delta\Pi^d(4\ \mathrm{GeV}^2)$	$\Delta\Pi^d(8\ \mathrm{GeV}^2)$
Number fits	18432	36864	4608	4608	9216
median	-0.0026	-4.6	-0.165	-0.046	-0.0076
total error	5 (−18.93%)	7 (−14.86%)	12 (−7.33%)	4 (−8.37%)	12 (−16.12%)
statistical error	3 (−12.16%)	4 (−8.01%)	5 (−3.04%)	1 (−1.43%)	8 (−9.97%)
systematic error	4 (−14.52%)	6 (−12.51%)	11 (−6.67%)	4 (−8.25%)	10 (−12.66%)
M_π fit	<1	<1	<1	<1	<1
$M_{\pi,\mathrm{IB}}$ fit	<1	<1	<1	<1	<1
M_Ω fit	<1	<1	<1	<1	<1
$M_{\Omega,\mathrm{IB}}$ fit	<1	<1	<1	<1	<1
$M_{\Omega,\mathrm{phys}}$	<1	<1	<1	<1	<1
$M_{ss,\mathrm{phys}}$	<1	<1	<1	<1	<1
kernel	3	<1	<1	<1	2
anomalous dim.	1	<1	10	3	1
tree-level	off	off	off	off	off
taste imp.	off	2	off	off	off
cuts	1	<1	2	<1	<1
A_0 on/off	on	on	on	on	on
A_2 on/off	2	<1	on	on	2
A_{2l} on/off	off	off	off	off	off
A_4 on/off	off	off	off	off	off
B_0 on/off	off	off	off	off	off
C_0 on/off	<1	<1	off	off	on
D_0 on/off	on	on	on	on	on
D_a on/off	off	3	on	on	on
D_l on/off	<1	off	off	off	off
D_s on/off	off	off	off	off	off
E_0 on/off	on	on	on	on	on
E_a on/off	off	off	on	on	6
E_l on/off	on	off	<1	<1	off
E_s on/off	off	off	off	off	off
F_0 on/off	on	on	on	on	on
F_a on/off	off	off	off	off	off
G_0 on/off	on	on	on	on	on
G_a on/off	off	off	off	off	off

TABLE 4.18: Same as Tab. 4.10 but for the disconnected contribution. Remember that the errors are to be understood on the last digits of the central value.

from our four-flavour lattice $\hat{\Pi}(Q^2)$. The connected light contribution accounts for ~ 80% of the total running of α up to $8\,\mathrm{GeV}^2$ (remember that $\Delta^{(5)}_{\mathrm{had}}(Q^2) = e^2\hat{\Pi}^{(5)}(Q^2)$). At low energies, from $Q^2 = 0$ to $Q^2 = 1\,\mathrm{GeV}^2$, it contributes almost 90%. The charm contribution becomes of course more important at higher energies, whereas the disconnected contribution is more important at long distances. Finite size corrections are in general small and give a correction within the total error. On the final five-flavour running from $Q^2 = 0$ to $Q^2 = 8\,\mathrm{GeV}^2$ we have already reached a sub-percent relative precision of ~ 0.6%.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^d_{00-04}(1\ \mathrm{GeV}^2)$	$\hat{\Pi}^d_{04-\infty}(1\ \mathrm{GeV}^2)$
iso	−0.0015(2)(4)[4]	−3.3(2)(5)[5]
qed	−0.0535(13)(14)[19]	−4.9(11)(1)[11]
qed-vv	−0.0518(8)(13)[15]	−4(1)(0)[1]
qed-vs	−0.0007(7)(0)[7]	0.2(5)(0)[5]
qed-ss	−0.0010(6)(1)[6]	−0.9(6)(0)[6]
sib	0.00382(10)(16)[19]	−0.92(12)(29)[32]

$O(Q^2)\cdot 10^4$	$\Delta\Pi^d(2\ \mathrm{GeV}^2)$	$\Delta\Pi^d(4\ \mathrm{GeV}^2)$	$\Delta\Pi^d(8\ \mathrm{GeV}^2)$
iso	−0.116(4)(11)[12]	−0.032(0)(4)[4]	−0.0034(3)(4)[5]
qed	−0.289(15)(4)[16]	−0.144(8)(3)[8]	−0.064(6)(12)[13]
qed-vv	−0.285(13)(4)[13]	−0.133(6)(3)[7]	−0.061(6)(12)[13]
qed-vs	0.000(8)(0)[8]	−0.0022(31)(2)[31]	0.0012(12)(1)[12]
qed-ss	−0.004(7)(1)[7]	−0.0085(32)(2)[32]	−0.0031(12)(1)[12]
sib	−0.0224(14)(9)[17]	−0.00177(9)(9)[12]	0.00179(14)(6)[15]

TABLE 4.19: Same as Tab. 4.11 but for the IB decomposition of the disconnected contribution. Note that the errors are to be understood on the last digits of the central value.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^d(1\,\mathrm{GeV}^2)$	$\hat{\Pi}^d(2\,\mathrm{GeV}^2)$	$\hat{\Pi}^d(4\,\mathrm{GeV}^2)$	$\hat{\Pi}^d(8\,\mathrm{GeV}^2)$	$\Delta\Pi^d((8,1)\,\mathrm{GeV}^2)$
median	-4.6	-4.8	-4.8	-4.8	-0.218
total error	7 (14.85%)	7 (14.35%)	7 (14.21%)	7 (14.19%)	12 (5.57%)
statistical error	5 (8.00%)	4 (7.73%)	4 (7.66%)	4 (7.65%)	6 (2.61%)
systematic error	6 (12.52%)	6 (12.09%)	6 (11.97%)	6 (11.95%)	11 (4.92%)

TABLE 4.20: Same as Tab. 4.12, but for the disconnected contribution to $\hat{\Pi}(Q^2)$. Note that the errors are to be understood on the last digits of the central value.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^{(4)}(1\,\mathrm{GeV}^2)$	$\hat{\Pi}^{(4)}(2\,\mathrm{GeV}^2)$	$\hat{\Pi}^{(4)}(4\,\mathrm{GeV}^2)$	$\hat{\Pi}^{(4)}(8\,\mathrm{GeV}^2)$
median	408	551	706	872
total error	5 (1.32%)	5 (0.98%)	5 (0.76%)	5 (0.61%)
statistical error	0 (0.07%)	0 (0.05%)	0 (0.04%)	0 (0.03%)
systematic error	5 (1.32%)	5 (0.97%)	5 (0.76%)	5 (0.61%)
$O(Q^2)\cdot 10^4$	$\Delta\Pi^{(4)}(2\,\mathrm{GeV}^2)$	$\Delta\Pi^{(4)}(4\,\mathrm{GeV}^2)$	$\Delta\Pi^{(4)}(8\,\mathrm{GeV}^2)$	$\Delta\Pi^{(4)}((8,1)\,\mathrm{GeV}^2)$
median	143.28	154.3	166.4	464.0
total error	33 (0.23%)	4 (0.26%)	5 (0.28%)	8 (0.18%)
statistical error	19 (0.13%)	2 (0.16%)	3 (0.18%)	6 (0.13%)
systematic error	26 (0.18%)	3 (0.20%)	4 (0.21%)	5 (0.12%)

TABLE 4.21: Euclidean four-flavour HVP function $\hat{\Pi}^{(4)}(Q^2)$ up to $Q^2 = 8\,\mathrm{GeV}^2$, in steps of powers of two, together with the corresponding discrete Adler function. These results for $\hat{\Pi}^{(4)}(Q^2)$ were obtained by summing the individual contributions to $\hat{\Pi}(1\,\mathrm{GeV}^2)$ and to the discrete Adler function in Tabs. 4.10, 4.13, 4.16 and 4.18, following the sampling procedure outlined in Sec. 4.3.2.4 in order to preserve possible correlations. The errors are to be understood on the last digits of the central value and we apply the PDG conventions for the rounding, which we base on the total error [2].

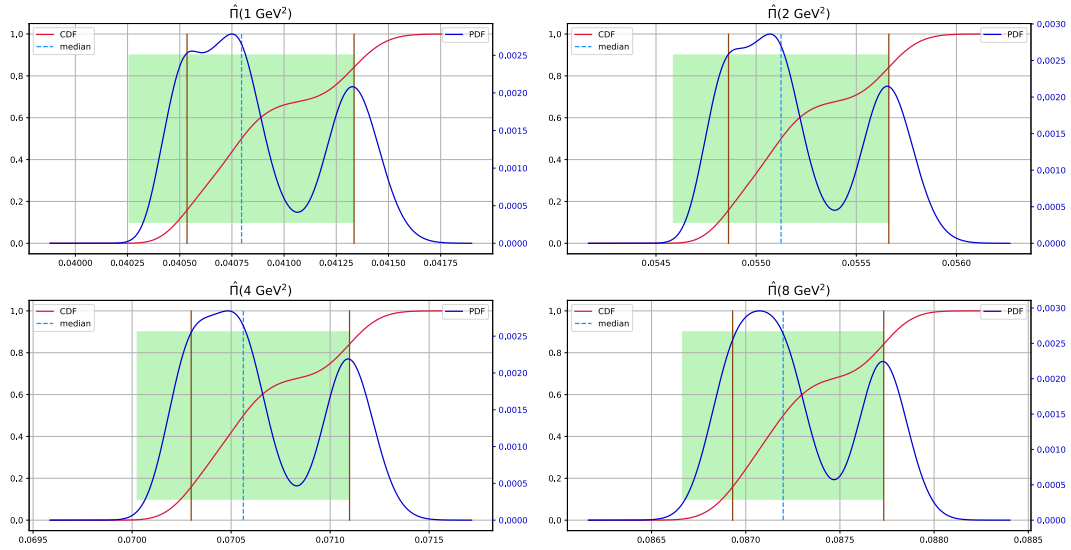


FIGURE 4.24: Same as Fig. 4.17, but for the sum of connected light, strange, charm and disconnected contribution to $\hat{\Pi}(1\,\mathrm{GeV}^2)$ (upper left), $\hat{\Pi}(2\,\mathrm{GeV}^2)$ (upper right), $\hat{\Pi}(4\,\mathrm{GeV}^2)$ (lower left), $\hat{\Pi}(8\,\mathrm{GeV}^2)$ (lower right), respectively.

$O(Q^2)\cdot 10^4$	$\hat{\Pi}^{(4)}(1\,\mathrm{GeV}^2)$	$\hat{\Pi}^{(4)}(2\,\mathrm{GeV}^2)$	$\hat{\Pi}^{(4)}(4\,\mathrm{GeV}^2)$	$\hat{\Pi}^{(4)}(8\,\mathrm{GeV}^2)$
iso	407(0)(5)[5]	551(0)(5)[5]	705(0)(5)[5]	871(0)(5)[5]
qed	−7.0(14)(4)[14]	−6.6(14)(4)[15]	−6.2(15)(4)[15]	−5.8(15)(4)[15]
qed-vv	−4.5(10)(2)[10]	−4.3000(10)(2)[11]	−4.0(10)(2)[11]	−3.7(10)(2)[11]
qed-vs	−0.4(7)(0)[7]	−0.5(7)(0)[7]	−0.5(7)(0)[7]	−0.5(7)(0)[7]
qed-ss	−2(1)(0)[1]	−1.9(9)(3)[10]	−1.7(10)(3)[10]	−1.6(10)(3)[11]
sib	1.11(24)(25)[35]	1.21(24)(25)[35]	1.25(24)(25)[35]	1.27(24)(25)[35]
$O(Q^2)\cdot 10^4$	$\Delta\Pi^{(4)}(2\,\mathrm{GeV}^2)$	$\Delta\Pi^{(4)}(4\,\mathrm{GeV}^2)$	$\Delta\Pi^{(4)}(8\,\mathrm{GeV}^2)$	$\Delta\Pi^{(4)}((8,1)\,\mathrm{GeV}^2)$
iso	143.15(19)(26)[32]	154.3(2)(3)[4]	166.3(3)(4)[5]	463.7000(6)(5)[8]
qed	0.33(12)(15)[19]	0.43(8)(9)[12]	0.41(7)(5)[9]	1.15(23)(17)[29]
qed-vv	0.21(5)(4)[6]	0.33(2)(6)[6]	0.34(0)(12)[12]	0.85(8)(10)[12]
qed-vs	−0.091(24)(2)[24]	−0.056(14)(1)[14]	−0.026(7)(1)[7]	−0.17(4)(0)[4]
qed-ss	0.21(9)(17)[19]	0.2(1)(1)[1]	0.1(1)(1)[1]	0.47(20)(15)[25]
sib	0.101(7)(2)[7]	0.0379(13)(23)[26]	0.0131(6)(3)[7]	0.153(9)(2)[9]

TABLE 4.22: Isospin-breaking decomposition of the Euclidean four-flavour HVP function $\hat{\Pi}^{(4)}(Q^2)$ up to $Q^2 = 8\,\mathrm{GeV}^2$, in steps of powers of two, together with the corresponding discrete Adler function. These results for $\hat{\Pi}^{(4)}(Q^2)$ were obtained by summing the individual contributions to $\hat{\Pi}(1\,\mathrm{GeV}^2)$ and to the discrete Adler function in Tabs. 4.11, 4.14, 4.16 (only for the isospin-symmetric contribution) and 4.19, following the sampling procedure outlined in Sec. 4.3.2.4 in order to preserve possible correlations. The errors are to be understood on the last digits of the central value and we apply the PDG conventions for the rounding, which we base on the total error [2]

4.3.4 Comparison with previous results

In Fig. 4.25 we compare our results for the five-flavour HVP function $\hat{\Pi}^{(5)}(Q^2)$ with the recent lattice determination by the Mainz collaboration [8], with two earlier lattice determinations of our collaboration [6, 7] and with a data-driven determination using data from the KNT collaboration from Ref. [14]. The results from our collaboration from 2017 [6] do not contain isospin-breaking corrections. We add our determinations from Tab. 4.22 with 100% uncertainties since our IB separation schemes slightly differ. Our results include the contribution from the bottom quark listed in Tab. 4.23. In Ref. [8] the authors provide a parametrization of their

$O(Q^2)\cdot 10^4$	O^{light}	O^{strange}	O^{charm}	O^{disc}	$\infty -$ ref	O^{bottom}	$O^{(5)}$
$\hat{\Pi}(1\,\text{GeV}^2)$	355[5]	41.6[4]	16.30[15]	−4.6[7]	3.10[22]	0.34[5]	411[5]
$\hat{\Pi}(2\,\text{GeV}^2)$	465[5]	60.4[4]	30.79[26]	−4.8[7]	3.24[25]	0.7[1]	555[5]
$\hat{\Pi}(4\,\text{GeV}^2)$	574[5]	80.8[4]	55.6[4]	−4.8[7]	3.31[25]	1.33[19]	710[5]
$\hat{\Pi}(8\,\text{GeV}^2)$	681[5]	101.70[35]	94.2[6]	−4.8[7]	3.36[25]	2.6[4]	878[5]
$\Delta\Pi(2\,\text{GeV}^2)$	110.12[32]	18.81[8]	14.49[12]	−0.165[12]	0.135[12]	0.34[5]	143.75[35]
$\Delta\Pi(4\,\text{GeV}^2)$	109.06[34]	20.45[8]	24.85[18]	−0.046[4]	0.08[1]	0.7[1]	155.1[4]
$\Delta\Pi(8\,\text{GeV}^2)$	107.0[4]	20.48[14]	38.57[25]	−0.0076[12]	0.04[1]	1.27[19]	167.7[5]

TABLE 4.23: Euclidean five-flavour HVP function $\hat{\Pi}^{(5)}(Q^2)$ up to $Q^2 = 8\,\text{GeV}^2$, in steps of powers of two, together with the corresponding discrete Adler function. We give separately the light, strange, charm and disconnected contributions which are the final results presented in the last sections. We also list the finite-size corrections computed in Sec. 4.3.2.3. The bottom quark contribution was obtained using the work of the HPQCD collaboration [13]. The last column contains the sum of all contributions. Note that the sum of the light, strange, charm and disconnected contributions was obtained using the sampling procedure in Sec. 4.3.2.4 while the bottom and FV corrections were simply added, and their errors were added in quadrature. The errors are to be understood on the last digits of the central value and we apply the PDG conventions for the rounding [2].

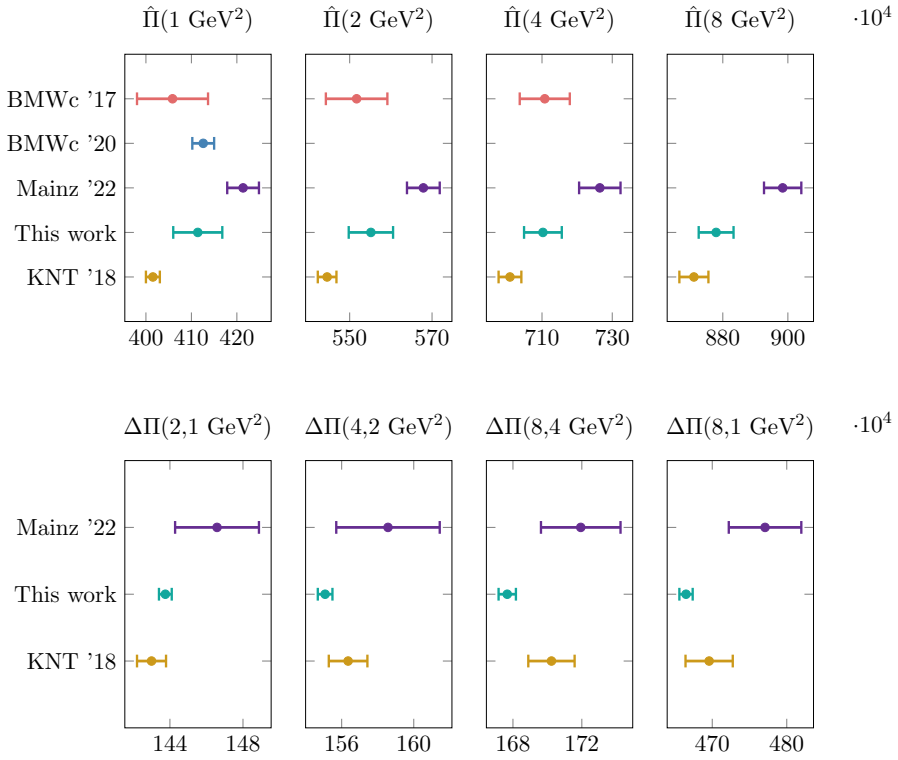


FIGURE 4.25: Compilation of results for the vacuum polarization function $\hat{\Pi}(Q^2)$ and its step scaling $\Delta\Pi(2Q^2, Q^2)$ in units of 10^4 for various values of Q^2 in GeV^2. The violet results are a recent lattice determination by the Mainz collaboration [8]. The authors in Ref. [8] provide a parametrization of their results based on Padé-approximants, including a correlation matrix. The yellow marks are the results from a data-driven determination based on the R-ratio data provided by KNT [14] from threshold to $Q = 11.199$ GeV and perturbative QCD to $\mathcal{O}(\alpha_s^4)$ using the package rhad [15] for the remaining integral. We also compare to two earlier determinations by our collaboration [6, 7].

results based on Padé approximants including a correlation matrix. In our comparison we do not include the results of the first lattice calculation of the hadronic contribution to the running of α by Burger et al. [226] who used twisted mass fermions. Indeed, that pioneering calculation is performed with heavier than physical pion masses and neglects the disconnected and isospin-breaking contributions. Moreover, the authors consider directly the bare HVP function $\Pi(Q^2)$ at non-zero lattice momenta ($Q = n2\pi/L$, $m2\pi/T$), forcing them to perform a model-dependent extrapolation in Q^2 to determine the subtraction constant $\Pi(0)$. All of these features make a

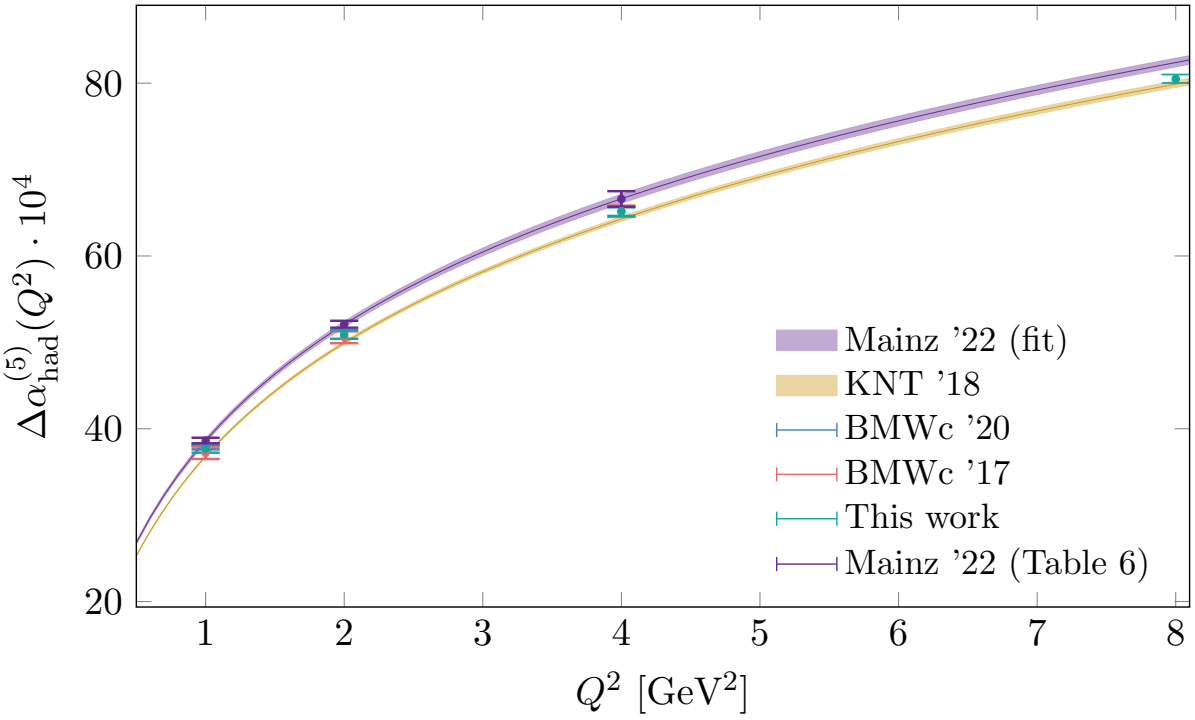


FIGURE 4.26: Comparison for the Euclidean running of $\Delta\alpha^{(5)}_{\rm had}(Q^2)$ up to $Q^2 = 8\,\text{GeV}^2$ in units of 10^4 obtained in this work, two earlier results from our collaboration [6, 7], a recent lattice determination by the Mainz collaboration [8] and a data-driven determination based on the R-ratio data provided by KNT [14] from threshold to $Q = 11.199$ GeV and perturbative QCD to $\mathcal{O}(\alpha_s^4)$ using the package rhad [15] for the remaining integral. The violet data points correspond to the values given in Tab. 6 of Ref. [8] while the curve was obtained using the parametrization based on Padé-approximants which the authors also provide.

reliable estimate of systematic errors difficult. Reassuringly, we agree well with our previous determinations, albeit with a larger error than the preliminary result of $\hat{\Pi}(1\,\text{GeV}^2)$ in Ref. [7]. We discuss below how we plan to reduce the uncertainty on this quantity. We agree in general quite well with the recent result from Mainz: the difference on $\hat{\Pi}(1\,\text{GeV}^2)$ and on each of the finite steps in the discrete Adler function lies between one and two sigma. However, our results lie systematically below those obtained by the Mainz collaboration. This difference may stem from the fact that they have not accounted for the logarithmically-enhanced discretization effect in the continuum extrapolations. As explained in detail in Sec. 4.3.2.1, the logarithmic term becomes important at small values of a and the function turns around. While the agreement in individual discrete Adler-function steps is good, once these are added up to obtain $\Pi(8\,\text{GeV}^2) - \Pi(1\,\text{GeV}^2)$, our result lies several units below the Mainz result.

The data-driven result is obtained by integrating R-ratio data (see Eq. (4.18)) from Ref. [14] from threshold to $Q = 11.199$ GeV, taking correlations into account via the correlation matrix that was also provided by the authors. For the remaining integral from $Q = 11.199$ GeV to ∞ we integrate the R-ratio obtained in perturbative QCD to $\mathcal{O}(\alpha_s^4)$ that we calculate using the Fortran-package rhad [15]. At $\mathcal{O}(\alpha_s^2)$ rhad includes the full mass dependence, at $\mathcal{O}(\alpha_s^3)$ quartic mass corrections and at $\mathcal{O}(\alpha_s^4)$ quadratic mass corrections. As input, we use the pole masses $m_c = 1.67 \pm 0.07$ GeV and $m_b = 4.78 \pm 0.06$ GeV and $\alpha(M_Z) = 0.1179 \pm 0.010$ [2]. We estimate the error due to the truncation of the perturbative calculation by performing the calculations at $\mathcal{O}(\alpha_s^3)$ and $\mathcal{O}(\alpha_s^4)$ and add the difference in quadrature. To take the uncertainties on $\alpha(M_Z)$ and on the pole masses into account, we perform the calculation for quantity-error and quantity+error and add

half the difference in quadrature to the uncertainty on $\hat{\Pi}^{(5)}(Q^2)$, assuming that the errors on the values of those quantities are not correlated. The central value is given by the calculation with the central values of the input.

While we find a slight excess compared to the data-driven approach for $\hat{\Pi}(1\,\mathrm{GeV}^2)$ of ~ 2.5%, we agree well for the discrete steps above $1\,\mathrm{GeV}^2$. Consequently, for $\hat{\Pi}(8\,\mathrm{GeV}^2)$ we already agree well within errors with the data-driven approach. Let us also remark that while our uncertainty is still larger than the uncertainty on the data-driven result for $\hat{\Pi}(1\,\mathrm{GeV}^2)$, our total errors on the discrete Adler function at larger energies are half as large.

As the plots show, the large errors on our determination of the running from 0 to Q^2 are fully determined by those on $\hat{\Pi}(1\ \mathrm{GeV}^2)$, where we are currently limited by the systematic error originating from the continuum extrapolation of the long-distance window of the light contribution. While taste-violations significantly distort the pion spectrum and lead to large cutoff effects we nevertheless included the fits to these data for a conservative error estimation, resulting in a relatively large systematic uncertainty at these low energies. A more thorough investigation of taste-violations and the addition of a smaller lattice spacing are underway. We anticipate the error on $\hat{\Pi}(1\,\mathrm{GeV}^2)$ to shrink significantly with the inclusion of an additional lattice spacing and a better understanding of the taste corrections. Our results become more precise as we go to higher values of Q^2, where we reach a sub-percent precision.

From the results in Fig. 4.25, the shift in the electromagnetic coupling is straightforwardly obtained via

$$\Delta\alpha^{(5)}_{\mathrm{had}}(Q^2) = e^2\hat{\Pi}^{(5)}(Q^2)\ . \tag{4.128}$$

In Fig. 4.26 we plot the hadronic contribution to the running of the electromagnetic coupling as a function of Euclidean Q^2 up to $8\,\mathrm{GeV}^2$. We plot our result along with the one from the data-driven determination based on R-ratio data from Ref. [15] combined with $\mathcal{O}(\alpha_s^4)$ perturbative QCD, from two earlier determinations from our collaboration [6, 7] and from the recent lattice determination from the Mainz collaboration [8]. For the Mainz determination, we also include the results the authors provide in Tab. 6 in Ref. [8]. While the latter are quite similar to those obtained from the parametrization for $\hat{\Pi}(1\,\mathrm{GeV}^2)$ and $\hat{\Pi}(2\,\mathrm{GeV}^2)$, the error on $\hat{\Pi}(4\,\mathrm{GeV}^2)$ is significantly larger. This is most probably due to the fact that the parametrization is a fit to all of their data (without taking into account correlations), thus reducing the error on individual values. In turn, this slightly increases the disagreement with our results.

While our results show no additional excess in the running at higher Q^2 compared to the data-driven determination, the Mainz lattice results tend to have an excess running in each of these intervals. Consequently, our result agrees with the data-driven determination at $Q^2 = 8\,\mathrm{GeV}^2$ within errors while the Mainz result finds an excess even at higher energies.

4.4 Conclusion

The value of the electromagnetic coupling at the Z-scale, $\alpha(M_Z^2)$, is an important input for precision tests of the standard model. The precision in the theoretical determination of this quantity is currently limited by the uncertainties in the hadronic contribution in the low-energy

region. While the traditional approach to calculate the hadronic running of α in this non-perturbative regime relies on a dispersion relation and requires experimental input from $e^+e^- \to$ hadrons cross-section data, lattice QCD provides an ideal tool to access this non-perturbative regime from first principles. In the first part of this thesis we report on an ab-initio lattice QCD calculation of the hadronic running of α in the space-like region up to $8\,\mathrm{GeV}^2$ that is obtained from 25 simulations with $2+1+1$ flavours of staggered fermions whose masses are tuned to their physical values, and with five lattice spacings down to 0.064 fm.

To avoid mixing systematic effects that may be very different in nature, we perform a flavour decomposition and split the running into long- and short-distance contributions at low energies. At higher virtualities we calculate the running in discrete steps of powers of two in the space-like momentum-transfer squared.

We include leading-order strong-isospin-breaking and electromagnetic effects in our calculations and perform a decomposition into iso-symmetric and isospin-breaking components. This is the first lattice study of the running of the electromagnetic coupling where these effects are studied in detail with lattice simulations.

To determine the running up to scales of a few GeV requires controlling large discretization artefacts and logarithmically-enhanced cutoff effects. For this purpose we design a number of improvement schemes, based on lattice perturbation theory, which we apply to our lattice data prior to performing continuum extrapolations. These improvements significantly reduce cutoff effects and allow us to take the continuum limit in a controlled fashion. Nevertheless, continuum-limit results obtained from corrected and from uncorrected data agree well. Moreover, the coefficients of the logarithmically-enhanced discetization effects, that we obtain in fits to unimproved data, are similar to those that we obtain in LO staggered perturbation theory. Thus, we consider both sets of fits in our analyses to quote conservative uncertainties in our final results. The charm contribution is, due to that quark's large mass, particularly plagued by discretization effects. To constrain its continuum extrapolation, we implement 48 different discretizations of the charm electromagnetic current and use a subset of five in the combined global fits that give our final results for this contribution.

At low momenta, finite-size effects effects become important. We calculate the necessary finite-volume corrections by performing a dedicated lattice study in a large volume of size $L_{\mathrm{big}}^3 \times T_{\mathrm{big}} = (10.752\ \mathrm{fm})^4$ and use NNLO chiral perturbation theory for the small residual extrapolation to infinite volume. At short distances and in the charm, strange and disconnected contribution finite-size effects can be neglected.

Our final result for the running of α up to $8\,\mathrm{GeV}^2$, $\Delta\alpha_{\mathrm{had}}^{(5)}(Q^2 = 8\,\mathrm{GeV}^2) = (80.0 \pm 0.5) \cdot 10^{-4}$, has already reached the same level of precision as the phenomenological determinations, see for instance Tab. VIII of Ref. [16]. For the contribution coming from the low-energy part of the running, $\hat{\Pi}(1\ \mathrm{GeV}^2)$, we are currently limited by the precision on the light contribution, which also constitutes the dominant contribution at that scale. At long distances, taste-violations in the pion spectrum introduce large discretization effects in this quantity. To remove these we use the so-called SRHO model above $t \geq 1.0$ fm, which describes well the discretization errors seen in our simulations, even down to $t = 0.4$ fm. Nevertheless, because we would like to confirm our understanding of taste-breaking effects before quoting very small error-bars, we estimate the

uncertainty on $\hat{\Pi}(1\ \mathrm{GeV}^2)$ very conservatively by also including fits to uncorrected data in our final estimate.

We compare our results to a recent lattice determination by the Mainz collaboration [8] who use different fermion and gluon discretizations (Wilson gauge action and $\mathcal{O}(a)$-improved Wilson fermions) in their simulations. For each individual step in the running our results agree with theirs within 1-2 standard deviations. However, our results lie systematically below those obtained by Mainz, resulting in a lower value for $\hat{\Pi}(8\,\mathrm{GeV}^2)$. These differences may be due to the fact that unlike Mainz we either include, in our parametrizations for the continuum extrapolations of uncorrected lattice results, logarithmically-enhanced discretization effects that are known to be present, or we improve the lattice results, before extrapolating them, with corrections computed to leading-order in staggered perturbation theory. Moreover, while we perform our calculations with the conserved staggered vector current, Mainz uses local-local and local-conserved correlation functions that must be renormalized, another possible source of systematic uncertainty.

We also compare our result to a data-driven determination based on R-ratio data from Ref. [14]. While we find a slight excess in the low-energy region, we agree well in the discrete Adler steps at higher energies. Hence, for $\Delta\alpha^{(5)}_{\mathrm{had}}(8\,\mathrm{GeV}^2)$ we already agree within errors with this data-driven determination. This is similar to what was found in the preliminary analysis in Ref. [7].

In a final, outlook chapter at the end of this thesis (Chap. IV), we discuss our strategy for reducing the uncertainties of the results presented here and possible extensions of our work to improve the running of α all the way up to M_Z.

Part III

Going beyond with axion-mediated dark matter

5 Dark Matter

Falls sich dies bewahrheiten sollte, würde sich also das überraschende Resultat ergeben, dass dunkle Materie in sehr viel grösserer Dichte vorhanden ist als leuchtende Materie.
(If this should be verified, it would lead to the surprising result that dark matter exists in much greater density than luminous matter.)
Zwicky 1933

5.1 A Brief Overview of Cosmology

In this section we will give a short overview of the basic assumptions, equations and terminology in cosmology [269]. We will encounter most of the functions and variables again in the following chapters. The framework together with the thermal history outlined below forms the basis of what is commonly called the standard model of cosmology.

The Friedmann-Lemaître-Robertson-Walker-metric describes a maximally symmetric space-time, i.e. a universe that is spatially homogeneous and isotropic on large scales. This is known as the cosmological principle. Modern cosmology is built on this metric and this principle. The invariant line element of the Friedmann-Lemaître-Robertson-Walker-metric can be written as

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \left(\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\right) \tag{5.1}$$

with χ the radial coordinate, $a(t)$ the scale factor and

$$S_k(\chi) = \begin{cases} \sin\chi & k = +1 \\ \chi & k = 0 \\ \sinh\chi & k = -1 \end{cases} \tag{5.2}$$

for a closed ($k = 1$), flat ($k = 0$), and open ($k = -1$) universe, respectively. Plugging the FLRW metric into Einstein's field equations

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ , \tag{5.3}$$

one obtains the Friedmann equations,

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho \qquad \text{1st Friedmann equation} \tag{5.4}$$

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2 + k}{a^2} = -8\pi G p \qquad \text{2nd Friedmann equation} \tag{5.5}$$

which describe the evolution of the universe under the cosmological principle. ρ is the total energy density, p the pressure and G Newton's gravitational constant. We use natural units. One also defines the *critical density*, the *density parameters for species i* and the *reduced density associated to curvature*, respectively, by

$$\rho_c(t) \equiv \frac{3H^2}{8\pi G}, \qquad \Omega_i \equiv \frac{\rho_i}{\rho_c} = \frac{8\pi G}{3H^2}\rho_i, \qquad \Omega_k \equiv -\frac{k}{H^2 a^2} \quad \overset{\text{1. Friedmann}}{\Rightarrow} \quad \sum_k \Omega_k = 1. \tag{5.6}$$

We have introduced the Hubble function as $H = \dot{a}/a$ characterizing the rate of expansion of the universe. As we will see in the following section, the density parameter Ω_{DM} of Dark Matter is measured by the Planck telescope and gives us the DM relic density today [270].

In the following chapters we will have to distinguish between particle species. To account for this it is convenient to work with effective parameters. In the early universe, curvature is negligible and radiation dominates. The first Friedmann equation then reads

$$H(T)^2 = \frac{8\pi}{3} G \rho_{\text{tot}}(T) \qquad \text{with} \qquad \rho_{\text{tot}}(T) = \frac{\pi^2}{30} g_* T^4 \tag{5.7}$$

and we have introduced the *effective number of relativistic degrees of freedom* (the contribution of non-relativistic degrees of freedom can be neglected) at temperature T as

$$g_*(T) = \sum_{i\in \text{ boson}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i\in \text{ fermion}} g_i \left(\frac{T_i}{T}\right)^4. \tag{5.8}$$

The factor 7/8 comes from the difference between Bose and Fermi statistics. For example, when all SM particles are still relativistic, we obtain $g_b = 28$ and $g_f = 90$, giving $g_* = g_b + 7/8 g_f = 106.75$ as the SM effective number of relativistic degrees of freedom. Since the SM particles were in thermal equilibrium in the early universe the total energy density $\rho_{\text{tot}}(T)$ is given by its equilibrium expression for relativistic particles. However, we take into account species that have decoupled from the plasma by introducing their temperatures T_i. We will give more details on equilibrium relations in Sec. 5.4.

The total entropy in a comoving volume sa^3 is conserved under equilibrium conditions and the total entropy density can hence be expressed as

$$s = \sum_i \frac{\rho_i + P_i}{T_i} \quad \overset{\mathrm{d}(a^3 s)=0}{\equiv} \quad \frac{2\pi^2}{45} g_{*s}(T) T^3, \tag{5.9}$$

where we have defined the *effective number of degrees of freedom in entropy* g_{s*} as

$$g_{*s}(T) = \sum_{i\in \text{ boson}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{i\in \text{ fermion}} g_i \left(\frac{T_i}{T}\right)^3 \tag{5.10}$$

Figure 5.1 shows $\sqrt{g_*(T)}$ and $g_{*s}(T)$ for the SM particles.

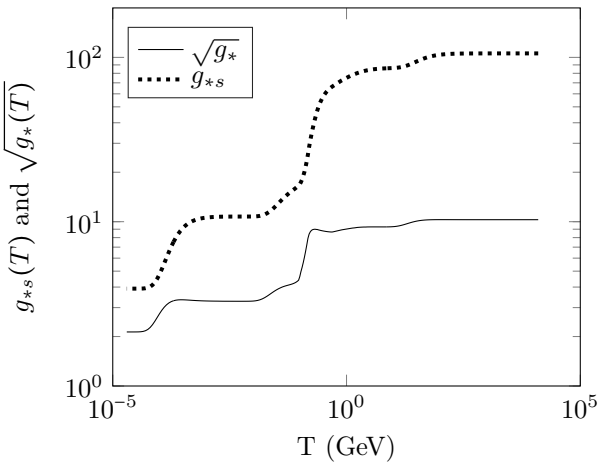


FIGURE 5.1: Effective number of degrees of freedom $\sqrt{g_*(T)}$ and effective number of degrees of freedom in entropy $g_{*s}(T)$ as a function of temperature for SM particles. Note that as the universe evolves in time it cools down (time evolution from right to left).

5.2 A brief history of the universe: gathering evidence for Dark Matter

As the authors in [271] state: *Most publications and presentations on this topic – whether at a technical or a popular level – either ignore the long history of this field or condense it into a brief anecdotal account, typically centered around the work on galaxy clusters by Fritz Zwicky in the 1930s and on galactic rotation curves by Vera Rubin in the 1970s.* I apologize in advance for doing exactly this; this thesis treating two very different subjects, I will only sketch the evidence for dark matter in an oversimplified summary. Instead, I will give a brief overview of the standard cosmological history, as we understand it today. This will also give me the chance to introduce several observables and notations that we will encounter again in the course of this thesis. For a more complete treatment of the evidence and the history of dark matter I refer the reader to some excellent references [91, 271–279]. Some of the events listed below are hypothesized (inflation and baryogenesis), others are predicted by our current understanding of the SM but not yet experimentally validated (electroweak phase transition, QCD phase transition). At pertinent points we will further mention the evidence for dark matter.

Let me briefly remark that this is not the first time people inferred the existence of some "dark" matter only from gravitational interaction, a great example being for instance the discovery of Neptune by Le Verrier and Adams in 1846 who (independently) observed the orbits of Uranus, compared them to predictions from Newtonian gravity and detected some anomalies. At that time, Neptune was dark matter.

Fritz Zwicky is often identified as the "discoverer" of what we call dark matter today. In his paper "Die Rotverschiebung von extragalaktischen Nebeln" from 1933 he investigated the measurements of radial velocities of galaxies in the nearby Coma cluster [280]. By applying the virial theorem he found that the mean density inside the cluster should be about 400 times higher than the mean density obtained from observations of visible matter. The radial velocities being much higher than expected for the observed visible mass, the visible mass would not prevent the galaxies

from escaping the cluster. His conclusion was that - if this turned out to be true - there must be additional non-luminous, "dark" matter [280].

In the first half of the 20th century only a few astrophysicists pointed to the problem of missing matter. However, in the second half with growing observational possibilities more indications could be gathered – from radial velocities of gas around a galaxy, over distorted images of far-off objects, to the dynamics of intracluster media which became detectable with the advent of X-Ray astronomy, manifesting the model of "Dark Matter" in astrophysics. A second, important step towards a general acceptance came from modern cosmology and its model of cosmological structure formation. The standard model of Cosmology is crucially based on a dark component.

Inflation Inflation was developed in the late 70's and early 80's by Alan Guth, Andrei Linde and Alexei Strobinsky as a solution to several fine-tuning problems in the conventional Big Bang theory, such as the *flatness*-problem (if the the universe is not fine-tuned initially as exactly flat, the density evolves away from the critical density) and the *horizon*-problem (causally disconnected regions are homogeneous). Inflation resolves these problems, allowing the universe to arise from generic initial conditions, by postulating an early period of accelerated expansion which changes the early universe's causal structure. This is usually realized by some scalar field (the inflaton) which dominated the energy density in the universe during inflation, slowly rolling down its potential. Inflation ends when it reaches the minimum of the potential where it oscillates and decays into SM particles. This process is called reheating and happened at the reheating temperature, T_{RH} – note however that reheating is not an instantaneous process. Inflationary models constrain $T_{\rm RH} \lesssim 10^{16}$ GeV whereas it could be as low as $T_{\rm RH} \gtrsim \mathcal{O}(\text{MeV})$, constrained by Big Bang Nucleosynthesis (see below) [281, 282].

Baryogenesis Baryogenesis aims to answer the question why our universe contains lots of matter but very little anti-matter. To set up the matter-antimatter asymmetry the so-called *Sakharov conditions* must be satisfied (baryon number B and CP symmetry must be broken, a deviation from thermal equilibrium is required) [283, 284]. SM physics cannot account for the observed matter-antimatter asymmetry in the universe [285]. Many models for baryogenesis have been proposed but none has been experimentally confirmed.

Electroweak phase transition We already briefly discussed the physics of the electroweak phase transition in Chap. 2. We saw that electroweak symmetry $SU(2)_L \times U(1)_Y$ is spontaneously broken to $\mathrm{U}(1)_{\rm em}$, giving masses to other particles via the Higgs mechanism. The form of the Higgs potential is of course put in by hand. However, using the techniques of thermal field theory, it is possible to explain, based on SM physics, how and when the symmetry breaking occurred. In the finite temperature theory framework the one-loop effective Higgs potential as a function of temperature can be calculated. One finds for the temperature dependence of the effective Higgs potential $V_{\rm eff}$ that for $T > T_c$ the theory is $SU(2)_L \times U(1)_Y$ symmetric, i.e. the minimum of $V_{\rm eff}$ is zero, while for $T \lesssim T_c$ the minimum at zero becomes unstable. Numerically, one finds $T_c \approx 163$ GeV.

Until today it is not clear of what order the electroweak phase transition was, but it is believed to be a smooth crossover [286].

QCD phase transition As we discussed in the first part of this thesis, QCD confines at low energies and quarks and gluons form colour-neutral bound states (hadrons). The QCD phase transition happened around temperatures of $T \sim 160$ MeV [287]. Above this temperature, hadrons are "melted" and constitute a quark-gluon plasma.

Below the confinement scale, perturbation theory breaks down and we cannot use the standard QCD Lagrangian to do perturbative calculations. In the first part of this thesis we used lattice QCD to study these low energy effects (however in vacuum). Another possibility is to build an effective field theory, such as chiral perturbation theory, where the relevant degrees of freedom are hadrons instead of quarks and gluons. These methods will become important later in this thesis, when we study the phenomenological consequences and detection strategies of axion-like particles.

Lattice QCD calculations at finite temperature strongly suggest that the QCD phase transition is in fact a crossover [287, 288].

Neutrino decoupling Neutrinos are kept in equilibrium with the SM bath via the weak interactions. They hence remained in chemical equilibrium as long as interactions were rapid enough compared to the Hubble expansion rate, $\Gamma \gtrsim H$. As we will see in Sec. 5.4, the interaction rate can be parametrized as $\Gamma \sim \langle \sigma v \rangle n$ where σ is the cross section of the neutrinos with SM particles, v their velocity and the angled brackets indicate a phase space average. For relativistic particles, $\Gamma \sim G_F^2 T^5$ with $G_F \approx 1.17 \times 10^{-5}$ GeV^{-2} Fermi's constant and since $H \sim T^2/M_{\rm pl}$, we have

$$\frac{\Gamma}{H} \sim G_F^2 M_{\rm pl} T^3 \sim \left(\frac{T}{\rm MeV}\right)^3 .$$

Hence, at a temperature $T \sim 1$ MeV neutrinos decoupled from the SM bath and afterwards moved freely. Since they decoupled earlier than photons (the photon decoupling happening at $T \sim 13.6$ eV and providing us with the CMB, see below), a direct observation of the neutrino background would give us crucial insights about the universe at earlier times. However, due to their weak interactions an observation of the neutrino background with current technology is not yet possible.

After neutrinos decoupled, their temperature simply got redshifted as $T_\nu \propto a^{-1}$, and neutrinos continued to share the same temperature as the SM bath. A departure from the photon temperature occurred during the next event.

Electron-positron annihilation Once the temperature dropped below the mass of the electron, it became more and more difficult for photons to annihilate into an e^+e^- pair. However, the backreaction $e^+e^- \to \gamma\gamma$ continued to be efficient, increasing the photon temperature relative to the neutrino temperature due to this energy injection. Using the effective number of degrees of freedom in entropy before and after e^+e^--annihilation freeze-out, one finds that after e^+e^--annihilation freeze-out the neutrino temperature scaled as

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma$$

relative to the photon temperature. From this relation one can infer the total radiation density

$$\rho_R = \frac{\pi^2}{30} g_\gamma T_\gamma^4 + \frac{\pi^2}{30}\frac{7}{8} g_\nu T_\nu^4 = \frac{\pi^2}{30}\left\{2 + \frac{7}{4} N_{\rm eff}\left(\frac{4}{11}\right)^{4/3}\right\} T_\gamma^4 \tag{5.11}$$

(the $\frac{7}{8}$ coming from the fact that neutrinos are fermions), where we have introduced $N_{\rm eff}$, the effective number of neutrino species. Planck data constrains the value $N_{\rm eff} = 3.00^{+0.57}_{-0.53}$ at 95 % C.L. from Planck TT+lowE (note that theory predicts $N_{\rm eff} \approx 3.046$ as neutrinos were not instantaneously decoupled at e^+e^--annihilation) [270]. Hence, measurements of $N_{\rm eff}$ provide a powerful tool to test new physics scenarios, as we will discuss further in Sec. 7.5.3.

Big Bang Nucleosynthesis Big bang nucleosynthesis describes the formation of the light elements (as D, ^{3}He, ^{4}He, Li). At temperatures above ~ 1 MeV, protons and neutrons were maintained in thermal equilibrium by the weak interactions,

$$n + \nu_e \leftrightarrow p + e^- \,, \qquad n + e^+ \leftrightarrow p + \bar{\nu}_e \,.$$

The neutron-to-proton relative number densities were hence given by their equilibrium values,

$$\frac{n_n}{n_p} \sim \exp\left(-\frac{m_n - m_p}{T}\right) \,,$$

with $m_n - m_p = 1.293$ MeV the small mass difference between neutrons and protons. Hence, for temperatures far above this value, $n_n \approx n_p$, whereas below $T \sim 1$ MeV the n_n/n_p decreases, until neutrinos decouple at $T \sim 0.8$ MeV and the neutron-to-proton ratio freezes out at the value 1/5.

Around $T \sim 0.7 - 0.05$ MeV, an important fraction of neutrons start decaying and nuclei of light elements are being produced by a chain of reactions, p + n → D, D + p → ^{3}He, D + D → ^{4}He, leaving us with a final helium to hydrogen mass fraction $X_4 \equiv 4n_{\rm He}/n_{\rm H} \sim 1/4$. In practice, a complicated system of coupled Boltzmann equations has to be solved in order to make precise predictions about the abundances of light elements. These predictions can then be compared to the measured primordial abundances through astrophysical observations. The standard BBN scenario agrees very well with observations, hence providing us with powerful constraints on the possibility of BSM particles (except for the measurement of ^{7}Li which disagrees with the value extracted from CMB data – the so-called *Lithium-problem* – which is hence often omitted as a BBN constraint [270]).

Finally, while evidence for dark matter coming from e.g. rotation curves or gravitational lensing could be explained by faint objects made of known matter, this is not the case for that coming from BBN. The exact light element abundances depend on the photon-to-baryon ratio, $\eta = n_\gamma/n_b$. Since n_γ can be obtained from the CMB temperature, BBN constrains $0.021 \leq \Omega_b h^2 \leq 0.024$ [2, 289]. Hence, baryons cannot make up for all the matter content in the universe. This is also shown in one of the plots in Fig. 5.2, where the primordial abundances of ^{4}He, D, ^{3}He and ^{7}Li as predicted by the standard model of BBN are depicted. The yellow boxes indicate the measured light element abundances.

Matter radiation equality At $T_{\rm eq} \sim 0.8$ MeV, the energy density in matter won over the energy density in radiation. Density perturbations evolved from then on differently, namely proportional to the scale factor. Baryonic perturbations can only start growing after photon decoupling. Hence, there is a tight interplay between $\Omega_m h^2$, the amount of matter in the universe today, which influences $T_{\rm eq}$, $\Omega_b h^2$ today and structure formation.

Recombination and last scattering: the Cosmic Microwave Background (CMB) Above $T \sim 1$ eV, photons maintained kinetic equilibrium with electrons via Compton scattering, $e^-\gamma \to e^-\gamma$, and the universe consisted of a plasma of photons, free electrons and nuclei, an ionised plasma. Below $T \sim 13.6$ eV, the ionisation energy of hydrogen, the photons started to cease to have enough energy to ionize hydrogen (however, the high-energy tail of the photon distribution still did). At around $T \sim 0.32$ eV electrons and nuclei could combine and formed neutral atoms (*recombination*). At some point, there were no longer enough free electrons around to maintain equilibrium with the photons and matter and radiation decoupled (*last scattering*, at $T \sim 0.26$ eV). The photons subsequently continued their path through space rather freely, preserving this primordial density map and providing us with a treasure trove of crucial information about the early history of the universe. The cosmic microwave background (CMB) is the most perfect naturally existing black body ever observed, with wavelengths of a few centimetres which is observable in all directions - an isotropic and homogeneous (temperature) distribution. Since then, the form of the spectrum of this radiation has not changed, but it became redshifted due to the ongoing expansion. Today, the spectrum is simply a rescaled version of the distribution function at decoupling. The temperature of the CMB we measure today is around 2.7 K.

Although the CMB temperature is globally isotropic and homogeneous, small anisotropies of $\mathcal{O}(10^{-4}$ K) which have become more and more precisely determined during the various missions can reveal information about structure formation. More precisely, from the CMB power spectrum, important information about the matter composition of our universe can be obtained:

$$\frac{\Delta T(\theta,\phi)}{\bar{T}} \equiv \frac{T(\theta,\phi)-\bar{T}}{\bar{T}} = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm} Y_l^m(\theta,\phi) \tag{5.12}$$

The angular power spectrum is given by the expansion parameters of the temperature fluctuations in spherical harmonics, $C_l = \left\langle |a_{lm}|^2 \right\rangle$. While baryonic matter feels the pressure of photons, dark matter collapses freely. These effects show up in the characteristic peaks of the CMB angular temperature power spectrum – for an excellent overview see [290] –

- Sachs-Wolfe effect: variations in the graviational potential lead to temperature variations (at low l)
- Acoustic peaks: competition between gravity and pressure leads to oscillations in the photon-baryon fluid (at intermediate l)
- Silk damping tail: diffusion of photons damps fluctuations (at high l)

The matter composition (namely a baryonic component and a dark matter component only interacting gravitationally with photons) of the universe strongly influences positions and/or

amplitudes of these peaks. Currently, the existing models best reconstructing the CMB need DM with an amount [270]

$$\left.\begin{aligned}\Omega_b h^2 &= 0.0224 \pm 0.0001\\ \Omega_m h^2 &= 0.1430 \pm 0.0011\end{aligned}\right\}\Omega_{DM} h^2 = 0.1200 \pm 0.001, \tag{5.13}$$

where Ω_m is the density parameter of the overall matter content, Ω_b the one of baryons and Ω_{DM} the one of dark matter (cf. section 5.1). Hence, dark matter is the dominant matter component in our universe.

Ω_m and Ω_Λ can also be obtained from measurements of type-1a supernovae (so-called "standard candles"). The best fit models of the data consistently agree with CMB data.

Reionization After the formation of the first stars, photons had again enough energy to ionize hydrogen, although not enough to re-establish equilibrium and the universe remained transparent. However, reionization indicates the period of the lighting of the first stars.

Structure formation After decoupling, the initially small primordial density fluctuations grew more than 10 orders of magnitude and led to the formation of stars, galaxies, galaxy clusters and so on. As we already mentioned, this cannot be achieved by only baryonic matter and the way these perturbations grew depends on the matter composition: DM begins clumping at matter-radiation equality while baryons only cluster after photons have decoupled. Hence, evidence for dark matter nowadays also comes from the methods of cosmological N-body simulations which model the physics of structure formation. These simulations made it possible to accurately reconstruct the formation of large-scale structures based on the current ΛCDM model.

Today Throughout this section we have been gathering evidence for the existence of dark matter, based on our current understanding of the cosmological history. Today we have several additional tools from astrophysics at hand, that have led to strong evidence for a missing matter component:

- **Gravitational lensing** is the art of determining the mass distribution of very heavy objects in the sky by observing how light emitted from various objects bends under the influence of the induced gravitational field. This tedious technique allows to efficiently reconstruct mass density profiles. One of the main breakthroughs of observational evidence of dark matter are the results from weak lensing of the bullet galaxy cluster. The bullet cluster technically consists of two colliding galaxy clusters that go through several separation and collision phases until they eventually merge to one cluster. Fig. 5.2 contains a plot of the distribution of galaxies, X-rays and a reconstruction of the mass distribution of all components obtained with weak lensing techniques. One would expect that after the collision of these clusters, the stars should not be much affected by the crash, interacting only gravitationally. However, the gas that surrounds these stars should interact electromagnetically, leaving the gas behind the stars at the collision point. This is indeed what has been observed. However, when determining the mass points of the system by

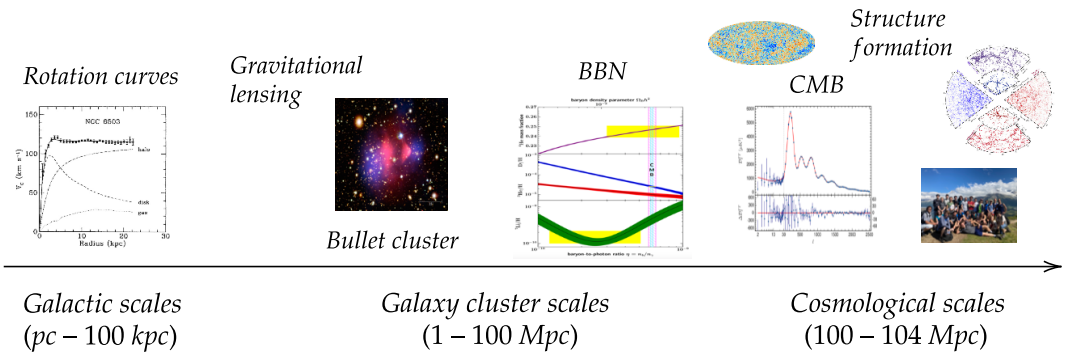


FIGURE 5.2: Compilation of some of the evidence for dark matter. Today we have evidence for dark matter at scales from galactic to clusters of galaxies to cosmological.

weak lensing techniques, one observes that they lie essentially in the center of the separated galaxies. Yet, one would instead expect them to be in the center of the gas, since the gas in galaxies accounts for a much bigger fraction of matter than stars. Again, there must be some missing mass that additionally is - if at all - interacting very weakly and that is therefore not slowed down by the collision. The bullet cluster is quite often seen as the main indication for the existence of dark matter; existing modified gravity theories failing here as these models so far do not distinguish between forms of matter. A nice feature of this observation is that it provides us with an estimate on the upper limit of the self-interaction cross section of dark matter, $\sigma_{\chi\chi}/m_\chi \lesssim 1$ cm^2/g [91].

- **Galactic rotation curves** refer to the circular velocities of stars and gas in the galactic disks. These curves can be obtained by comparing the spectra of galaxies to the standard neutral hydrogen spectrum from which the velocity around the centre as a function of the distance to the centre can be deduced. If one assumes that the galaxy is cylindrical and by applying basic Newtonian mechanics, one finds a relation between the radial mass distribution $M(r)$ of the galaxy and the radial velocity $v_c(r)$,

$$v_c(r) = \sqrt{\frac{GM(r)}{r}} \tag{5.14}$$

with G Newton's gravitational constant and $M(r)$ the enclosed mass. From Newtonian mechanics, $v_c(r)$ should fall off as $1/\sqrt{r}$ ("Keplerian fall-off"). However, none of the observed radial velocities around galaxies shows this fall-off at all, on the contrary, the velocity distribution is constant at large radii (see for example [275]), leading again to the conclusion that either the law of gravity has to be modified or that some missing mass is responsible. In addition, this missing mass should have a distribution $\rho(r)$ according to

$$\rho(r) = \frac{1}{4\pi r^2}\frac{\mathrm{d}M(r)}{\mathrm{d}r} = \frac{v_c^2}{4\pi G r^2}, \tag{5.15}$$

i.e. a linear increase of the dark matter halo mass [91].

Today the evidence for dark matter has converged to an anomaly at all scales, remarkably agreeing on its abundance. It ranges from galactic (1 pc-100 kpc) to clusters of galaxies (1-100

Mpc) to cosmological scales (100 - 104 Mpc), as sketched in Fig. 5.2 [1]. The standard model of cosmology has been extremely successful in describing the thermal history of our universe. However, the exact nature of the "dark components", dark matter and dark energy, remains to be understood. As we underlined several times above, while the existence of dark matter is established only by gravitational means, there is great evidence that the missing mass must be non-baryonic – the most accepted solution being a (or several) particle(s) beyond the standard model. In this thesis, we will investigate the possibility for an unknown particle making up the "dark matter" and will hence use the term "dark matter" in this context.

5.2.1 What we know about the missing mass

What do we know about DM? If we assume that DM is a particle and we place ourselves in the standard ΛCD model, we can gather information about its properties. The DM particle(s) should

- **have a relic density** of $\Omega_{\rm DM}h^2 = 0.120 \pm 0.001$, measured by the Planck telescope [270].
- **be (almost) neutral**, i.e. not interact too strongly with QED or QCD, otherwise we would have already detected it. However, a small electrical charge ($\epsilon \sim 10^{-6} \times e$ for DM masses $\sim 1 - 60$ MeV) is not excluded [291].
- **be stable** over cosmological time scales, more concretely, it needs to have a lifetime longer than the age of the universe, $\tau_{\rm DM} > 4 \times 10^{17}$ s. From the CMB and structure formation stronger bounds can be inferred, $\tau_{\rm DM} \gtrsim 10^{19}$ s ($\tau_{\rm DM} \gtrsim 10^{25} - 10^{29}$ s if the DM decays into photons or baryons) [292].
- **be cold**, i.e. not have high relative velocities. Hot dark matter, such as neutrinos, would have led to a top-down structure formation (first large superclusters of galaxies formed which fragmented into smaller clusters, galaxies and so on) whereas observations indicate that galaxies are older than clusters, hence strongly favouring a bottom-up formation.
- **not self-interact too strongly.** We will come back to this point in more detail in Sec. 5.3.
- **be massive.** However, it is fair to say that we do not have much of a clue about what the mass of the DM should be: its allowed mass ranges over nearly 80 orders of magnitude. If DM is comprised of bosons, its mass is constrained at the lower end by its de-Broglie wavelength which should not exceed the size of the smallest galaxies (simply because it has to fit in), leading to $m_{\rm DM} \gtrsim 10^{-22}$ eV [293]. If the DM particle is a fermion, the lower bound is more constraining, $m_{\rm DM} \gtrsim 0.13$ keV at 95% C.L. (Tremaine-Gunn bound [294, 295]), therefore excluding SM neutrinos. This bound comes from Pauli's exclusion principle limiting the number of fermions that can be compressed into a halo ("phase-space packing"). The upper limit is set by the Planck-scale, $m_{\rm DM} < 10^{19}$ GeV.

[1] Finally, we are evidence for dark matter! This is of course a bit far-fetched but I take the opportunity to sneak in a picture of me and my friends at the Les Houches summer school on dark matter.

5.2.2 Why the standard model is not enough

Following Ockham's razor one is tempted to look for a particle which could serve as a DM candidate in the established theory, the SM. However, neither quarks, nor the Higgs or the EW gauge bosons, nor charged leptons can do the trick: The quark content is bound by experiments, the Higgs and the EW gauge bosons are too short-lived and charged leptons interact electromagnetically and would cause a strongly electrically charged universe [276, 277]. What is left are neutrinos which once were seen as an attractive DM candidate, the interactions being of weak nature as predicted by the freeze-out/WIMP mechanism. But when neutrinos are produced in the hot and early universe they have a very small mass and thus behave ultra-relativistically. After decoupling, they free-stream and thereby spoil structure formation (more precisely, hot DM as neutrinos require top-down structure formation, whereas the bottom-up structure formation provided by cold DM shows better agreement with observation). Moreover, the amount of neutrinos in the universe is bound by experiments today [2]. This leads us to the conclusion that there must be physics beyond the standard model (BSM).

5.3 Detection strategies: shake it, make it, break it, brake it?

After having outlined the gravitational evidence for DM, this chapter will give a summary of the various detection strategies of DM, the key hypothesis being that the DM couples to the SM. To test the DM-SM interactions one can think of three complementary methods, depicted in figure 5.3. Firstly, DM interacting with SM particles leads us to the idea to build detectors on Earth where DM particles should in principle constantly arrive. Since we are expecting weakly interacting particles we need massive detectors in order to increase the sensitivity. This method is called *direct detection* (shake it!). Secondly, in very dense regions in the cosmos, such as centres of galaxies, we expect DM particles to collide and annihilate or to decay into SM particles. We can therefore look for primary or secondary annihilation, decay products coming from regions in the galactic halo where we expect abundant DM particles and see if we can detect anomalous signals. This approach is called *indirect detection* (break it!). Last but not least, by colliding SM particles at particle colliders DM could be produced and lead to a signal in the form of missing energy or displaced vertices (make it!) [2]. We will review each of these below.

We included two additional signatures in Fig. 5.3: Measurements of the relic density provide us with a constraint that any DM model featuring a specific DM production mechanism has to verify (bake it!). Finally, the possibility that DM self-interacts could help resolving (some of) the small-scale structure problems in the standard ΛCDM scenario. However, the self-interaction strength is constrained by a variety of observations from merging clusters or halo shapes (brake it?).

Lastly, let me already mention that since in our model the DM particle interacts only very feebly with the SM particles and additionally via a pseudoscalar mediator, the couplings between the DM particle and the SM are too small to lead to an observable signal at current and near future

[2] The nice descriptions for the detection strategies (make it, break it, shake it) come from Jesse Thaler.

direct, indirect detection or collider experiments (for the production mechansisms freeze-in and decoupled freeze-out that we focused on). We will verify this explicitly in Sec. 7.5.4.2. In this section, I will therefore only briefly outline the basic principles of these main DM detection strategies.

Shake it! With a local DM density of $\rho_0 = 0.3$ GeV/cm^3 DM particles should constantly hit the Earth. The basic principle is that the DM particles recoil, excite or ionize the nuclei (or the electrons) of some target material resulting in heat (phonons), light (photons) or charge (ionization) as signals. In most attempts the emphasis lies on detecting interactions through elastic collisions with the target nuclei since radioactivity gives more or less the same signal as inelastic scattering processes (γ-rays) and therefore presents a strong background [296]. The key challenge is therefore to distinguish the signal from potential background events, which is also why most of the existing laboratories are located deep down under the Earth's surface. On the other hand, since DM is (if at all) weakly interacting with SM particles, the target should be big to augment the chance for an interaction. While many direct detection experiments target WIMPs with masses around the electroweak scale, there have been huge advancements in probing Sub-GeV DM with electron recoil [297, 298].

In order to compare theory to experimental data we parametrize the interaction by the differential scattering rate for DM scattering on nuclei. This differential scattering rate parametrizes the number of events at a given energy per unit time and mass of the detector material. In general, the differential scattering rate $\mathrm{d}R_T/\mathrm{d}E_R$ in direct detection is then given by [277, 299]

$$\frac{\mathrm{d}R_T}{\mathrm{d}E_\mathrm{R}} = \frac{\rho_\chi}{m_\chi}\frac{C_T}{m_T}\int_{v\geqslant v_\mathrm{min}^{v_\mathrm{esc}}(E_\mathrm{R})} d^3v\ f(\vec{v},t)\ v\ \frac{d\sigma_T}{dE_\mathrm{R}}(E_\mathrm{R},\vec{v})\ , \tag{5.16}$$

where ρ_χ is the local DM density, m_χ is the DM mass, m_T and C_T are the target's mass and mass fraction, respectively, $f(\vec{v},t)$ is the DM velocity distribution in the Earth's rest frame, convoluted with $d\sigma_T/dE_\mathrm{R}$, the differential DM nucleus cross section. For the DM velocity distribution one usually assumes a Maxwell-Boltzmann distribution in the galactic restframe. v_min is the minimal DM velocity needed for the target to scatter with recoil energy E_R and depends on the momentum transfer and the reduced DM-nucleus mass. v_esc is the escape velocity above which the DM is no longer gravitationally bound. Integrating the left-hand-side of eq. (5.16)

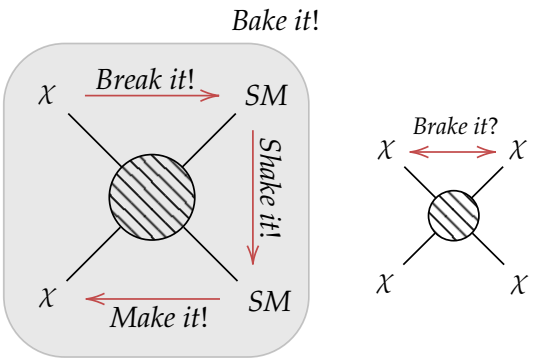


FIGURE 5.3: Different mechanisms for detection of Dark Matter: Shake it (direct detection), make it (collider searches), break it (indirect detection), bake it (relic density), brake it (self-interactions). Some of the nice descriptions come from Jesse Thaler.

over the recoil energy gives the event rate. The particle physics part entering the expression for the differential scattering rate is the differential cross section for the DM-nucleus interaction.

One can further divide the DM-nucleus cross section into a spin-dependent and spin-independent contribution. If the DM couples to the nucleon number A – this is the case for scalar or vector couplings between the DM and the quarks – the cross-section is spin-independent. For axial-vector couplings, however, the DM couples to the nucleon spin, giving rise to a spin-dependent cross-section. The spin-independent cross-section is coherently enhanced, $\sigma_{N\chi} \propto A^2$, significantly increasing the sensitivity as compared to the spin-dependent case where the spins add incoherently and scale as $\sigma_{N\chi} \propto J(J+1)$ with J the spin. In particular, for nuclei with an equal number of protons and neutrons, $n_{\text{protons}} = n_{\text{neutrons}}$, the spin-dependent cross-section vanishes [277]. However, in general, describing the differential cross section in Eq. (5.16) is complicated and highly model dependent. Hence, in order to compare experimental data to the wealth of possible interactions in a model-independent way, the common approach is to formulate the differential cross section in the framework of effective field theories [300–304]. We will come back to this point in Sec. 7.5.4 when we calculate direct detection constraints on the model we study in this thesis.

As we discussed above, direct detection experiments put constraints on the scattering cross section $\sigma_{\chi N}$ depending on the DM mass. Figure 5.4 summarizes recent bounds from various direct detection experiments. The turquoise band shows the results from XENONnT which targeted 100 GeV WIMPS. So far, no DM candidate has been detected but a wide parameter space could be excluded[3]. The excluded regions severely constrain models in which the DM particle is produced via thermal freeze-out. This has led to a paradigm shift towards various other possibilities for dark matter candidates, with masses above 1 TeV or below 10 GeV, and with other production mechanisms. In particular, in the model we study in Chap. 7, we will focus on scenarios where the DM-SM couplings are too weak to allow for DM generation via standard freeze-out and the DM is thermally decoupled from the SM bath. Note that direct detection experiments can even be sensitive to freeze-in, see for instance [320, 321].

Make it! Collider experiments set bounds on the parameter space of DM by missing energy analyses. To produce a signal at a collider, the DM particles need to interact sufficiently strongly with the SM particles. The focus lies therefore on WIMP-like particles. In order to decide whether an event at hadron colliders has the potential to be saved and analyzed, one needs a trigger. For dark matter searches, these are events with missing energy above a certain threshold. To detect a missing energy event, another particle has to be produced, such that we can observe a missing transverse three-momentum (we only know that the three-momentum in the transverse plane has to add up to zero). Due to difficulties in reconstructing the entire recoil, only events with $E_{\text{T,missing}} \gtrsim 100$ GeV can safely be distinguished from background. In order to trigger on a missing transverse momentum, most searches typically look for mono-jet, mono-photon or mono-Z (so-called mono-X) signatures in the initial state, depicted in Fig. 5.5. These events are

[3]There was however reported an excess below ~ 7 keV in the XENON1T DM experiment which favoured certain BSM scenarios such as a solar axion model [305]. In the XENONnT DM experiment, no excess above background was found, excluding their previous BSM explanations of the XENON1T excess [306]. DAMA/LIBRA have been reporting annual modulations over 14 cycles with a significance of 12.9σ for recoil energies of $2-6$ keV [307]. However, this result is in contradiction with other direct detection experiments, in particular with recent results from COSINE-100 and ANAIS with sensitivity to the same processes that find results compatible with no annual modulation and in the case of ANAIS, incompatibe with DAMA/LIBRA at 3.3σ [308–311].

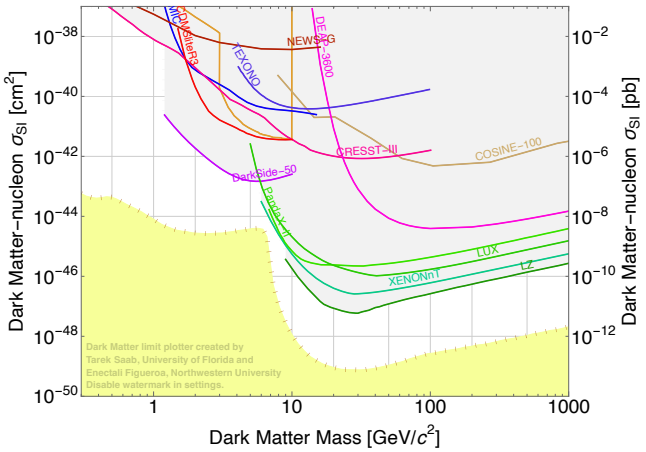


FIGURE 5.4: Experimental upper limits on the spin-independent DM-nucleon cross-section, $\sigma_{\rm SI}$, as a function of the DM mass from various selected direct detection experiments [310, 312–318]. The yellow shaded region is the neutrino "fog" background expectation (here for Xenon). The plot was created using the Dark Matter Limit Plotter [319].

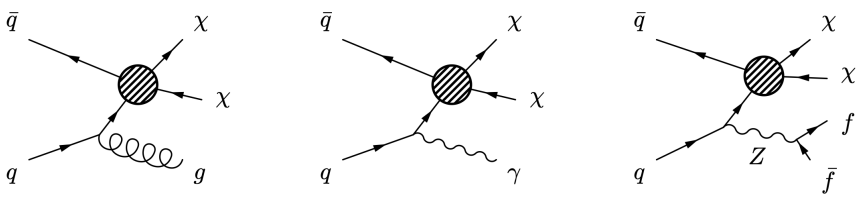


FIGURE 5.5: Different mono-X channels at hadron colliders such as the LHC: mono-jet, mono-photon or mono-Z productions. The mono-X productions in the initial state serve as a trigger for missing transverse energy searches. Feynman diagrams taken from [274].

then often interpreted in terms of either model-dependent UV complete theories, such as the minimal supersymmetric model (MSSM), simplified models, such as axial-vector, vector, scalar or pseudo-scalar mediators coupled to a certain species of DM, or effective field theories, by the experimental collaborations. For certain simplified models, collider searches at for instance ATLAS at the LHC are able to set more stringent constraints on the spin-dependent DM-neutron cross-section [322]. This is depicted in Fig. 5.6, which shows recent exclusion limits from ATLAS on the spin-independent WIMP–nucleon scattering cross-section for a leptophilic vector mediator model (left) and on the spin-dependent WIMP–neutron cross-section for a leptophobic axial-vector mediator model (right) compared to limits from direct detection experiments. However, as mentioned above, these limits rely heavily on simplified DM models with specific values for the couplings (see caption).

Note that future planned electron colliders as FCC-ee or a linear collider (ILC/Japan) that we mentioned in Sec. 4.1 would also significantly increase the sensitivity to DM particles with electroweak interactions [274]. Mono-photon searches at BaBar are able to set exclusion limits on light DM with masses of a few hundred MeV and ~ 10 GeV [51, 323].

For now, these experiments are not very sensitive to standard freeze-in DM since the production cross-section is suppressed by the tiny couplings $g_{\rm SM-DM} \sim 10^{-11}$, see Sec. 5.4. However, specific simplified models, where the FIMP is produced by the decay of an electrically charged and/or coloured parent particle interaction can also lead to interesting signatures such as heavy stable

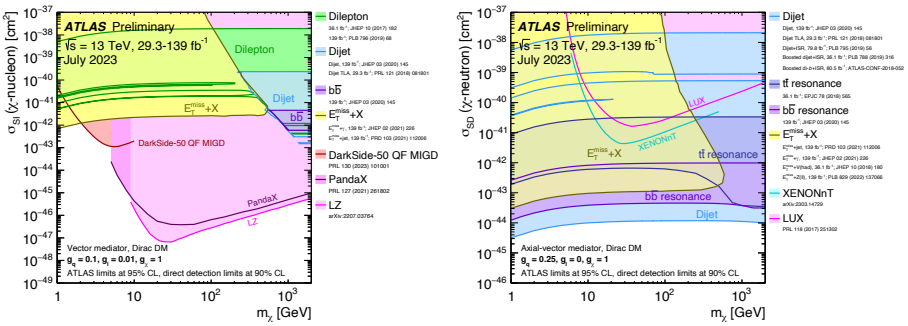


FIGURE 5.6: Comparison of the constraints from direct-detection experiments with inferred LHC limits in the context of simplified models. The mediator mass is fixed by the DM mass. LHC limits are shown at 95% CL and direct-detection limits at 90% CL. *Left:* Constraints on the spin-independent WIMP–nucleon scattering cross-section for a leptophilic vector mediator simplified model, with a DM coupling $g_\chi = 1$, quark coupling $g_q = 0.1$, and lepton coupling $g_l = 0.01$. *Right:* Constraints on the spin-dependent WIMP–neutron cross-section in the context of the leptophobic axial-vector mediator simplified model, with a DM coupling $g_\chi = 1$, quark coupling $g_q = 0.25$, and no coupling to leptons. Figures taken from [322], see the reference for more details.

charged particle signatures, disappearing tracks, displaced vertices and displaced leptons at the LHC [324].

For light mediators such as dark photons or axions, beam-dump and fixed-target experiments serve as very sensitive collider probes. We will come back to beam-dump experiments in more detail in Sec. 7.5.2.1.

Break it! The idea behind indirect detection is to look for DM-SM interactions happening in the cosmos instead of in a detector on Earth. One looks for either DM annihilation, DM decay or DM oscillations, originating from ultralight DM, into SM particles or for their secondary effects. Likewise the energy transfer from DM-SM scattering can have detectable effects, in particular on cosmological observables and constrains, for instance, the lifetime of DM. An advantage of indirect detection is that the whole universe is the laboratory, which means that we have a large volume at our disposal and a long time to integrate over; making use of astrophysics telescopes, which were originally built for other purposes. In addition, the energies in stars and other astrophysical objects are much larger than can be attained in accelerators on Earth. However, the backgrounds can be large and complicated to model. A compilation of recent indirect detection constraints can be found in [292, 325–327].

Photons and neutrinos as annihilation or decay products have the advantage that they subsequently travel along approximately straight lines pointing towards the original direction of the primary particle (the DM). One hence needs to look in directions where one expects a high DM signal. A nice feature of the theory behind indirect detection is that the particle flux can be separated into an astrophysics and a particle physics part (which again is separated into SM and BSM physics). The astrophysical part (the so-called "J-factor" or "D-factor") has to be calculated only once, and different models can be plugged in afterwards [277, 292].

For charged particles such as electrons or positrons the spectrum gets distorted and (analytical or numerical) models have to be developed to take this effect into account [277].

Note that even certain freeze-in scenarios could be observable via indirect detection [328].

Brake it? In specific DM models, DM can self-interact (e.g. in the model we will discuss in Chap. 7 the mediator can generate DM-DM elastic scattering). While the standard ΛCDM model of cold collisionless DM is extremely successful on large scales (on distances $\mathcal{O}(\text{Mpc})$), N-body simulations and observations have revealed some tensions with small scale structures [329–331]: [4]

- *Cusp-core* DM only simulations predict that the density profile of DM increases towards the center (a "cuspy" radial DM density profile)$\rho_{\text{DM}} \sim 1/r$) – well modelled by the NFW profile – while observations from e.g. rotation curves of disc galaxies show a flatter profile ($\rho_{\text{DM}} \sim$ const.) in dwarf galaxies. These tensions are most striking in dwarf and low-surface brightness galaxies as these are DM dominated [333, 334].
- *Diversity problem* Structure formation for collisionless DM is expected to be a self similar process (i.e. all halos should have the same density profile) but for disk galaxies with the same maximal circular velocity the core densities vary by a factor $\mathcal{O}(10)$.
- *Missing satellites problem* Structure formation is a hierarchical process (smaller halos merge, form larger halos). Some of the smaller halos survive the merger process and we thus expect rich substructures (cold DM halos should contain many subhalos). In the milkyway we should have $\mathcal{O}(100 - 1000)$ subhalos with $v_{\text{max}} \sim 10 - 30$ km/s which can host galaxies which is predicted by simulations. Only 10 dwarf spheroidal galaxies were discovered.
- *Too-big-to-fail* From simulations we expect the most luminous satellites to host the most massive subhalos. However, the subhalos are much denser than expected, they are too-big-to-fail.

Many self-interacting DM models can explain these tensions over a large range of scales (dwarf galaxies, low and high surface brightness spiral galaxies, galaxy clusters). Self-interactions hence help transport heat to the inner regions, thermalizing it and resulting in a velocity dispersion that is more uniform with radius. At large distances these self interactions would not have any observable effect as the scattering rate is proportional to the DM density. There are, interestingly, observational constraints on self-interacting dark matter which are roughly of the same size as the strength needed to solve these small scale problems. We already mentioned the Bullet cluster in the introduction: From weak lensing we see that the centre of mass is not located at the collision point but where the stars are. If DM was strongly self interacting it would stay behind at the collision point (it would "brake"). Further constraints come from the shape of the density profile (self-interactions lead to a more spherical profile while some cores exhibit a rather elliptical profile), the escape velocity of DM in smaller halos (it would lead to a dissipation of smaller halos), from the thermodynamics of galaxies and from supermassive black holes which would grow faster for collisional DM [331].

[4] However, including baryonic effects in simulations might also help reconcile simulations and observations [332].

5.4 Bake it!

In this section we will be concerned with the theoretical framework for describing the production and evolution of the DM density in the early universe. The formalism will be very important for the study of the DM model in Chap. 7 (featuring various DM mechanisms that can set the DM relic density), hence requiring a thorough introduction to the subject. Broadly, the landscape of DM production mechanisms can be classified into *thermal* and *non-thermal* DM candidates. Thermal DM candidates were in thermodynamic equilibrium with the SM at some point in the early universe while non-thermal candidates comprise all other possible mechanisms. As we stressed in the introduction, the lack of any conclusive experimental evidence for the WIMP suggests moving away from this standard paradigm and considering alternative possibilities. In particular, thermal models require large couplings to the SM particles to maintain equilibrium and are hence constrained by direct, indirect and collider experiments. For an overview of the plethora of dark matter production mechanisms that have been introduced and considered in the literature see for instance the nice lectures by Joshua Ruderman [335].

5.4.1 Boltzmann equation

To calculate the evolution and the present day abundance of dark matter we have to take an excursion to statistical physics and introduce the Boltzmann equation. This equation describes the time evolution of the number of particles of a certain species in the universe under a series of equilibrium assumptions. The Boltzmann equation can be derived from the Liouville equation via the Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) expansion. The BBGKY expansion introduces a s-particle probability density function which allows one to derive an implicit hierarchical equation for f_s, by considering only two-particle interactions [336, 337]. By truncating the hierarchy at first order and neglecting correlations between the particles one obtains the Boltzmann equation

$$\left(\frac{\partial f}{\partial t} + \frac{1}{m} p \nabla_q + F \cdot \nabla_p\right) f(q,p,t) = \mathcal{C}[f] \ . \tag{5.17}$$

The right hand side is a collision operator which embodies our ignorance of the full resolution of the collision between particles. The Boltzmann equation is an evolution equation for the phase space distribution $f(\vec{x},\vec{p},t)$, which "counts" how many particles are in a phase space volume $\mathrm{d}^3x\,\mathrm{d}^3p$ at $\vec{x},\vec{p}$ at a certain time t: $\mathrm{d}N = g/(2\pi)^3\,\mathrm{d}^3x\,\mathrm{d}^3p f(\vec{x},\vec{p},t)$ with g the number of internal degrees of freedom and $(2\pi)^3$ the size of a unit cell of phase space volume (since we set $\hbar = 1$).

To now apply the Boltzmann equation in cosmology, i.e. in the framework of general relativity, we realize that the Liouville operator in its covariant form is given by [269]

$$\hat{L}_{\text{covariant}} = p^\alpha \frac{\partial}{\partial x^\alpha} - \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \frac{\partial}{\partial p^\alpha} \ , \tag{5.18}$$

resulting in a covariant version of the Boltzmann equation,

$$\hat{L}_{\text{covariant}}[f] = \mathcal{C}_{\text{covariant}}[f] \ . \tag{5.19}$$

We will make the form of the collision operator explicit in the following. From now on, we will drop the subscript "covariant". The evolution of the phase space distribution for inert particles (i.e. a system of free particles), taking into account the expansion of the universe, is described by the left-hand side of the equation and interactions with other particles will change the right-hand side. Hence, a system of free particles satisfies $\hat{L}[f] = 0$. On a FLRW background metric, i.e. in an isotropic and homogeneous universe, $f(\vec{x}, \vec{p}, t) = f(p(t), t)$, with p the modulus of the spatial momentum components, and the Liouville operator takes the form

$$\hat{L}[f] = E\left(\frac{\partial}{\partial t} - Hp\frac{\partial}{\partial p}\right) f \,. \tag{5.20}$$

Collision operator Let us assume a scattering process of the form

$$\chi + a + b + \ldots \leftrightarrow i + j + k + \ldots$$

and let us assume we are interested in the time evolution of the phase space distribution of the particle species χ (ultimately the DM phase space distribution). As we mentioned above, the collision operator encodes the information on the interactions of the χ-particles with other particle species, such as (inverse) decays, annihilations, elastic scatterings etc. We will not attempt to derive the form of the collision operator from first principles but we will state its expression and advocate its plausibility. In its general form it reads [338]

$$\begin{aligned} C\left[f_\chi\right] = & \frac{1}{2}\int d\Pi_a d\Pi_b \ldots d\Pi_i d\Pi_j d\Pi_k \ldots (2\pi)^4 \delta^{(4)}\left(p_\chi + p_a + p_b + \ldots - p_i - p_j - p_k - \ldots\right) \\ & \times \Big[-\left|\mathcal{M}_{\chi+a+b+\ldots\to i+j+k+\ldots}\right|^2 f_\chi f_a f_b \ldots (1 \mp f_i)\left(1 \mp f_j\right)(1 \mp f_k)\ldots \\ & + \left|\mathcal{M}_{i+j+k+\ldots\to\chi+a+b+\ldots}\right|^2 f_i f_j f_k \ldots \left(1 \mp f_\chi\right)(1 \mp f_a)(1 \mp f_b)\ldots \Big] \end{aligned} \tag{5.21}$$

Several pieces enter the expression of the collision operator:

- The Lorentz-invariant phase space measure

  $$\mathrm{d}\Pi_\alpha = \frac{g_\alpha}{(2\pi)^3}\frac{\mathrm{d}^3 p_\alpha}{2E_\alpha}$$

  and the phase space distribution f_α since in order for the process to happen, the incoming particles have to meet each other.

- The δ-function ensures energy and momentum conservation.

- $|\mathcal{M}|^2$ is the matrix element squared, averaged over the initial and final state spins of the process, i.e. the probability for the particles to interact. This is the particle physics piece.

- The process $\chi + a + b + \ldots \to i + j + k + \ldots$ *reduces* the number of χ particles, hence the minus sign, whereas the backreaction increases its number, hence the plus sign.

- The factors $1 \pm f_\alpha$ for bosons and fermions, respectively take into account Bose enhancement or Pauli blocking for the final state particles.

We can further make some often well justified assumptions which allow us to simplify the collision operator. Assuming T-invariance (i.e. CP-invariance which will always be satisfied in the scenarios we study) we have

$$|\mathcal{M}|^2 \equiv \left|\mathcal{M}_{\chi+a+b+\ldots\to i+j+k+\ldots}\right|^2 = \left|\mathcal{M}_{i+j+k+\ldots\to\chi+a+b+\ldots}\right|^2 .$$

Assuming $f_\alpha \ll 1$ we can neglect quantum statistical factors and set $(1 \pm f_\alpha) \simeq 1$. Incorporating these simplifications in Eq. (5.21) we obtain

$$\begin{aligned} C\left[f_\chi\right] = & \frac{1}{2}\int d\Pi_a d\Pi_b \ldots d\Pi_i d\Pi_j d\Pi_k \ldots (2\pi)^4 \delta^{(4)}\left(p_\chi + p_a + p_b + \ldots - p_i - p_j - p_k - \ldots\right) \\ & \times |\mathcal{M}|^2 \left[f_i f_j f_k \ldots - f_\chi f_a f_b \ldots\right] . \end{aligned} \tag{5.22}$$

Equilibrium relations Our early universe was hot and dense and was therefore a thermal environment. If the interaction processes are rapid enough compared to the typical timescale under which the conditions are changing, particles are in equilibrium. One distinguishes between two types of equilibria [335]:

Kinetic equilibrium: Kinetic equilibrium is established if the particles involved in a reaction scatter efficiently enough to exchange energy among them and hence to maintain the same temperature. In this case, the particles' phase space distributions can be described by Bose-Einstein or Fermi-Dirac equilibrium distributions

$$f_{\mathrm{BE/FD}}(p,t) = \frac{1}{e^{(E(p)-\mu(t))/T(t)} \mp 1} . \tag{5.23}$$

Neglecting quantum statistical factors one obtains the Maxwell-Boltzmann distribution

$$f_{\mathrm{MB}}(p,t) = e^{-(E(p)-\mu(t))/T(t)} . \tag{5.24}$$

Chemical equilibrium: Particles are in chemical equilibrium if the number changing reactions $\chi + a + b + \ldots \leftrightarrow i + j + k + \ldots$ are fast enough, such that $\mu_\chi + \mu_a + \mu_b + \ldots = \mu_i + \mu_j + \mu_k + \ldots$. In an expanding universe, chemical equilibrium is established as long as the interaction rate Γ is bigger than the expansion rate of the universe, $\Gamma \gtrsim H$ and several reactions take place per Hubble time. In this case we can apply the principle of detailed balance of which we will in the following often make use, $f\chi^{\mathrm{eq}} f_a^{\mathrm{eq}} f_b^{\mathrm{eq}} \ldots = f_i^{\mathrm{eq}} f_j^{\mathrm{eq}} f_k^{\mathrm{eq}} \ldots$.

From the phase space distribution we can also calculate its moments, such as the particles' number density, energy density and pressure distributions [338]

$$\begin{aligned} n_\chi(t) &= g_\chi \int \frac{\mathrm{d}^3\mathrm{p}}{(2\pi)^3} f_\chi(t,p) , \\ \rho_\chi(t) &= g_\chi \int \frac{\mathrm{d}^3\mathrm{p}}{(2\pi)^3} f_\chi(t,p) E_\chi , \\ P_\chi(T) &= \frac{g_\chi}{(2\pi)^3} \int d^3p \; f_\chi(p,T) \frac{|p|^2}{3E} . \end{aligned} \tag{5.25}$$

If we neglect quantum statistical factors and assume $\mu = 0$, we can plug the Maxwell-Boltzmann distribution (5.24) in the above expressions such that the particles' number density, energy

density and pressure equilibrium distributions are given by

$$
\begin{aligned}
n_\chi^{\rm eq}(T) &= \frac{g_\chi}{2\pi^2}\, m^2\, T K_2\left(m_\chi/T\right)\,,\\
\rho_\chi^{\rm eq}(T) &= \frac{g_\chi}{2\pi^2}\, m_\chi^2\, T\left(m_\chi K_1\left(m_\chi/T\right) + 3T K_2\left(m_\chi/T\right)\right)\,,\\
P_\chi^{\rm eq}(T) &= \frac{g_\chi}{2\pi^2} m_\chi^2 T^2 K_2\left(m_\chi/T\right) = T n_\chi^{\rm eq}(T)\,,
\end{aligned}
\tag{5.26}
$$

respectively, for a given particle of mass m_χ and degrees of freedom g_χ, where $K_1(x)$ and $K_2(x)$ are the first and second order Bessel functions. It is particularly interesting to study the expression for the equilibrium number density in two limiting cases, the non-relativistic limit $T \ll m_\chi$ and the relativistic limit, $T \gg m_\chi$,

$$
n_\chi^{\rm eq}(T) = \begin{cases} \frac{g_\chi}{\pi^2} T^3 & m_\chi \ll T \\ g_\chi \left(\frac{m_\chi T}{2\pi}\right)^{3/2} e^{-m_\chi/T} & m_\chi \gg T \end{cases}
\tag{5.27}
$$

For temperatures well below the mass of the particle, its number density is exponentially suppressed (this is often referred to as Boltzmann suppression).

We will often make the assumption that the phase space distribution of the SM particles is described by the equilibrium distribution (5.23) or (5.24). In this case we can multiply by g_χ and integrate both sides of the Boltzmann equation over the particles' momenta to obtain an evolution equation for the particles' number density. The Liouville operator becomes

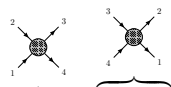
$$
g_\chi \int \frac{{\rm d}^3 p}{(2\pi)^3} {\rm L}\left[f_\chi\right] = \frac{1}{a^3}\frac{\rm d}{{\rm d}t}(a^3 n_\chi) = \frac{{\rm d}n_\chi}{{\rm d}t} + 3H n_\chi\,,
\tag{5.28}
$$

where again the second term takes into account the expansion of the universe. For simplicity, let us restrict ourselves to $2 \to 2$ processes, generalizing the following discussion to decays is straightforward, see Sec. 7.2. A generalization to $2 \to n$ processes with many-body final states is feasible but requires some more thought, see for instance [339] and also Sec. 7.3.3. Finally, the integrated Boltzmann equation takes the general expression

$$
\frac{{\rm d}n_\chi}{{\rm d}t} + 3H n_\chi = \int {\rm d}\Pi_\chi\, {\rm d}\Pi_2\, {\rm d}\Pi_3\, {\rm d}\Pi_4 \times (2\pi)^4 \delta^{(4)}\left(p_\chi + p_2 - p_3 - p_4\right) |\mathcal{M}|^2 \left[f_3 f_4 - f_\chi f_2\right]\,.
\tag{5.29}
$$

Assuming that particles 3 and 4 are part of a thermal bath at temperature T and re-equilibrate quickly after being produced, we can replace $f_3 f_4$ by $f_3^{\rm eq} f_4^{\rm eq}$ and we can make use of the principle of detailed balance, $f_3^{\rm eq} f_4^{\rm eq} = f_\chi^{\rm eq} f_2^{\rm eq}$. Then the collision term can be rewritten as

$$
g_\chi \int \frac{{\rm d}^3{\rm p}}{(2\pi)^3} {\rm C}\left[f_\chi\right] = -\langle \sigma_{\chi 2\to 34} v_{\rm møl}\rangle \Big(\overbrace{n_\chi n_2}^{\rm Fig.\ 1} - \overbrace{n_\chi^{\rm eq} n_2^{\rm eq}}^{\rm Fig.\ 2}\Big)\,,
\tag{5.30}
$$



with

$$
\langle \sigma v_{\rm møl}\rangle = \frac{\int \sigma v_{\rm møl} e^{-E_\chi/T} e^{-E_2/T}\, {\rm d}^3 p_\chi\, {\rm d}^3 p_2}{\int e^{-E_\chi/T} e^{-E_2/T}\, {\rm d}^3 p_\chi\, {\rm d}^3 p_2}
\tag{5.31}
$$

and where the Møller velocity is defined by

$$v_{\text{møl}} \equiv F/(E_\chi E_2) = \frac{(p_\chi \cdot p_2)^2 - m_\chi^2 m_2^2}{E_\chi E_2} = \sqrt{(\vec{v}_\chi - \vec{v}_2)^2 - (\vec{v}_\chi \times \vec{v}_2)^2} \tag{5.32}$$

and

$$F(m_\chi, m_2, s) = \frac{\sqrt{(s - (m_\chi + m_2)^2)(s - (m_\chi - m_2)^2)}}{2} \,. \tag{5.33}$$

We have assumed Maxwell-Boltzmann distributions and have approximated $1 \pm f_\alpha \simeq 1$. In the following we will omit the subscript "møl". Under these assumptions, all but one integral on the right hand side in Eq. (5.29) can be solved analytically and Paolo Gondolo and Graciela Gelmini obtained in their seminal work [23]

$$\langle \sigma_{\chi 2 \to 34} v \rangle = \frac{1}{2T K_2(m_1/T)\, K_2(m_2/T)} \int_{s_{\min}}^{\infty} \sigma_{\chi 2 \to 34}(s) \frac{F(m_\chi, m_2, s)^2}{m_\chi^2 m_2^2 \sqrt{s}} K_1(\sqrt{s}/T)\, \mathrm{d}s \,, \tag{5.34}$$

where $\sigma(s)$ is the total unpolarized cross section for the process $12 \to 34$ in terms of the centre-of-mass energy $\sqrt{s}$, K_1 and K_2 the usual modified Bessel functions and the lower limit of the integral is given by $s_{\min} = \max\left((m_\chi + m_2)^2, (m_3 + m_4)^2\right)$.[5] The integral in s has to be performed numerically. The reaction rate for a given process in terms of the thermally averaged cross section is $\langle \sigma v \rangle\, n^{\text{eq}}$ and the condition for chemical equilibrium becomes

$$\langle \sigma v \rangle\, n^{\text{eq}} \gtrsim H \,. \tag{5.36}$$

It is often advantageous to write the Boltzmann equation in terms of the inverse temperature $z \equiv m_\chi/T$[6] and in terms of the so-called comoving number density,

$$Y_\chi \equiv \frac{n_\chi}{s} \,,$$

with s the total entropy density of the universe, see Eq. (5.9). Using entropy conservation of the radiation bath and neglecting the temperature dependence of the effective degrees of freedom

[5] A quick note about factors of two: this expression comes from the construction of the thermally averaged cross section in terms of the usual total unpolarized cross section. Since we average over the degrees of freedom in the initial state to calculate the cross section, we need to include the correct number of degrees of freedom in the number densities. Furthermore, for processes with two identical particles in the initial state, we need to include an additional factor of 1/2. The phase-space integral over-counts the configurations of identical particles. The cross section already takes this into account, since there we integrate over the final state particles and divide by a factor of 2 for identical final state particles. The difference here is that we need to do the same for identical initial state particles since we also integrate over their phase spaces (while we do not for the cross section). One can check that using this definition detailed balance in terms of the thermally averaged cross section reads as usual

$$\langle \sigma_{ij \to kl} v \rangle\, n_i^{\text{eq}} n_j^{\text{eq}} = \langle \sigma_{kl \to ij} v \rangle\, n_k^{\text{eq}} n_l^{\text{eq}} \,. \tag{5.35}$$

[6] Note that for numerical solutions of the Boltzmann equation it can in addition be helpful to formulate the equation in terms of $\log_{10} z$. We will come back to numerical tricks in Sec. 7.3.3.

$\mathrm{d}g_*/\mathrm{d}T \sim 0$, the Boltzmann equation then takes the form [7]

$$\frac{\mathrm{d}Y_\chi}{\mathrm{d}z} = -\frac{s}{Hz}\langle\sigma_{\chi 2\to 34}v_{\text{møl}}\rangle\left(Y_\chi Y_2 - Y_\chi^{\text{eq}}Y_2^{\text{eq}}\right) . \tag{5.38}$$

5.4.2 Freeze-out and WIMPs

Now that we are armed with the tools required to calculate the evolution of the DM relic density we will study the standard paradigm for DM production, freeze-out. One assumes that in the early hot and dense universe reactions were fast enough and all particle species, including the dark matter, were in thermal equilibrium and shared the same temperature.

To understand the basic physical mechanism, let us for a moment consider a simple toy model, where we assume that the DM particle interacts with some SM particle species via $\chi\chi \leftrightarrow$ SM SM and let us assume that this is indeed the only relevant DM number changing interaction. In this case, the Boltzmann equation for the comoving DM number density reads

$$\frac{\mathrm{d}Y_\chi}{\mathrm{d}z} = -\frac{s}{Hz}\langle\sigma_{\chi\chi\to\text{SMSM}}v\rangle\left(Y_\chi^2 - Y_\chi^{\text{eq}2}\right) . \tag{5.39}$$

where we have replaced the distributions of the SM particles by their equilibrium distributions and we have made use of the principle of detailed balance: The first term on the right hand side corresponds to SM bath particles annihilating into DM (thus increasing the number of DM particles), and the second term corresponds to DM annihilations (the inverse process), decreasing n_χ. In this simple toy model, DM could be produced by two different mechanisms, namely freeze-out and freeze-in. For relatively large couplings, such that condition (5.36) is satisfied, the SM and the DM will interact strongly enough to reach equilibrium in the early universe and the comoving number density Y_χ as a function of the inverse temperature stays constant. At some point the temperature drops below the mass of the DM particle; it becomes more and more difficult to produce the DM and the production process gets suppressed. The DM particles continue to annihilate into SM particles – in particular, if nothing happened we would be left with no DM today. At some point, however, the universe has expanded so much that the DM particles are too dilute to find each other; they decouple from the plasma and we are left with a constant number of DM particles. Unfortunately, there are no analytic solutions to this equation; it has to be solved numerically. Its numerical solution for varying reaction rates is plotted in Fig. 5.7. However, by making some simplifications, we can obtain a back-of-the-envelope analytic solution for the DM relic density and get a feeling for the behavior of Eq. (5.40). Performing the change of variables $z = T/m_\chi$ as above and neglecting the temperature dependence of the effective degrees of freedom g_*, the Boltzmann equation can be written as [269]

$$\frac{\mathrm{d}Y_\chi}{\mathrm{d}z} = \frac{\lambda}{z^2}(Y_\chi^{\text{eq}2} - Y_\chi^2) \qquad \text{(Riccati-equation)} , \tag{5.40}$$

where we have defined $\lambda \equiv \sigma_0 s_1/H_1$. We have made several assumptions:

[7] Note that in the computations we perform in Chap. 7 we take this temperature dependence into account. In more generality, one can replace the time derivative by a derivative with respect to the radiation temperature via the relation

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -HT\left(1 + \frac{1}{3}\frac{\mathrm{dlog}\, g_{*s}}{\mathrm{dlog}\, T}\right)^{-1} . \tag{5.37}$$

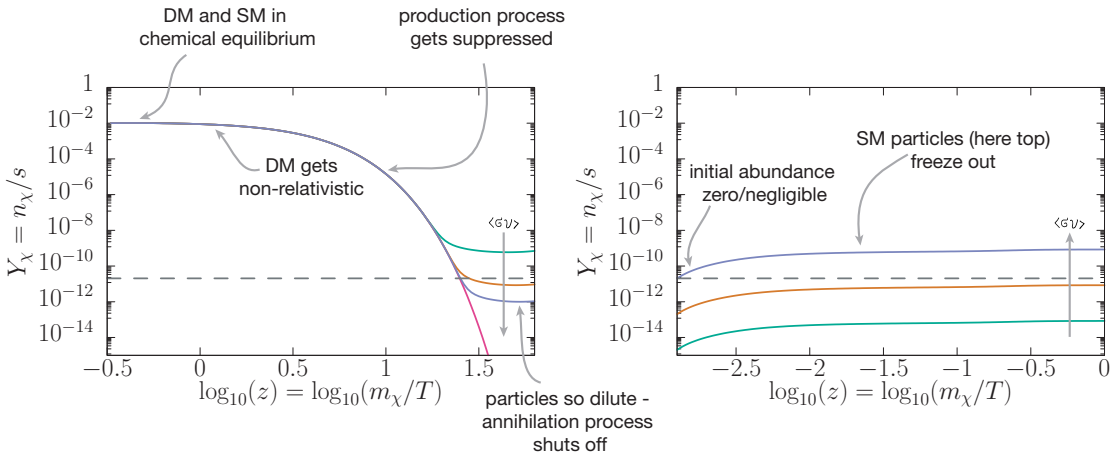


FIGURE 5.7: Freeze-out (left) versus freeze-in (right) for different reaction rates. The figure was produced assuming the SM fermions produce the DM with a coupling proportional to their masses.

- We assume the universe is radiation dominated, i.e. we have $H(z) = H(z=1)/z^2$ and $s = s(z=1)/z^3$, see Sec. 5.1.
- $\langle\sigma v\rangle$ will depend on the velocity of the DM and we have expanded $\langle\sigma v\rangle$ in powers of z $\langle\sigma v\rangle \sim \sigma_0 z^{-n}$ (where $n = 0$ corresponds to s-wave, $n = 1$ to p-wave etc. annihilation). By considering only the s-wave piece, i.e. $\langle\sigma v\rangle \sim \sigma_0$, λ becomes a constant.
- As above, we have neglected the temperature dependence of the effective degrees of freedom g_*.

From Fig. 5.7 we see that for $z \gg 1$, Y_χ will be much larger than $Y_\chi^{\rm eq}$ ($Y_\chi^{\rm eq}$ becomes Boltzmann suppressed) and we can approximate at late times

$$\frac{\mathrm{d}Y_\chi}{\mathrm{d}z} \approx -\frac{\lambda Y_\chi^2}{z^2} \overset{\text{Integrate}}{\Rightarrow} \frac{1}{Y_{\chi,0}} - \frac{1}{Y_{\chi,f}} = \frac{\lambda}{z_f} \tag{5.41}$$

with $z_f = T_f/m_\chi$ the temperature-mass ratio at freeze-out and $Y_{\chi,0}$ the DM abundance today. Further, $Y_{\chi,f} \gg Y_{\chi,0}$, and therefore

$$Y_{\chi,0} \simeq \frac{z_f}{\lambda}\,. \tag{5.42}$$

Estimating $z_f \sim 10-20$, one finally obtains [8]

$$\Omega_\chi h^2 \simeq 0.3\left(\frac{z_f}{10}\right)\sqrt{\frac{g_*}{100}}\frac{10^{-37}\ \mathrm{cm}^2}{\langle\sigma v\rangle} \tag{5.43}$$

which in fact gives the right DM relic density for a cross-section of the order of the weak interaction $G_f^2 \sim 10^{-10}$ GeV^{-2} and a mass similar to the weak scale. This result is called the *WIMP miracle* and the corresponding DM candidate a WIMP (Weakly Interacting Massive Particle) [269]. Commonly studied WIMP candidates are neutralinos that come from supersymmetric extensions of the SM. Eq. (5.43) might tell us something – a priori the value of the weak scale is not provided in deriving Eq. (5.43) – or it might just be a numerical coincidence. Incidentally, assuming the DM is a strongly interacting particle (like a dark pion or a dark proton) one finds a similar relation, the *SIMP miracle* [339].

[8] For a Dirac fermion the yield would be multiplied by an additional factor of 2.

The WIMP hypothesis is however severely challenged due to the lack of any experimental evidence despite decades of experimental efforts. In fact, as we saw in Sec. 5.3 (e.g. Fig. 5.4), direct, indirect and collider searches have ruled out a substantial portion of the parameter space. This outcome has led to a move away from the standard WIMP produced via thermal freeze-out and has led people to investigate in alternative production mechanisms [50, 51, 340]. In particular, much investigation has been put in the freeze-in mechanism and its associated FIMP (feebly interacting massive particle) [19, 52] which is a particularly simple DM genesis scenario.

5.4.3 Freeze-in

Since the WIMP has not been detected so far one could wonder if alternative scenarios requiring weaker couplings are also possible. In the freeze-in mechanism, the DM particles interact so feebly with the SM particles that they never reach equilibrium [9]. If one assumes that the inflaton decays preferably into SM particles (i.e. the DM abundance is zero or negligibly small after reheating) then the DM will be gradually produced by scattering of SM particles in the thermal plasma without ever reaching thermal equilibrium, until the temperature drops below the SM particle mass(es) producing the DM. At this point, the abundance of SM particles gets Boltzmann suppressed, $n_{SM} \propto e^{-m_{SM}/T}$, DM producing processes decouple and the number of DM particles in the universe remains constant; it *freezes in*. The DM abundance is then simply determined by the number of decays or annihilations of SM particles into DM particles.

5.4.3.1 Infrared dominated freeze-in

Let us look again at our simple toy model, Eq. (5.39). As we will see, in the case of a renormalizable DM-SM interaction with g_χ a dimensionless coupling, DM production is maximized at low temperatures (IR dominated) and hence insensitive to UV physics, as in the thermal freeze-out scenario. Let us estimate the DM relic density in this case: Since Y never reaches the equilbrium value, $Y \ll Y_{\rm eq}$ and we can omit the Y^2 term. It is then convenient to introduce the reaction density [21]

$$\gamma \equiv \langle \sigma\, v \rangle \left(n_\chi^{\rm eq} \right)^2 \tag{5.44}$$

in terms of which Eq. (5.39) becomes (let us perform the change of variables $T = m_\chi/z$ to make the temperature dependence explicit)

$$\frac{{\rm d}Y}{{\rm d}T} = -\frac{\gamma}{sHT}\,. \tag{5.45}$$

Therefore, the yield today at $T_0 \approx 0$ is given by the integrated reaction density,

$$Y_0 = -\int_{T_0}^{T_{\rm RH}} \frac{\gamma}{sHT}\, {\rm d}T\,. \tag{5.46}$$

The reaction density can be recast into [21]

$$\gamma = \int \sigma v\ {\rm e}^{-E_1/T} {\rm e}^{-E_2/T}\ {\rm d}^3p_1\ {\rm d}^3p_2 = 2\pi^2 T \int {\rm d}s\ \sigma\left(s - 4m^2\right)\sqrt{s} K_1(\sqrt{s}/T)\,. \tag{5.47}$$

[9] Generically, for freeze-in one requires couplings which are $\lesssim 10^{-7}$ to prevent equilibration with the SM bath.

At temperatures far above the masses of the involved particles, we have $\sigma(s) \sim g_\chi/s$ with g_χ some dimensionless coupling, which results in $\langle \sigma\, v\rangle \sim T^{-2}$ and $\gamma \sim T^4$ by dimensional analysis (since $n_{\rm eq} \sim T^3$ for $T \gg m$, $s \sim T^3$ and $H \sim T^2$), hence the integrand decays like T^{-2}. At temperatures below $\max(m_\chi, m_{\rm SM})$ the integrand is exponentially Boltzmann-suppressed. We can therefore roughly estimate the order of magnitude of the dark matter yield due to freeze-in from the saddle point of the integrand:

$$Y_0 \approx C\, \frac{\gamma}{SH}\bigg|_{T=\max(m_{\rm SM},m_\chi)} \sim \frac{g_\chi}{\max(m_{\rm SM}, m_\chi)}\,, \tag{5.48}$$

where C is a factor of a few. The freeze-in process is infrared-dominated, because it is driven by renormalizable interactions with dimensionless couplings.

Solving Eq. (5.46) approximately as we did in the last section one obtains [52, 278]

$$\Omega h_X^2 \approx \frac{2 m_X Y_X}{3.6\times 10^{-9}{\rm GeV}} = \frac{1.01\times 10^{24}}{g_*^S\sqrt{g_*^\rho}}\, g_\chi^2\,, \tag{5.49}$$

implying

$$g_\chi \simeq 10^{-11}\left(\frac{m}{m_\chi}\right)^{1/2}\left(\frac{\Omega_\chi h^2}{0.12}\right)^{1/2}\,. \tag{5.50}$$

If the DM particle and the SM particle producing it have similar masses, the required value for the coupling between the SM and the DM particle is $\mathcal{O}(10^{-11})$, implying that in general indeed $\langle \sigma v\rangle\, n_{\rm eq} \ll H$ throughout DM production and the DM particles never reach equilibrium [10]. The freeze-in process is infrared-dominated, because it is driven by effective interactions with dimensionless couplings. This is advantageous as one does not need to make assumptions about the physics at high temperatures (UV quantities), such as the reheating temperature after inflation. In models with higher-dimensional couplings between the DM and the SM particles, this conclusion would not hold. In the presence of such operators, the contribution to freeze-in would be UV-dominated and the result would depend on the reheating temperature. This scenario can be described by the so-called ultraviolet freeze-in and explicitly depends on the reheating temperature [52, 341].

The evolution of Y for different reaction rates for the (IR dominated) freeze-in versus the freeze-out mechanism is depicted in Fig. 5.7. Via both mechanisms the observed relic density can be generated. The essential difference is that in the freeze-out mechanism Y is inversely proportional to the reaction rate, while it is proportional to $\langle \sigma v\rangle$ in the freeze-in mechanism. Indeed, this is what we expect: During freeze-in Y is determined by twice the "number" of pair creations, during freeze-out Y follows the equilibrium number density until it decouples and stays constant. Since $Y_{\rm eq}$ decreases with decreasing temperature, the earlier $\langle \sigma v\rangle\, n_{\rm eq}$ drops below H the more DM particles will be left over.

5.4.3.2 Ultraviolet dominated freeze-in

As we saw in the previous section, if DM freezes in via renormalizable operators, the DM relic density is infrared dominated, receiving its largest contribution at temperatures $T \sim \max(m_{\rm DM}, m_{\rm prod})$

[10] Again, this can easily be generalized to SM particles decaying into DM. In this case the required coupling is approximately a factor 10 smaller [19].

where $m_{\rm prod}$ is the mass of the particles producing the DM [19]. The final relic abundance is hence independent of UV physics and, in particular, of the reheating temperature at which DM production from SM particles started. In the case of higher dimensional operators, however, the DM freeze-in is UV dominated (i.e. its production is maximised at high temperatures) and depends on the reheating temperature [19, 341]. As we will briefly see in the following, this statement can already be understood from simple dimensional analysis. The idea is the following: In general, one would expect the distinct sectors of the low energy theory to be connected by the UV physics via higher dimensional operators. Hence, one might argue that the IR freeze-in scenario represents a fairly special construction (one would normally expect dimensionless parameters to be $\mathcal{O}(1)$).

Let us compare the dimensional analysis in the previous paragraph to what happens if the scattering process is mediated by a higher dimensional operator (for simplicity, let us consider an example where the DM is produced via a 5-dimensional operator in a $2 \to 2$ scattering process). This could for example be an operator $\mathcal{O} \sim 1/\Lambda\ \varphi\bar{\psi}\psi\phi$, with φ a scalar DM particle and ψ and ϕ fermionic and scalar SM particles, respectively, arising in an effective field theory with cutoff Λ [11]. In this case the matrix element scales like

$$\mathcal{M} \sim \frac{\sqrt{s}}{\Lambda}\ ,$$

where $\sqrt{s}$ is the characteristic energy scale of the process. Hence, the cross section behaves as $\sigma(s) \sim g_\chi s$, yielding $\gamma \sim T^6$ and

$$Y_0 \sim g_\chi T_{RH}\ .$$

The DM yield is UV dominated and produced immediately at reheating. Generalising this simple estimate to a $(n+4)$-dimensional operator $\frac{1}{\Lambda^n}\phi_1 \dots \phi_{n+3}\varphi$ (with ϕ_i scalar particles part of the thermal bath and φ the DM particle), DM can be produced by processes $\phi_1\phi_2 \to \phi_3 \dots \phi_{n+3}\varphi$ with many-body final states. Then the matrix element squared and the differential Lorentz invariant phase space for n-body final states scale, from dimensional analysis, approximately as

$$|M|_n^2 \sim \frac{1}{\Lambda^{2n}}\ , \qquad \text{DLIPS}_{(n+2)} \sim s^n\ ,$$

respectively . Using these approximate scaling relations, the integrals on the right hand side of the Boltzmann equation can be solved analytically and, neglecting combinatorial factors, one obtains for the DM yield [341]

$$Y_{(n)} \simeq \frac{90}{1.66 \times (2\pi)^9 \sqrt{g_*^\rho g_*^S}} \frac{1}{2n-1} \left(\frac{n!(n+1)!}{\pi^{2n-1}}\right)\left(\frac{M_{\rm Pl} T_{\rm RH}^{2n-1}}{\Lambda^{2n}}\right). \tag{5.51}$$

Two remarks are now in order. Firstly, from Eq. (5.51) we find

$$Y_{(2)} \sim \frac{T_{RH}^3}{\Lambda^4}\ , \qquad Y_{(3)} \sim \frac{T_{RH}^5}{\Lambda^6}\ , \qquad \dots$$

i.e. if several higher dimensional operators with the same UV cutoff Λ are present, the DM yield is dominated by the operator with the smallest mass dimension. This is true for cases where all DM producing processes are driven by higher dimensional operators. However, secondly, if one

[11] Of course, the reheating temperature has to below the UV scale for the effective theory to be valid.

field involved in the higher dimensional operator acquires a vev then the expansion around the vacuum gives terms with couplings of different mass dimension. In the case of a 5-dimensional operator this results in a UV-dominated contribution and an IR-dominated contribution from an effectively renormalizable operator in the vev expansion. Depending on the vev and the reheating temperature, the DM yield is then either UV or IR dominated. We will see explicitly in Chap. 7 that this is precisely what happens in our model where UV contributions arising from the mixing with the Higgs sector become important for reheating temperatures above the Higgs vacuum expectation value.

To set the stage for the model we study in this thesis, where the interaction between the SM and the DM particles are mediated by an axion-like particle, we will continue by giving a brief introduction to axions, which were proposed as a very natural solution to the strong CP problem.

6 Axions and axion-like particles

In Sec. 2.3.4 we discussed how the solution of the $\mathrm{U}(1)_A$ problem introduced a new problem in the SM: the strong CP problem [1]. A very prominent and theoretically tempting solution to the strong CP problem are axions which naturally promote the $\bar{\theta}$ to a dynamical variable as the expectation value of a spinless quantum field [53–56]. Axions arise as Nambu-Goldstone bosons by introducing a new scalar field and by imposing the existence of a global symmetry. The mechanism will be shortly introduced in the following. For a detailed pedagogical derivation of the axion solution highlighting many physical properties see the original paper by Peccei and Quinn [53] or the toy model in [17].

6.1 The axion solution to the strong CP problem

As we saw in Sec. 2.3.4, the phases in the fermion mass matrix and the θ-parameter in front of the $G\tilde{G}$ term are intimately related via the axial anomaly. The Peccei-Quinn solution to the strong CP problem makes use of this relation. In particular, Robert Peccei and Helen Quinn noticed that if there is an additional global $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry which is anomalous under strong interactions, then one can use a field redefinition to rotate the physical $\bar{\theta}$-term away. This can be achieved by introducing chiral fermions which transform under this symmetry and by ensuring that the symmetry is exact except for the anomaly. In particular, we cannot add a mass term, otherwise we would get an additional piece proportional to the mass and the $\bar{\theta}$ term would again be physical. Unfortunately, there is no anomalous axial symmetry in the SM which is not also explicitly broken by the mass term.

There is however another way to implement this (anomalous) symmetry, namely with a boson. The main idea is to add a new spin-0 field with effective Lagrangian [18, 342–344]

$$\mathcal{L}_a = \frac{1}{2}\left(\partial_\mu a\right)^2 + \mathcal{L}\left(\partial_\mu a, \psi\right) + \frac{g_s^2}{32\pi^2}\frac{a}{f_a}G^a_{\mu\nu}\tilde{G}^{a\mu\nu} \tag{6.1}$$

to the SM Lagrangian, with f_a a dimensionful parameter. The interaction between the boson and the gluonic fields mimics the $\bar{\theta}$-term.

[1] To introduce this chapter with an anecdote: The axion was named by Frank Anthony Wilczek – an american physicist and Nobel prize winner – after a detergent for its ability of "whitewashing" the SM from its strong CP stain.

Under a rotation of angle α this effective Lagrangian possesses a quasi shift symmetry $a(x) \to a(x) + \alpha f_a$ which leaves the action invariant up to a term

$$\delta S = \frac{\alpha}{32\pi^2} \int \mathrm{d}^4x \; G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \; , \tag{6.2}$$

i.e. it shifts exactly like an anomalous term would shift (by a constant times $G\tilde{G}$), so this shift symmetry is anomalous (although here at tree-level). This field redefinition can then be used to remove any pre-existing $G\tilde{G}$-term from the Lagrangian. The physics is of course invariant under this field redefinition and all observables are automatically independent of $\bar{\theta}$. As we will see later, what really happens is that low-energy QCD effects dynamically set the parameter to the correct value, hence removing any strong CP-violating term. In addition, the Vafa-Witten theorem makes sure that in a theory without this $\bar{\theta}$ term there is no spontaneous CP breaking, and we get a CP-conserving theory [345].

Clearly, the effective Lagrangian in Eq. (6.1) can emerge from various different UV-completions. In order for the PQ solution to work, a few main ingredients are needed [346]:

- A new global $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry, spontaneously broken at a high scale f_a. This ensures the low-energy theory has a Goldstone boson, the axion.
- $\mathrm{U}(1)_{\mathrm{PQ}}$ is anomalous with respect to QCD ($\partial_\mu J^\mu_{\mathrm{PQ}} \sim G\tilde{G}$). This can be achieved by assigning chiral charges to some coloured fermions, i.e. $Q(q_L) \neq Q(q_R)$. The net effect is the last term in Eq. (6.1), the associated Goldstone boson has a coupling to gluons.
- Finally, non perturbative QCD effects will induce $\mathcal{L}_a = \mathcal{L}_{\chi PT}\left(\partial_\mu a, \pi, \eta, \eta', \ldots\right) + V_{\mathrm{eff}}\left(\bar{\theta} + \frac{a}{f_a}, \pi, \eta\right)$ with the minimum of the potential at $\bar{\theta} + \frac{\langle a \rangle}{f_a} = 0$, hence dynamically restoring the CP-symmetry.

Typically, in this effective setup one also expects other higher dimensional couplings. In order for the effective theory to still solve the strong CP problem, we can add terms which do not spoil this shift symmetry, such as derivative couplings times any current made of SM fields and higher dimensional couplings to SM gauge fields [18],

$$\mathcal{L}_a \subset \frac{1}{2}(\partial_\mu a)^2 + \frac{a+\bar{\theta}}{f_a}\frac{g_s^2}{32\pi^2} G^a_{\mu\nu}\tilde{G}^{a\mu\nu} + g_{a\gamma}\frac{a}{f_a}\frac{\alpha_{\mathrm{em}}}{8\pi} F_{\mu\nu}\tilde{F}^{\mu\nu} + \sum_f c_f \frac{\partial_\mu a}{2f_a} j^\mu_{(a)} - \bar{q}\mathcal{M}_q q \; , \tag{6.3}$$

with $j^\mu_{(a)} = \bar{q}_f \gamma^\mu \gamma_5 q_f$ and we have explicitly kept the pre-existing $\bar{\theta}$ term from the SM Lagrangian [2]. Now, one can perform a (chiral) reparametrization of the fermion fields

$$q \to \exp\left(i\gamma_5 \frac{a + f_a\bar{\theta}}{2f_a}\boldsymbol{Q}_a\right) q \; , \tag{6.4}$$

where $\boldsymbol{Q}_a$ is a generic matrix and we require $\mathrm{Tr}(\boldsymbol{Q}_a) = 1$. This reparametrization has four effects:

- The transformation being chiral, the fermion measure is not invariant and the Jacobian of the transformation introduces a term which, since $\mathrm{Tr}(\boldsymbol{Q}_a) = 1$, exactly cancels the CP-violating $G\tilde{G}$ term.

[2] Note that one also expects couplings to the weak bosons, see Sec. 6.3, but they should be suppressed at low energies.

- The fermionic kinetic term induces a contribution to the derivative couplings $c_f \to c_f - Q_{a,f}$.
- The quark mass matrix gets dressed with an exponential of the axion field $\mathcal{M}_a = \exp\left(i\frac{a}{f_a}\frac{\gamma_5}{2}\right)\mathcal{M}_q \exp\left(i\frac{a}{f_a}\frac{\gamma_5}{2}\right)$.
- Assuming that QED is also anomalous under $\mathrm{U}(1)_{\mathrm{PQ}}$, the coupling to photons gets modified $g_{a\gamma} \to g_{a\gamma} - \frac{2N_c}{8\pi^2 f_a}\mathrm{Tr}\left(\boldsymbol{Q}_a \boldsymbol{Q}_f^2\right)$ with $\boldsymbol{Q}_f = \mathrm{diag}(Q_f)$ and Q_f the electric charge of the fermion f.

We will see below that the low energy minimum of the axion potential corresponds to a theory without a $G\tilde{G}$ term.

There are various axion models, the original one proposed by Peccei and Quinn with only two Higgs doublets as ingredients, however, was quickly experimentally ruled out by rare Kaon decays [347]. In the original PQ model the vacuum expectation values v_1 and v_2 of the two Higss doublets spontaneously break $SU(2)_L \times \mathrm{U}(1)_Y \times \mathrm{U}(1)_{\mathrm{PQ}} \to \mathrm{U}(1)_{\mathrm{em}}$ with $v_1^2 + v_2^2 = v^2$ and $\tan\beta = v_1/v_2$. Hence, all axion parameters are fixed from flavour conservation by the axion decay constant $f_a \propto v \sim 246$ GeV with v the scale of electroweak symmetry breaking [348]. A few years later, two main types of so-called *invisible axion* models were proposed, the Kim-Shifman-Vainshtein-Zakharov (KSVZ) [349, 350] and the Dine-Fischler-Srednicki-Zhitnitsky (DFSZ) [351, 352] models. They could evade the stringent constraints by decoupling the vev of the scalar f_a from the electroweak symmetry breaking scale v. We will briefly introduce the DFSZ-type axion in the next paragraph.

6.2 DFSZ model

The Lagrangian in eq. (6.3) is non-renormalizable and requires an ultraviolet completion. Here we will sketch an ultraviolet completion of the DFSZ (Dine-Fischler-Srednicki-Zhitnitsky) type by which our model will be inspired [351, 352]. In the DFSZ model, the axion couples to the SM via the Higgs sector. It is based on a two-Higgs-doublet model with two Higgs fields Φ_1 and Φ_2

$$\Phi_1 = \begin{pmatrix} H_1^+ \\ H_1^0 \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix},$$

and with an additional global $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry and an additional scalar ϕ. The scalar will aquire a vev at a scale $f_a \gg v$, with v the electroweak symmetry breaking scale. The interaction Lagrangian between the Higgs doublets and the scalar is

$$\mathcal{L}_{\Phi_1\Phi_2\phi} = \lambda_1 \Phi_1^\dagger \Phi_1 \phi^\dagger \phi + \lambda_2 \Phi_2^\dagger \Phi_2 \phi^\dagger \phi + \left[\lambda_3 (\Phi_1^\dagger \Phi_2)\phi^2 + \mathrm{h.c.}\right], \tag{6.5}$$

which is invariant under $\mathrm{U}(1)_{\mathrm{PQ}}$ provided one chooses the charges of the scalars appropriately,

$$\Phi_1 \to \exp(i\alpha_1\theta)\,\Phi_1, \quad \Phi_2 \to \exp(i\alpha_2\theta)\,\Phi_2, \quad \phi \to \exp(i\alpha\theta)\phi,$$

with $2\alpha - \alpha_1 + \alpha_2 = 0$. If $\alpha \neq 0$ and by coupling the SM fermions to the Higgs doublets via Yukawa couplings of type II [353], the SM fermions must be charged under $\mathrm{U}(1)_{\mathrm{PQ}}$ where

$\mathrm{U}(1)_{\mathrm{PQ}}$ is necessarily chiral, $\alpha_1 \neq \alpha_2$ and hence $\mathrm{U}(1)_{\mathrm{PQ}}$ is anomalous. Further, one chooses a typical mexican hat potential for the additional scalar

$$V_\phi = \mu^2 \phi^\dagger \phi + \frac{\lambda}{2}\left(\phi^\dagger \phi\right)^2 \tag{6.6}$$

with μ^2 and λ chosen appropriately such that $\mathrm{U}(1)_{\mathrm{PQ}}$ is spontaneously broken and ϕ obtains a vacuum expectation value $\langle\phi\rangle = f_a/\sqrt{2}$. The parameters in the Higgs potential must satisfy the measured 125 GeV mass of the Higgs and a vacuum expectation value $v_{EW} = \sqrt{\langle\Phi_1\rangle^2 + \langle\Phi_2\rangle^2}$ [354], where one defines $\tan(\beta) = \langle\Phi_1\rangle / \langle\Phi_2\rangle$. As we advertised in the previous section, the most straight-forward manner is to parametrize the additional scalar in the linear representation after SSB. Then, the SM fermions couple at tree-level to the axion via a Yukawa-like coupling $(\chi_P^f m_f/f_a)\bar{f}\gamma_5 f$, where $\chi_P^u = 1/\tan\beta$, $\chi_P^d = \chi_P^e = \tan\beta$ and m_f is the fermion mass [17, 351, 352].

6.3 A note on anomalies

In the effective Lagrangian for the model in this study, see Sec. 7.2, we include only tree-level couplings of the axion to SM fermions and do not include any coupling of the axion to gauge bosons via $a\,\mathrm{tr}F\widetilde{F}$ terms. This is motivated by a paper by Quevillon and Smith [17]. I would like to use this note to summarize their findings and to advertise the specific UV completion we choose.

The nature of the couplings in front of the $G\tilde{G}$ and $F\tilde{F}$-terms in Eq. (6.3) is often ascribed to the anomaly of $\mathrm{U}(1)_{\mathrm{PQ}}$ in the fermionic current: the Goldstone boson couples to its symmetry current, $f_a p^\mu = \langle 0|J^\mu_{\mathrm{PQ}}|a(p)\rangle$ which itself is anomalous, $\partial^\mu J^{\mathrm{PQ}}_\mu \sim g_s^2 \mathcal{N}_C G\tilde{G} + e^2 \mathcal{N}_{\mathrm{em}} F\tilde{F}$ ($\mathcal{N}_C$ and $\mathcal{N}_{\mathrm{em}}$ are the QCD and QED anomaly coefficients, respectively). Applying this argument, one seems to immediately be able to identify the couplings of axions to gauge bosons as induced by heavy fermions – without computations. However, as pointed out by Quevillon and Smith in [17], this identification is wrong. The appearance of the anomaly in the decay of axion to gauge bosons is traced back to a reparametrization of the fermion fields in the polar representation for the scalar field in the spontaneously broken theory,

$$\Phi_{\mathrm{polar}}(x) = \frac{1}{\sqrt{2}}(v_a + \sigma^0(x))e^{-ia(x)/v_a}\ , \qquad \Phi_{\mathrm{linear}}(x) = v_a + \sigma^0(x) + ia(x)\ ,$$

with v_a the vev of the scalar. The reparametrization in the polar representation induces derivative couplings between the axion and fermions, contributing to the axion-gauge boson decay via triangle graphs. In fact, the tree-level couplings between axion and gauge bosons arise only in the polar representation and only serve to cancel the contribution from an anomalous part arising in the triangle amplitude between the axions and the fermions. In the linear representation no anomalies show up in the calculation of the decay amplitudes. In the case of axion-photon couplings (or axion-gluon couplings) this can be understood in terms of the axial-current Ward identity,

$$\partial_\mu\left(\bar{f}\gamma^\mu\gamma_5 f\right) - e^2\mathcal{N}_{\mathrm{em}} F_{\mu\nu}\tilde{F}^{\mu\nu} = 2im_f\bar{f}\gamma_5 f\ . \tag{6.7}$$

Finally, both representations give the same physical decay amplitude, as expected, since the physics should not depend on the chosen representation [3].

This remark will become important later once we get to the UV completion of our model in Chap. 7. While in generic axion models the polar representation is convenient (also because it makes the shift symmetry of the axion under a $\mathrm{U}(1)_{\mathrm{PQ}}$ transformation apparent [355] as it is the only field transforming under the non-linearly realized PQ symmetry, $a \to a + \theta$), as Quevillon and Smith argue, this representation leads to a proliferation of parameters. In fact, the non-renormalizable effective Lagrangian naturally arising in generic axion models in the polar representation for the scalar field would be

$$\mathcal{L}_{NR}^{\mathrm{eff}} = \frac{1}{2}\partial_\mu a^0 \partial^\mu a^0 - \frac{\partial_\mu a^0}{2v_a} \sum_{f=u,d,e,\nu} \left(\chi_V^f \bar{\psi}_f \gamma^\mu \psi_f + \chi_A^f \bar{\psi}_f \gamma^\mu \gamma_5 \psi_f \right) \tag{6.8}$$

$$+ \frac{a^0}{16\pi^2 v_a} \left(g_s^2 \mathcal{N}_C^{eff} G_{\mu\nu}^a \tilde{G}^{a,\mu\nu} + g^2 \mathcal{N}_L^{eff} W_{\mu\nu} \tilde{W}^{\mu\nu} + g'^2 \mathcal{N}_Y^{eff} B_{\mu\nu} \tilde{B}^{\mu\nu} \right), \tag{6.9}$$

with v_a some high scale and $\mathcal{N}_{C,L,Y}^{eff}$ and $\chi_{V,A}^f$ a priori free Wilson coefficients. It can then be shown that this Lagrangian is phenomenologically equivalent to the renormalizable Lagrangian

$$\mathcal{L}_R^{\mathrm{eff}} = \frac{1}{2}\partial_\mu a^0 \partial^\mu a^0 - i \sum_{f=u,d,e} \frac{m_f}{v_a} \chi_P^f a^0 \bar{\psi}_f \gamma_5 \psi_f, \quad \text{with } \chi_P^f = \chi_A^f \tag{6.10}$$

typically arising if the linear representation is chosen. Both Lagrangians are equivalent if one requires anomaly cancellation and the validity of the naive Ward identities ($\partial_\mu V^\mu = 0$ and $\partial_\mu A^\mu = 2imP$). The renormalizable effective Lagrangian in Eq. (6.10) is thus the most general implementation of the PQ (or DFSZ) axion model.

6.4 Properties of the axion at low energies

To study the low energy properties of the axion, let us come back to the more general setup in Sec. 6.1. In particular, let us restrict ourselves to the axion coupling to gluons in Eq. (6.3), namely $(a+\bar{\theta})/f_a G\tilde{G}$. The axion propagator receives quantum corrections by the interactions to QCD, see Fig. 6.1, and gives the axion a mass. By dimensional analysis, and since the axion-gluon coupling is suppressed by $1/f_a$, the mass of the axion should scale as $\mathcal{O}\left(\Lambda_{\mathrm{QCD}}^4/f_a^2\right)$, where Λ_{QCD} is the typical non-perturbative QCD-scale. Strictly speaking, at temperatures below the QCD phase transition, the axion is therefore not a Goldstone boson since the symmetry is explicitly broken by the anomaly and Goldstone's theorem does not apply (it leads to a tilting of the mexican hat potential). This tilting is exactly what solves the strong CP problem, which gives a preferred CP conserving value for the axial component and the fact that the axion is not an exact Goldstone boson is, after all, the whole point of the axion solution of the strong CP problem. However, as $m_a \ll \Lambda_{\mathrm{QCD}}$, the effect of the anomaly can be seen as a small perturbation and we have spontaneous symmetry breaking of an approximate symmetry, much as in the case of the pions where the symmetry is explicitly broken by the small quark masses. f_a is experimentally

[3] While the misidentification of axion-gauge-boson decays being induced by the anomaly remains without consequences for axion couplings to gluons and photons, for chiral gauge theories this shortcut leads to wrong results: $a \to \gamma Z, ZZ, W^+W^-$ are in fact independent of the anomaly coefficients due to additional anomalies which occur for chiral theories [17].

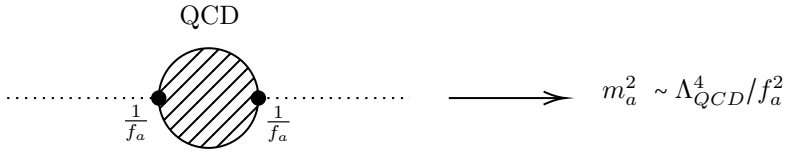


FIGURE 6.1: Axion-QCD interactions give the axion a mass of $\mathcal{O}\left(\Lambda^4_{\mathrm{QCD}}/f_a^2\right)$ (by dimensional analysis).

bounded from below, hence the mass of the axion has to be small, typically below $\mathcal{O}(\mathrm{eV})$. This hierarchy between the mass of the axion and the QCD scale allows the proper use of effective field theories, i.e. an expansion in hierarchies of different scales. This permits us to determine the axion properties comparably precisely and an effective theory containing only the three pseudo Nambu-Goldstone bosons of the broken $SU(2)_A$ (the three pions) and the axion works well.

Therefore, at low energy in the XPT EFT framework, the theory can effectively be described in terms of the axion and the pseudo Nambu-Goldstone bosons of the broken symmetries which become the relevant degrees of freedom. Let us assume that we have integrated out the heavier quarks and only look at the 2-flavour theory of the two lightest quarks u and d.

Starting from the Lagrangian in Eq. 6.3 with only the axion-gluon coupling we perform the field redefinition Eq. 6.4 and remove the $G\tilde{G}$ term. The effective chiral Lagrangian [4] including the axion then reads [18]

$$\mathcal{L}_{\chi PT+a} = f_\pi^2 \,\mathrm{Tr}\left(\partial_\mu U \partial^\mu U^\dagger\right) - B_0 f_\pi^2 \,\mathrm{Tr}\left(U\mathcal{M}_a^\dagger + \mathcal{M}_a U^\dagger\right) , \tag{6.11}$$

with $f_\pi = 92.07\mathrm{MeV}$ the pion decay constant and

$$U = \exp\left(\frac{i}{f_\pi}\begin{pmatrix} \pi^0 & \sqrt{2}\pi^- \\ \sqrt{2}\pi^+ & -\pi^0 \end{pmatrix}\right)$$

a matrix containing the neutral and the two charged pion fields, where $\mathcal{M}_a$ is the quark mass matrix dressed with the axion field. Since the chiral Lagrangian for pions is an expansion in the light quark matrix and since the dependence on the axion is encoded in the quark mass matrix, we obtain an analytical dependence on the axion field. Choosing $Q_a = \mathrm{diag}(1/2, 1/2)$ and Taylor-expanding the exponential in $a/f_a \ll 1$ in the pion ground state $\pi_0 = 0$ gives an effective potential for the axion of the form [18]

$$V(a, \pi_0) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4m_u m_d}{(m_u + m_d)^2}\sin^2\left(\frac{a + \bar{\theta} f_a}{2 f_a}\right)} , \tag{6.12}$$

which has a minimum at $a/f_a = -\bar{\theta}$ (actually an infinite number of minima $a/f_a + \bar{\theta} = 2\pi n$). The coefficient in front of the $G\tilde{G}$-term is consequently dynamically determined by the state of minimal energy of the axion field. This solves the strong CP problem in a very elegant way, spontaneously relaxing the CP-violating term to zero and thereby explaining its smallness and the non-observation of strong CP violation.

[4] For a general introduction to chiral perturbation theory see for instance [66, 134, 356–358].

Expanding the potential further in a/f_a one obtains an expression for the mass of the axion,

$$m_a^2 = \frac{m_u m_d}{(m_u + m_d)^2} \frac{m_\pi^2 f_\pi^2}{f_a^2} \ , \tag{6.13}$$

which scales indeed as $\mathcal{O}\left(\Lambda_{\rm QCD}^4/f_a^2\right)$. Including higher order corrections one finds

$$m_a \simeq 5.70(7)\ \mu\text{eV}\left(\frac{10^{12}\ \text{GeV}}{f_a}\right)$$

Note that the coupling to gluons generically also introduces an effective coupling to photons (e.g. via a quark loop)

$$\frac{\alpha_{\rm em}}{8\pi}\frac{g_{a\gamma}}{f_a} aF\tilde{F} \ ,$$

with a model independent contribution $g_{a\gamma} = 1.92(4)$ – it is suppressed by $\alpha_{\rm em}$ but still sizeable at low energies when QCD confines. This coupling allows for the conversion of axions to photons in the background of an electromagnetic field which is an important mechanism for numerous axion searches.

Axions themselves could be candidates for dark matter and could have been produced in the early universe through the misalignment mechanism, decay of topological defects or thermally, in which case they would constitute hot dark matter. Indeed, they are one of the oldest DM candidates. Since in our specific model we focus on axion-like particles with masses in the MeV-GeV range which do not constitute the dark matter but mediate the interactions with the SM, we will not introduce the theoretical groundwork related to axion DM, or for detection strategies of the QCD axion with masses below $\mathcal{O}$(MeV). A very nice review can for instance be found in [359].

6.5 Axion-like Particles

Many extensions of the SM give rise to particles which do not necessarily solve the strong CP problem but can inherit similar properties and are therefore called "axion-like particles" (ALPs). If they do solve the strong CP problem they are often referred to as the "QCD axion". As axions, ALPs come from an extension of the standard model by a new global U(1) symmetry that is spontaneously broken by a hidden Higgs mechanism [360].

$$\mathcal{L}_{ALP} = \frac{1}{2}\partial_\mu a \partial^\mu a - g_{ag}\frac{a}{f_a}\frac{\alpha_s}{8\pi}G^a_{\mu\nu}\tilde{G}^{a\mu\nu} - g_{a\gamma}\frac{a}{f_a}\frac{\alpha}{8\pi}F_{\mu\nu}\tilde{F}^{\mu\nu} + \frac{1}{2}\frac{g_{af}}{f_a}\partial_\mu a \bar{f}\gamma^\mu\gamma^5 f \tag{6.14}$$

We have now set the theoretical grounds to finally discuss the ALP-mediated dark matter model in the following chapter which is the main subject of this study.

7 Axion-like particles as mediators for dark matter

7.1 Introduction

The last two decades have seen a surge in the theoretical exploration of dark matter (DM) models, and brought a wealth of new data to constrain them. As we saw in Sec. 5.3, the possibility that dark matter directly couples to standard model (SM) states via renormalizable interactions is by now severely constrained. For example, we saw that LHC data and direct detection experiments rule out vast parts of the parameter space for standard electroweak-scale WIMPs produced via thermal freeze-out [50, 51]. This has led to a paradigm shift towards various other possibilities for dark matter candidates, with masses above 1 TeV or below 10 GeV, and with other production mechanisms. In particular, models where the DM-SM interactions are mediated by additional particles have attracted much attention.

Among such models are those with a fermionic dark matter candidate and a pseudoscalar mediator a, interacting with the SM via a Lagrangian of the form

$$\mathcal{L} = a \sum_f C_f \frac{m_f}{f_a} \bar{f} i\gamma_5 f + a\, C_\chi \frac{m_\chi}{f_a} \bar{\chi} i\gamma_5 \chi + \dots \tag{7.1}$$

Here f stands for any SM fermion, and χ is the DM particle. One possible motivation for this type of interaction is that it allows the efficient suppression of direct detection cross sections while still allowing for large enough DM-SM interactions to create the observed relic abundance via thermal freeze-out. However, as we saw in Sec. 5.4, freeze-out is of course by no means the only mechanism for dark matter production. In fact, one of the oldest dark matter candidates is itself a pseudoscalar state which is extremely light and extremely weakly coupled, namely a non-thermally produced QCD axion. At intermediate values of the couplings, dark matter could be produced by the freeze-in mechanism: Assuming that its abundance is initially zero after reheating, dark matter will be gradually produced by scattering processes in the thermal plasma without ever reaching thermal equilibrium, until these processes decouple and the DM abundance remains constant [19, 52].

In this chapter we study a dark matter model with interactions of the type of Eq. (7.1) but for a range of couplings which are too small to give the correct relic density through freeze-out. Our motivation is unrelated to direct detection constraints (which are irrelevant for such small couplings) but are rather guided by top-down reasoning: If there exists a light pseudoscalar a,

it might well be an axion-like particle (ALP), by which we mean the pseudo-Goldstone boson of an approximate $\mathrm{U}(1)_{\mathrm{PQ}}$ global symmetry which is spontaneously broken at a high scale f_a; this provides a reason for it to be light. It is natural to have a emerge from an extended Higgs sector in some UV completion of the SM, which could explain the flavour-preserving couplings to SM states of Eq. (7.1). In this case, besides these effectively renormalizable couplings, one also expects dimension-5 couplings of the form

$$\mathcal{L} = a \sum_f C_f \frac{y_f}{\sqrt{2} f_a} h \bar{f} i \gamma_5 f + \dots \tag{7.2}$$

where h is the SM Higgs boson, which will turn out to be important for the dark matter abundance in parts of the parameter space.

If there are additional fermions χ charged under $\mathrm{U}(1)_{\mathrm{PQ}}$, they would also couple to the ALP. If there are no additional fermions charged under both $\mathrm{U}(1)_{\mathrm{PQ}}$ and under the standard model, there exists a field basis in which no dimension-5 couplings $aF\widetilde{F}$ between the ALP and the SM gauge bosons are induced above the electroweak scale. The only relevant couplings between the SM and the ALP are then those of Eqs. (7.1) and (7.2). Moreover, since there is no evidence for new physics close to the electroweak scale, the scale f_a should be large, $f_a \gg$ TeV. Hence the ALP couplings to fermions may be too small to allow for dark matter production via freeze-out, and other production mechanisms should be studied.

It is tempting to try to identify a in this scenario with a (DFSZ-like) QCD axion itself. However, if a QCD axion were massive enough to be short-lived on cosmological scales (thus mediating the SM-DM interactions, rather than being part of the dark matter sector itself), it would be subject to multiple astrophysical and laboratory constraints, see e.g. [18] for an overview. While it may be possible to circumvent these constraints by extensive model-building, in this paper we will not attempt to do so but rather assume that the a mass is mainly due to some explicitly $\mathrm{U}(1)_{\mathrm{PQ}}$-breaking effect other than the anomaly. The usual QCD axion relation between m_a and f_a therefore no longer holds, and a does not contribute to solving the strong CP problem.

To summarise our findings, we will show that the observed dark matter abundance can be produced in this model by several distinct mechanisms, depending on the values of the ALP-SM and ALP-DM couplings as well as on the ALP and DM masses. It can be produced via ALP-mediated freeze-in from the scattering of SM particles, via freeze-in from the scattering of ALPs, via freeze-out of ALP-DM interactions which kept the DM in equilibrium with the ALP, or finally via standard freeze-out. We will numerically study a number of example scenarios with regard to the most relevant cosmological, astrophysical and collider constraints.

7.2 Particle content, interactions and relevant processes

Our model contains a single Dirac fermion χ and a pseudoscalar ALP a. The Lagrangian is

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\mathrm{SM}} &+ \frac{1}{2}\partial_\mu a \partial^\mu a + \bar{\chi}(i\partial\!\!\!/ - m_\chi)\chi + i\sum_f \frac{C_f}{f_a}\left(m_f + \frac{y_f}{\sqrt{2}}h\right) a \bar{f}\gamma_5 f \\ &- \frac{1}{2}m_a^2 a^2 + i\frac{C_\chi}{f_a} m_\chi\, a\bar{\chi}\gamma_5\chi\,,\end{aligned} \tag{7.3}$$

where f is any standard model fermion with mass m_f and Yukawa coupling y_f and h is the Higgs boson. This Lagrangian is valid below the electroweak symmetry breaking scale. In App. F we present a possible ALP-model for the effective Lagrangian (7.3). The purpose was to provide an explicit UV completion for the simple effective model (7.3) with ALP masses in the MeV-GeV range and a coupling to the DM particles. The ALP-model is inspired by the DFSZ-axion. As briefly discussed in Sec. 6.3, from Ref. [17], we see that there is no need to include both the 5-dimensional gauge terms and fermion mass dependent pseudoscalar terms if there are no heavy vector fermions integrated out. If the fermion content is only that coming from the SM, then the anomaly cancellation requirements automatically imply that the contributions from dimension-5 terms cancel to leave only dimension-4, mass dependent terms. Hence, we have not included any coupling of the ALP to gauge bosons via $a\,\mathrm{tr}F\widetilde{F}$ terms. Including them would be perfectly possible but would lead to a proliferation of parameters, so we restrict our analysis to more minimal models where there are no heavy fermions charged under both the SM gauge group and $\mathrm{U}(1)_{\mathrm{PQ}}$. Nevertheless, effective $a\gamma\gamma$ and agg vertices are induced at one loop by (finite) SM fermion triangle graphs.

We will often refer to the hidden sector (HS), by which we mean the ensemble of ALPs and DM particles. For future convenience, we define the effective axion couplings by

$$g_{aff} = \frac{C_f}{f_a}\,, \qquad g_{a\chi\chi} = \frac{C_\chi}{f_a}\,, \tag{7.4}$$

where we will later refer to $g_{a\chi\chi}$ as the hidden sector coupling, and to g_{aff} as the connector coupling. In this model, dark matter is stabilized by a global $\mathrm{U}(1)$ χ-number symmetry which also ensures that it can only be produced in $\chi\bar{\chi}$ pairs. A variant with a Majorana dark matter candidate stabilized by a $\mathbb{Z}_2$ symmetry could also be viable.

From the above Lagrangian we can deduce the various means by which:

- the dark matter can interact with standard model particles, e.g. through interactions with fermions mediated by the pseudoscalar, $f\bar{f} \leftrightarrow \chi\bar{\chi}$ or $f\bar{f} \leftrightarrow h\chi\bar{\chi}$, or via a loop diagram with gauge bosons, $\gamma\gamma \leftrightarrow \chi\bar{\chi}$;
- the ALP can interact with the SM, e.g. via $q\bar{q} \leftrightarrow ga$, $qg \leftrightarrow qa$, or the (inverse) decay $f\bar{f} \leftrightarrow a$, or $f\bar{f} \leftrightarrow ha$;
- the ALPs and the dark matter can interact, i.e. $aa \leftrightarrow \chi\bar{\chi}$ or $\chi a \leftrightarrow \chi a$.

These different interactions, where those responsible for DM and ALP generation are summarised in Tab. 7.1, give rise to a rich phenomenology. Different DM generation mechanisms are at play in different regions of the parameter space. The study of these mechanisms was pioneered in a different model by [21], and more recently refined in [22].

As detailed in the next sections, the mechanisms at play in our model are the following[1]:

- For sufficiently small $g_{a\chi\chi}$ and g_{aff}, and assuming that the HS relic density is negligible at reheating, dark matter is produced by freeze-in. Depending on the couplings, the

[1]Note also that if the ALP is coupled to a strongly-interacting hidden sector, the DM can be produced via an additional, qualitatively different production mechanism, the so-called SIMP mechanism [339].

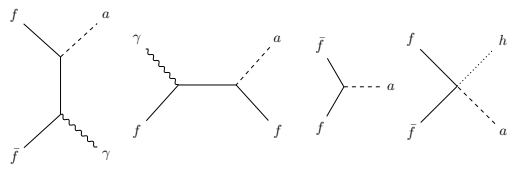
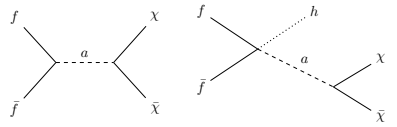
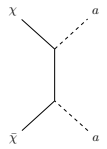
Interaction	Processes	Scaling
SM ↔ ALPs		g_{aff}^2
SM ↔ χ		$(g_{aff} \cdot g_{a\chi\chi})^2$
ALPs ↔ χ		$g_{a\chi\chi}^4$

Table 7.1: Different sector interaction processes and their scaling with the hidden sector and connector couplings. Note that the diagrams $f\bar{f} \leftrightarrow \gamma a$ and $f\gamma \leftrightarrow fa$ shown are equivalent to those for $qq \leftrightarrow ga$, $qg \leftrightarrow qa$, $\bar{f}\gamma \leftrightarrow \bar{f}a$ and $\bar{q}g \leftrightarrow \bar{q}a$ and therefore the latter are not shown explicitly.

dominant process may be direct freeze-in $f\bar{f} \to \chi\bar{\chi}$ mediated by an off-shell ALP, or ALP production from SM states followed by $aa \to \chi\bar{\chi}$ (where the ALP may or may not be in equilibrium with the SM). These processes are infrared dominated, since they are induced by (effectively) renormalizable operators. For large reheating temperatures, the dominant process for direct freeze-in will be $2 \to 3$ scattering $f\bar{f} \to h\chi\bar{\chi}$, which is suppressed by phase space but ultraviolet-dominated, i.e. sensitive to the reheating temperature.

- For intermediate g_{aff}, ALPs will be produced abundantly from the SM thermal bath but will not reach thermal equilibrium with the SM particles. However, if $g_{a\chi\chi}$ is sufficiently large, the hidden sector will form a separate thermal bath at a lower temperature, in which DM is produced by freeze-out. We study this mechanism of freeze-out from a thermally decoupled dark sector, or to put it simply, decoupled freeze-out (DFO), in detail.

- For larger g_{aff} and $g_{a\chi\chi}$, the entire dark sector will thermalize with the SM, and DM can be produced via standard freeze-out. Since we are interested in the small-coupling regime, and since the standard freeze-out phase is already well studied in the literature, we will not analyse it in detail.

7.3 The phase diagram of production mechanisms

7.3.1 General Boltzmann equations

In order to study the various DM generation mechanisms we start from the covariant form of the Boltzmann equation for the Friedmann-Lemaître-Robertson-Walker metric, where the momenta of the particles have been integrated over. The SM particles are in equilibrium with the photon

bath at temperature T, allowing us to replace the distributions of SM particles by the equilibrium distributions. We neglect quantum statistical factors, assuming Maxwell-Boltzmann statistics:

$$f_{\rm eq}(p,T) = e^{-E/T}, \tag{7.5}$$

such that the particles' number density, energy density and pressure equilibrium distributions are given by Eq. (5.26). r

More generally, the hidden sector particles may be in kinetic equilibrium at a temperature T' which is different from T:

$$f(p,T') = \frac{n(T')}{n_{\rm eq}(T')} e^{-E/T'} . \tag{7.6}$$

Kinetic equilibrium is not reached in the freeze-in phase, where the scattering among hidden sector particles is not efficient (see [20] for a detailed analysis of freeze-in in a related model).

The Boltzmann equations governing the evolution of the a and χ number densities are given by

$$\begin{aligned}
\frac{{\rm d}n_\chi}{{\rm d}t} + 3Hn_\chi &= \sum_f \langle \sigma_{\chi\bar\chi\to f\bar f} v\rangle \left(\left(n_\chi^{\rm eq}\right)^2 - n_\chi^2\right) + \langle \sigma_{aa\to\chi\bar\chi} v\rangle n_a^2 - \langle \sigma_{\chi\bar\chi\to aa} v\rangle n_\chi^2 \\
&\quad + \sum_{i,j,k} \left(\langle \sigma_{\chi\bar\chi\to ijk} v\rangle \left(\left(n_\chi^{\rm eq}\right)^2 - n_\chi^2\right) + \langle \sigma_{\chi\bar\chi i\to jk} v^2\rangle n_i^{\rm eq} \left(\left(n_\chi^{\rm eq}\right)^2 - n_\chi^2\right)\right) , \\
\frac{{\rm d}n_a}{{\rm d}t} + 3Hn_a &= \sum_{i,j,k} \langle \sigma_{ia\to jk} v\rangle \left(n_a^{\rm eq} n_i^{\rm eq} - n_a n_i^{\rm eq}\right) + \langle \Gamma_a\rangle \left(n_a^{\rm eq} - n_a\right) \\
&\quad - \langle \sigma_{aa\to\chi\bar\chi} v\rangle n_a^2 + \langle \sigma_{\chi\bar\chi\to aa} v\rangle n_\chi^2 .
\end{aligned} \tag{7.7}$$

Here i,j,k are SM particles which are involved in the relevant processes, as listed in Tab. 7.1. For example, if all fermion couplings C_f are of the same order and at temperatures where all SM particles are relativistic, the dominant contribution for ALP production comes from $i = g$, $j = \bar k = t$. The cross sections for the processes entering Eq. (7.7) are given in App. G. Most of the processes appearing in Eq. (7.7) are $2 \to 2$ processes, and enter via the typical thermally averaged cross section $\langle \sigma v\rangle$, defined by[2]

$$\langle \sigma_{12\to 34} v\rangle = \frac{C}{2TK_2(m_1/T)\,K_2(m_2/T)} \int_{s_{\rm min}}^\infty \sigma(s) \frac{F(m_1,m_2,s)^2}{m_1^2 m_2^2 \sqrt{s}} K_1(\sqrt{s}/T)\, {\rm d}s , \tag{7.8}$$

where $\sigma(s)$ is the cross section for the process $12 \to 34$ as a function of the squared centre-of-mass energy s. Here C is an additional factor of 1/2 for the case of identical particles in the initial state ($C = 1$ for non-identical initial state particles). We have further used the abbreviation

$$F(m_1,m_2,s) = \frac{\sqrt{(s-(m_1+m_2)^2)(s-(m_1-m_2)^2)}}{2} \tag{7.9}$$

and the lower limit of the integral is given by $s_{\rm min} = \max\left((m_1+m_2)^2, (m_3+m_4)^2\right)$. Note that, where appropriate, we have made use of detailed balance,

$$\langle \sigma_{ij\to kl} v\rangle n_i^{\rm eq} n_j^{\rm eq} = \langle \sigma_{kl\to ij} v\rangle n_k^{\rm eq} n_l^{\rm eq} . \tag{7.10}$$

[2] We have not explicitly stated the temperature dependence of the thermally averaged cross sections in Eq. (7.7) as this depends on the production regime considered. For example, in the DFO mechanism the hidden sector interactions take place at the common temperature of the hidden sector which is different from that of the SM, see below in section 7.3.3.

For the (inverse) decays which contribute, we define the thermally averaged decay rate to be

$$\langle \Gamma_a \rangle = \Gamma_a \frac{K_1(m_a/T)}{K_2(m_a/T)} \,. \tag{7.11}$$

where Γ_a is the decay rate of the pseudoscalar a. We provide details of the ALP decay rate in the various decay channels in App. G.4. For the calculation of the DM relic density we restrict ourselves to (inverse) decay into fermions in the perturbative regime, see Eq. (G.9).[3]

The Hubble rate in Eq. (7.7) is given in terms of the total energy density of the universe $\rho = \rho_{\rm SM} + \rho_{\rm HS}$, via

$$H = \left(\frac{8}{3}\pi G \rho\right)^{1/2} \,, \tag{7.12}$$

where G is the gravitational constant. The energy density of the visible sector is

$$\rho_{\rm SM} = g_{*\rho,{\rm SM}}(T) \frac{\pi^2}{30} T^4 \,, \tag{7.13}$$

with $g_{*\rho,{\rm SM}}(T)$ the SM effective degrees of freedom in energy. The hidden sector energy density differs according to the production regime as we will explain in more detail in section 7.3.3. For the SM effective degrees of freedom in energy and in entropy, denoted by $g_{*\rho}$ and g_{*s}, respectively, we adopt the recent results from an improved analysis, covering a wide range of temperatures [361]. This analysis is based on state-of-the-art results of perturbative and non-perturbative calculations of thermodynamic quantities in the SM. In particular, for an improved description of the QCD phase transition they adopt results from the "Budapest-Wuppertal"-collaboration [225] in $2+1+1$ flavour lattice QCD. We use the fit functions provided in this paper in our analysis.

Note that by solving this set of differential equations we only calculate the abundance of χ particles and later set $n_{\rm DM} = n_\chi + n_{\bar{\chi}} = 2n_\chi$. In addition, we point out that later in this section we will find that it is convenient to make a change of variables from t to $z = m_\chi/T$, and the differential equations will be written as a function of z instead of t.

7.3.2 Freeze-in

The freeze-in mechanism can generate the observed DM abundance for the case that both g_{aff} and $g_{a\chi\chi}$ are sufficiently small, assuming that the initial abundance of DM particles and ALPs is zero or negligible. Both a and χ will be produced from interactions amongst SM particles in the thermal bath, but for small g_{aff} and $g_{a\chi\chi}$ the DM will not thermalize, neither with the photons nor with the ALPs. The production stops once the temperature of the particles involved drops below the DM mass scale, since the abundances become Boltzmann-suppressed and the production rate becomes negligible compared to the Hubble expansion rate. The co-moving dark matter number density is effectively frozen from that point on, while the ALPs will eventually decay back to standard model states.

[3] For energies in the range $\sim 0.5-1.2$ GeV various hadronic decay channels open up. However, when we present results on the DM relic density, we are either in the perturbative regime or apply a hard cut on the decay width below $T \leq 600$ MeV (see also the discussion in section 7.6.4 and in App. G.4).

The interplay of the different sectors allows for three different freeze-in regimes, namely freeze-in of the DM directly from SM particles, freeze-in from the ALPs and sequential freeze-in [20, 22]. All three regimes are characterised by a value of $g_{a\chi\chi}$ which is too small for the hidden sector particles to establish thermal equilibrium among each other. Depending on the range in g_{aff}, the ALPs may or may not be in equilibrium with the SM, as described in the following.

7.3.2.1 Freeze-in from SM particles

In this regime, the DM is produced directly from the SM fermions via an s-channel ALP (see Tab. 7.1). In order to generate the correct relic density via this mechanism, one requires $g_{a\chi\chi}$ to be small, such that $\langle\sigma_{aa\to\chi\bar{\chi}}v\rangle \propto g_{a\chi\chi}^4$ is small enough for $aa \to \chi\bar{\chi}$ to be negligible compared to $f\bar{f} \to (h)\chi\bar{\chi}$. In this regime we find that ALPs are in thermal equilibrium with the SM, i.e. $n_a = n_a^{\mathrm{eq}}(T)$.

In order to calculate the relic abundance we require the appropriate Boltzmann equation for this region, which we obtain by adapting Eq. (7.7) accordingly. We first define the dark matter yield $Y = Y_\chi$ as usual by

$$Y = \frac{n_\chi}{s} \tag{7.14}$$

where $s = s_{\mathrm{SM}} + s_{\mathrm{HS}}$ is the total entropy density of the universe. The entropy density in the freeze-in regime is given by

$$s = \frac{2\pi^2}{45} g_{*s}(T) T^3 \,, \tag{7.15}$$

with the effective degrees of freedom in entropy given by $g_{*s}(T) = g_{*s,\mathrm{SM}}(T) + g_{*s,\mathrm{a}}(T)$ and we can neglect the entropy density of the DM particles.

The ALP abundance plays no role in this freeze-in scenario, so we neglect all terms in Eq. (7.7) involving n_a. Since the DM is never in equilibrium, $Y \ll Y_{\mathrm{eq}}$, we can neglect the back-reaction as usual. Concerning the contribution to the DM yield from $2 \to 2$ scattering, it is convenient to introduce the reaction density for the interaction $ij \to kl$ [21]

$$\gamma_{ij\to kl} = \langle\sigma_{ij\to kl}\, v\rangle n_i^{\mathrm{eq}} n_j^{\mathrm{eq}} \tag{7.16}$$

in terms of which the Boltzmann equation Eq. (7.7) for the yield from freeze-in becomes

$$\frac{\mathrm{d}Y}{\mathrm{d}T} = -\sum_f \frac{\gamma_{\chi\bar{\chi}\to f\bar{f}}}{3Hs^2}\frac{\mathrm{d}s}{\mathrm{d}T} + (2 \leftrightarrow 3 \text{ terms})\,. \tag{7.17}$$

Here we have replaced the z dependence by a dependence on T. The corresponding infrared-dominated contribution to the yield today at $T \approx 0$ is therefore obtained by integrating the reaction density,

$$Y_{0,\mathrm{IR}} = -\sum_f \int_0^\infty \frac{\gamma_{\chi\bar{\chi}\to f\bar{f}}}{3Hs^2}\frac{\mathrm{d}s}{\mathrm{d}T}\,\mathrm{d}T\,. \tag{7.18}$$

An important point concerns the relative magnitude of the contributions from $2 \to 2$ interactions and $2 \to 3$ interactions. In a hypothetical model with only renormalizable couplings between the ALP and the SM fermions, as defined by Eq. (7.1), freeze-in would proceed only via $f\bar{f} \to \chi\bar{\chi}$ scattering, which is infrared dominated. However, as we stated previously, in realistic models

where the ALP-fermion couplings are generated from the ALP mixing with the standard model Higgs, one also obtains the dimension-5 couplings of Eq. (7.2). These couplings induce $2 \to 3$ interactions which are phase-space suppressed, but nevertheless cannot be neglected in general,[4] because they are ultraviolet dominated and their contributions therefore scale as the reheating temperature $T_{\rm RH}$. For instance, if $T_{\rm RH}$ is large enough to neglect all masses of the involved particles, then the dark matter yield from such a process can be estimated as [341]

$$Y_{0,\rm UV} \approx \frac{135}{(2\pi)^9} \frac{1}{1.66\, g_*^{3/2}} g_{aff}^2 g_{a\chi\chi}^2\, y_f^2\, m_\chi^2\, T_{\rm RH}\, M_P\,. \tag{7.19}$$

A thorough analysis of ALP production from UV freeze-in, allowing for all possible dimension-5 couplings between the ALP and the SM to be present, was conducted in [362], and models similar to ours have been studied in both their UV and IR freeze-in phases in [363]. We will eventually find that, for reheating temperatures $T_{\rm RH} \gtrsim 200$ GeV, UV freeze-in via $2 \to 3$ scattering gives significant contributions to the dark matter yield; see App. G for details of the computation.

7.3.2.2 Freeze-in from ALPs

In this regime, the ALPs are in thermal equilibrium with the SM, and produce the DM via the t-channel process $aa \to \chi\bar{\chi}$. Obtaining the observed relic density via this mechanism requires a larger $g_{a\chi\chi}$ compared to the case in section 7.3.2.1, and a correspondingly smaller g_{aff} to avoid overproduction.

Since the ALPs are in thermal equilibrium with the SM particles, this regime is conceptually very similar to the previous one. We just need to replace the reaction density $\gamma_{\chi\bar{\chi}\to f\bar{f}}$ in Eq. (7.18) by the appropriate $\gamma_{\chi\bar{\chi}\to aa}$:

$$Y_0 = -\int_0^\infty \frac{\gamma_{\chi\bar{\chi}\to aa}}{3Hs^2} \frac{{\rm d}s}{{\rm d}T}\, {\rm d}T. \tag{7.20}$$

In this regime, there are no direct sizeable UV-dominated contributions to the DM yield. However, UV-dominated $2 \to 2$ processes such as $f\bar{f} \to ah$ will contribute to the ALP abundance and will therefore reduce the value of g_{aff} at which the ALPs reach equilibrium with the SM. We determine this boundary by numerically solving the Boltzmann equation for the ALPs and checking whether $n_a \gtrsim n_a^{\rm eq}$. For IR-dominated contributions, the yield is essentially insensitive to the upper integration limit (provided it is chosen above the freeze-out temperature of the top quark). Since the UV-dominated contributions depend on the reheating temperature, the appropriate integration limit is at $z_{RH} = m_\chi/T_{RH}$. As always we assume that the universe has reheated into SM particles only, i.e. $n_a(T_{RH}) \approx 0$. Note that $\mathcal{O}(1)$ corrections to the UV contributions can be expected, due to the uncertainty associated with the details of the reheating mechanism.

7.3.2.3 Sequential freeze-in

In the sequential freeze-in regime, dark matter is again produced from ALPs via $aa \to \chi\bar{\chi}$. The difference with the previous regime is that the ALPs are not thermalized. Nevertheless, the ALP abundance is sufficient to obtain the correct DM relic density.

[4] Note on the other hand that the $f\bar{f}h \to \chi\bar{\chi}$ interactions included in Eq. (7.7) can safely be neglected.

Calculating the DM relic abundance is non-trivial, since the ALPs are not even in kinetic equilibrium, and therefore the unintegrated Boltzmann equations governing the individual momentum modes should be employed for a precise quantitative analysis. This was done in a study dedicated to the freeze-in regime in Ref. [20] for a related model. Due to time limitations, we choose to follow Ref. [22], where a simplified analysis is proposed employing the integrated Boltzmann equations

$$\begin{aligned} \frac{\mathrm{d}n_a}{\mathrm{d}t} + 3Hn_a &= \sum_{i,j,k} \gamma^{\mathrm{eq}}_{ia\to jk} + \langle \Gamma_a \rangle \, n_a^{\mathrm{eq}}(T) \,, \\ \frac{\mathrm{d}n_\chi}{\mathrm{d}t} + 3Hn_\chi &= \gamma_{\chi\bar{\chi}\to aa} \,, \end{aligned} \tag{7.21}$$

where $\gamma_{aa\to\chi\bar{\chi}} = \langle \sigma_{aa\to\chi\bar{\chi}} v \rangle(T)\, n_a^2$, and n_a is determined from the first equation. The second term on the RHS of the first equation describes the production of the ALP via inverse decays. We have neglected the term $-\gamma_{\chi\bar{\chi}\to aa}$ in the first equation since $n_a^{\mathrm{eq}}(T) \gg n_a$. The temperature of the ALPs is taken to be the temperature T of the SM bath.

This gives a rough estimate of the parameter values leading to the observed relic density; however we stress that this method is relying on several rather crude approximations. Specifically, we assume that the axions are in kinetic equilibrium in order to be able to define a temperature, and we then equate this temperature with that of the photons. It should be possible to obtain a more precise result following the method of Ref. [20], but this is beyond the scope of our work.

7.3.3 Decoupled freeze-out (DFO)

For a sufficiently large hidden sector coupling $g_{a\chi\chi}$, the interactions between the HS particles will be strong enough for them to thermalize. If the ALP-SM couplings g_{aff} are sufficiently small, then the hidden sector will nevertheless be thermally decoupled from the SM thermal bath, at a temperature $T' \ll T$. This regime is somewhat more complicated to analyse than the freeze-in regime, due to an increased number of interactions playing a role, and due to the non-trivial interplay between the hidden sector number densities and the temperature T'. We will therefore proceed to explain some of the technical aspects in more detail.

As in the case of freeze-in, we assume the initial number density of DM particles and of ALPs to be negligible. The DM and ALPs will then freeze in until the abundances are sufficient to allow the particles to thermalize. This happens once the reaction rate for $aa \leftrightarrow \chi\bar{\chi}$ scattering exceeds the expansion rate of the universe,

$$\Gamma_{aa\to\chi\chi} \equiv \langle \sigma_{aa\to\chi\bar{\chi}}\, v \rangle n_a^{\mathrm{eq}}(T') > H \,. \tag{7.22}$$

While $aa \leftrightarrow \chi\bar{\chi}$ is the dominant process for DM and ALP annihilation and production, the DM and ALP number densities will approximately track their equilibrium values $n_\chi^{\mathrm{eq}}(T')$ and $n_a^{\mathrm{eq}}(T')$, respectively. After T' falls below the mass of the χ and a, we allow for the possibility that n_χ and n_a deviate from the equilibrium distributions. The number densities n_χ and n_a and the hidden-sector temperature T' are then given by the solution of a system of three coupled differential equations, as explained in the following.

Solving the energy transfer Boltzmann equation The temperature T' of the hidden sector is therefore clearly a crucial ingredient in the calculation of the relic density in this regime. It depends on how much energy is transferred from the SM to the hidden sector. The temperature T' of the hidden sector is calculated from the hidden-sector energy density ρ' via the equation of state. To determine ρ', we start from the Boltzmann equation for the phase-space distribution $f(p,t)$ of either of the hidden sector particle species:

$$\left(\partial_t - H\,p\,\partial_p\right) f(p,t) = \frac{1}{E(p)} C[f(p,t)]\,. \tag{7.23}$$

Here we have used that, by isotropy and homogeneity, $f(p,t)$ can only depend on the modulus p of its 3-momentum and on time. $C[f]$ is the collision operator and E is the energy. With the energy density given by Eq. (5.26), we can integrate Eq. (7.23), writing

$$\frac{\partial \rho'}{\partial t} + 3H\left(\rho' + P'\right) = \int \frac{d^3p}{(2\pi)^3} C[f(p,t)] \tag{7.24}$$

where we have integrated by parts and used $P' = \frac{1}{3}\langle p \frac{\partial E}{\partial p}\rangle$. The integrated collision operator for a $1\,2 \to 3\,4$ process is

$$\int \frac{d^3p}{(2\pi)^3} C[f] = g_1 g_2 \int \frac{d^3p_1}{(2\pi)^3}\frac{d^3p_2}{(2\pi)^3} f_1(p_1) f_2(p_2) v_{\text{Møl}}\, \mathcal{E}(\vec{p}_1,\vec{p}_2) \tag{7.25}$$

where $v_{\text{Møl}}$ is the usual Møller velocity and the energy transfer rate $\mathcal{E}$ is, in terms of the matrix element $\mathcal{M}$ and the transferred energy ΔE_{tr},

$$\mathcal{E}(\vec{p}_1,\vec{p}_2) = \frac{1}{2E_1 2E_2 v_{\text{Møl}}} \int \prod_{i=3,4} \frac{d^3p_i}{(2\pi)^3}\frac{1}{2E_i} |\mathcal{M}|^2 (2\pi)^4 \delta^{(4)}(p_1+p_2-p_3-p_4)\Delta E_{tr}\,. \tag{7.26}$$

Solving Eq. (7.24) enables us to obtain ρ' as a function of the temperature of the visible sector T. In the reannihilation region described in Ref. [364], Eq. (7.24) was applied to a model with fermionic dark matter coupled to both the SM photon and a hidden-sector photon. [5] In that case, the hidden sector is populated predominantly via $f\bar{f} \to \chi\bar{\chi}$ processes, and the mass degeneracies of the particles involved allow to analytically simplify the integrated collision operator along the lines of [23], leaving a single integral to be evaluated numerically. In our model, the $2 \to 2$ processes $f\bar{f} \to a\gamma$, $\gamma f \to af$, $q\bar{q} \to ag$, $gq \to aq$ and hermitian conjugates must also be taken into account, as well as (inverse) decay processes $f\bar{f} \to a$. To treat the $2 \to 2$ processes, some work is needed to cast the integrated collision operator into a form amenable to numerical integration. For a process $1\,2 \to 3\,4$, neglecting the back-reaction, the integrated collision operator is given in Eq. (7.25) in terms of the energy transfer rate $\mathcal{E}(\vec{p}_1,\vec{p}_2)$ for an initial state with momenta $\vec{p}_1$ and $\vec{p}_2$, and of the Møller velocity

$$v_{\text{Møl}} = \frac{\sqrt{(s-(m_1+m_2)^2)(s-(m_1-m_2)^2)}}{2E_1E_2} \equiv \frac{F}{E_1E_2}\,. \tag{7.27}$$

[5] Note that while our DFO region is very similar to the reannihilation region of Ref. [364], it differs from the latter in two ways: the first is that in our case the energy transfer from the SM to the dark sector proceeds via the SM to ALP transition rather than SM to DM, see figure 7.7. The second is that during freeze-out, the production of DM from the SM does not become the dominant production mechanism, and as a result we do not observe a bump in the DM comoving number density as opposed to the case in Ref. [364].

Assuming the momentum distributions to be Maxwell-Boltzmann-like, the collision term becomes

$$\int \frac{d^3p}{(2\pi)^3} C[f] = g_1 g_2 \int \frac{d^3p_1}{(2\pi)^3} \frac{d^3p_2}{(2\pi)^3} e^{-(E_1+E_2)/T} v_{\text{Møl}} \mathcal{E}(\vec{p}_1, \vec{p}_2) \tag{7.28}$$
$$= \frac{g_1 g_2}{32\pi^4} \int dE_+ dE_- ds\, e^{-E_+/T} F\, \mathcal{E}(\vec{p}_1, \vec{p}_2)$$

where, following Ref. [23], in the second line we have replaced $E_+ = E_1 + E_2$, $E_- = E_1 - E_2$ and $s = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2$. The energy transfer rate $\mathcal{E}(\vec{p}_1, \vec{p}_2)$ can be written as in Eq. (7.26). For the transferred energy ΔE_{tr}, there are two cases which we need to distinguish: For $f\bar{f} \to \chi\bar{\chi}$, $\Delta E_{tr} = E_+$, and for $f\bar{f} \to a\gamma$, $\gamma f \to af$, $q\bar{q} \to ag$, $gq \to aq$ (and the corresponding hermitian conjugate processes) we have $\Delta E_{tr} = E_3$. It turns out that in the former case, as studied in Ref. [364], the identical particles in the initial state and ΔE_{tr} being independent of the final state result in several simplifications in $\mathcal{E}(\vec{p}_1, \vec{p}_2)$, absent for the latter. In order to demonstrate the difference between these two cases let us now perform the integration in p_4 (where in the following discussion we will use the notation $p_i = |\vec{p}_i|$):

$$\mathcal{E}(\vec{p}_1, \vec{p}_2) = \frac{1}{4F} \int \frac{d^3p_3}{(2\pi)^3} \frac{1}{2E_3 2E_4} |i\mathcal{M}|^2 (2\pi)\delta(E_1 + E_2 - E_3 - E_4)\Delta E_{tr}$$

where $E_4 = (m_4^2 + |\vec{p}_1 + \vec{p}_2 - \vec{p}_3|^2)^{1/2}$. The next steps depend on the value of ΔE_{tr}. Clearly if $\Delta E_{\text{tr}} = E_+$, it does not affect the integration over p_3 and Ω, resulting in the simple relation $\mathcal{E}(\vec{p}_1, \vec{p}_2) = \sigma(\vec{p}_1, \vec{p}_2)E_+$. However, when $\Delta E_{\text{tr}} = E_3$ integrating in p_3 and Ω is not so straightforward. For the latter case, let us consider the argument of the delta function which we will call $g(p_3) \equiv E_1 + E_2 - E_3 - E_4$:

$$g(p_3) = E_+ - E_3 - (m_4^2 + E_+^2 - s + p_3^2 - 2p_1 p_3\, c_{13} - 2p_2 p_3\, c_{23})^{1/2}$$

where $c_{ij} = \cos\theta_{ij}$, and we have defined θ_{ij} to be the angle between the momenta of particles i and j. Since we perform the integration in c_{13}, we need to replace θ_{23} by $\theta_{13} + \theta_{12}$, where $c_{12} = (E_+^2 - s - p_1^2 - p_2^2)/(2p_1 p_2)$. In order to obtain the solution p_3^0 of the equation $g(p_3) = 0$ analytically, and to speed up the computation, we are obliged to make the assumption $E_3 = p_3$. This assumption is easily justifiable for the dominantly contributing case $f = t$ (and results in an effect of at most $\sim 1.4\%$), and we have checked that $q\bar{q} \to ga$ and $f\bar{f} \to \gamma a$ dominate over $f\gamma \to fa$ and $qg \to qa$. We then find

$$\mathcal{E}(\vec{p}_1, \vec{p}_2) = \frac{1}{8\pi F} \int p_3^2 dp_3 dc_{13} \frac{1}{4E_4} |i\mathcal{M}|^2 \frac{\delta(p_3 - p_3^0)}{|g'(p_3)|_{p_3 \to p_3^0}} \tag{7.29}$$

where we have replaced ΔE_{tr} by E_3. The result for the integrated collision operator is then obtained by inserting Eq. (7.29) in Eq. (7.28) and performing the integration numerically, setting $E_3 = p_3$.

For the case of $f\bar{f} \to \chi\bar{\chi}$, and $\mathcal{E}(\vec{p}_1, \vec{p}_2) = \sigma(\vec{p}_1, \vec{p}_2)E_+$, the integrated collision operator can easily be simplified to [364]:

$$\int \frac{d^3p_3}{(2\pi)^3} C[f_3] = \frac{g_1 g_2}{32\pi^4} \int ds\, \sigma(s)\, (s - 4m^2)\, s T K_2\left(\frac{\sqrt{s}}{T}\right) \tag{7.30}$$

where we have adopted the limits of integration [23, 364]: $s > (m_1 + m_s)^2$, $E_+ > \sqrt{s}$ and $-\frac{2F}{s}\sqrt{E_+^2 - s} < E_- + E_+ \frac{m_1^2 - m_2^2}{s} < \frac{2F}{s}\sqrt{E_+^2 - s}$.

The case of inverse decays, $1\,2 \to X$, is not identical, and deserves to be studied as it has not previously been required, starting with

$$\begin{aligned}\mathcal{E}(\vec{p}_1, \vec{p}_2) =& \frac{1}{2E_1 2E_2 v_{Mol}} \int \frac{d^3 p_X}{(2\pi)^3} \frac{1}{2E_X} |i\mathcal{M}|^2 (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_X) \Delta E_{tr} \\ =& \frac{1}{2E_1 2E_2 v_{Mol}} \frac{1}{2E_X} |i\mathcal{M}|^2 (2\pi) \delta(E_1 + E_2 - E_X) E_+ \,. \end{aligned} \tag{7.31}$$

We can then insert this expression in Eq. (7.28), and find for $m_1 = m_2 = m$

$$\begin{aligned}\int \frac{d^3 p_3}{(2\pi)^3} C[f_3] =& \frac{g_1 g_2}{32\pi^3} \int dE_+ \, e^{-E_+/T} \, |i\mathcal{M}|^2 \, \frac{2F\, E_+}{m_X^2} \sqrt{E_+^2 - m_X^2} \, \Theta(m_X^2 - 4m^2) \\ =& T \frac{g_1 g_2}{32\pi^3} m_X \sqrt{m_X^2 - 4m^2} \, |i\mathcal{M}|^2 \, K_2\left(\frac{m_X}{T}\right) \Theta(m_X^2 - 4m^2) \end{aligned} \tag{7.32}$$

which in terms of the decay width of particle X can be expressed as

$$\int \frac{d^3 p_3}{(2\pi)^3} C[f_3] = T \frac{g_1 g_2}{2\pi^2} \Gamma_X \, m_X^3 \, K_2\left(\frac{m_X}{T}\right) \Theta(m_X^2 - 4m^2) \,.$$

Determining the hidden sector equation of state As explained in the previous paragraph, by first calculating the energy which is transferred from the visible to the hidden sector as in Eq. (7.24) we can subsequently obtain T' from the equation of state of the hidden sector. In this equation we require the energy density and pressure of the hidden sector, which in general will be given by the sum of the energy density and pressure distributions of the ALPs and the DM,

$$\rho' + P' = \rho_a + \rho_\chi + P_a + P_\chi \,. \tag{7.33}$$

The hidden sector temperature becomes meaningful only once the hidden sector particles thermalize. Assuming an initially radiation dominated universe in both the visible and hidden sectors, $P' = \rho'/3$ and $\rho \propto T^4$, and changing variables using $\frac{\partial}{\partial t} \approx -HT\frac{\partial}{\partial T}$, we find that Eq. (7.24) becomes

$$\frac{\partial \rho'}{\partial t} + 4H\,\rho' = -H\left(T\frac{\partial \rho'}{\partial T} - 4\rho'\right) = -HT\rho\frac{\partial}{\partial T}\left(\frac{\rho'}{\rho}\right) = \int \frac{d^3 p}{(2\pi)^3} C[f(p,t)] \,. \tag{7.34}$$

As long as the hidden sector particles are relativistic, solving this equation for ρ' will provide us with the temperature of the hidden sector via Eq. (5.26). This will in fact be what we need as an initial condition for T', and to be more precise it can be obtained by iteratively solving

$$\frac{\rho'}{\rho} = \frac{\rho_a^{\rm eq}(T') + \rho_\chi^{\rm eq}(T')}{\frac{\pi^2}{30} g_{\rm eff,SM}(T) T^4} \,. \tag{7.35}$$

However, as we will shortly explain in more detail, as soon as the ALP and DM number densities deviate from equilibrium, they remain tightly entangled with the hidden sector temperature and cannot be solved independently. In addition, since the hidden sector particles are massive, we cannot assume radiation domination at all times, i.e. from thermalization until freeze-out. We

will therefore have to solve a system of three coupled differential equations in n_χ, n_a and T'. This is facilitated by rewriting Eq. (7.24) as

$$z\frac{\mathrm{d}\rho'}{\mathrm{d}T'}\frac{\mathrm{d}T'}{\mathrm{d}z} = -3(\rho' + P') + \frac{1}{H}\int \frac{d^3p}{(2\pi)^3} C[f(p,t)] \,. \tag{7.36}$$

Naturally, the evolution of the DM and ALP number densities depends on T'. As we will see, the equation of state and consequently the dynamics of the hidden sector temperature T', in return, will depend on the hidden sector number densities. In particular, we will have to distinguish between three main regions for the hidden sector equation of state:

- Once thermal equilibrium is achieved between the ALP and the DM, this state is maintained at a common temperature T' for $T' > m_\chi, m_a$. Since we assume Maxwell-Boltzmann statistics, the ALP and DM energy density and pressure in the equation of state are given by the equilibrium expressions in Eq. (5.26),

$$\rho' = \rho_a^{\mathrm{eq}}(T') + \rho_{\mathrm{DM}}^{\mathrm{eq}}(T') \qquad P' = P_a^{\mathrm{eq}}(T') + P_{\mathrm{DM}}^{\mathrm{eq}}(T') \tag{7.37}$$

  and the evolution of the hidden sector temperature is given by Eq. (7.36).

- Once the hidden sector temperature drops below the mass of the DM, $T' \lesssim m_\chi$, the DM equilibrium distribution becomes Boltzmann suppressed, as the production from the ALPs becomes less efficient. Therefore we allow for the possibility that the DM number density will start to diverge from the equilibrium distribution:

$$\rho_\chi = \frac{\rho_\chi^{\mathrm{eq}}(T')}{n_\chi^{\mathrm{eq}}(T')}\, n_\chi \qquad P_\chi = \frac{P_\chi^{\mathrm{eq}}(T')}{n_\chi^{\mathrm{eq}}(T')}\, n_\chi = T' n_\chi \,, \tag{7.38}$$

  where we make the assumption that the DM phase space distribution is proportional to the Maxwell-Boltzmann distribution, and that the proportionality factor is a function of temperature only. Note that in general decoupling takes place later after $T' \approx m_\chi/10$ [365] and that the point where we change expressions is therefore chosen conservatively.

- Once the ALP number density becomes Boltzmann suppressed around $T' \lesssim m_a$, the ALP number density, too, diverges from $n_a^{\mathrm{eq}}(T')$ and will eventually decouple. The hidden sector equation of state is then given by

$$\rho' + P' = \frac{\rho_\chi^{\mathrm{eq}}(T')}{n_\chi^{\mathrm{eq}}(T')}\, n_\chi + \frac{\rho_a^{\mathrm{eq}}(T')}{n_a^{\mathrm{eq}}(T')}\, n_a + T'\left(n_\chi + n_a\right) \,. \tag{7.39}$$

Note that we have verified that kinetic equilibrium is maintained in the hidden sector throughout freeze out due to the scattering process $\chi a \to \chi a$.

Solving the system of coupled differential equations The arguments given above suggest that a full numerical solution of the ALP and DM number densities together with the temperature T' seems to be in order. The potentially relevant a-number changing processes are (inverse) decays into SM particles and $2 \to 2$ scattering processes, as well as $aa \leftrightarrow \chi\bar{\chi}$. The relevant processes for DM production and annihilation are also $aa \leftrightarrow \chi\bar{\chi}$, as well as a-mediated pair

production $f\bar{f} \to \chi\bar{\chi}$. The quantities n_χ, n_a and T' are finally obtained as the solutions of a system of three coupled differential equations, namely Eq. (7.36) in conjunction with the equation of state, and the Boltzmann equations

$$\begin{aligned} Hz\frac{\mathrm{d}n_\chi}{\mathrm{d}z} + 3Hn_\chi = &\sum_f \left\langle \sigma_{\chi\bar{\chi}\to f\bar{f}} v \right\rangle (T) n_\chi^{\mathrm{eq}}(T)^2 + \sum_{i,j,k} \left\langle \sigma_{\chi\bar{\chi} i\to jk} v^2 \right\rangle n_i^{\mathrm{eq}} \left(n_\chi^{\mathrm{eq}}\right)^2 \\ &+ \left\langle \sigma_{aa\to\chi\bar{\chi}} v \right\rangle (T') n_a^2 - \left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle (T') n_\chi^2 \end{aligned} \tag{7.40}$$

$$\begin{aligned} Hz\frac{\mathrm{d}n_a}{\mathrm{d}z} + 3Hn_a = &\left\langle \Gamma_a \right\rangle n_a^{\mathrm{eq}}(T) + \sum_{i,j,k} \left\langle \sigma_{ia\to jk} v \right\rangle (T) n_a^{\mathrm{eq}}(T) n_i^{\mathrm{eq}}(T) \\ &- \left\langle \sigma_{aa\to\chi\bar{\chi}} v \right\rangle (T') n_a^2 + \left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle (T') n_\chi^2 \end{aligned} \tag{7.41}$$

with the initial condition for T' provided by the solution of Eq. (7.35). Here i, j, k are SM particles involved in the a-number changing processes, see Tab. 7.1. We have again neglected the backreaction of the hidden sector abundance on the SM bath because $T' \ll T$, so $n_a^{\mathrm{eq}}(T') \ll n_a^{\mathrm{eq}}(T)$ and $n_\chi^{\mathrm{eq}}(T') \ll n_\chi^{\mathrm{eq}}(T)$.[6] To facilitate the numerical solution of this stiff set of equations we make an approximation for small values of $z = m_\chi/T$ by replacing n_χ by its equilibrium distribution $n_\chi^{\mathrm{eq}}(T')$ in Eq. (7.41). This approximation is highly justified as long as the SM source terms are negligible and as long as kinetic equilibrium is maintained. Since we do not know a priori how these effects will influence the dynamics of our system of coupled equations we use this simplification as long as we can be certain that the ALP and the DM are still in kinetic equilibrium and as long as the source terms are negligible with respect to the hidden sector interaction. More precisely, we use $n_\chi^{\mathrm{eq}}(T')$ as long as $T' > m_\chi$ (which corresponds to $z' > 1$). We further check that $\langle \sigma_{\mathrm{SM}\to\mathrm{HS}} v \rangle n_{\mathrm{HS}}^2 \lesssim \langle \sigma_{\chi\chi\to aa} v(T') \rangle n_\chi^2$ which turns out to always be the case. As we will show in section 7.6, elastic scattering $a\chi \leftrightarrow a\chi$ and $a\bar{\chi} \leftrightarrow a\bar{\chi}$ remains active and keeps the hidden sector particles in kinetic equilibrium, thus the point where we change equations is chosen conservatively.

7.4 Finite temperature effects

We incorporate finite-temperature corrections in our analysis in light of the observation in Ref. [20] that these have a large impact on the relic density in the case of sequential freeze-in. Such corrections arise due to the fact that perturbation theory breaks down in the presence of an additional scale, the temperature of the plasma. Calculations therefore require the resummation of diagrams, which can be implemented in a simple way, following Ref. [20], on adopting the hard thermal-loop (HTL) approximation [21], where only loops involving soft momenta $\sim gT \lesssim T$ are resummed. Within this approximation, it turns out that the masses of fermions and gauge bosons appearing in the dispersion relations can simply be replaced by temperature dependent quantities. For the gauge bosons this amounts to adding a mass to the propagators which is the thermal Debye mass (neglecting differences between the transverse and longitudinal modes),

[6] This simplification becomes important also from a technical point of view since we are solving the Boltzmann equations numerically. Adding negligible terms to the already stiff equations leads to instabilities. Note also that we do not solve the equations numerically in terms of $Y = n/s$ because we would have to deal with derivatives $\mathrm{d}h_{\mathrm{eff}}/\mathrm{d}T$ which caused severe numerical instabilities.

which for the case of gluons is given by [366]

$$m_g^2(T) = \frac{4\pi\alpha_s(2\pi T)T^2}{3}\left(n_c + \frac{n_f(T)}{2}\right) , \tag{7.42}$$

with n_f the number of active flavours in the plasma, and for photons [366]

$$m_\gamma^2(T) = \frac{4\pi\alpha_{\rm em}T^2 n_{\rm ch}(T)}{3} , \tag{7.43}$$

with $n_{\rm ch}(T)$ the number of electromagnetically charged particles in the plasma.

For the case of fermions we also add a thermal mass correction to the propagators, neglecting the difference between particle and hole states [367]. For the leptons this takes the form [368]

$$m_l^2(T) = \frac{4\pi\alpha_{\rm em}(2\pi T)T^2 q_f^2}{8} \tag{7.44}$$

and for quarks

$$m_q^2(T) = \frac{4\pi T^2(\alpha_s(2\pi T)/6 + \alpha_{\rm em}(2\pi T)q_f^2)}{8} . \tag{7.45}$$

In order to capture the effect of thermal corrections on interaction vertices we renormalize the coupling constants at the scale of the first Matsubara mode, $\omega = 2\pi T$, making use of the renormalization group equations in vacuum [367]. Note that we perform this running for the electromagnetic, strong couplings and Yukawa couplings, although for the latter we found that this resulted in a negligible effect on the results.

7.5 Testing the model: current constraints on ALPs

In this section we review the most important existing and future experimental constraints on our model. ALPs can significantly impact astrophysical, cosmological and collider processes, and the various constraints on ALPs have widely been discussed in the literature (see [369] for a review). In section 7.6 we will explore the regions of parameter space of our model where the correct relic density is obtained via different production mechanisms. We will therefore focus on those collider, astrophysical and cosmological constraints which are relevant in these regions.

While the discussion has been general so far, to discuss the constraints and for the subsequent numerical analysis we concentrate on the simplest case of strictly flavour-universal axion-fermion couplings, i.e. universal values of C_f and g_{aff} in Eqs. (7.3) and (7.4).

7.5.1 Astrophysical constraints

7.5.1.1 Horizontal branch stars

The conditions inside horizontal branch (HB) stars in globular clusters would allow a sizeable production of ALPs inside the stellar core. HB stars lie in the horizontal branch in the Hertzsprung-Russell diagram, i.e. they contain little metal and are in the state of a stable helium burning core and a hydrogen burning shell. Red Giants (RG) evolve to HB stars after their core becomes

so hot and dense that the helium ignites. The interaction of ALPs with particles in the core of HB stars could alter the stellar evolution of these stars and in order to avoid conflicts with observations, constraints must be placed on the ALP parameter space. The constraints coming from HB stars are two-fold: If ALPs interact very weakly, they mostly escape freely, draining energy from the star. This puts an upper bound on the coupling. On the other hand, for larger couplings, they will be trapped inside the source and radiate energy which provides a lower bound.

Energy Loss Argument The first scenario we envisage is a very weakly interacting ALP, which would stream out freely of the hot core and accelerate the cooling of the star. A branch in the Hertzsprung-Russell diagram corresponds to a certain stage in the stellar evolution, depending on the type of nuclear fuel being burnt. Stars inside globular clusters travel along the different branches of the Hertzsprung-Russell diagram during their stellar evolution. Hence, the number of stars inside a certain branch is proportional to the time the star "lives" in a certain branch. The R-parameter which is the ratio of the number of HB stars over RG stars,

$$R = \frac{N_{\rm HB}}{N_{\rm RG}} = \frac{\tau_{\rm HB}}{\tau_{\rm RG}} \,, \tag{7.46}$$

is a suitable quantity to assess the impact of new particles on HB stars inside globular clusters: An additional energy loss mechanism by ALPs would manifest itself in a contraction and heating of the core, which would mainly have an effect on the nuclear fuel consumption (helium), hence reducing the star's Helium burning lifetime and therefore the number of HB stars. The Helium burning lifetime is related to the energy emitted per unit time and mass averaged over a typical HB core, $\langle\varepsilon\rangle$, via its luminosity. One usually assumes that the energy emission caused by additional particles should not exceed the energy emission from Helium burning, $\langle\varepsilon_a\rangle \lesssim \langle\varepsilon_{3\alpha}\rangle \approx 100$ erg g^{-1} s^{-1} [25]. The dominant production mechanism for ALPs with a coupling to electrons in HB stars is the Compton process. Here, we consider three possible ALP production mechanisms, the Compton process, bremsstrahlung and the loop induced Primakoff process. For the calculation of the energy emitted by ALPs we use Eq. (12) in Ref. [370] for the Compton process, expression (56c) given in Ref. [371] for the bremsstrahlung process, and Eq. (A2) in Ref. [372] for the Primakoff process where we replace the axion-photon coupling by expression J.5, making the approximation of zero momentum transfer.

ALP opacity While in most models ALPs have very weak couplings, one could on the other hand imagine that the new particle has such a strong coupling that it is trapped inside the stellar core. In this scenario the ALP will scatter and decay inside the star, transporting heat between different regions and thereby contributing to radiative energy transfer. Again, observations indicate that models of stellar structure work well and that a new source of radiative energy transfer should therefore not exceed that due to photons, parametrized by the so-called Rosseland mean opacity. The photon opacity in HB stars is typically $\kappa_\gamma \sim 0.5$ cm^2/g. We include four sources for the ALP's effective opacity, namely (loop induced) decay into photons and the inverse Primakoff process (using Eqs. (A.2) and (A.6) from Ref. [372], replacing $g_{a\gamma\gamma}$ by expression J.5), the Compton process and bremsstrahlung (using Eq. (3.8) in Ref. [25]).

7.5.1.2 Supernova SN1987A

The SN1987A neutrino observations in 1987 in the Large Magellanic cloud provide us with another strong constraint. The neutrino flux coming from the core collapse, which lasted a few seconds, was measured. The presence of weakly-coupled particles in supernovae would provide an additional cooling mechanism. By comparing to data from SN1987A, bounds on the ALP coupling can therefore be derived. Here, following Ref. [24, 373] we will adopt the "Raffelt criterion" to obtain these bounds, i.e. we demand that the luminosity emitted due to the ALP is less than that due to the neutrinos, $L_\nu = 3 \cdot 10^{52}$ ergs/s [73]. The cooling time should also be in agreement with the data. The high temperatures and density provided by SN1987A creates a conducive environment for the production of weakly-interacting particles such as the ALP. As the core temperature is around 30 MeV, and taking the Boltzmann tail into consideration, we could imagine that ALPs of masses up to $\mathcal{O}(100\text{ MeV})$ could be generated. As our ALPs only interact with fermions at tree level, the dominant production mechanism of the ALPs is via bremsstrahlung. In order to calculate the production rate, we first need to relate the quark Lagrangian to the nuclear Lagrangian:

$$\mathcal{L} \supset -i \sum_N \frac{m_N C_N}{f_a} a \bar{N} \gamma^\mu \gamma_5 N \,, \tag{7.47}$$

where we sum over the proton and neutron, $N = p, n$ with masses m_N and coupling to the ALP C_N respectively. Fortunately, the relation between the ALP-quark couplings and the ALP-nucleon couplings is well known, and the state of the art results can be found in Ref. [374]. We then obtain $C_p = C_n = 0.43\, C_f$. Corrections to the diagrammatic calculation of the nuclear-scattering bremsstrahlung cross section are obtained using the results for the spin-flip current at N^{3}LO in chiral perturbation theory. Three multiplicative factors reproducing these corrections were provided in Ref. [24], and are included in our analysis:

- The factor $\gamma_f = 1/(1 + \frac{n_B \sigma_{np\pi}}{2\omega})$ acts as a cut-off preventing scattering at arbitrarily low energies, where n_B is the baryon number density, ω is the energy, and $\sigma_{np\pi} \simeq 15$ is the nucleon-nucleon scattering cross section when the pion is massless.
- The factor γ_p accounts for the change in phase space due to the non-zero pion mass, and is obtained from s defined in Eq. (49) of Ref. [375], which we multiply by a factor $1/(1 - e^{-x})$ to account for the detailed-balance condition, following Ref. [24].
- Finally γ_h contains the ratio of the mean free path of the weakly-coupled particle calculated at higher order in χPT to that at Born level, and is given by r_{Y_e} given in Eq. (5) of Ref. [376].

Putting these factors together with the Born-level expression, the decay rate is given by:

$$\Gamma_a = \sum_N \frac{Y_N^2 C_N^2}{8 f_a^2} \frac{n_B^2 \sigma_{np\pi}}{\omega} \gamma_f \gamma_p \gamma_h \sqrt{1 - \frac{m_a^2}{\omega^2}} \,, \tag{7.48}$$

where Y_N is the mass fraction of the nucleon N (we adopt $Y_p = 0.3$ [373]). Making use of this decay width, on integrating over the volume V and the ALP phase-space we obtain the

luminosity using [24]

$$L_a = \int_0^{R_\nu} dV \int \frac{d^3 k_a}{(2\pi)^3}\, \omega\, e^{-\omega/T}\, \Gamma_a \exp\left(-\int_0^{R_{\rm far}} dr \left(\Gamma_a + \Gamma_{all}^{\rm abs}\right)\right) \tag{7.49}$$

where k_a is the ALP momentum. The integration limits involve two radii: R_ν is the neutrinosphere radius, $\simeq 40$ km, beyond which most neutrinos free stream until arriving at Earth, and the far radius is given by $R_{\rm far} \simeq 100$ km. Further the contribution to the absorptive width of the decay of the ALP to leptons is included explicitly via

$$\Gamma_{a\ell\ell}^{\rm abs} = \frac{m_a}{8\pi} \sum_\ell \Theta(m_a - 2m_\ell)\, \frac{m_\ell^2}{f_a^2} \sqrt{1 - \frac{4m_\ell^2}{m_a^2}}\,. \tag{7.50}$$

In order to calculate the luminosity, we further require the temperature and density profile of the proto-neutron star. These profiles are subject to large uncertainties, which have been estimated in the past by comparing the different profiles proposed in the literature. We adopt the profiles introduced in Ref. [373], where a detailed study and comparison with other models for the hidden photon case is available. The comparison of the luminosity in the QCD axion and ALP case can be found in Ref. [24].

7.5.2 Collider constraints

In this section we review the existing and future experimental constraints on our model. ALPs can significantly impact astrophysical, cosmological and collider processes. The various constraints on ALPs have widely been discussed in the literature (see [369] for a review). In the next section we will explore the regions of parameter space of our model where the correct relic density is obtained via different production mechanisms. In this work, we will therefore focus on those astrophysical, cosmological and collider constraints, which are relevant in these regions.

Let us remark that we focus on constraints on the pseudoscalar mediator since direct and indirect detection experiments are out of reach for the small couplings considered in this work. On the contrary, diverse experimental efforts on ALPs possibility to test our model.

7.5.2.1 SLAC E137 electron beam dump experiment

Electron and proton beam-dump experiments are among the accelerator-based experiments where the ALPs in our model could be produced. Within the parameter region of interest a strong constraint comes from the E137 beam dump experiment at SLAC [377]. A highly energetic incident beam of electrons (~20 GeV) was dumped on Aluminium plates surrounded by cooling water, the target material. It was followed by an absorber hill of $D = 179$ m thickness and a decay volume of $L = 204$ m. At the end of the decay volume an electromagnetic calorimeter of transverse size (2 m × 3 m) and (3 m × 3 m) — during Run 1 and Run 2, respectively — was placed, able to detect photons and charged particles.

Calculating the number of events In the SLAC E137 electron beam dump experiment no events were detected [377]. This puts an experimental upper bound on the event rates $\mathcal{N}_s$

$$\mathcal{N}_s < \mathcal{N}_s^{\text{up}} = 2.996 \text{ for } \mathcal{N}_{s,\text{det}} = 0 \text{ at } 95\% \text{ CL}\,. \tag{7.51}$$

Following [360, 369, 377] we calculate the theoretically expected event rate to obtain a bound on the ALP's coupling and mass. The theoretical number of events can be obtained by an integral over all possible production cross sections folded with the probability to detect the ALP and the differential track-length distribution of the shower particles in the beam, multiplied by the total number of electrons dumped. To obtain the theoretical number of events in a specific decay channel we additionally multiply by the corresponding branching ratio, yielding

$$\mathcal{N}_{\text{theory}}(a \to XX) = \mathcal{N}_{e,\text{ inc}} \sum_{i,Z} P_Z \int \mathrm{d}E\; T_i(E)\, p(E)\, \sigma_Z^{i\to ak}(E)\; \mathcal{B}(a \to XX)\,. \tag{7.52}$$

$\mathcal{N}_{e,\text{ inc}}$ is the total number of electrons dumped (~ 30 Coulomb), P_Z is the number density of atoms per cm^3 of atomic number Z and $\sigma_Z^{i\to ak}$ are the cross sections for the different production processes with i the initial particle. T_i is the corresponding track-length distribution of the initial incoming particles, $i \in \{\gamma, e^+, e^-\}$. $p(E)$ is the probability for the ALP to decay visibly. The differential track-length distributions for the secondary shower particles can be estimated in the Weizsäcker-Williams approximation which allows one to relate the incident electrons in the initial state to the distributions of photons, electrons and positrons which are produced in electromagnetic shower cascades. Intuitively, the differential track-length $T(E)$ is the length a particle with given energy E will travel. One then directly considers the interactions of these particles with the target [378]. We extract the track-length distributions for photons, electrons and positrons in the beam dump (Fig. 14 of Ref. [377]) which have been calculated using the SLAC EGS program [379]. This spares us from finding an analytic expression for the Weizsäcker-Williams approximation. Through interactions of the shower particles with the target material, the ALPs in our model are predominantly produced by the Primakoff process ($Z\gamma \to Za$), by bremsstrahlung ($Ze^\pm \to Ze^\pm a$) and by (non-)resonant positron annihilation ($e^+e^- \to \gamma a$ and $e^+e^- \to a$). The Feynman diagrams for these processes are shown in Fig. 7.1. We can sum over the production channels since all other final state particles produced in the process are absorbed in the shielding. The ALP subsequently decays and its decay products can be detected in the electromagnetic calorimeter. Therefore, in order to obtain the theoretically expected number of events, we include the branching fraction for the decay channel $a \to \gamma\gamma$ (loop induced) and $a \to e^+e^-$, which opens up for $m_a > 2m_e$. The various decay channels of the ALP in our model are summarised in appendix G.4.

Following [360] we include the probability $p(E)$ for the ALP to be detected. Taking the experimental layout into account, we can envisage five scenarios where the ALP could have decayed invisibly: Either none or only one of the decay particles reaches the detector or both of them reach the detector but the opening angle is too small and the two final state particles are indistinguishable. For instance, if the ALP is highly boosted and its lifetime is long, it will certainly decay behind the detector. We will thus have to take the ALP decay probability into account to obtain a final result for the theoretically expected number of events. Details on the calculation of the probability for the ALP to decay visibly can be found in App. I. We assume that most of the ALPs are produced in the cooling water [369].

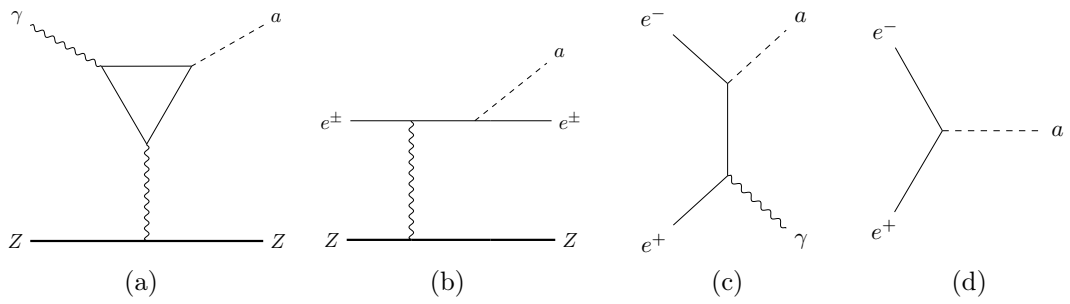


FIGURE 7.1: Feynman diagrams for the four different production mechanisms of ALPs in our model in the E137 beam dump experiment: (a) Primakoff production (b) bremsstrahlung (both off a target nucleus of atomic number Z) (c) non-resonant e^+e^- annihilation (d) resonant e^+e^- annihilation.

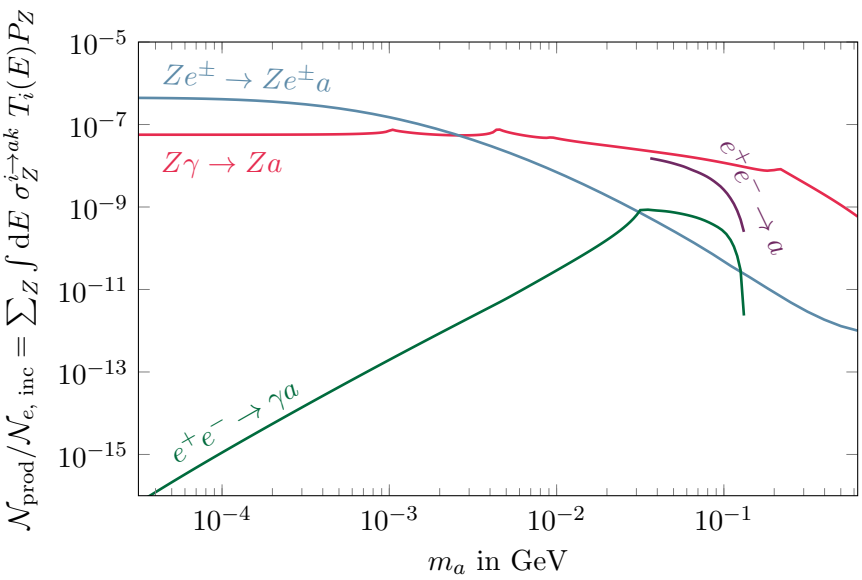


FIGURE 7.2: ALP yield per incident electron $\mathcal{N}_{\rm prod}/\mathcal{N}_{e,\,\rm inc} = \sum_Z \int {\rm d}E\ \sigma_Z^{i\to ak}\ T_i(E) P_Z$ for the various production processes of the ALP in our model in electron beam dump experiments, namely Primakoff production ($Z\gamma \to Za$), bremsstrahlung ($Ze^\pm \to Ze^\pm a$), non-resonant ($e^+e^- \to \gamma a$) and resonant ($e^+e^- \to a$) positron annihilation.

Details about the ALP production processes As mentioned above, the relevant production processes for our model in electron beam dump experiments are the Primakoff process, bremsstrahlung and (non-)resonant positron annihilation (see Fig. 7.1). In the Primakoff process, depicted in Fig. 7.1a), a secondary photon emits an ALP in the vicinity of the nucleus. Since the ALP in our model does not couple to photons at tree level, the process is induced at one-loop level, with a loop function depending on the momentum transfer. Details about the cross section and the loop calculation relevant for this process can be found in Apps. H.1 and J.1, respectively. In bremsstrahlung (see Fig. 7.1b), the electron is scattered off the mass shell of the nucleus and returns to the mass shell by emission of an ALP. Lastly, the ALP can be produced by (non-)resonant positron annihilation, see Figs. 7.1c) and 7.1d). The cross sections for bremsstrahlung and positron annihilation can be found in Apps. H.2 and H.3.

It is interesting to look at the interplay of the different contributions as a function of the ALP mass. From Eqs. (H.1) and (H.10) we see that for models where the ALP couples to both electrons and photons at tree level the Primakoff and the bremsstrahlung processes scale approximately

as

$$\sigma_Z^{\gamma\to a} \equiv \sigma(Z\gamma \to Za) \sim \alpha_{\rm em} g_{a\gamma\gamma}^2 \quad \text{and} \tag{7.53}$$

$$\sigma_Z^{e\to ae} \equiv \sigma(Ze^\pm \to Ze^\pm a) \sim \alpha_{\rm em}^2 g_{aff}^2 \frac{m_e^2}{m_a^2} \,, \tag{7.54}$$

for $m_e \ll m_a \ll E_i$, with E_i the energy of the initial particle. This means that bremsstrahlung is suppressed by a factor $\sigma_Z^{e\to ae}/\sigma_Z^{\gamma\to a} \sim \alpha_{\rm em} m_e^2/m_a^2$ (see also Ref. [380]). Interestingly, in our model, due to the loop suppression in the Primakoff process, see Eq. (J.5), the cross sections scale as

$$\sigma_Z^{e\to ae}/\sigma_Z^{\gamma\to a} \sim m_e^2/(m_a^2 \alpha_{\rm em}) \tag{7.55}$$

and the Primakoff process becomes dominant only for larger ALP masses. This can be observed in figure 7.2 where we plot the ALP yield per incident electron from the various production processes depicted in figure 7.1,

$$\mathcal{N}_{\rm prod}/\mathcal{N}_{e,\,\rm inc} = \sum_Z \int {\rm d}E\; \sigma_Z^{i\to ak}\; T_i(E)\; P_Z \,, \tag{7.56}$$

as a function of the ALP mass m_a, for an ALP-electron coupling $g_{aff} = 1$ GeV^{-1} and ignoring the detection probability of the ALP. The range of accessible positron energies limits the range of producible ALPs to those with masses $m_{a,\rm min} \sim 35$ MeV for the resonant case and $m_{a,\rm max} \sim 135$ MeV for production by both resonant and non-resonant positron annihilation. For small ALP masses non-resonant positron annihilation scales approximately as $\sigma^{e\to a\gamma} \sim m_a^2$. While in the original analysis [377], ALP production by non-resonant positron annihilation was included, in later analyses positron annihilation is often neglected altogether since the positrons arise as secondary particles in the beam. Yet, the number of secondary positrons can be large enough for positron annihilation to become important for large ALP masses. To be precise, it turns out that it indeed dominates over the often considered bremsstrahlung for ALP masses around 50 – 100 MeV and clearly cannot be neglected in models where the Primakoff process is highly suppressed, see also Ref. [381] for a related discussion in the case of dark photons. In our case, however, the ALP coupling to photons arises from fermion loops and the suppression of photoproduction is such that positron annihilation is non-negligible, but not large enough for it to become dominant.

7.5.2.2 Exotic Higgs decays at the LHC

Production of the ALP in Higgs decays would be mainly via $h \to aa$ (for $m_a \leq m_h/2$) or $h \to Za$ (for $m_a < m_h - m_Z \simeq 85$ GeV). While it is possible to write extra effective operators for the haa as well as the hZa coupling, we assume for this work that the only new terms are those in the Lagrangian in Eq. (7.3). Therefore both these Higgs decay channels are mediated by a fermion loop and dominated by the top-quark contribution.

Higgs exotic decay searches at the LHC currently include $h \to Z(a \to gg)$ [382] and $h \to aa$ with (1) both $a \to b\bar{b}$ [383], (2) both $a \to \ell^+\ell^-$ [384], and (3) $aa \to b\bar{b}\mu^+\mu-$ [385]. A detailed study of these reveals (see App. J for further details on the sensitivity of each of these) that

current searches are not sensitive to coupling regimes that would correspond to $g_{aff} \lesssim 1/\text{GeV}$. We therefore can safely ignore limits from these searches.

7.5.2.3 Production in decays of mesons

For intermediate mass ranges ($2m_e \lesssim m_a \lesssim 5$ GeV), the best chance for detecting an ALP could be via decays of heavy mesons. Dedicated experiments provide the possibility to obtain rather precise measurements of branching fractions and thereby constrain couplings to very small values. We consider constraints on the emission of an ALP in rare B and K flavour-changing neutral current processes, proceeding at the quark level via $b \to sa$ or $s \to da$. Depending on the decay mode and the lifetime of the ALP, such decays could be constrained by $B \to K^{(*)}\ell^+\ell^-$ and $K \to \pi\ell^+\ell^-$ or $B \to K^{(*)}\nu\bar{\nu}$ and $K \to \pi\nu\bar{\nu}$. For ALP masses smaller than $2m_e$, the only decay possible is to photons, and the lifetime of the ALP is long enough that it decays outside the detector volume of current experiments. In this case, it would show up as an invisible decay mode of the said meson with the accompanying products identifiable.

The main experimental constraints that we consider in this work are therefore:

- $B^+ \to K^+X(\to \mu^+\mu^-)$ for long-lived scalar X (LHCb [386]), the 95% C.L. upper limits on the branching ratio are given as a function of the lifetime of X in the range 0.1 to 1000 ps.
- $B^0 \to K^{*0}X(\to \mu^+\mu^-)$ where X is a scalar particle with mass in the range 214 to 4350 MeV (LHCb [387]), the 95% C.L. upper limits on the branching ratio are given as a function of the mass and lifetime of X. The limit is of the order 10^{-9} over the majority of this range.
- $B^0 \to K^{(*)}X(\to \mu^+\mu^-)$ at fixed target experiments, limits can be extracted from CHARM results as described in Ref. [388].
- $K^+ \to \pi^+\nu\bar{\nu}$ from NA62 [389], where 90% C.L. upper limits are given for the $K^+ \to \pi^+X$ branching ratio, where X is a long-lived scalar or pseudoscalar particle decaying outside the detector, for lifetimes longer than 100 ps.
- $K^\pm \to \pi^\pm e^+e^-$, where NA48/2 [390] provides a 90% C.L. upper limit (here we assume the lifetime should be less than 10 ns).
- $K^\pm \to \pi^\pm X(\to \mu^+\mu^-)$ for long-lived X (NA48/2 [391]) , the 90% C.L. upper limits on the branching ratio are given as a function of the lifetime of X in the range 100 ps to 100 ns.

We currently do not apply the $K \to \pi\gamma\gamma$ (which only dominates near 100 MeV) and $B_s \to \mu^+\mu^-$ (which only dominates for non-minimal flavour violation and for $m_a \sim 3-10$ GeV). Further the $B \to K\nu\bar{\nu}$ from BaBar [392] is not included as the relevant parameter space is covered by the $K^+ \to \pi^+\nu\bar{\nu}$ search.

In order to obtain the constraints on our parameter space, we first calculate the $B \to K^{(*)}a$ and $K \to \pi a$ branching ratios. For the former, the expression can be found in Refs. [393] and [394]. Form factors for the $B \to K$ transition are taken from Ref. [395] and for $B \to K^*$ from Ref. [396]. For the Kaon decay and $K \to \pi a$, we follow Ref. [397], taking into account the octet enhancement in non-leptonic Kaon decays.

For each point in parameter space, we then verify the ALP lifetime, which allows us to impose the appropriate limit, and then for the charged lepton channels multiply by the branching ratio for the necessary ALP decay, as given in App. G.4. For the $B \to K^{(*)}X(\to \mu^+\mu^-)$, $K^+ \to \pi^+ X$ (where X is long lived) and $K^\pm \to \pi^\pm X(\to \mu^+\mu^-)$ channels, the experimental constraints are directly given for the contribution of a new particle X, as a function of the mass and lifetime of X, such that we do not require the SM branching ratio. These limits can easily be translated from the $m_a - \tau_a$ to the (m_a, g_{aff})-plane. For $K^+ \to \pi^+ e^+ e^-$ however we need to include the SM contribution, for which we adopt the chiral perturbation theory result from Ref. [398], taking into account the associated theoretical uncertainty. The constraints from the CHARM experiment [399, 400] are extracted in Ref. [388], by first calculating the B meson production spectra and then convoluting it with the branching ratio for $B \to K^{(*)}a$. In order to be detected, the ALP produced must then decay to muons within the detector volume. Making use of the probability of this occurring, a limit on the ALP coupling to fermions g_{aff} was obtained for a given mass m_a, for further details about this procedure see Ref. [388].[7]

7.5.3 Cosmological constraints

If a substantial number of ALPs are produced in the very early universe, they can affect the successful predictions for big bang nucleosynthesis (see for example [372, 401, 402]). Whether or not this is a concern depends on the ALP abundance, mass, and lifetime.

If the ALPs and the DM are still relativistic during the time of BBN (occuring at a photon temperature between $T \sim 1$ MeV and $T \sim 10$ keV), these additional relativistic degrees of freedom will contribute to the energy density of the universe, cf. Eq. (7.13), which increases the speed at which it expands (see Eq. (7.12)). A faster expansion causes an earlier freeze-out of neutrons, manifesting itself in a change of the neutron-to-proton ratio and changing the abundances of Helium-4 and Deuterium. The number of additional relativistic degrees of freedom around the time of BBN, $\Delta N_{\rm eff}^{\rm BBN}$, is thus constrained by measurements of these light element abundances [403]. In most of the scenarios we envisage, however, the hidden sector particles are frozen out before the QCD phase transition, and the hidden sector temperature is generally lower than the photon temperature.

Heavier ALPs with masses larger than $\sim 1-10$ MeV can still be constrained if they are abundant and long-lived. Out-of-equilibrium decays of ALPs after neutrino decoupling will heat up the plasma and therefore decrease the effective number of neutrinos $\Delta N_{\rm eff}^{\rm CMB}$ observed in the CMB. Since neutrino decoupling takes place just before BBN, comparably strong constraints on the ALP lifetime can also be inferred from $\Delta N_{\rm eff}^{\rm BBN}$ [401].

The combined constraints from $\Delta N_{\rm eff}$ rule out axion-like particles with either masses below $\mathcal{O}(10$ MeV) or with lifetimes above $\mathcal{O}(0.01$ s) when assuming that the ALPs couple predominantly to photons [401, 402]. However, this is not the case in our model.

Finally, for heavier ALPs, electromagnetic showers produced in ALP decays during or after BBN can destroy the newly created nuclei and thus directly alter the light element abundances, see [28, 404] for recent studies. For ALP masses above the GeV scale, hadronic showers give rise to

[7] We are very grateful to the authors of Ref. [388] for providing us with the bounds they obtained in the $m_a - g_{aff}$ plane via private communication.

additional constraints, excluding abundant hadronically decaying particles with lifetimes down to about $\mathcal{O}(0.1\ \mathrm{s})$. This is because cascade hadrons can scatter off background protons, which once again increases the neutron-to-proton ratio [405]; see [406] for a recent numerical analysis.

7.5.4 Constraints on the DM

7.5.4.1 DM relic density constraint

Our model should explain the DM relic density measured today and thus has to satisfy the relic density constraint. The DM relic density today measured by Planck is [270]

$$\Omega_{\mathrm{DM}} h^2 = \frac{\rho_{\mathrm{DM}}}{\rho_{\mathrm{crit}}/h^2} = 0.120(1)\ . \tag{7.57}$$

Throughout the following sections, when calculating the DM relic densities for the different DM production mechanisms, we solve the Boltzmann equations until the comoving χ number density $Y_\chi = n_\chi/s$ stays constant. We then solve for the couplings which lead to the observed DM relic density today, where we assume that χ and its antiparticle $\bar{\chi}$ make up all the DM. Once the number changing interactions stop, the DM number density will become redshifted due to the ongoing expansion of the universe. Using entropy conservation in a comoving volume, we relate the number densities after freeze-out (of either the DM, see section 7.3.3 or the bath particles annihilating into DM, see section 7.3.2) and the measured density parameter of DM today by

$$\Omega_\chi h^2 \equiv \frac{\rho_\chi}{\rho_{\mathrm{crit}}/h^2} = \frac{m_\chi n_{\chi,1}}{\rho_{\mathrm{crit}}/h^2}\frac{s_0}{s_1} = \frac{m_\chi Y_\chi s_0}{\rho_{\mathrm{crit}}/h^2}\ , \tag{7.58}$$

with $\rho_{\mathrm{crit}}/h^2 = 1.053672(24) \times 10^{-5}$ GeV cm^{-3} and $s_0 = 2891.2$ cm^{-3} [2] and where we have used the subscript "0" for quantities today and the subscript "1" for quantities after freeze-out, when we stop the simulation. Note that for our model $\Omega_\chi = \Omega_{\mathrm{DM}}/2$ (see section 7.2).

7.5.4.2 (No) constraints from dark matter phenomenology

We have verified explicitly that, as one would expect, the couplings between the DM particle and the SM are too small to lead to an observable signal at current and near future direct or indirect detection experiments, in both the freeze-in and the DFO region. In particular, the present limits from AMS02 on the DM annihilation cross section into SM particles [407, 408] would need to improve by several orders of magnitude to become relevant for our model. The same is true for present limits on dark matter-nucleon interactions in effective field theory for direct detection experiments [299]. Below we will detail the calculations which led to this conclusion.

Indirect detection Constraints from indirect detection experiments provide stringent bounds on Dark Matter models, particularly from the detection of diffuse gamma rays from Dwarf spheroidal galaxies. While the annihilation cross section in our model is suppressed by the fact that the couplings we consider are small, we nonetheless preferred to verify whether the current limits result in any exclusion of the parameter space we consider. At tree level the annihilation

of the DM would proceed via the ALP mediated diagram to SM fermions $\chi\bar{\chi} \to f\bar{f}$:

$$\langle\sigma(\chi\bar{\chi} \to f\bar{f})v\rangle \sim \frac{g_{aff}^2 g_{a\chi\chi}^2 m_f^2 m_\chi^4 N_c \sqrt{1-\frac{m_f^2}{m_\chi^2}}}{2\pi\left(m_a^2 - 4m_\chi^2\right)^2} , \tag{7.59}$$

involving the factor $g_{aff}^2 g_{a\chi\chi}^2$. The current limits from AMS02 on the annihilation cross section can be found where the annihilation is 100% into leptons [407] or hadrons [408], where limits are of the order of $\langle\sigma v\rangle \lesssim 10^{-28} - 10^{-26}$ cm^3 s^{-1}, and from these results we can estimate the limit on our model. We therefore can confirm that for the couplings in the freeze-in and the DFO regime annihilation cross sections lie $\mathcal{O}(10)$ orders of magnitude below the current constraints.

Direct detection Here we adopt the constraints from Ref. [299] on our model. The authors work in a non-relativistic effective field theory basis to analyse direct detection experiments for various possible DM-nucleon interactions, making the approximation of a heavy and a light mediator as compared to the typical momentum transfer $|\vec{q}|$ between the nucleon and the DM. In this approximation they calculate the scattering cross section as a function of a constant reference cross section at some arbitrary $|q_{\text{ref}}|$, where $|q_{\text{ref}}| = 100$ MeV was chosen. They subsequently provide bounds on this reference cross-section. For our Lagrangian (7.3), the relevant interaction is a pseudoscalar-pseudoscalar interaction with a local interaction term $\bar{\chi}\gamma_5\chi\bar{N}\gamma_5 N$ leading to a scattering amplitude

$$\mathcal{A}_{\text{PS-PS}} \propto \frac{g_\chi^{\text{PS}} f_n^{\text{PS}}}{\left(|\vec{q}|^2 + m_a^2\right)} \bar{\chi}\gamma^5\chi\bar{N}\gamma^5 N ,$$

χ and N are the DM and nucleon spinors, respectively. The differential cross section for the light mediator case ($m_a \ll |q_{\text{ref}}|$) then reads

$$\frac{d\sigma_{T,\text{ light}}^{\text{PS-PS}}}{dE_R} = \sigma_{\text{ref}}^{\text{PS-PS}} \frac{m_T}{2\mu_T^2 v^2} \frac{\mu_T^2}{\mu_N^2} \frac{4}{3} \frac{J_T+1}{J_T} \left[\left\langle S_{p,T}\right\rangle + \frac{f_n^{\text{S}}}{f_p^{\text{S}}} \left\langle S_{n,T}\right\rangle\right]^2 F_{\text{SD},T}^2 , \tag{7.60}$$

whereas the differential cross section for heavy mediators $m_a \gg |q_{\text{ref}}|$ can be obtained from the previous one via

$$\frac{d\sigma_{T,\text{ heavy}}^{\text{PS-PS}}}{dE_R} = \frac{|\vec{q}|^4}{\left(|\vec{q}|^2 + m_a^2\right)^2} \left(\frac{d\sigma_{T,\text{ light}}^{\text{PS-PS}}}{dE_R}\right) . \tag{7.61}$$

The authors also estimate the sensitivity of various future potential experimental configurations. Here we compare the constraints from figs. (3) and (4) from Ref. [299] on a PS-PS interaction between the DM and the nucleon for the values of our couplings which give the correct relic density in the freeze-in and the DFO regions (cf. secs. 7.6.3 and 7.6.4), adopting the expression for the ALP-nucleon coupling from section 7.5.1.2, $C_n = C_p = 0.43\ C_f$. We then find

$$\sigma_{\text{ref}} = \begin{cases} \frac{\mu_N^2}{16\pi}\left(0.43 g_{a\chi\chi} g_{aff}\right)^2 & \text{for } m_a \ll |q_{\text{ref}}| \\ \frac{\mu_N^2}{16\pi}\left(\frac{0.43 g_{a\chi\chi} g_{aff} q_{\text{ref}}^2}{m_a^2}\right)^2 & \text{for } m_a \gg |q_{\text{ref}}| \end{cases} . \tag{7.62}$$

Again, the couplings in the freeze-in and in the DFO region are too small to lead to an observable signal at current and near future experiments.

Self interactions The standard ΛCDM scenario of cold collisionless DM has proven to be extremely successful at large distances. However, predictions from DM-only N-body simulations have revealed certain discrepancies in structure formation at small scales when compared to observations [331]. DM self-interactions were proposed to reconcile predictions from simulations with observations at galactic scales [331, 409, 410]. Since our DM candidate has self-interactions via the axion couplings, it is interesting to ask whether it could resolve any of the small-scale structure problems plaguing the standard ΛCDM scenario. Along with possible hints for self-interacting DM come a variety of constraints from merging clusters (among them the Bullet cluster) or halo shapes, see Ref. [331] for a summary. Most of the possible hints suggest a self-interaction cross section per DM mass of the order of $\sigma/m_\chi \sim 0.1-1\ \mathrm{cm}^2/\mathrm{g}$. Interestingly, existing constraints turn out to be weaker or roughly of the same order as the cross sections needed to alleviate the tensions. Since the hidden sector particles are in equilibrium in the DFO region, the interaction strength might become large enough to lead to an observable signal in our model.

To quantify the self interaction strength in our model, we consider the momentum transfer cross section as proposed by [411] [8]

$$\sigma_\mathrm{T} = \int \mathrm{d}\Omega (1 - |\cos\theta|) \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \,. \tag{7.63}$$

We find that, in the weakly coupled regime and at small velocities where the Born approximation is valid, the cross sections are too small to lead to sizeable interactions. In the non-relativistic strongly coupled regime for very light mediators, i.e. for

$$\frac{g_{a\chi\chi}^2 m_\chi^3}{4\pi m_a} > 1 \,,$$

the cross section has been claimed to be Sommerfeld-enhanced [412–414]: If two DM particles move slowly enough they can exchange a light mediator multiple times. This leads to a long range self interaction. Sommerfeld-enhanced processes are therefore non-perturbative and have to be resummed. However, as we will see in section 7.6.4, the largest possible DM-ALP coupling in the DFO region is of the order of $g_{a\chi\chi} \sim 2.5 \cdot 10^{-2}\ \mathrm{GeV}^{-1}$, such that if we keep the ratio $m_\chi/m_a = 10$ fixed, this condition is never met. Additionally, as shown recently in Ref. [415] (see also Ref. [416]), the potential arising from a pseudoscalar exchange does not lead to any significant Sommerfeld enhancement at small velocities. Hence, the DM particles in our model do not self-interact sufficiently to explain the discrepancies in structure formation observed at small scales.

7.6 Results

In this section we will first present the results of the relevant constraints on our model, providing bounds on the ALP-fermion coupling, g_{aff}, as a function of the ALP mass, m_a. As introduced in section 7.5, we focus on constraints from the electron beam dump experiment E137 at SLAC, from the supernova SN1987A, from horizontal branch stars and flavour constraints from heavy meson

[8] Alternatively, the viscosity cross-section $\sigma_\mathrm{V} = \int \mathrm{d}\Omega \sin^2\theta \mathrm{d}\sigma/\mathrm{d}\Omega$ weighting forward and backward scattering equally is sometimes used.

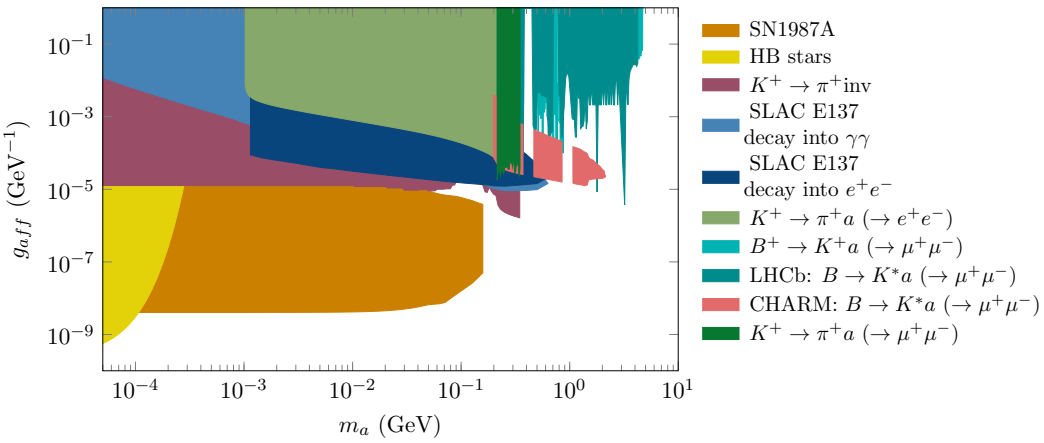


FIGURE 7.3: Full constraint plot on the ALP-fermion coupling g_{aff} as a function of the ALP mass m_a for an ALP which only couples to SM fermions at tree level. We show collider and astrophysical constraints discussed in section 7.5.

decays. Armed with the tools discussed in section 7.2 we subsequently study the generation of DM in the early universe, calculating the dark matter relic density generated for a large range of ALP-SM and ALP-DM couplings for a specific set of masses. This requires solving the Boltzmann equations in various different production regimes. We focus on the regions of parameter space needed to reproduce the observed DM abundance in the freeze-in and in the DFO regions, i.e. in DM genesis scenarios where the DM and the SM remain out of equilibrium, for a large range of ALP and DM masses. Finally, we compare these regions to the constraints on the ALP to study the phenomenological implications of our model. Throughout this section, when calculating the DM relic density, we fix the ratio $m_\chi/m_a = 10$.

7.6.1 Collider and astrophysical constraints

Fig. 7.3 summarises the experimental bounds on ALPs discussed in section 7.5 we obtained for the model considered in this work, see Eq. (7.3), where the ALP only couples to SM fermions at tree level.

We have considered a number of collider constraints coming from SLAC E137, NA62, NA48, LHCb and CHARM. For the electron-beam dump at SLAC we solve $\mathcal{N}_{\text{theory}}(m_a, g_{aff}) = \mathcal{N}_s^{\text{up}} = 2.996$, with $\mathcal{N}_{\text{theory}}$ given by Eq. (7.52), for different values of m_a. The ALP can be produced by the Primakoff process, by bremsstrahlung or by electron-positron annihilation. It can decay into two photons or, for $m_a > 2m_e$, into an electron-positron pair. For even larger ALP masses, various other decay channels open up, cf. Fig. G.1. The original analysis [377] considers separately 1) Primakoff production with decay into either two photons or into e^+e^-, 2) bremsstrahlung production with decay into e^+e^- and 3) non-resonant positron annihilation with subsequent decay into e^+e^-. In the present work, we provide a complete analysis without making assumptions on dominant production and/or decay channels. In order to provide a combined bound, we sum over all possible production processes, since the only observable final-state particles are the decay products of the ALP (additional final-state particles from the production process are absorbed in the beam dump absorber). In Fig. 7.3 we depict in light and dark blue the constraints from a decay of the ALP into two photons and an electron-positron pair, respectively. As explained in

more detail in section 7.5, a too large ALP-fermion coupling would lead to an overproduction of ALPs, whereas ALPs with too small couplings are long-lived and would escape the detector. The upper bound thus stems from the fact that the detection probability of ALPs decaying too early becomes exponentially suppressed, leading to the typical nose-like shape in both SLAC beam dump constraints. Comparing our results to Fig. 19 of Ref. [57], we find that the combination of several production processes excludes a slightly larger region of parameter space. This stems from the dominant Primakoff production process for larger ALP masses, cf. Fig. 7.2. Since the loop interaction is kinematically suppressed, the bounds on the ALP-fermion coupling g_{aff} for an ALP decaying into two photons are shifted towards larger couplings compared to the constraints on a tree-level ALP-photon coupling, see for instance Fig. 2 in Ref. [369].

Flavour constraints also have an important impact on the parameter space considered. The calculation of the bounds from rare B and K decays shown in Fig. 7.3 was discussed in section 7.5. Rare B decays provide access to the parameter space for masses up to $m_a \sim m_B - m_K^{(*)}$. The LHCb bounds from $B^+ \to K^+ X(\to \mu^+\mu^-)$ are shown in cyan, and from $B^0 \to K^{*0} X(\to \mu^+\mu^-)$ in dark cyan, for a long-lived scalar X, as indicated. We further show the CHARM bound coming from the $B^0 \to K^{(*)} X(\to \mu^+\mu^-)$ analysis in pink. Together, these form a powerful probe of g_{aff}, excluding almost completely couplings down to $\mathcal{O}(10^{-5})$ GeV^{-1}, as pointed out in Ref. [388]. Coming to rare K decays, we show the NA48/2 limits on $K^\pm \to \pi^\pm e^+ e^-$ in light green and from $K^\pm \to \pi^\pm X(\to \mu^+\mu^-)$ for long-lived X in dark green as well as the NA62 limit from $K^+ \to \pi^+ \nu\bar{\nu}$ in purple. These three bounds are highly powerful and complementary probes of the parameter space, providing complete coverage down to $g_{aff} \sim 10^{-5}$ for $m_a \lesssim m_K - m_\pi$. Note that the uneven nature of certain constraints from B and K meson decays is a direct artefact of the experimental limits, and the gaps in the constraints are due to the fact that in regions where m_a corresponds to the mass of certain mesons, a reliable limit cannot be obtained.

The astrophysical constraints from SN1987A and from HB stars in globular clusters are depicted in orange and yellow respectively. The most stringent constraint relevant for the freeze-in region is from SN1987A. This restricts the coupling g_{aff} to be less than $\sim 4 \cdot 10^{-9}$ GeV^{-1} for ALP masses up to $\mathcal{O}(100)$ MeV. The only constraint which probes lower couplings is that from HB stars, but this is only relevant for very small values of m_a. As described in section 7.5.1.2, the bound is obtained, following Ref. [24, 373], by demanding that the luminosity emitted due to the ALP is less than that due to the neutrinos, $L_\nu = 3 \cdot 10^{52}$ ergs (i.e. the "Raffelt condition"). The upper line in the constrained region comes from the fact that if the coupling is too large the ALPs would not escape from the Supernova, and the lower line comes from the fact that if the coupling is too low, fewer ALPs would be produced. Note that our bound is obtained by following the calculation in Ref. [24, 373], which involves several updated nuclear physics calculations, as well as including the energy dependence of the optical depth such that the energy loss at large couplings is correctly accounted for, and the bound therefore differs by an order of magnitude from previous work. We have verified that multiplying or dividing this luminosity by a factor two would result in a change in the limit on the coupling of less than 50%.

The constraint coming from HB stars excludes a wide region of parameter space for small ALP masses (1 eV $< m_a <$ 10 keV) and small couplings $g_{aff} \lesssim 3 \times 10^{-9}$ GeV^{-1} (yellow), by requiring $\langle \epsilon_a(m_a, g_{aff}) \rangle \lesssim 100$ erg g^{-1} s^{-1}. Not surprisingly, since the dominant production mechanism in HB stars for ALPs with a coupling to fermions is the Compton process and since the Primakoff process in our model is loop-suppressed, we essentially obtain the same bound as in Fig. 4 in

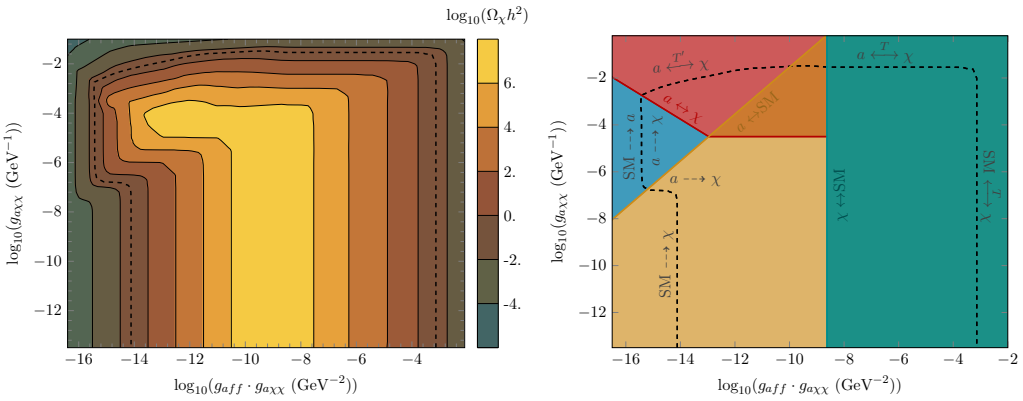


FIGURE 7.4: Hidden sector coupling $g_{a\chi\chi}$ as a function of the product of the hidden sector and the ALP-fermion coupling, $g_{a\chi\chi} \cdot g_{aff}$, for m_χ = 10 GeV and m_a = 1 GeV. The dashed line corresponds to the combination of couplings which give the correct DM relic density $\Omega_{\rm DM}h^2 = 0.12$ as measured by the Planck telescope. *Left:* Contour plot of the DM relic density as a function of $g_{a\chi\chi} \cdot g_{aff}$ and g_{aff}, giving the so-called "Mesa" phase diagram: For our choice of masses six different DM genesis scenarios are possible, namely freeze-in from the SM, freeze-in from ALPs, sequential freeze-in, decoupled freeze-out (DFO), freeze-out from the hidden sector and freeze-out from the SM. *Right:* Different phases of the diagram, corresponding to equilibrium relations between the three sectors. The equilibrium relations essentially determine the production mechanism.

Ref. [25], which the authors calculated considering the Compton process only. It was pointed out in Ref. [25] that radiation is not the dominant mechanism for heat transfer inside HB stars and that indeed convection is more efficient. Let us stress that this analysis is a crude estimate of the bound for our model and therefore most likely subject to large uncertainties. Making dedicated simulations is beyond the scope of this paper, in particular since we are mainly interested in ALPs with masses ranging in the MeV-GeV range for which the number densities are Boltzmann-suppressed in HB stars. For more recent analyses of constraints on the ALP-photon coupling from HB stars see for instance Refs. [417–419]. A global analysis from HB stars, red giants and white dwarfs suggests a non-zero ALP-SM coupling (*HB-hint*) [420]. Other strong astrophysical constraints on an ALP-electron coupling come from the delay of Helium ignition in red giants, see for example [421, 422].

It is worthwhile mentioning that we have not made any assumptions about the branching ratios or preferred decay modes for the calculation of the bounds on our model. Since the ALP only couples to fermions at tree level, the branching ratios for the various production and decay modes follow naturally.

7.6.2 Phase diagram

As described in detail in section 7.2, the interplay between the three different sectors in our model, namely the SM particles, the ALP and the DM, gives rise to various DM production mechanisms. Solving the Boltzmann equations in a large range of hidden sector and connector couplings and plotting the final abundance as a function of these couplings gives what in [21] (and later in [22] for the massive mediator case) was dubbed a "Mesa" phase diagram. For our model, this phase diagram is shown in the left panel of Fig. 7.4 for an ALP with mass m_a = 1 GeV and the DM with mass m_χ = 10 GeV. The dotted line indicates the combination of couplings

$g_{a\chi\chi}$ and g_{aff} such that the final abundance corresponds to the observed DM density.[9] To aid the understanding of what follows, we also plot in the right panel of Fig. 7.4 the different phases of the diagram, corresponding to equilibrium relations between the three sectors, as these essentially determine the production mechanism. Let us have a look at the various regimes in more detail:[10]

SM -→ χ
$(a \overset{T}{\longleftrightarrow} SM)$

Freeze-in from SM. Starting from the bottom-left (small $g_{a\chi\chi}$ and $g_{a\chi\chi} \cdot g_{aff}$) of the phase diagram in Fig. 7.4, the dominant process is $f\bar{f} \to \chi\bar{\chi}$ and, for reheating temperatures above $T_{RH} \gtrsim 200$ GeV, ultraviolet-dominated $2 \to 3$ interactions. DM is produced from scattering of SM fermions in the thermal bath without ever reaching equilibrium. From Eq. (7.18) we realise that the final relic density is proportional to $(g_{a\chi\chi} \cdot g_{aff})^2$ (cf. Tab. 7.1). The size of each term individually is a free parameter, however, g_{aff} is large enough to ensure a-SM equilibrium. Hence, we expect a vertical line in the $(g_{a\chi\chi} \cdot g_{aff}, g_{a\chi\chi})$ phase diagram.

a -→ χ
$(a \overset{T}{\longleftrightarrow} SM)$

Freeze-in from ALPs. Going upwards in the phase diagram, i.e. increasing the value of the hidden sector coupling, while still keeping g_{aff} large enough for efficient ALP-SM scattering, collisions of thermal ALPs become more likely than collisions of thermal SM particles. Following the same arguments as above, Y_χ scales like $g^4_{a\chi\chi}$ and is therefore independent of g_{aff}. Hence, we expect a horizontal line in the $(g_{a\chi\chi} \cdot g_{aff}, g_{a\chi\chi})$ phase diagram, appearing as a plateau in Fig. 7.4.

SM -→ a
a -→ χ
$(a \cdots \chi)$

Sequential freeze-in. The relic abundance is obtained by a chain of sequential reactions $\sum_{i,j,k} ij \to ak$ (with $\{i, j, k\}$ SM particles) followed by $aa \to \chi\bar{\chi}$. Hence, $Y_\chi \propto \langle \sigma_{aa\to\chi\bar{\chi}} v \rangle n_a^2$, i.e. $Y_\chi \propto (g_{a\chi\chi})^4$. Similarly, $n_a \propto \sum_{i,j,k} \langle \sigma_{ia\to jk} v \rangle n_a^{\rm eq}(T) n_i^{\rm eq}(T)$, i.e. $n_a \propto (g_{aff})^2$. Finally, $Y_\chi \propto (g_{aff} \cdot g_{a\chi\chi})^4$ and we expect a vertical line, but shifted to the left because the ALPs and the DM are now both out of equilibrium. Note that our calculation in this region is subject to uncertainties (see section 7.3.2.3). On performing a full analysis as in Ref. [20] we would expect the line in Fig. 7.4 to move to the right since as our approximation may overestimate the production of ALPs up to a factor of two. Consequently, a larger ALP-fermion coupling would be needed in order to produce the observed DM relic density.

$a \overset{T'}{\longleftrightarrow} \chi$
(SM -→ a)
(SM -→ χ)

DFO. As explained in section 7.3.3, the ALP-SM and the SM-DM processes enter the final relic density only weakly via T' (see also section 7.6.4 for an extended discussion). The dependence on g_{aff} is mild, hence the almost horizontal line in the phase diagram. For increasing g_{aff}, more hidden sector particles are produced and the strength among the hidden sector particles has to slightly increase to ensure a later freeze-out of $\chi-$ particles.

[9] Note that we plot the abundance as a function of $g_{a\chi\chi}$ and the product of $g_{a\chi\chi} \cdot g_{aff}$ as these are the proportionality factors for the competing DM production processes, cf. Tab. 7.1.

[10] This is extensively discussed in Ref. [22], and we will therefore only briefly comment on the different shapes of the regions.

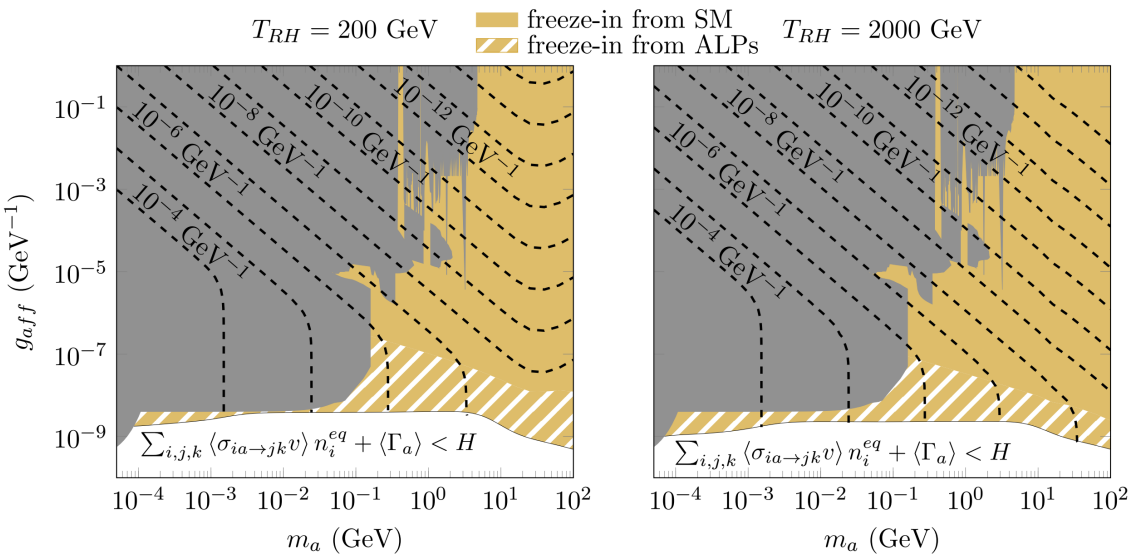


FIGURE 7.5: ALP-fermion coupling g_{aff} as a function of m_a for lines of constant hidden-sector coupling $g_{a\chi\chi}$ (plotted in dashed black) which reproduce the observed DM relic density via freeze-in from SM and freeze-in from ALPs as indicated for reheating temperatures T_{RH} = 200 GeV (left) and T_{RH} = 2000 GeV (right). We fixed the ratio m_χ/m_a = 10. The lower line indicates the value of g_{aff} for which the ALPs and the SM reach equilbrium. In grey, we have included the relevant constraints on our ALP model on the connector coupling g_{aff} in this parameter region (cf. Fig. 7.3).

$a \overset{T}{\longleftrightarrow} \chi$
$(a \overset{T}{\longleftrightarrow} SM)$

Thermal freeze-out from ALPs. Going further to the right in the phase diagram, i.e. further increasing g_{aff}, the ALPs and the SM and then also the SM and DM will equilibrate and all three sectors will share a common temperature. In the right half of the phase diagram DM will therefore be produced by freeze-out. However, g_{aff} is still small enough for $aa \leftrightarrow \chi\bar{\chi}$ to be the more efficient process. The behaviour is as in DFO except that the weak dependence on g_{aff} disappears completely, since the hidden and the visible sector share the same temperature.

SM $\overset{T}{\longleftrightarrow} \chi$
$(a \overset{T}{\longleftrightarrow} SM)$

Thermal freeze-out from SM. This is the usual thermal freeze-out from $2 \to 2$ DM-SM scattering which in our model is set by $f\bar{f} \leftrightarrow \chi\bar{\chi}$ scattering and hence by the product $(g_{a\chi\chi} \cdot g_{aff})^2$. As in freeze-in from SM we expect a vertical line, except that the final relic abundance depends now inversely on $(g_{a\chi\chi} \cdot g_{aff})^2$.

7.6.3 Freeze-in region

Having discussed collider and astrophysical constraints on the mediator in our model in the previous section, we now investigate the phenomenological implications of a feebly interacting DM particle. The freeze-in region is defined by the couplings of the DM to both the ALPs and the visible sector being so tiny that the DM particles never reach thermal equilibrium. Their initial abundance being zero or negligible, the DM particles are gradually produced by out-of-equilibrium scattering of particles in the thermal bath. The ALPs, on the other hand, are in thermal equilibrium with the SM in the conventional freeze-in scenario. (We do not consider

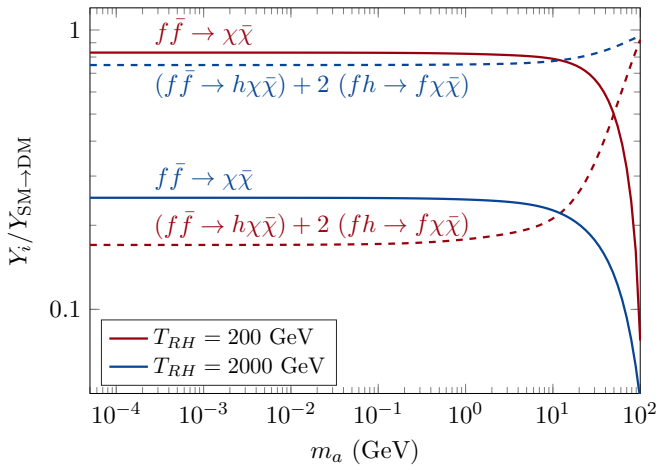


FIGURE 7.6: Relative contributions of infrared-dominated interaction $f\bar{f} \to \chi\bar{\chi}$ (solid) and ultraviolet-dominated $2 \to 3$ interactions $(f\bar{f} \to h\chi\bar{\chi}) + 2(fh \to f\chi\bar{\chi})$ (dashed) to the final χ-abundance as a function of the ALP mass in the freeze-in from SM regime for reheating temperatures $T_{RH} = 200$ GeV (red) and $T_{RH} = 2000$ GeV (blue). We fixed the ratio $m_\chi/m_a = 10$.

DM production via sequential freeze-in as its description is subject to theoretical uncertainties, as described in section 7.3.2.3.) Hence the reaction rates satisfy

$$\sum_{i,j,k} \left\langle \sigma_{ia\to jk} v \right\rangle n_i^{\mathrm{eq}}(m_\chi) + \langle \Gamma_a \rangle > H(m_\chi) \qquad \text{(equilibrium of ALP-SM)}\,,$$

$$\left.\begin{aligned} &\sum_{i,j,k} \left\langle \sigma_{\chi\bar{\chi}[k]\to ij} v^{[2]} \right\rangle [n_k^{\mathrm{eq}}(m_\chi)]\; n_\chi^{\mathrm{eq}}(m_\chi) \\ &\left\langle \sigma_{\chi\bar{\chi}\to aa} v \right\rangle n_\chi^{\mathrm{eq}}(m_\chi) \end{aligned}\right\} < H(m_\chi) \qquad \text{(DM out of equilibrium)}\,, \tag{7.64}$$

where, in the second line, the symbols in square brackets pertain to $2 \to 3$ interactions. Depending on the relative size of the couplings g_{aff} and $g_{a\chi\chi}$, DM generation by either $ij \to \chi\bar{\chi}[k]$ (where i, j, k are SM particles) or $aa \to \chi\bar{\chi}$ scattering will be more efficient. This results in two distinct regimes which we label *freeze-in from SM particles* (cf. section 7.3.2.1) and *freeze-in from ALPs* (cf. section 7.3.2.2). In these regimes, the DM relic abundance is given by Eqs. (7.18) and (7.20), respectively. Since the relic density is proportional to the couplings entering the matrix element squared of the dominant process, it suffices to factorize the couplings and solve for those couplings satisfying the relic density constraint in Eq. (7.57).

As discussed in section 7.3.2.1, ultraviolet-dominated contributions to the DM relic density from a-mediated $2 \to 3$ scattering processes $f\bar{f} \to h\chi\bar{\chi}$, $fh \to f\chi\bar{\chi}$ and $\bar{f}h \to \bar{f}\chi\bar{\chi}$ become important for reheating temperatures above a few hundred GeV. These contributions introduce a dependence on the reheating temperature since the final relic abundance scales with T_{RH}, see Eq. (7.19). Here we consider two representative scenarios: For $T_{RH} = 200$ GeV, UV-dominated processes are practically negligible; the DM abundance will therefore be independent of the exact value of T_{RH} if it is chosen even lower (but above the temperature of top quark freeze-out). By contrast, for $T_{RH} \sim 2000$ GeV, the result is fully dominated by the UV-dominated processes. This behaviour can be observed in Fig. 7.6 where we compare the relative contribution of the IR- and UV-dominated processes to the final DM abundance as a function of the ALP mass in the freeze-in from SM scenario, i.e. where $aa \to \chi\bar{\chi}$ scattering is negligible. We keep the ratio $m_\chi/m_a = 10$ fixed. Indeed, for $T_{RH} = 200$ GeV, the DM abundance is set by $f\bar{f} \to \chi\bar{\chi}$ scattering, except for large ALP masses where the $2 \to 3$ interactions always dominate. By contrast, for

reheating temperatures of the order of a few TeV, it is the ultraviolet contributions that set the final DM abundance.

In Fig. 7.5 we depict the results for these two scenarios in a large range of ALP masses, together with the constraints derived in the previous section. We plot the value for the ALP-SM coupling g_{aff} as a function of the ALP mass m_a for lines of constant hidden sector coupling $g_{a\chi\chi}$ (dashed black lines). Inside the yellow region DM production from SM fermion scattering is more efficient (freeze-in from SM particles), and inside the yellow hatched region production from ALP scattering (freeze-in from ALPs). These freeze-in regimes span a vast region of parameter space in the (m_a, g_{aff})-plane. We cut the plot at the upper theoretical bound $C_f \lesssim 4\pi$ which results in $g_{aff} = C_f/f_a \lesssim 1$ GeV^{-1} since $f_a \gg v$ with v the electroweak scale, cf. App. F. Below the lower boundary the χ-particles are produced via sequential freeze-in or decoupled freeze-out (DFO) because the ALPs and the visible sector cease to share the same temperature.

The point where the ALPs equilibrate with the SM plasma is affected by $\mathcal{O}(1)$ finite-temperature corrections and by UV-dominated processes entering the equilibrium conditions Eq. (7.64), notably $f\bar{f} \to ah$. These can enhance ALP production at earlier times. As a consequence, the simple conditions in Eq. (7.64) are only approximately true and in fact, equilibration can be attained for some $T \gtrsim m_\chi$. To obtain the lower boundary in Fig. 7.5, we solve the Boltzmann equation for the ALP number density numerically and determine the couplings for which $n_a(T) = n_a^{\text{eq}}(T)$ for some $T_{RH} \gtrsim T \gtrsim m_\chi$.

In general, the ALP-SM interaction strength g_{aff} must be comparably large if thermal equilibrium is to be reached, and is therefore likely in reach of upcoming experiments. For $g_{aff} \gtrsim 10^{-6} - 10^{-5}$ GeV^{-1} and m_a below a few GeV, the parameter space is ruled out by constraints from rare B and K decays and the electron beam dump experiment at SLAC. For masses below ~ 0.2 GeV, the remaining parameter space is covered by the SN constraint, see also Fig. 7.3. However, we remark that the relic density constraint inhibits large hidden sector couplings: $g_{a\chi\chi}$ has to be as small as $\sim 10^{-15}$ GeV^{-1} for the largest connector couplings and for large ALP masses as in this parameter region DM production is controlled by $f\bar{f} \to \chi\bar{\chi} + (2 \to 3$ terms) and therefore by the product of couplings $g_{aff} \cdot g_{a\chi\chi}$ (see Tab. 7.1).

When the DM is dominantly produced by freeze-in from ALPs, we expect the contours in constant $g_{a\chi\chi}$ in the (m_a, g_{aff})-plane to be vertical because the relic density is independent of the connector coupling and proportional to $(g_{a\chi\chi} m_\chi)^4$. In the freeze-in from SM scenario the relic density is both proportional to $m_\chi^2 \propto m_a^2$ and to g_{aff}^2. For $T_{RH} = 200$ GeV and DM masses above the mass of the top quark, this dependence is weakened by the Boltzmann suppression of the particles predominantly producing the DM via $t\bar{t} \to \chi\bar{\chi}$, leading to an upwards bending of lines of constant hidden sector coupling. Here, interactions have to become stronger; between reheating and freeze-out of the top there is barely enough time to produce the heavy DM particle. For the ultraviolet-dominated case, the lines of constant $g_{a\chi\chi}$ get shifted towards smaller values of g_{aff} since more DM is produced at earlier times. Not surprisingly, the freeze-in from the mediator scenario is independent of the two reheating temperatures since $aa \to \chi\bar{\chi}$ is infrared dominated and the ALPs belong to the SM bath. We would like to add that, depending on the temperature, finite-temperature corrections can influence the DM production from fermion scattering by up to $\mathcal{O}(100\%)$, resulting in a change in the final relic density of $\sim 10\%$ for $T_{RH} = 200$ GeV and $\sim 5\%$ for $T_{RH} = 2000$ GeV.

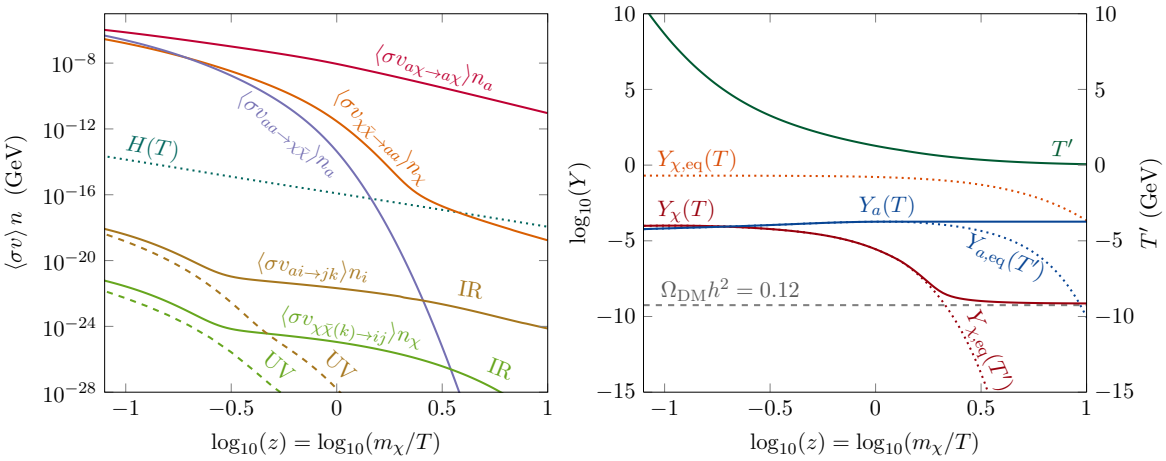


FIGURE 7.7: Illustrative example of the evolution of the number densities and the reaction rates in the DFO scenario for $m_\chi = 10$ GeV and $m_a = 1$ GeV and for a set of couplings which give the measured dark matter relic abundance: $g_{aff} = 9.44 \cdot 10^{-12}$ GeV^{-1} and $g_{a\chi\chi} = 1.06 \cdot 10^{-2}$ GeV^{-1}. *Left:* Evolution of the different terms competing on the right hand side of the Boltzmann equations for n_χ and n_a as a function of $\log_{10}(z) = \log_{10}(m_\chi/T)$. We also plot the evolution of the reaction rate for elastic scattering among the hidden sector particles, $a\chi \to a\chi$. *Right:* The evolution of the comoving number density of χ (red) and a (blue) as a function of $\log_{10}(z) = \log_{10}(m_\chi/T)$. The equilibrium distributions $Y_{\chi,\mathrm{eq}}(T')$ (dotted red) and $Y_{a,\mathrm{eq}}(T')$ (dotted blue) and $Y_{\chi,\mathrm{eq}}(T)$ (dotted orange) are also shown for reference. We plot along the evolution of the hidden sector temperature T' (green).

Note that, in principle, ALPs with masses $m_a \lesssim$ MeV which are still relativistic during big bang nucleosynthesis are severely constrained by cosmology, see section 7.5. We do not include these bounds in our analysis since the relevant parameter space is already excluded by collider experiments or astrophysics. For larger masses, lifetimes above $\tau \sim 0.01$ s can be excluded [401]; yet in the freeze-in regime the ALP-fermion coupling strength is large such that lifetimes are < 0.01 s.

In summary, we find that for feebly interacting dark matter, which has tiny couplings to the mediator, the correct relic density can be obtained in a large region of unexcluded parameter space, likely within the reach of future experiments.

7.6.4 DFO region and cosmological constraints

As the hidden sector coupling $g_{a\chi\chi}$ increases, the interactions among the ALPs and the DM become frequent enough to bring them into thermal equilibrium, however at a temperature T' which is distinct from that of the photons, T. These now constitute a dark sector which is thermally decoupled from the visible sector. Energy is gradually transferred to the hidden sector by rare scatterings of SM particles into both the ALPs and the DM. The DM relic density is set by $aa \leftrightarrow \chi\bar\chi$ interactions and the mechanism resembles ordinary freeze-out but occurs at a different temperature, i.e. by *decoupled freeze-out (DFO)*. To obtain the relic density in the DFO regime we proceed as outlined in section 7.3.3 and section 7.5.4.1. The calculation is more involved than for freeze-in, since now a stiff system of coupled differential equations (7.36), (7.40) and (7.41) has to be solved, which we have accomplished using the stiff solver package dvode [423]. We track the evolution of $Y_\chi \equiv n_\chi/s$ until the DM particles freeze out and Y stays constant (more precisely, we require the relative change in Y_χ to be smaller than $\varepsilon = 5 \cdot 10^{-4}$). As in the case of freeze-in, here too, UV contributions to SM-DM and SM-ALP interactions can introduce

a dependence on the reheating temperature. However, these only enter indirectly via the energy transfer from the SM. We have checked this explicitly and find that for reheating temperatures below $\mathcal{O}$(TeV) they do not significantly influence the final DM relic abundance. To be more specific, we find that a variation of T_{RH} = 200 GeV and T_{RH} = 2000 GeV leads to a change of ~ 22% in the final DM relic abundance and a change ~ 7% in the required value of $g_{a\chi\chi}$. Hence, the DFO scenario retains very little sensitivity to the ultraviolet for reheating temperatures below a few TeV, and in the following we can neglect these contributions. However, we do add finite-temperature corrections to both ALP and DM production from the SM, which result in $\mathcal{O}(1)$ corrections, mainly originating from thermal mass effects.

In the left panel of Fig. 7.7 we present an illustrative example of the evolution of the various reaction rates at play in the DFO regime, together with the Hubble rate, for m_χ = 10 GeV and m_a = 1 GeV and for a set of couplings which give the measured dark matter relic abundance ($g_{aff} = 9.44 \cdot 10^{-12}$ GeV^{-1} and $g_{a\chi\chi} = 1.06 \cdot 10^{-2}$ GeV^{-1}). For reactions faster than the Hubble rate, i.e. lines above the turquoise dotted Hubble line, the particles involved achieve equilibrium. Once the reaction rates drop below the Hubble rate, the particles become so dilute that they no longer collide sufficiently often and the process shuts off. We also distinguish between the infrared- (solid) and ultraviolet- (dashed) dominated SM-ALP and SM-DM interactions. A few observations can be made: Firstly, $aa \leftrightarrow \chi\bar{\chi}$ interactions are the dominant DM number changing interactions throughout the evolution. The mechanism which sets the relic abundance is therefore conceptually very similar to the simpler freeze-out scenario, however governed by T' rather than T. Secondly, neither the SM-ALP nor the SM-DM reaction rates ever exceed the Hubble rate, and the hidden sector is indeed decoupled. Nevertheless, the SM slowly transfers energy to the hidden sector both via SM-ALP and SM-DM interactions, with the dominant contribution coming from SM-ALP interactions. Lastly, elastic scattering $a\chi(\bar{\chi}) \leftrightarrow a\chi(\bar{\chi})$ stays active until after freeze-out of the χ-particles, as assumed in section 7.3.3. For the same parameters, in the right panel of Fig. 7.7 the evolution of the comoving number densities of the DM and the ALP are shown, as well as the hidden sector temperature T'. We also plot the equilibrium distributions $Y_{\chi,\mathrm{eq}}(T')$ and $Y_{a,\mathrm{eq}}(T')$ as dotted lines, and, for comparison, $Y_{\chi,\mathrm{eq}}(T)$. As in the simpler thermal freeze-out scenario, Y_a and Y_χ follow the equilibrium distributions, $Y_{a,\mathrm{eq}}(T')$ and $Y_{\chi,\mathrm{eq}}(T')$, respectively. For rapid interactions the right hand side of the Boltzmann equation vanishes and Y_χ stays constant.[11] The ALPs and the DM separate from the equilibrium distributions at the same time, around $T' \simeq m_\chi/10$, since no other particles are around to maintain frequent number changing interactions. Since $T' < T$ the inverse SM freeze-out temperature can be as small as $z \sim 2-3$.

In Fig. 7.8 we show the region of parameter space where the correct relic density is obtained for a large range of masses and couplings, again keeping the ratio m_χ/m_a = 10 fixed, together with the constraints derived in the previous section. We have checked that the observations made in the previous paragraph hold for the whole set of parameters depicted in Fig. 7.8. The region in the (m_a, g_{aff})-plane is shaped by a number of conditions:

- **Relic density** For connector couplings, g_{aff}, which are too small, the hidden sector does not become sufficiently populated. Although $g_{a\chi\chi}$ might be large enough to establish

[11] Due to the relation between T and T', affected by the rate of energy transfer to the hidden sector, Y_χ and Y_a ostensibly increase in the plot; this is because the plot shows the evolution of the comoving number densities as a function of the SM temperature T.

equilibrium between the hidden sector particles, $n_\chi^{\rm eq}(T')$ can never reach the amount of DM density observed today. This is indicated by the lower boundary.

- $\boldsymbol{a} \leftrightarrow$ **SM** On the other hand, if the connector coupling g_{aff} is too large, the interactions between the hidden sector and the SM become strong enough to establish thermal equilibrium. Depending on the hidden sector coupling, the DM is then either produced by freeze-in (see section 7.6.3) or thermal freeze-out. We remark that the numerical solution close to the transition between the freeze-in and the DFO regime is challenging and we chose this upper boundary conservatively.

- $\boldsymbol{g}_{a\chi\chi}$ The DM-mediator interaction is (cf. Eq. (7.3))

$$C_\chi \frac{m_\chi}{f_a} a\bar{\chi}\gamma_5\chi \equiv g_{a\chi\chi} m_\chi a\bar{\chi}\gamma_5\chi\,. \tag{7.65}$$

  Our effective theory is valid only below the scale f_a. Thus, the reheating temperature T_{RH} should be below this scale to safely ignore UV contributions. On the other hand, T_{RH} has to be higher than m_χ. We consequently need a hierarchy $f_a \gtrsim T_{RH} \gtrsim m_\chi$, i.e. small $g_{a\chi\chi} = C_\chi/f_a$, and this is the reason why the DM (the ALP for a fixed mass ratio) should not be too heavy.

- **QCD** Finally, we employ a perturbative description of the strong interactions, only convergent at high enough energies. The DM abundance should therefore be set by interactions happening at temperatures before the QCD phase transition. In practice, we set the upper boundary labelled "QCD" by requiring that the χ-particles freeze out at temperatures above the threshold $T_{\rm pert} = 600$ MeV.

The shape of the contours of constant $g_{a\chi\chi}$ can be understood as follows. By increasing the strength of the ALP-fermion coupling, more energy will be transferred to the hidden sector and heat up the hidden sector particles. This will increase the amount of DM particles, $n_\chi^{\rm eq}(T')$. In order to satisfy the relic density constraint (7.57), the interaction strength among the hidden sector particles has to increase to keep $\chi\bar{\chi} \to aa$ active for longer, such that the χ-particles can continue to annihilate. The DM-ALP interaction is proportional to the product $(g_{a\chi\chi} \cdot m_\chi)^4$. For increasing m_a, $g_{a\chi\chi}$ has to decrease to compensate the effect of increasing m_χ. We find that in this scenario g_{aff} enters only via SM $\to a$ processes because the energy transfer and DM production from the connector process $f\bar{f} \to \chi\bar{\chi}$ is always subdominant. The dependence on g_{aff} is therefore very mild. Furthermore, we observe that $g_{a\chi\chi}$ has to be large compared to the freeze-in scenario to maintain equilibrium among the hidden sector particles.

It turns out that it is extremely challenging to test the DFO region with collider searches for ALPs: in this scenario the visible sector and the mediator barely talk to each other, such that the sensitivity of experiments probing the relevant mass range, in particular the SN bound and the bounds from rare B and K decays, would have to improve by several orders of magnitude. However, the tiny coupling between the SM particles and the ALP makes the ALP relatively long lived, and since the ALPs are abundantly produced along with the DM particles (cf. Fig. 7.7) their decay can have important implications for the cosmological history during and after big bang nucleosynthesis, as briefly discussed in section 7.5. Of course, the imprint ALPs will leave on cosmological observables strongly depends on its decay products and its lifetime. In the range of ALP masses we consider, the dominant decay channel varies, see Fig. G.1. For ALPs with

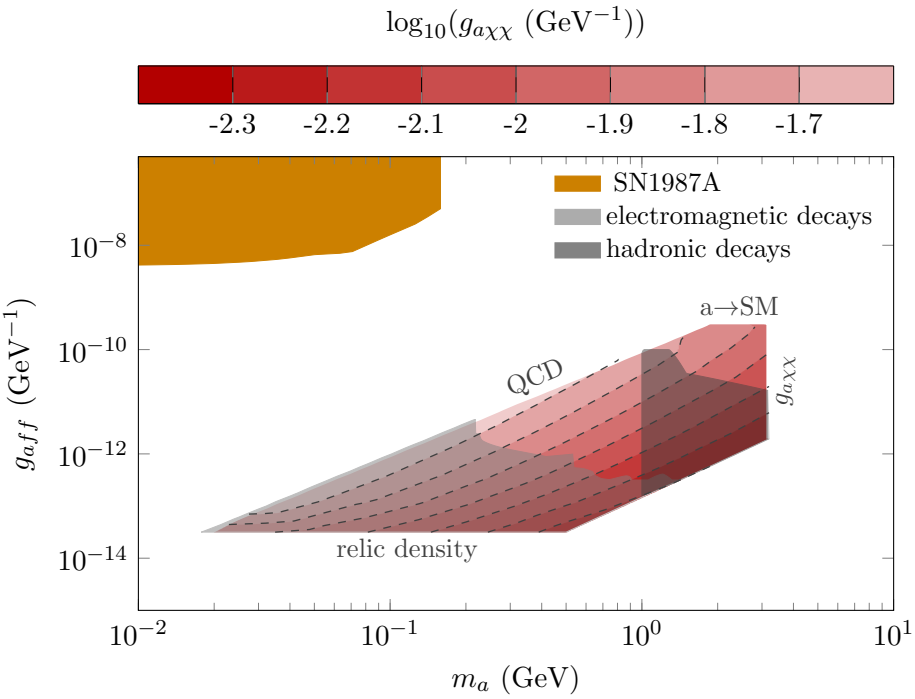


FIGURE 7.8: Contour plot of the hidden-sector couplings $g_{a\chi\chi}$ in the (m_a, g_{aff})-plane which give the observed DM relic density via decoupled freeze-out (DFO). We fixed the ratio $m_\chi/m_a = 10$. We have included the relevant constraints on our ALP model on the connector coupling g_{aff} in this parameter region (cf. Fig. 7.3). We additionally plot the region excluded by cosmological constraints of additional particles decaying electromagnetically (light grey) and decaying hadronically (dark grey). For explanations about the boundaries of the DFO region and the constraints from electromagnetic and hadronic decays see the main text.

masses $m_a \lesssim 2m_\mu$, i.e. ALPs which dominantly decay into photons and e^+e^-, the constraints are very similar. Here, we apply the bounds from Ref. [28] on very long-lived ALPs in the sub-GeV mass range. Generically speaking, they exclude sufficiently abundant particles decaying electromagnetically with lifetimes $\tau_a \sim 10^3 - 10^5$ s. The bound labelled "electromagnetic decays" in Fig. 7.8 was obtained by applying the bounds on the ALP's lifetime from Figs. 4 and 5 in Ref. [28], interpolating between the mass of the decaying particle and the dominant decay channel of the ALP. For masses in the range $2m_\mu \lesssim m_a \lesssim 1$ GeV, the ALP dominantly decays into muons. In principle, the applicable constraints are the electromagnetic ones here – see also the discussion in Ref. [372] – and the lifetime is short enough for them not to matter. For hadronic decays the bounds become more severe, cf. section 7.5, and lifetimes above $\tau \sim 0.1$ s can be excluded. In Ref. [406] bounds are provided on hadronically decaying particles with masses in the GeV-TeV range. The smallest mass for which results are available is 30 GeV. We apply the corresponding bound to our model, extrapolating from the given shapes that the bounds will remain approximately constant for lower masses. We have checked that, taking into account the branching ratios of the various decay channels, the energy injection from ALPs is sufficient in the DFO region for these constraints to apply. However, as outlined in section 7.5, photo-dissociation is clearly not the only possible scenario. For instance, ALPs decaying into photons can re-equilibrate with the SM, a scenario which was studied in depth in Ref. [401], excluding much shorter lifetimes. However, as studied in Ref. [404], since the temperature of our hidden sector T' is in general well below T, we expect these bounds to be alleviated in our case. In order to obtain precise cosmological bounds a detailed analysis of the evolution of the ALP number

density and its imprint on cosmological observables would hence be necessary, yet beyond the scope of this paper.

We conclude this section by remarking that, as mentioned above, the upper left boundary labelled "QCD" arises due to a theoretical limitation. It might be interesting to describe the processes involved at low energies using non-perturbative techniques in order to explore the remaining parameter space. In doing so, we should be able to close the gap between the "QCD" line and ALP-SM equilibrium condition in Fig. 7.5.

7.7 Conclusion

To conclude, we study a fermionic DM model with a pseudoscalar (ALP) mediator coupled to SM and DM fermions at tree level. In this model, different production mechanisms can set the relic abundance, depending on which interactions are most important. In fact, we find that, similar to the case of a fermionic DM model with a massive dark photon mediator studied in Ref. [22], six different production mechanisms are possible for the set of ALP and DM masses we have considered, namely freeze-in from SM particles, freeze-in from mediators, sequential freeze-in, decoupled freeze-out (DFO), freeze-out from mediators and freeze-out from the SM. We numerically solve the general set of coupled Boltzmann equations for a range of ALP-fermion and ALP-DM couplings to obtain the phase diagram for a fixed ALP and DM mass (m_a = 1 GeV, m_χ = 10 GeV). We focus on two cases where the hidden sector, comprised of the DM and ALP, does not thermalise with the visible sector: freeze-in (from the SM or from ALPs) and DFO.

In the freeze-in and DFO regimes, we determine the relic density by solving the relevant Boltzmann equations. This is most challenging in the DFO region, as discussed in sec. 7.3.3, where a set of three coupled differential equations must be solved (the Boltzmann equations for the ALP and DM number densities and the Boltzmann equation governing the energy transfer from the SM to the dark sector), and a variety of processes contribute to ALP and dark matter production. We derive for the first time the energy transfer Boltzmann equation for the case where there is a single hidden sector particle in the final state, and present a new strategy for determining the hidden sector temperature from the evolution of the energy density. Throughout our work we take into account thermal effects by implementing temperature-dependent corrections to the SM masses and couplings, as these can alter the results at the $\mathcal{O}(1)$ level, in particular in the case of the production of ALPs. The dominant contribution stems from the increased masses of the in medium SM particles.

In a model where the ALP-fermion couplings arise from an extended Higgs sector in the UV, both infrared-dominated and ultraviolet-dominated processes contribute to freeze-in, and either of them may dominate depending on the hierarchy of the reheating temperature and the electroweak scale. In particular, we find that UV-dominated $2 \to 3$ interactions play a role for the production via freeze-in from the SM for reheating temperatures $T_{RH} \gtrsim 200$ GeV. We therefore choose to study one typical infrared-dominated (T_{RH} = 200 GeV) and one ultraviolet-dominated scenario (T_{RH} = 2 TeV). In the DFO region all ultraviolet-dominated processes enter only at the subleading level via the energy transfer to the hidden sector, such that the final DM abundance is largely insensitive to the ultraviolet contributions for reheating temperatures below a few TeV, and these are therefore neglected.

We also pay particular attention to calculating the experimental bounds on the relevant parameter space, including collider, astrophysics, cosmology and DM constraints. Some of the most powerful constraints come from collider experiments, particularly beam dump constraints coming from SLAC E137 and flavour constraints notably from NA62 and LHCb, together constraining masses up to the B meson mass and the g_{aff} coupling down to $\mathcal{O}(10^{-6}\ \text{GeV}^{-1})$. For the beam dump experiment we improve previous analyses by including all dominant ALP production processes, namely the Primakoff process, bremsstrahlung and (non-)resonant positron annihilation. In particular, we show that positron annihilation can become the dominant process for larger ALP masses in models where the ALP coupling to photons is highly suppressed and in this case clearly cannot be neglected. For the scenario considered, where the only tree-level couplings of the ALP are to fermions, we added the different production scenarios since they cannot be distinguished by experiment. Our analysis excludes a large part of the parameter space in the (m_a, g_{aff})-plane and puts upper and lower bounds on the coupling. For the flavour constraints, we implement the state-of-the-art bounds for the most constraining channels from NA62, NA48/2, LHCb and CHARM, which together rule out a large section of parameter space, down to $g_{aff} \sim \mathcal{O}(10^{-5}\ \text{GeV}^{-1})$ for masses up to $m_a \sim m_{K^+} - m_{\pi^+}$, and in the range $g_{aff} \sim \mathcal{O}(10^{-4}\ \text{GeV}^{-1})$ to $\mathcal{O}(10^{-2}\ \text{GeV}^{-1})$ up to $m_a \sim m_B - m_{K^{(*)}}$.

Out of all the constraints considered, astrophysics provides the most stringent bounds on our parameter space. In particular SN1987A and HB stars strongly probe the parameter space for $m_a \lesssim \mathcal{O}(0.1\ \text{GeV})$, extending down to $g_{aff} \sim \mathcal{O}(10^{-9}\ \text{GeV}^{-1})$. For the former constraint, we adapt the state-of-the-art analysis [24, 373] to the scenario we consider, including N^{3}LO corrections in chiral perturbation theory. While we do consider DM direct detection, indirect detection and self-interactions, due to the pseudoscalar nature of the mediator, as expected, none of these provide relevant constraints.

The tiny coupling between the SM and the mediator which is required to reproduce the observed DM relic density in the DFO region is responsible for making the new sector so difficult to probe. It therefore seems that current and near future experiments cannot access this region of parameter space, yet cosmology can provide a powerful tool to constrain secluded dark sectors. Here we carry out an estimate of the consequences that the presence of ALPs would have on cosmology for the DFO region: Applying the constraints from Refs. [28] and [406], we exclude very long lived ($\tau_a \gtrsim 10^3 - 10^5$ s) electromagnetically decaying ALPs and short lived ($\tau_a \gtrsim 0.1$ s) hadronically decaying ALPs as their decay products would photodissociate light elements.

In sections 7.6.3 and 7.6.4 we present the regions of parameter space in the (m_a, g_{aff})-plane where the correct relic density is obtained, with contour lines showing the required value of $g_{a\chi\chi}$. For freeze-in, as seen in Fig. 7.5, we find a large unexcluded region, extending to $g_{aff} \sim 10^{-9}\ \text{GeV}^{-1}$ for $10^{-4}\ \text{GeV} < m_a < 10^2\ \text{GeV}$. For larger values of g_{aff} and m_a, smaller values of $g_{a\chi\chi}$ are required. Probing this region, for the largest values of $g_{a\chi\chi}$, would require the flavour constraints to improve substantially, or low-mass resonance searches at ATLAS or CMS. For the DFO regime, as seen in Fig. 7.8, the correct density is obtained for relatively low values of $g_{aff} \sim 10^{-13} - 10^{-10}\ \text{GeV}^{-1}$ and large values of $g_{a\chi\chi}$ ($\sim 10^{-2}\ \text{GeV}^{-1}$). Again it appears that cosmology offers the only means of probing such small values of g_{aff}. Moreover, it turns out the DFO region is disfavoured by the standard BBN scenario.

Part IV

Outlook

In the first part of this thesis we presented a precision calculation of the hadronic contribution to the running of the electromagnetic coupling based on lattice QCD simulations within the SM of particle physics. In the second part, we discussed the study of a dark matter model as an extension to this standard model. To conclude, let us briefly comment on the various possible improvements and future prospects of both projects.

While the data-driven approach to determine the hadronic vacuum polarization has so far provided the most precise determinations, it requires experimental input from $e^+e^- \to$ hadrons cross-section data. Lattice QCD as an ab-initio tool to calculate these hadronic contributions has now reached a comparable precision. Our result for the hadronic running of α in the space-like region up to $8\,\mathrm{GeV}^2$ is on the same level of precision as recent data-driven calculations. However, to compete with the experimental precision on electroweak observables expected from future collider experiments, the current precision will have to be halved. We now discuss how we plan to improve our current precision even further.

We have already reached sub-percent precision for the discrete steps in the step scaling function at short distances above $Q^2 = 1\,\mathrm{GeV}^2$ in the dominant light and strange connected contribution. The precision on our final result, $\hat{\Pi}^{(5)}_{\mathrm{had}}$, is limited by the result of the long-distance contribution to $\hat{\Pi}^l(1\,\mathrm{GeV}^2)$. Taste-breaking effects complicate the continuum extrapolation of this quantity by introducing a large curvature. While the SRHO formalism can be used to reduce discretization errors efficiently, we would like to undertake a more in-depth investigation of those effects using other low-energy effective models, such as NNLO chiral perturbation theory. Another possibility would be to break the computation of this quantity up into at least two additional windows. In the intermediate t-range between $0.4 - 1.0$ fm, where the integrand of $\hat{\Pi}^l(1\,\mathrm{GeV}^2)$ is peaked, the leading effects of taste-breaking, associated with 16 taste pairs of pions, remain small enough to allow for a controlled continuum extrapolation without improvement based on effective models. Most importantly, however, we plan to include two new ensembles bracketing the physical point at an even smaller lattice spacing $a \approx 0.048$ fm, which should greatly reduce the error originating from the continuum extrapolation. Finally, while the continuum extrapolations from fits to uncorrected data and data corrected using LO staggered PT agree well for short-distance quantities and the coefficient of the logarithmically-enhanced discretization term that we obtain from fits to uncorrected data is similar to the one that we computed analytically in LO staggered perturbation theory, this additional lattice spacing should also help constrain those discretization effects which become important at small a.

We also plan to compare our non-perturbative running results to those obtained in perturbation theory, with the objective of finding precise agreement over a significant range of virtualities. Such a matching onto perturbation theory would then allow to run α perturbatively all the way up to the Z-boson mass. An implementation of the running in four-flavour perturbative QCD to $\mathcal{O}(\alpha_s^4)$ is underway. However, it may be necessary to push our non-perturbative calculations to higher virtualities to convincingly show a matching to perturbative running at the sub-percent level. If those virtualities are large on the scale of our present lattice cutoffs, new techniques will have to be developed. These questions will be addressed in future work.

Up to now in this thesis, we have not explored the close relationship between the hadronic running of α and the leading hadronic contribution to the anomalous magnetic moment of the muon, $a_\mu^{\mathrm{LO,had}}$. Both are determined by the hadronic vacuum polarization, but emphasize very different

kinematical regimes. Such a discussion has become particularly important in light of BMW's result for $a_\mu^{\mathrm{LO,had}}$, which is larger than the result obtained using the data-driven approach [7].

On the lattice, $a_\mu^{\mathrm{LO,had}}$ can be obtained by integrating the HVP over Euclidean momenta, convoluted by a Q^2-dependent kernel that is peaked around $(m_\mu/2)^2 \sim (50\ \mathrm{MeV})^2$. Taking an average of the latest independent lattice determinations of the intermediate window contribution to $a_\mu^{\mathrm{LO,had}}$ [7, 37–40], one arrives at a difference of 4.2σ with the data-driven result. Since both the data-driven and lattice approaches aim at computing the same quantity, it is crucial to understand the origin of the disagreement in $a_\mu^{\mathrm{LO,had}}$. The two approaches are mathematically related by a Laplace transform. The running of α and the (discrete) Adler function are suitable observables to pinpoint the energy range in which the largest disagreement appears.

Since the lattice result for $a_\mu^{\mathrm{LO,had}}$ is larger than that of the reference approach, one may ask whether this higher value implies an increase in the tensions in EW global fits [45–49]. As we saw in the last section, the recent determination of the running of α by the Mainz group also indicates a slight excess in the low energy region. Their result for $\alpha_{\mathrm{had}}^{(5)}(M_Z)$ is however not in significant tension with EW fits [8]. It would be interesting to use our result for $\Delta\alpha_{\mathrm{had}}^{(5)}(8\,\mathrm{GeV}^2)$ as a prior in global electroweak fits to determine the Higgs mass indirectly and compare this value to the measured Higgs mass.

Our current lattice sizes and spacings allow us to obtain the running of α up to $\sim 8\ \mathrm{GeV}^2$. Our final aim is to obtain the running up to the mass of the Z-boson and $\alpha(M_Z^2)$. If matching to perturbative QCD expression can be achieved at such virtualities, we will employ perturbation theory at $\mathcal{O}(\alpha_s^4)$. On a longer timescale, we are further investigating strategies to obtain the running all the way up to the Z-scale non-perturbatively using step-scaling techniques [424]. Several members of the BMW collaboration have performed a first study in the two-dimensional Schwinger model (QED_2) to test the step scaling strategy which now has to be extended to QCD [425].

Finally, on the experimental side, future possible direct measurements of the effective electromagnetic coupling in the space-like region from $\mu - e$ scattering (MUonE-experiment) would give direct access to the Euclidean HVP [426, 427]. These measurements would in fact provide a third approach that is independent of both the reference and the lattice approach.

Likewise, our work on ALP mediated DM has raised several questions which we would like to tackle. First we would like to improve the accuracy of our calculation in the sequential freeze-in region, by solving the unintegrated Boltzmann equation. It would also be interesting to compare our results in the DFO region to those obtained using the unintegrated Boltzmann equations. Furthermore, it might be worthwhile to use non-perturbative methods to investigate the DFO mechanism in the region above the QCD boundary. Finally, for our work, we took into account thermal effects by implementing temperature-dependent corrections to the SM masses and couplings. It would be interesting to undertake a more involved analysis of possible additional finite temperature effects which might play an important role.

We further presented brand-new calculations of constraints on the ALP, in particular for SN, beam-dump and flavour constraints which are crucial for the parameter space of interest. Large regions in both the freeze-in and DFO regimes remain unexplored, and while the former could be probed by prospective collider experiments, the latter would be difficult to access. However

as the standard model of cosmology is very successful it does not leave much room for new relativistic degrees of freedom during the time of BBN. Cosmological bounds are at present the only way to probe the DFO region, therefore an updated more detailed analysis of the effect of these on our parameter space is crucial. It is worth noting that the sensitivity of the CMB stage IV observations on $\Delta N_{\rm eff}$ is expected to improve by more than one order of magnitude and will hence either strengthen existing constraints or hint towards the existence of new degrees of freedom [428]. In this work we focus on a Dirac fermion dark matter particle but, as previously mentioned, a Majorana fermion is an equally well-motivated possibility which could merit a detailed study, although one should expect qualitatively similar results. Finally, the potential sensitivity of future experiments is of utmost interest, and in this view it should be worthwhile to assess the sensitivity of planned future flavour experiments and upgrades, beam dump experiments and cosmology measurements.

Appendices

Appendices: running of α

A Compilation of current precision

Group, year, Ref.	$\Delta\alpha^{(5)}_{\rm had}(M_Z)$	Remarks
Geshkenbein+Morgunov (1995) [429]	0.02780(6)	$\mathcal{O}(\alpha_s)$ resonance model
Swartz (1996) [430]	0.02752(46)	data-driven, include correlations
Krasnikov, Rodenberg (1998) [431]	0.02737(39)	data-driven and semi-analytical ansatz for $\sqrt{s} > 2.3$ GeV
Kühn+Steinhauser (1998) [210]	0.02775(17)	data-driven and pQCD full $\mathcal{O}(\alpha_s^2)$, for $\sqrt{s} > 1.8$ GeV
Erler (1999) [205]	0.02779(20)	data-driven and pQCD to $\mathcal{O}(\alpha_s^3)$ for $\sqrt{s} > 1.8$ GeV unsubtracted dispersion relation
Groote *et al.* (1998) [432]	0.02776(41)	QCD sum rule with less data $\mathcal{O}(\alpha_s^3)$
Martin (2000) [433]	0.02738(20)	includes new BES data, $\mathcal{O}(\alpha_s^3)$ pQCD
Troconiz+Yndurain (2005) [434]	0.02749(12)	pQCD for $s > 2\text{GeV}^2$
Burkhardt+Pietrzyk (2005) [435]	0.02758(35)	data-driven (new BES data) pQCD for $\sqrt{s} > 12$ GeV
HMNT (2006) [436]	0.02768(22)	data-driven
Jegerlehner (2008) [203]	0.027594(219) 0.027515(149)	data-driven/pQCD to $\mathcal{O}(\alpha_s^4)$ Adler function ($\sqrt{s_0} = 2.5$ GeV)
DHMZ (2010) [437]	0.02742(10)	pQCD from $1.8 < \sqrt{s} < 3.7$ GeV and for $\sqrt{s} > 5$ GeV
Jegerlehner (2011) [438]	0.027498(135)	Adler function ($\sqrt{s_0} = 2.5$ GeV) alphaQED package
HLMNT (2011) [439]	0.027626(138)	data-driven pQCD for $\sqrt{s} = 2.6 - 3.7,\ > 5$ GeV
Bodenstein (2012) [440]	0.02757(08)	heavy quark contributions in pQCD
Jegerlehner (2017) [441]	0.027738(158)	includes $\gamma - \rho$ mixing, corrected τ data and perturbative Adler function
KNT (2018) [14]	0.027611(111)	new data combination method pQCD for $\sqrt{s} > 11.2$ GeV
DHMZ (2019) [4]	0.02761(10)	pQCD from $1.8 < \sqrt{s} < 3.7$ GeV and for $\sqrt{s} > 5$ GeV
KNT (2019) [5]	0.027609(112)	data-driven with updates in R-ratio data and pQCD for $\sqrt{s} > 11.2$ GeV
Mainz (2022) [8]	0.02773(15)	LQCD up to $Q^2 = 7\,\text{GeV}^2$ and $\mathcal{O}(\alpha_s^2)$ pQCD in Jegerlehner's scheme [438]

TABLE 7.2: Compilation of results for $\Delta\alpha^{(5)}_{\rm had}(M_Z^2)$ from different groups over almost twenty years. We also provide remarks about the analysis. Note that the last entry is the first full lattice QCD calculation of the hadronic running.

B Staggered leading order lattice perturbation theory

We discuss results at leading order in staggered perturbation theory. In the following, we drop the subscript "LO" for better readability.

B.1 Current correlators

Let us work out the free, staggered fermion expression for the current-current correlators in the TMR for all the different staggered tastes. At leading order in perturbation theory, we can work with naive quarks and, to obtain the staggered result, take the 16th root (16 since we use staggered quarks and take the 4th root of the fermion determinant). We recover the correct factors for staggered, QCD results in the final expressions. We use again the notation from Ref. [108] and work in lattice units, i.e. with $a = 1$. We restore the lattice spacing in the final expressions. Then, the generic form of a current with spin s and taste t is given by

$$J^t_{s,x} = \frac{1}{N_{\delta x}} \sum_{\delta x} \bar{\psi}_x B^\dagger_{\bar{t},x} \gamma_s \psi_{x+\delta_x} \ , \tag{B.1}$$

with $B^\dagger_{\bar{t},x} = (-1)^{\bar{t}\cdot x}\gamma_t^\dagger = e^{i\pi\bar{t}\cdot x}\gamma_t^\dagger$ the taste transformation, $\gamma_t = \prod_\mu \gamma_\mu^{t_\mu}$, and $\delta x \in \{-1, 0, 1\}$ such that $(\delta x = s + t) \bmod 2$, and $N_{\delta x}$ the number of δx. Then, the connected contribution to the current correlator from the origin is:

$$\begin{aligned} C(x) = \langle J^t_{s,x} J^t_{s,0}\rangle &= \frac{1}{N_{\delta x} N_{\delta x'}} \sum_{\delta x, \delta x'} \langle \bar{\psi}_x B^\dagger_{\bar{t},x} \gamma_s \psi_{x+\delta x} \bar{\psi}_0 B^\dagger_{\bar{t}_0} \gamma_s \psi_{\delta x'}\rangle \\ &= -\frac{1}{N_{\delta x} N_{\delta x'}} \sum_{\delta_x, \delta_{x'}} \mathrm{Tr}\left\{ B^\dagger_{\bar{t},x} \gamma_s M^{-1}_{x+\delta_x,0} B^\dagger_{\bar{t},0} \gamma_s M^{-1}_{\delta_{x'},x} \right\} \ , \end{aligned} \tag{B.2}$$

with $S(p)$ the momentum space and M^{-1}_{xy} the position space propagator, related via

$$M^{-1}_{xy} = \int_p e^{ip\cdot(x-y)} S(p) \ . \tag{B.3}$$

Let us define

$$F(p) = \frac{1}{N_{\delta x}} \sum_{\delta x} e^{ip\delta x} = \prod_\mu \frac{1}{N_{\delta x_\mu}} \sum_{\delta x_\mu} e^{ip_\mu \delta x_\mu} = \prod_\mu F_\mu(p) \ , \tag{B.4}$$

in terms of which Eq. (B.2) can be rewritten as

$$C(x) = -\int_{pq} e^{i(p+\pi\bar{t}-q)x} F(p) F(q)\, \mathrm{Tr}\left\{ \gamma_t^\dagger \gamma_s S(p) \gamma_t^\dagger \gamma_s S(q) \right\} \ , \tag{B.5}$$

where we have also inserted the expression of the taste transformation. To obtain the correlator in the time momentum representation we sum over spatial components,

$$\begin{aligned} C(x_0) = \sum_{\vec{x}} C(x) = -\int_{pq_0} & e^{i(p_0+\pi\bar{t}_0-q_0)x_0} F(p_0, \vec{p})\, F\left(q_0, \tilde{p} + \pi\vec{\bar{t}}\right) \\ & \times\ \mathrm{Tr}\left\{ \gamma_t^\dagger \gamma_s S(p_0, \vec{p})\, \gamma_t^\dagger \gamma_s S\left(q_0, \vec{p} + \pi\vec{\bar{t}}\right) \right\} \ . \end{aligned} \tag{B.6}$$

Now (see also App. A of Ref. [108]) $S\left(q_0,\vec{p}+\pi\vec{t}\right) = B^{\dagger}_{\bar{r},0} S\left(q_0,\vec{p}\right) B_{\bar{r},0}$ where we denote $\bar{r} = (0,\vec{\bar{t}})$ and $B_{\bar{r},0} = \gamma_r$ so

$$C\left(x_0\right) = -\int_{pq_0} e^{i\left(p_0+\bar{t}_0\pi-q_0\right)x_0} F\left(p_0,\vec{p}\right) F\left(q_0,\vec{p}+\pi\vec{t}\right) \times \mathrm{Tr}\left\{\gamma_r\gamma_t^{\dagger}\gamma_s S\left(p_0,\vec{p}\right)\gamma_t^{\dagger}\gamma_s\gamma_r^{\dagger} S\left(q_0,\vec{p}\right)\right\} \quad \text{(B.7)}$$

where we consider a spatial component of the EM current along the axis $\hat{\imath}$, $s = (0,\hat{\imath})$, and we average over the three spatial directions. Now, Eq. (B.4) yields

$$F_\mu(p) = \begin{cases} 1 & \text{if } (t+s)_\mu \bmod 2 = 0 \\ \cos p_\mu & \text{if } (t+s)_\mu \bmod 2 = 1 \end{cases}$$

and

$$F_\mu(q_0,\vec{p}+\pi\vec{t}) = \begin{cases} 1 & \text{if } (t+s)_\mu \bmod 2 = 0 \\ \cos q_0 & \text{if } \mu = 0 \ \text{ and } \ (t+s)_0 \bmod 2 = 1 \\ \cos p_j & \text{if } \mu = j, (t+s)_j \bmod 2 = 1 \ \text{ and } \ \bar{t}_j = 0 \\ -\cos p_j & \text{if } \mu = j, \ (t+s)_j \bmod 2 = 1 \ \text{ and } \ \bar{t}_j = 1 \end{cases} .$$

Let us first work out the time-local correlators. These have $(s_0+t_0) \bmod 2 = 0$ and since $s_0 = 0$, we have $t_0 = 0$ and $F\left(p_0,\vec{p}\right) = F(0,\vec{p})$ for all p. Then

$$C\left(x_0\right) = -(-1)^{\bar{t}_0 x_0}\frac{1}{3}\sum_i \int_{\vec{p}} \underbrace{F(0,\vec{p})F(0,\vec{p}+\pi\bar{t})}_{A} \times \underbrace{\mathrm{Tr}\left\{\gamma_r\gamma_t^{\dagger}\gamma_i S(x_0,\vec{p})\gamma_t^{\dagger}\gamma_i\gamma_r^{\dagger} S\left(-x_0,\vec{p}\right)\right\}}_{4B} . \quad \text{(B.8)}$$

Let us work out the free propagator and its time-momentum representation. The momentum space propagator is

$$\tilde{S}(p) = \tilde{M}^{-1}(p) = \frac{-i\gamma_\mu\hat{p}_\mu + m}{\hat{p}^2+m^2} \quad \text{with} \quad \hat{p}_\mu = \sin p_\mu . \quad \text{(B.9)}$$

So, the time-momentum propagator is given by

$$S(x_0,\vec{p}) = \int_{p_0} e^{ip_0 t}\tilde{S}\left(p_0,\vec{p}\right) \quad \text{(B.10)}$$

For $x_0 > 0$, let $z = e^{ip_0}$ so $\hat{p}_0 = \frac{1}{2i}\left(z-\frac{1}{z}\right)$. Performing the contour integration we find for the naive fermion propagator in time-momentum representation

$$\begin{aligned} S(x_0,\vec{p}) &= \int \mathrm{d}p_0 e^{ip_0x_0} M^{-1}(p_0,\vec{p}) \quad \text{(B.11)} \\ &= \frac{e^{-Et}}{\sinh(2E)}\left(\gamma_0\sinh(E) - i\vec{\gamma}\cdot\sin(\vec{p}) + m\right) - \frac{\left(-e^{-E}\right)^{x_0}}{\sinh(2E)}\left(\gamma_0\sinh(E) + i\vec{\gamma}\cdot\sin(\vec{p}) - m\right) . \end{aligned}$$

We write

$$S(x_0,\vec{p}) = \gamma_0 S_0\left(x_0,\vec{p}\right) + \gamma_i S_i\left(x_0,\vec{p}\right) + S_m\left(x_0,\vec{p}\right) , \quad \text{(B.12)}$$

where the sum over the spatial index i is implied. Note that $S\left(-x_0,\vec{p}\right) \equiv \hat{S}(x_0,\vec{p}) = -\gamma_0 S_0 + \gamma_i S_i + S_m$ and we denote

$$S_\alpha\left(x_0,\vec{p}\right) = \sigma_\alpha\left(x_0,\vec{p}\right)\left\{1+\zeta_\alpha(-1)^{x_0}\right\} , \quad \text{with} \quad \zeta_\alpha = \begin{cases} -1 & \text{for } \alpha = 0 \\ 1 & \text{for } \alpha = i,m \end{cases} \quad \text{(B.13)}$$

where

$$\sigma_0\left(x_0,\vec{p}\right)=\frac{\sinh E}{\sinh 2E}e^{-Ex_0},\quad \sigma_i\left(x_0,\vec{p}\right)=-\frac{i\hat{p}_i}{\sinh 2E}e^{-Ex_0},\quad \sigma_m\left(x_0,\vec{p}\right)=\frac{m}{\sinh 2E}e^{-Ex_0}\ . \tag{B.14}$$

Explicitly, we have for instance for the conserved current (ID V in Tab. 4.4) $s=(0,\hat{\imath})$, $t=(0,\vec{0})$ and hence $\bar{t}=(0,\vec{0})$. Then $A=\cos^2 p_i$ and $B=S_0^2+2S_i^2-\vec{S}^2+S_m^2$ and one obtains

$$C^{\mathrm{Vtl}}\left(x_0\right)=-\frac{8}{3}\sum_i\int_{\vec{p}}\cos^2 p_i\left\{\sigma_0^2+2\sigma_i^2-\vec{\sigma}^2+\sigma_m^2+(-1)^{x_0}\left(-\sigma_0^2+2\sigma_i^2\sigma^2+\sigma_m^2\right)\right\}\ . \tag{B.15}$$

The remaining staggered correlators for the time-local currents can be obtained along the same lines and we collect the expressions below. Now, in the case of time non-local currents we have, from Eq. (4.25),

$$\begin{aligned}C(x)&=\left\langle\left(J_{s,t}(x)\right)_{\mathrm{tnl}}\left(J_{s,t}(0)\right)_{\mathrm{tnl}}\right\rangle\\&=-\frac{1}{4N_{\delta\vec{x}}N_{\delta\vec{x}'}}\sum_{\delta\vec{x},\delta\vec{x}'}\mathrm{Tr}\left\{B^\dagger_{\bar{t},x}\gamma_s M^{-1}_{x+\hat{0}+\delta\vec{x},0}B^\dagger_{\bar{t},0}\gamma_s M^{-1}_{\hat{0}+\delta\vec{x}',x}+(-1)^{\bar{t}_0}B^\dagger_{\bar{t},x}\gamma_s M^{-1}_{x+\hat{0}+\delta\vec{x},\hat{0}+\delta x'}\gamma_s B^\dagger_{\bar{t},0}M^{-1}_{0,x}\right.\\&\left.+(-1)^{\bar{t}_0}\gamma_s B^\dagger_{\bar{t},x}M^{-1}_{x,0}B^\dagger_{\bar{t},0}\gamma_s M^{-1}M^{-1}_{\hat{0}+\delta\vec{x}',x+\hat{0}+\delta x}+\gamma_s B^\dagger_{\bar{t},x}M^{-1}_{x,\hat{0}+\delta\vec{x}'}\gamma_s B^\dagger_{\bar{t},0}M_{0,x+\hat{0}+\delta\vec{x}}\right\}\ .\end{aligned} \tag{B.16}$$

Let us define here

$$G(\vec{p})=\frac{1}{N_{\delta x}}\sum_{\delta\vec{x}}e^{i\vec{p}\cdot\delta_x}=\prod_i\frac{1}{N_{\delta x_i}}\sum_{\delta_{x_i}}e^{ip_i\delta_{x_i}}=\prod_i G_i(\vec{p}) \tag{B.17}$$

in terms of which Eq. (B.16) becomes

$$\begin{aligned}C(x_0)=\sum_{\vec{x}}C(x)=-\frac{(-1)^{\bar{t}_0x_0}}{2}\int_p&\underbrace{G(\vec{p})G(\vec{p}+\pi\vec{t})}_{A}\underbrace{\mathrm{Tr}\left\{B^\dagger_{\bar{t},0}\gamma_s S(x_0+1,\vec{p})B^\dagger_{\bar{t},0}\gamma_s\gamma_r^\dagger\hat{S}(x_0-1,\vec{p})\gamma_r\right\}}_{4B}\\&+\underbrace{(-1)^{\bar{t}_0}G(\vec{p})^2}_{C}\underbrace{\mathrm{Tr}\left\{B^\dagger_{\bar{t},0}\gamma_s S(x_0,\vec{p})\gamma_s B^\dagger_{\bar{t},0}\gamma_r^\dagger\hat{S}(x_0,\vec{p})\gamma_r\right\}}_{4D}\ .\end{aligned} \tag{B.18}$$

For instance, for the time non-local analogue of the conserved current (ID Vtnl in Tab. 4.5) we have $s=(0,\hat{\imath})$, $t=(1,\vec{0})$, hence $\bar{t}=(0,1,1,1)$ and $\bar{t}_0x_0=0$. We then obtain for the various terms in Eq. (B.18)

$$B=\cos^2 p_i=-A \tag{B.19}$$

$$\begin{aligned}C=-\Big\{&S_0\left(x_0+1\right)S_0\left(x_0-1\right)+2S_i(x_0+1)S_i\left(x_0-1\right)\\&-\vec{S}\left(x_0+1\right)\cdot\vec{S}\left(x_0-1\right)+S_m(x_0+1)S_m\left(x_0-1\right)\Big\}\end{aligned} \tag{B.20}$$

$$D=\left\{S_0^2\left(x_0\right)+2S_i^2\left(x_0\right)-\vec{S}^2\left(x_0\right)+S_m^2\left(x_0\right)\right\}\ . \tag{B.21}$$

Thus, the oscillating cross terms cancel in $AB+CD$ and we get

$$C^{\mathrm{Vtnl}}\left(x_0\right)=-\frac{8}{3}\sum_i\int_{\vec{p}}\cos^2 p_i\left\{\sigma_0^2+2\sigma_i^2-\vec{\sigma}^2+\sigma_m^2\right\} \tag{B.22}$$

which is the result of Eq. (B.15), without the oscillatory term, and, when integrated over a finite, fixed, physical interval in x_0, it has the same continuum limit. The remaining currents can be calculated analogously.

abbrev. J	spin$\otimes$taste	$f^J(p_i, p_j, p_k)$	B^J	$B^{J,TL}$
V	$\gamma_i \otimes 1$	$\cos(ap_i)^2$	$\hat{E}^2 - \hat{p}_i^2$	$\hat{p}_i^2$
IV	$\gamma_i \otimes \xi_i$	1	$\hat{E}^2 - \hat{p}_i^2$	$\hat{p}_i^2$
X	$\gamma_i \otimes \xi_i\xi_j$	$\sum_{j\neq i} \cos(ap_j)^2$	$\hat{E}^2 - \hat{p}_i^2$	$\hat{p}_i^2$
XI	$\gamma_i \otimes \xi_i\xi_j\xi_k$	$\cos(ap_j)^2 \cos(ap_k)^2$	$\hat{E}^2 - \hat{p}_i^2$	$\hat{p}_i^2$
XV	$\gamma_i \otimes \xi_j$	$\sum_{j\neq i} \cos(ap_j)^2 \cos(ap_i)^2$	$\hat{E}^2 - \hat{p}_i^2$	$\hat{p}_i^2$
XX	$\gamma_i \otimes \xi_j\xi_k$	$\prod_j \cos(ap_j)^2$	$\hat{E}^2 - \hat{p}_i^2$	$\hat{p}_i^2$
Vtnl (25)	$\gamma_i \otimes \xi_0$	$\cos(ap_i)^2$	$\hat{E}^2 - \hat{p}_i^2$	0
IVtnl (24)	$\gamma_i \otimes \xi_i\xi_0$	1	$\hat{E}^2 - \hat{p}_i^2$	0
Xtnl (30)	$\gamma_i \otimes \xi_i\xi_j\xi_0$	$\sum_{j\neq i} \cos(ap_j)^2$	$\hat{E}^2 - \hat{p}_i^2$	0
XItnl (31)	$\gamma_i \otimes \xi_i\xi_j\xi_k\xi_0$	$\cos(ap_j)^2 \cos(ap_k)^2$	$\hat{E}^2 - \hat{p}_i^2$	0
XVtnl (35)	$\gamma_i \otimes \xi_j\xi_0$	$\sum_{j\neq i} \cos(ap_j)^2 \cos(ap_i)^2$	$\hat{E}^2 - \hat{p}_i^2$	0
XXtnl (40)	$\gamma_i \otimes \xi_j\xi_k\xi_0$	$\prod_j \cos(ap_j)^2$	$\hat{E}^2 - \hat{p}_i^2$	0

TABLE 7.3: Expressions of the different terms in Eq. (B.23) for the various staggered currents. We use the labelling of Ref. [132].

By performing the calculations for the remaining staggered currents we obtain a general expression for the LO current-current correlators in the TMR for the different staggered tastes. To obtain staggered, QCD results for a quark of flavour f with electric charge $Q_f = eq_f$ we have to multiply the previous expressions by $q_f^2 N_c/16$. In the following we set $x_0 = t$ to use the notation from Sec. 4.3.1.4. Let us denote $\hat{p}_i = \sin(ap_i)/a$, $\hat{E} = \sinh(aE)/a$, $E = \operatorname{arcsinh}(a\sqrt{\sum_i \hat{p}_i^2 + m^2})/a$ and $\dot{E} = \hat{E}\sqrt{1 + (a\hat{E})^2}$. The different staggered current correlators then take the form

$$C^J(t) = \frac{N_c q_f^2}{3} \sum_i \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}\vec{p}}{(2\pi)^3} \frac{e^{-2Et} f^J\left(p_i, p_j, p_k\right)}{4\dot{E}^2} \left[B^J - (-1)^t B^{J,TL}\right] , \tag{B.23}$$

where the oscillatory term $(-1)^t B^{J,TL}$ only appears for the time-local currents. This expression can be decomposed into a contribution which approaches the continuum result in the limit $a \to 0$ and in case of the time-local currents a part which is oscillating around this solution. For the currents, we use the definitions from Ref. [132]. The results are summarized in Tab. 7.3. For instance, the expression for the conserved time-local current correlator V reads

$$C^{\mathrm{V}}(t, a) = \frac{N_c q_f^2}{3} \sum_i \int_{-\pi/a}^{\pi/a} \frac{d\vec{p}}{(2\pi)^3} \frac{\cos(ap_i)^2 e^{-2Et}}{4\dot{E}^2} \left[\hat{E}^2 - \hat{p}_i^2 \left(1 + (-1)^t\right)\right] . \tag{B.24}$$

The disconnected contributions vanish at leading order. There are also contact terms which are proportional to $\delta_{t,0}$ and therefore do not contribute to the TMR integral for our set of observables.

B.2 HVP and (discrete) Adler function

Here we give the results for our observables using the traditional kernel function $k_0(Q, t)$. The results using the lattice momentum $\hat{Q}$ can be derived straight-forwardly in a similar fashion. For simplicity, we use lattice units in this section, $a = 1$, and recover the lattice spacing in the final

expressions. The different kernel functions then read

$$\begin{aligned} k^{\Pi}(t,Q) &= \frac{\cos(Qt)-1}{Q^2} + \frac{t^2}{2} \\ k^{D}(t,Q) &= -12\pi^2\left(\frac{\cos(Qt)-1}{Q^2} - \frac{t\sin(Qt)}{2Q}\right) \\ k^{\Delta\Pi}(t,Q,Q_R) &= \frac{\cos(Qt)-1}{Q^2} - \frac{\cos(Q_R t)-1}{Q_R^2} \ . \end{aligned} \tag{B.25}$$

Our observables are obtained via the modified Fourier transform

$$O_l^J(Q^2,a) = 2\sum_{t=0}^{\infty} k_l^O(t,Q)C^J(t,a) \ , \tag{B.26}$$

i.e. by the sum over t of the correlator multiplied by some kernel function. Here, we calculate the result in infinite volume and infinite time. Inserting the results obtained in the previous section for $C^J(t)$ we obtain

$$O_l^J = 2q_f^2 N_c \frac{1}{3}\sum_i \int \frac{\mathrm{d}\vec{p}}{(2\pi)^3}\frac{f^J\left(p_i,p_j,p_k\right)}{4\dot{E}^2}\sum_{t=0}^{\infty} k_l^O(t,Q)e^{-2Et}\left[B^J - (-1)^t B^{J,TL}\right] \ . \tag{B.27}$$

In particular, operators with time-local currents IV, V, X, XI and XV take the form

$$O_l^J = 2q_f^2 N_c \int \frac{\mathrm{d}\vec{p}}{(2\pi)^3}\frac{f^J\left(p_i,p_j,p_k\right)}{4\dot{E}^2}\sum_t k_l^O(t,Q)e^{-2Et}\left[\hat{E}^2 - \hat{p}_i^2\left(1+(-1)^t\right)\right] \tag{B.28}$$

while for the time non-local operators IVnl, Vnl, Xnl, XInl and XVnl we simply have

$$O_l^J = 2q_f^2 N_c \int \frac{d\vec{p}}{(2\pi)^3}\frac{f^J\left(p_i,p_j,p_k\right)}{4\dot{E}^2}\sum_t k_l^O(t,Q)e^{-2Et}\left[\hat{E}^2 - \hat{p}_i^2\right] \ . \tag{B.29}$$

Hence, we have a contribution common to all current tastes and one particular to the time-local operators. We therefore define

$$A_l^O \equiv \sum_t k_l^O(t,Q)e^{-2Et} \ , \qquad A_l^{O,TL} \equiv \sum_t k_l^O(t,Q)e^{-2Et}(-1)^t \ . \tag{B.30}$$

To obtain the results for the various observables, this leaves a number of sums to calculate. We denote them by[12]

$$c_1 \equiv \sum_t \frac{\cos(Qt)}{Q^2}e^{-2Et} \qquad c_2 \equiv \sum_t \frac{\cos(Qt)}{Q^2}e^{-2Et}(-1)^t \tag{B.31}$$

$$e_1 \equiv \sum_t -\frac{1}{Q^2}e^{-2Et} \qquad e_2 \equiv \sum_t -\frac{1}{Q^2}e^{-2Et}(-1)^t \tag{B.32}$$

$$t_1 \equiv \sum_t \frac{t^2}{2}e^{-2Et} \qquad t_2 \equiv \sum_t \frac{t^2}{2}e^{-2Et}(-1)^t \tag{B.33}$$

$$s_1 \equiv \sum_t \frac{t\sin(Qt)}{2Q}e^{-2Et} \qquad s_2 \equiv \sum_t \frac{t\sin(Qt)}{2Q}e^{-2Et}(-1)^t \ . \tag{B.34}$$

In infinite volume and infinite time these sums can be calculated analytically by using a geometric series. We collect our results below.

[12] Again, we will restrict ourselves here to the kernel functions k_0 employing Q. The corresponding expressions for the kernel functions k_1 using $\hat{Q}$ can be obtained along the same lines.

HVP For the term common to all discretizations of the current we find

$$\begin{aligned} A^{\hat{\Pi}(Q^2)} &= (c_1 + e_1 + t_1) \\ &= \frac{a\dot{E}}{8(a\hat{E})^2}\left(\frac{1}{(a\hat{E})^2} - \frac{4}{(aQ)^2}\frac{\cos(aQ)-1}{-2(a\hat{E})^2+\cos(aQ)-1}\right). \end{aligned} \tag{B.35}$$

The one specific to time-local operators reads

$$\begin{aligned} A_{TL}^{\hat{\Pi}(Q^2)} &= (c_1 + e_1 + t_1) \\ &= \frac{(a\hat{E})^4}{8(a\dot{E})^3}\left(-\frac{4(\cos(aQ)-1)\left((a\hat{E})^2+1\right)}{(aQ)^2\left(2(a\hat{E})^2+\cos(aQ)+1\right)} - 1\right). \end{aligned} \tag{B.36}$$

Discrete Adler function For the one common to all discretizations of the current we obtain

$$\begin{aligned} A^{\Delta\Pi} &= c_1(Q) + e_1(Q) - [Q \to Q_R] \\ &= \frac{a\dot{E}}{(a\hat{E})^2}\left(\frac{1-\cos(aQ)}{2(aQ)^2\left(-2(a\hat{E})^2+\cos(aQ)-1\right)} - [Q \to Q_R]\right). \end{aligned} \tag{B.37}$$

For the term specific to time-local operators we get

$$\begin{aligned} A_{TL}^{\Delta\Pi} &= c_2(Q) + e_2(Q) - [Q \to Q_R] \\ &= \frac{(a\hat{E})^2}{a\dot{E}}\left(\frac{1-\cos(aQ)}{2(aQ)^2\left(2(a\hat{E})^2+\cos(aQ)+1\right)} - [Q \to Q_R]\right). \end{aligned} \tag{B.38}$$

Adler function The one common to all discretizations of the current reads

$$\begin{aligned} A^D &= e_1 + c_1 + s_1 \\ &= -12\pi^2\left(\frac{a\dot{E}\sin(aQ)}{2aQ\left(-2a^2\hat{E}^2+\cos(aQ)-1\right)^2} + \frac{(1-\cos(aQ))a\dot{E}}{2(a\hat{E})^2(aQ)^2\left(-2a^2\hat{E}^2+\cos(aQ)-1\right)}\right). \end{aligned} \tag{B.39}$$

For the one specific to time-local currents we obtain

$$\begin{aligned} A_{TL}^D &= e_2 + c_2 + s_2 \\ &= -12\pi^2\left(\frac{a\dot{E}}{2(aQ)^2}\left(\frac{4(a\hat{E})^2 - aQ\sin(aQ) + 2\cos(aQ) + 2}{\left(2(a\hat{E})^2+\cos(aQ)+1\right)^2} - \frac{(a\hat{E})^2}{(a\dot{E})^2}\right)\right). \end{aligned} \tag{B.40}$$

The final expressions are then obtained by numerically evaluating the integral

$$O_l^J = \frac{2q_f^2 N_c}{3}\sum_i \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}\vec{p}}{(2\pi)^3}\frac{f^J\left(p_i,p_j,p_k\right)}{4\dot{E}^2}\left[B^J A_l^O - B^{J,TL} A_{l,TL}^O\right]. \tag{B.41}$$

B.3 Coefficients of the leading logarithmically-enhanced cutoff effects

In this section we explicitly derive the coefficient of the logarithmically-enhanced cutoff effects for the various staggered currents in leading order lattice perturbation theory. We outline the derivation for the conserved current (current V, see Tab. 7.3) for two definitions of the kernel functions.

To summarize our findings, this coefficient is the same for all observables (Adler function, discrete Adler function and HVP). Similarly, the coefficient does not change for the time-local and time-non-local currents. The calculation can be performed in the massless limit, since in the massive case the same coefficient is found but an additional term arises which can however be absorbed in the $(aQ)^2$ coefficient.

Let us briefly outline the calculation and the various sources of logarithmically-enhanced cutoff effects in our observables $O_l^J(Q^2,a)$. All observables can be obtained from the modified Fourier transform

$$O_l^J(Q,t,a) = 2\int_0^\infty \mathrm{d}t\; k_l(t,Q,a)\; C^J(t,a)\;. \tag{B.42}$$

Upon expansion of $C^J(t,a)$ in powers of a we obtain

$$C^J(t,a) = G(t)\big|_{a^0} + a^2\left(G(t)\big|_{a^2} + G^J(t)\big|_{a^2}\right) + \mathcal{O}\left(a^4\right)\;, \tag{B.43}$$

where the first two terms, $G(t)\big|_{a^0}$ and $G(t)\big|_{a^2}$ are common to all staggered currents and the third term is a contribution specific to current J. This contribution stems from the specific prefactors $f^J(p_i,p_j,p_k)$ that we calculated in App. B.1. For example, for the conserved current V we have $f^J(p_i,p_j,p_k) = \cos(ap_i)^2$. In general, if one uses the kernel function $k_1(Q,t)$ with lattice momentum $\hat{Q}$, the kernel function contributes cutoff effects, too. Note that $k_2(Q,t)$, which we briefly introduced in Eq. (4.58), does not contribute to logarithmically-enhanced cutoff effects at $\mathcal{O}(a^2)$. However, they show up at $\mathcal{O}(a^4)$. Similarly, note that in general a whole tower of logarithmically-enhanced cutoff effects are generated. One can easily convince oneself by expanding the kernel function k_1 and k_2 to higher orders in a. Similarly, from a heuristic argument, one also expects a whole tower of such cutoff effects originating from the correlator. We give an intuitive argument below. These effects are not seen in our simulations and we do not calculate their coefficients. Hence, the cutoff effects to order a^2 for any of our observables in LO LPT read

$$O_l^J\left(Q^2,t,a\right)\Big|_{a^2} = k_l^O(t,Q,a)\Big|_{a^0}\left\{G(t)\big|_{a^2} + G^J(t)\big|_{a^2}\right\} + k_l^O(t,Q,a)\Big|_{a^2}\; G(t)\big|_{a^0} \tag{B.44}$$

Example calculation for $\Delta\Pi_0^{\mathrm{V}}(Q^2,Q_R^2)$ and $\Delta\Pi_1^{\mathrm{V}}(Q^2,Q_R^2)$ While our previous derivations were fairly general, let us now perform the calculation for a particular observable, current and kernel function to set the stage. Here we choose to derive the coefficient of the $a^2\ln a^2$ term for the discrete Adler function for the conserved current (J =V) using: 1) the traditional kernel function k_0 with Q and 2) the kernel function k_1 containing the lattice momentum $\hat{Q}$, because this example is particularly simple and instructive. Any other choice can be obtained along the same lines. Here we restrict ourselves to the massless limit, $m \to 0$. Performing the calculation

in the massive case one obtains the same logarithmic cutoff effect, albeit with an additional contribution that can be absorbed in the pure a^2 term.

For this specific example J =V the current correlator takes the form

$$C^{\mathrm{V}}(t,a) = \frac{N_c q_f^2}{3} \sum_i \int_{-\pi/a}^{\pi/a} \frac{d\vec{p}}{(2\pi)^3} \frac{\cos(ap_i)^2 e^{-2Et}}{4\dot{E}^2} \left[\hat{E}^2 - \hat{p}_i^2 \left(1 + (-1)^t\right)\right] , \tag{B.45}$$

and the kernel functions read

$$\begin{aligned} k_0^{\Delta\Pi}(Q,Q_R,t) &= \left(\left[\frac{\cos(Qt)-1}{Q^2}\right] - [Q \to Q_R]\right) , \\ k_1^{\Delta\Pi}(Q,Q_R,t) &= \left(\left[\frac{\cos(Qt)-1}{\hat{Q}^2}\right] - [Q \to Q_R]\right) , \end{aligned} \tag{B.46}$$

with $\hat{Q} = (2/a)\sin(aQ/2)$. In the following we set $N_c = 3$. First, let us expand the current correlator in powers of a. By carrying out the sum and the angular integrals we obtain,

$$G(t)\big|_{a^0} = -\frac{q_f^2}{4\pi^2} \int_0^{2\pi\ a} \mathrm{d}p\ p^2\ e^{-pt} \tag{B.47}$$

$$G(t)\big|_{a^2} = \frac{q_f^2}{16\pi^2} \int_0^{2\pi/a} \mathrm{d}p\ p^4\ e^{-pt} \left\{1 - \frac{4}{15}pt\right\} \tag{B.48}$$

$$G^V(t)\big|_{a^2} = \frac{q_f^2}{80\pi^2} \int_0^{2\pi/a} \mathrm{d}p\ p^4\ e^{-pt} . \tag{B.49}$$

Similarly, expanding $k_1^{\Delta\Pi}(Q,Q_R,t,a)$ for $a \to 0$ yields

$$k_1^{\Delta\Pi}(Q,Q_R,t,a) \xrightarrow{a\to 0} \left[\frac{\cos(Qt)-1}{Q^2} + \left\{\frac{1}{12}(\cos(Qt)-1)\right\} a^2\right] - [Q \to Q_R] + \mathcal{O}(a^4) . \tag{B.50}$$

The integrals in t are easily carried out. For the cutoff effects common to all staggered currents one finds

$$\Delta\Pi_0(Q,Q_R,t,a)\big|_{a^2}^{\mathrm{comm}} = \frac{q_f^2}{48\pi^2} \int_0^{2\pi/a} \mathrm{d}p\ p^4 \frac{3p^2 + 11Q^2}{15p(p^2+Q^2)^2} , \tag{B.51}$$

whereas the cutoff coefficient specific to current V is

$$\Delta\Pi_0(Q,Q_R,t,a)\big|_{a^2}^{\mathrm{V}} = \frac{q_f^2}{40\pi^2} \int_0^{2\pi/a} \mathrm{d}p\ \frac{p^4}{p(p^2+Q^2)} . \tag{B.52}$$

and is hence logarithmically divergent as $a \to 0$. However, this term still contains subdominant, non-logarithmic terms. To obtain the logarithmically-enhanced cutoff effect we only keep the dominant terms. Hence, at $\mathcal{O}(a^2)$ for $\Delta\Pi_0(Q,Q_R,t,a)$ (i.e. with kernel k_0) we find

$$\Delta\Pi_0(Q,Q_R,t,a)\big|_{a^2} \xrightarrow[m\to 0]{a\to 0} \frac{q_f^2}{\pi^2}\left(-\frac{1}{48} + \frac{1}{80}\right)\left[Q^2 \ln\left(\frac{aQ}{2\pi}\right)^2 - Q_R^2 \ln\left(\frac{aQ_R}{2\pi}\right)^2\right] . \tag{B.53}$$

Including also the cutoff effect originating from the kernel function k_1 we obtain

$$\Delta\Pi_1(Q,Q_R,t,a)\big|_{a^2}^{\mathrm{comm}} = \frac{q_f^2}{48\pi^2} \int_0^{2\pi/a} \mathrm{d}p\ \frac{2p^3Q^2}{(p^2+Q^2)^2} , \tag{B.54}$$

and hence

$$\Delta\Pi_1(Q,Q_R,t,a)\big|_{a^2} \xrightarrow[m\to 0]{a\to 0} \frac{q_f^2}{\pi^2}Q^2\left(-\frac{1}{48}+\frac{1}{80}+\frac{1}{48}\right)\left[Q^2\ln\left(\frac{aQ}{2\pi}\right)^2 - Q_R^2\ln\left(\frac{aQ_R}{2\pi}\right)^2\right], \tag{B.55}$$

where the first fraction in the paranthesis is common to all staggered current correlators, the second is specific to the conserved current and the last originates from the kernel function. Interestingly, the leading-order coefficient in front of the $\mathcal{O}(a^2)$ logarithmically-enhanced cutoff effect for $\Delta\Pi_1$ has an opposite sign to the one for $\Delta\Pi_0$, as now also $k_1(Q,t)$ receives $\mathcal{O}(a^2)$ corrections.

Notice that for instance O_1^{IV}, i.e. the local staggered current IV together with the kernel k_1 containing the lattice momentum $\hat{Q}$, is absent of any $a^2\ln(a^2)$ enhanced artefacts at LO in lattice perturbation theory. They show however up at $\mathcal{O}(a^4)$ but are small enough to be neglected.

We conclude this appendix by a brief remark concerning higher order logarithmic terms in the expansion. As already mentioned, by expanding the kernel function k_1 to higher orders in a, it follows trivially that a whole tower of $(aQ)^{2n}\ln(aQ)^2$, $n\geq 1$, terms is generated. Here we would like to argue that we expect the same to happen for k_0 due to higher order terms in the expansion of the correlator. Expanding the kernel functions to higher orders in t, one obtains terms $\sim (Qt)^{2(n+1)}/Q^2$ which combined with terms $\alpha_n(a/t)^{2n}$ from the expansion of the correlator are logarithmic in a. For instance, for $n=2$,

$$C(t) \underset{a\ll t}{\overset{t\to 0}{\sim}} \frac{1}{t^3}\left(\alpha_0+\alpha_1\frac{a^2}{t^2}+\alpha_2\frac{a^4}{t^4}+\cdots\right) \tag{B.56}$$

$$O\left(Q^2,a\right)\Big|_{a^4} \sim \int_a \mathrm{d}t\; Q^4t^6C(t) \sim a^4Q^4\int_a \mathrm{d}t\;\frac{t^6}{t^7} \sim a^4Q^4\ln a^2\,. \tag{B.57}$$

C Continuum results in leading order perturbation theory

C.1 Current correlator

Let us first look at the leading-order result in the continuum. The propagator can be calculated in the time momentum representation using contour integration. We find for the zero-momentum projected current correlator of a quark of flavour f with charge $Q_f = eq_f$

$$\begin{aligned} C(t) &= \frac{q_f^2N_c}{(2\pi)^3}\frac{2}{3}\int d\vec{p}\;\sum_i \frac{e^{-2Et}}{E^2}(E^2-p_i^2) \\ &= \frac{1}{\pi^2}\frac{N_c}{3}q_f^2\int_0^\infty dr\; r^2(2r^2+3m_f^2)\frac{e^{-2\sqrt{r^2+m_f^2}t}}{r^2+m_f^2}\,, \end{aligned} \tag{C.1}$$

with mass m_f, electric charge q_f and N_c number of colours.

C.2 HVP and (discrete) Adler function

For the Adler function of a quark of flavour f with charge $Q_f = eq_f$

$$\begin{aligned} D(Q^2) &= 24\pi^2 \int_0^\infty \mathrm{d}t \left(-\frac{t\,\sin(Qt)}{2Q} - \frac{\cos(Qt)-1}{Q^2} \right) \frac{1}{3} \sum_{i=1}^{3} C_{ii}(t) \\ &= N_c q_f^2 \left(-\frac{6m_f^2}{Q^2} - \frac{24 m_f^4 \log\left(\frac{1}{y} - \sqrt{\frac{1}{y^2}-1}\right)}{Q^3\sqrt{4m_f^2+Q^2}} + 1 \right), \end{aligned} \tag{C.2}$$

where $y = 2m_f/\sqrt{4m_f^2+Q^2}$. In the massless case $D(Q^2) \xrightarrow{m_f \to 0} N_c q_f^2$ which counts the number of degrees of freedom of a quark of flavour f multiplied by its electric charge. For the HVP we obtain

$$\hat{\Pi}(Q^2) = \frac{N_c q_f^2}{36\pi^2 Q^3} \left(6(2m_f^2 - Q^2)\sqrt{4m_f^2+Q^2} \log\left(\frac{1}{y} - \sqrt{\frac{1}{y^2}-1} \right) + 12 m_f^2 Q - 5Q^3 \right). \tag{C.3}$$

The discrete Adler function follows trivially from Eq. (C.3).

D Removal of an additional discretization error

Discretization errors for our short distance observables take the form

$$O\left(Q^2, a\right) = O\left(Q^2\right) \left\{ 1 + \Gamma (aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \mathcal{O}\left((aQ)^2\right) \right\}. \tag{D.1}$$

$O(Q^2)$ is the observable in the continuum and can itself be expanded in powers of $\alpha_s(Q^2)$ and Γ receives $\alpha_s(1/a)$ corrections:

$$\Gamma = \Gamma_{\mathrm{LO}} + \Gamma_{\mathrm{NLO}} \alpha_s\left(\frac{1}{a}\right) + \mathcal{O}\left(\alpha_s^2\right) \tag{D.2}$$

$$O\left(Q^2\right) = O_{\mathrm{LO}}\left(Q^2\right) + O_{\mathrm{NLO}}\left(Q^2\right) \alpha_s\left(Q^2\right) + \mathcal{O}\left(\alpha_s^2\right). \tag{D.3}$$

Expanding equation (D.1) in powers of $\alpha_s(Q^2)$ and $\alpha_s(1/a)$ we obtain

$$\begin{aligned} O\left(Q^2, a\right) = &\underbrace{O_{\mathrm{LO}}\left(Q^2\right) \left\{ 1 + \Gamma_{\mathrm{LO}} (aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \mathcal{O}(a^2,\ \text{no } \alpha_s) \right\}}_{O_{\mathrm{LO}}(Q^2, a^2)} \\ &\times \left\{ 1 + \frac{O_{\mathrm{NLO}}}{O_{\mathrm{LO}}}\left(Q^2\right) \alpha_s\left(Q^2\right) + \ldots \right\} \left\{ 1 + \Gamma_{\mathrm{NLO}} \alpha_s\left(\frac{1}{a}\right) (aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \ldots \right\} \end{aligned} \tag{D.4}$$

with $O_{\mathrm{LO}}(Q^2)$ and $O_{\mathrm{NLO}}(Q^2)$ the LO and NLO results in the continuum, respectively. $O_{\mathrm{LO}}(Q^2, a)$ is the result for the observable in LO lattice perturbation theory (LPT). Adding the LO continuum result and removing the result from LO LPT we remove all discretization errors at LO. We

are hence left with the following, improved $\bar{O}(Q^2, a)$

$$\begin{aligned}\bar{O}\left(Q^2,a\right) &\equiv O\left(Q^2,a\right) - \left(O_{\mathrm{LO}}\left(Q^2,a\right) - O_{\mathrm{LO}}\left(Q^2\right)\right) \\ &= O\left(Q^2\right) + O_{\mathrm{LO}}\left(Q^2\right)\left\{(aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 \left[\Gamma_{\mathrm{LO}} \frac{O_{\mathrm{NLO}}}{O_{\mathrm{LO}}}\left(Q^2\right)\alpha_s\left(Q^2\right) + \Gamma_{\mathrm{NLO}}\alpha_s\left(\frac{1}{a}\right)\right] + \ldots\right\}\end{aligned} \tag{D.5}$$

Since Γ_{LO} and $O_{\mathrm{NLO}}\left(Q^2\right)$ are known, we can define an additionally subtracted $O\left(Q^2,a\right)$ that will have yet formally smaller discretization errors,

$$\begin{aligned}\tilde{O}\left(Q^2,a\right) &= \bar{O}\left(Q^2,a\right) - (aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 \Gamma_{\mathrm{LO}} O_{\mathrm{NLO}}\left(Q^2\right)\alpha_s\left(Q^2\right) \\ &= O\left(Q^2\right) + \Gamma_{\mathrm{NLO}} O_{\mathrm{LO}}\left(Q^2\right)\alpha_s\left(\frac{1}{a}\right)(aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \ldots \\ &= O\left(Q^2\right)\left\{1 + A\left(Q^2\right)\alpha_s\left(\frac{1}{a}\right)(aQ)^2 \ln\left(\frac{aQ}{2\pi}\right)^2 + \ldots\right\}\end{aligned} \tag{D.6}$$

with

$$A\left(Q^2\right) = \Gamma_{\mathrm{NLO}} \frac{O_{\mathrm{LO}}\left(Q^2\right)}{O\left(Q^2\right)} = \Gamma_{\mathrm{NLO}}\left\{1 - \alpha_s\left(Q^2\right)\frac{O_{\mathrm{NLO}}}{O_{\mathrm{LO}}}\left(Q^2\right) + \mathcal{O}\left(\alpha_s^2\left(Q^2\right)\right)\right\} \tag{D.7}$$

so that the Q^2-dependence of $A\left(Q^2\right)$ only appears as a higher order in α_s in the expansion of $\bar{O}\left(Q^2,a\right)$. Now the one-loop running of α_s is given by

$$\alpha_s\left(\frac{1}{a}\right) = \frac{4\pi}{b_0 \ln(1/a\Lambda)^2} \tag{D.8}$$

with $b_0 = 11 - 2/3 n_f$ (see Eq. (2.58)). Hence,

$$\alpha_s\left(\frac{1}{a}\right)\ln(aQ)^2 \simeq \frac{4\pi\left\{\ln Q^2/\Lambda^2 - \ln\left(\frac{1}{a\Lambda}\right)^2\right\}}{b_0 \ln(1/a\Lambda)^2} = \left(\frac{\alpha_s\left(\frac{1}{a}\right)}{\alpha_s\left(Q^2\right)} - 1\right)\frac{4\pi}{b_0} \tag{D.9}$$

and there are no longer any log enhanced discretization errors, except for α_s suppressed terms, such as

$$\sim (aQ)^2 \ln(aQ)^2 \left\{\Gamma_{\mathrm{LO}} D_{\mathrm{NNLO}}\left(Q^2\right)\alpha_s^2\left(Q^2\right)\right\}$$

and higher orders, that should be small. All the other terms are regular $(aQ)^2$, $(aQ)^2\alpha_s\left(\frac{1}{a}\right)$ or higher order terms. At this point, logarithmically-enhanced discretization errors can probably be ignored.

E Continuum extrapolations and distributions

We collect below the remaining plots of continuum extrapolations and cumulative and probability distribution functions which we did not show in Sec. 4.3.3.

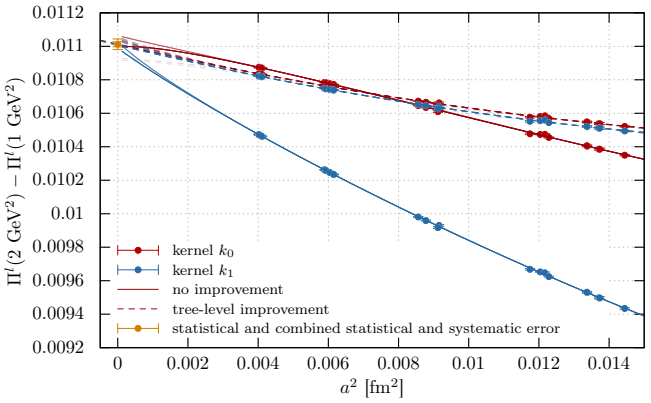


FIGURE E.1: Same as Fig. 4.14, but for $\Pi^l(2\,\mathrm{GeV}^2) - \Pi^l(1\,\mathrm{GeV}^2)$

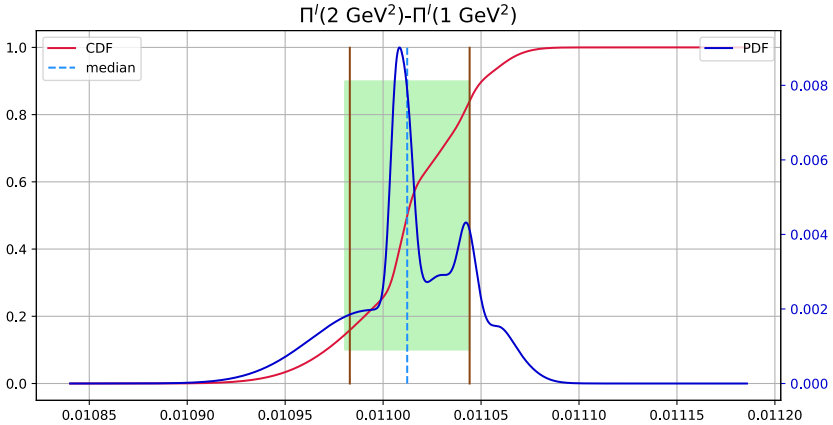


FIGURE E.2: Same as Fig. 4.17, but for $\Pi^l(2\,\mathrm{GeV}^2) - \Pi^l(1\,\mathrm{GeV}^2)$

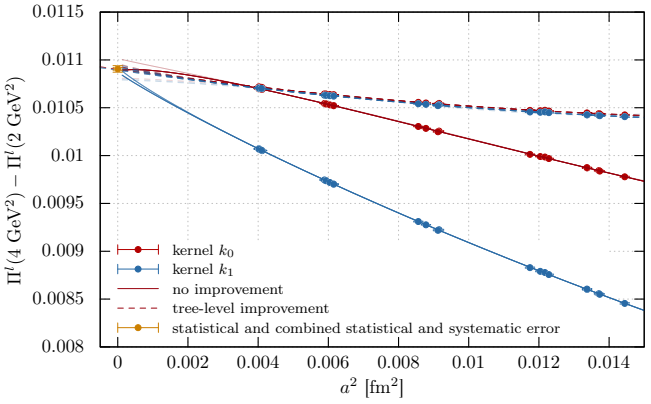


FIGURE E.3: Same as Fig. 4.14, but for $\Pi^l(4\,\mathrm{GeV}^2) - \Pi^l(2\,\mathrm{GeV}^2)$

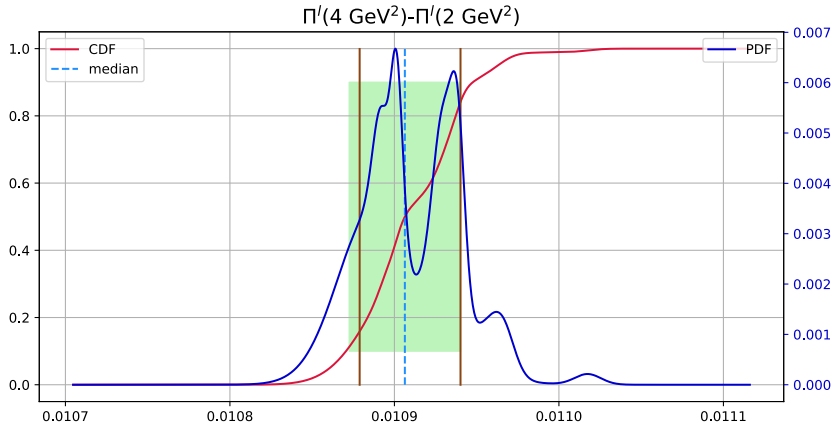


FIGURE E.4: Same as Fig. 4.17, but for $\Pi^l(4\,\mathrm{GeV}^2) - \Pi^l(2\,\mathrm{GeV}^2)$

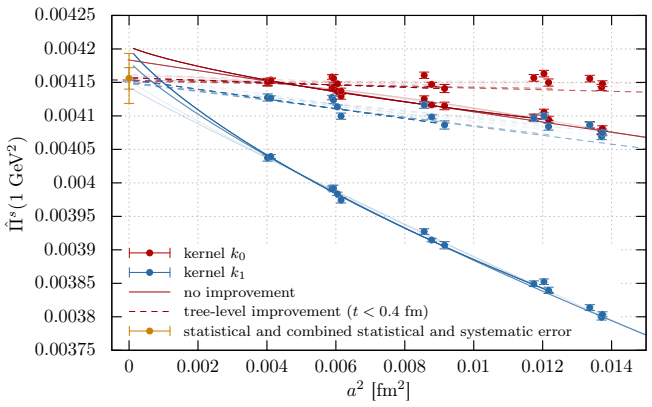


FIGURE E.5: Same as Fig. 4.14, but for $\hat{\Pi}^s(1\,\mathrm{GeV}^2)$

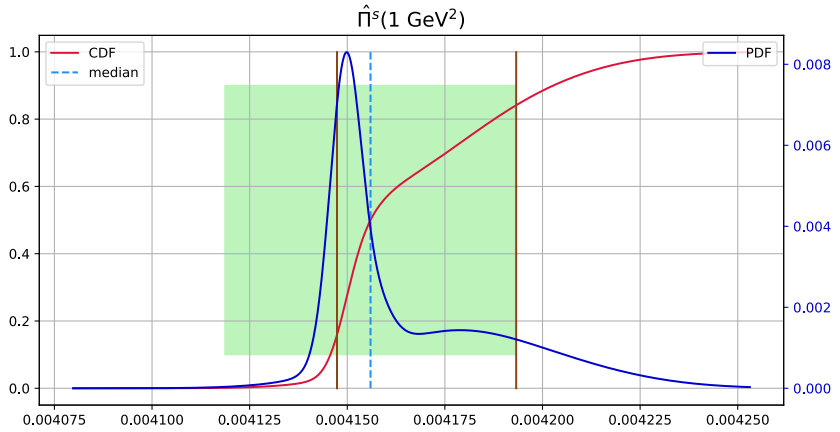


FIGURE E.6: Same as Fig. 4.17, but for $\hat{\Pi}^s(1\,\mathrm{GeV}^2)$

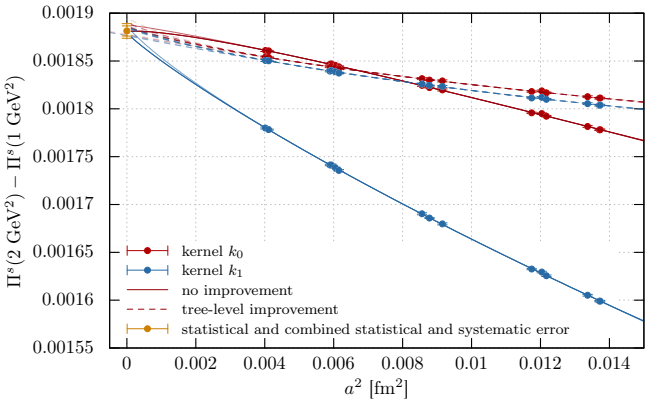


FIGURE E.7: Same as Fig. 4.14, but for $\Pi^s(2\,\mathrm{GeV}^2) - \Pi^s(1\,\mathrm{GeV}^2)$

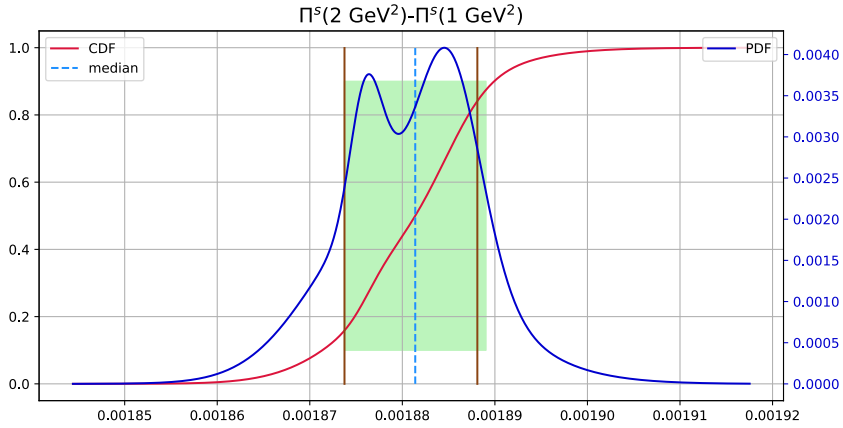


FIGURE E.8: Same as Fig. 4.17, but for $\Pi^s(2\,\mathrm{GeV}^2) - \Pi^s(1\,\mathrm{GeV}^2)$

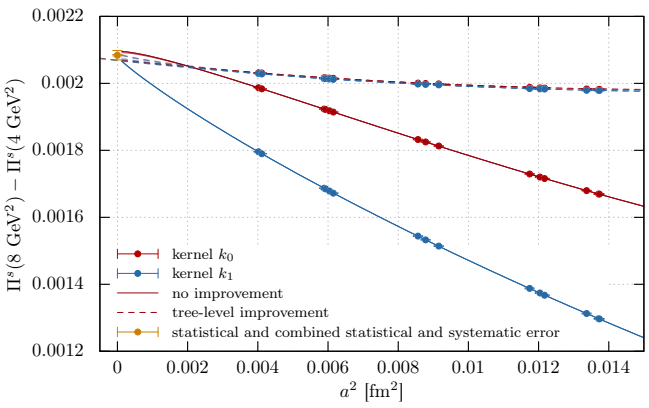


FIGURE E.9: Same as Fig. 4.14, but for $\Pi^s(8\,\mathrm{GeV}^2) - \Pi^s(4\,\mathrm{GeV}^2)$

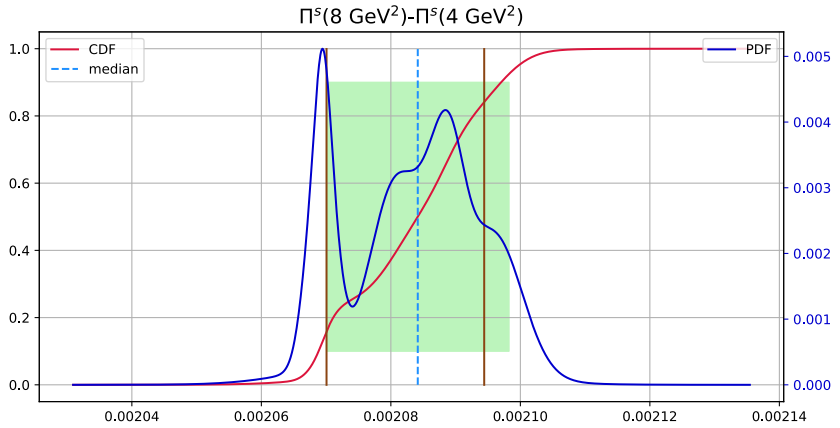


FIGURE E.10: Same as Fig. 4.17, but for $\Pi^s(8\,\mathrm{GeV}^2) - \Pi^s(4\,\mathrm{GeV}^2)$

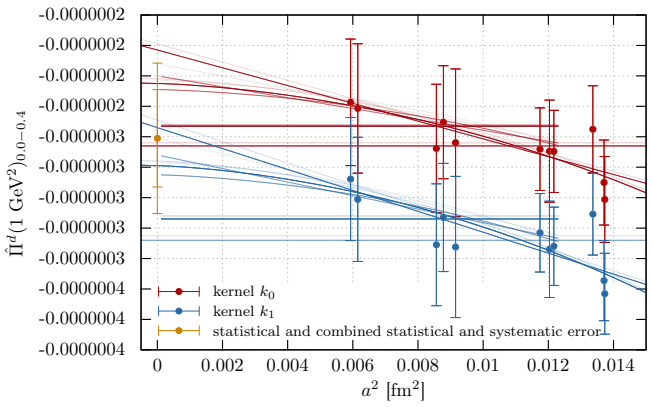


FIGURE E.11: Same as Fig. 4.14, but for $\hat{\Pi}^d(1\,\mathrm{GeV}^2, 0.0, 0.4)$

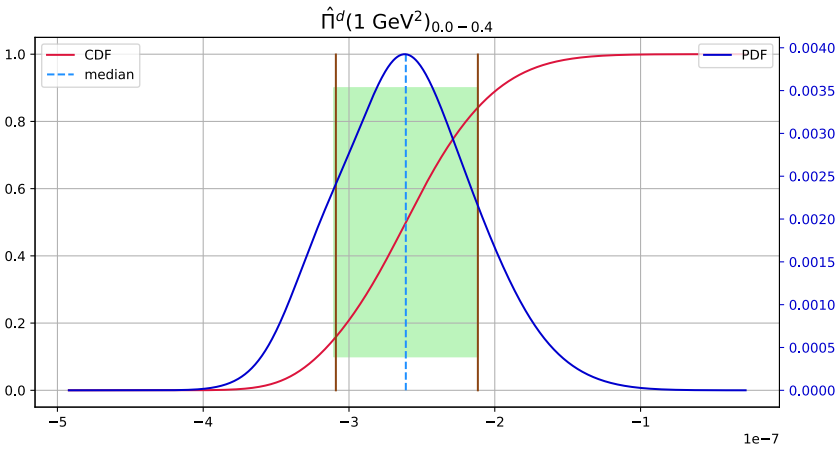


FIGURE E.12: Same as Fig. 4.17, but for $\hat{\Pi}^d(1\,\mathrm{GeV}^2, 0.4, \infty)$

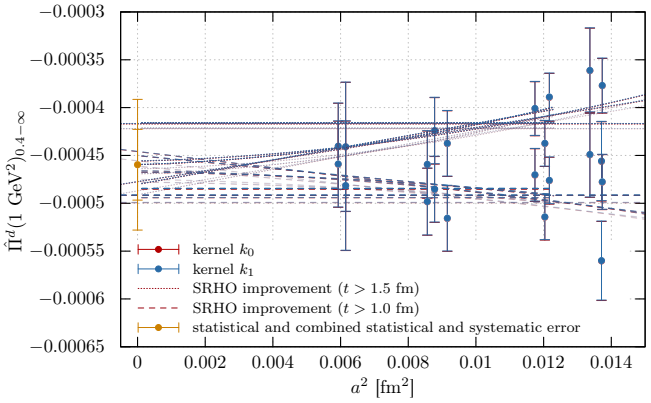


FIGURE E.13: Same as Fig. 4.14, but for $\hat{\Pi}^d(1\,\mathrm{GeV}^2, 0.4, \infty)$

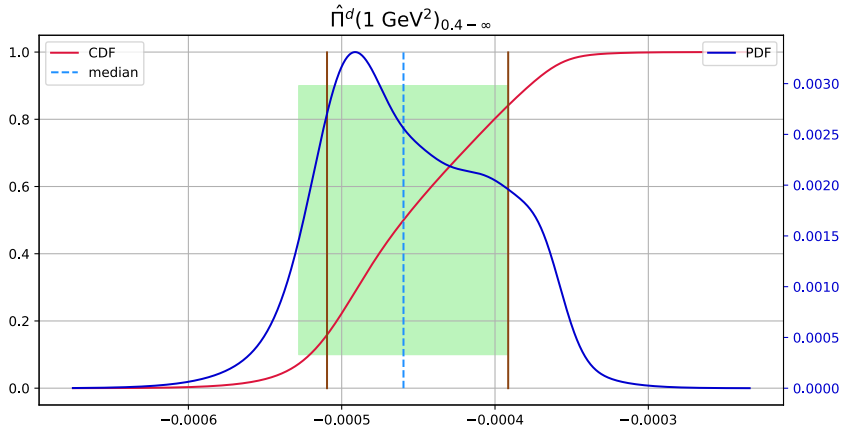


FIGURE E.14: Same as Fig. 4.17, but for $\hat{\Pi}^d(1\,\mathrm{GeV}^2, 0.4, \infty)$

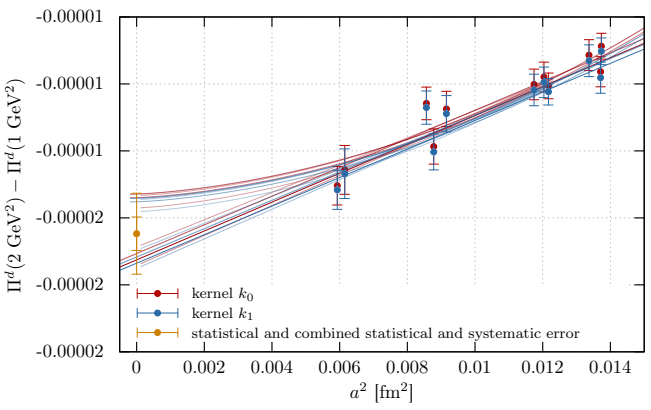


FIGURE E.15: Same as Fig. 4.14, but for $\Pi^d(2\,\mathrm{GeV}^2) - \Pi^d(1\,\mathrm{GeV}^2)$

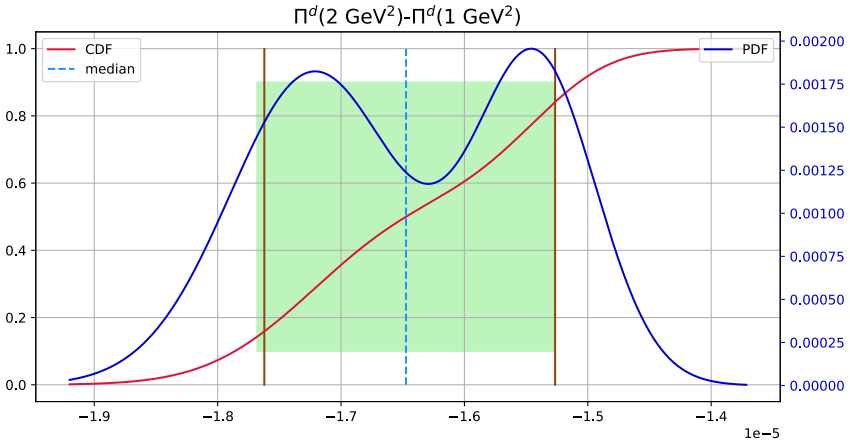


FIGURE E.16: Same as Fig. 4.17, but for $\Pi^d(2\,\mathrm{GeV}^2) - \Pi^d(1\,\mathrm{GeV}^2)$

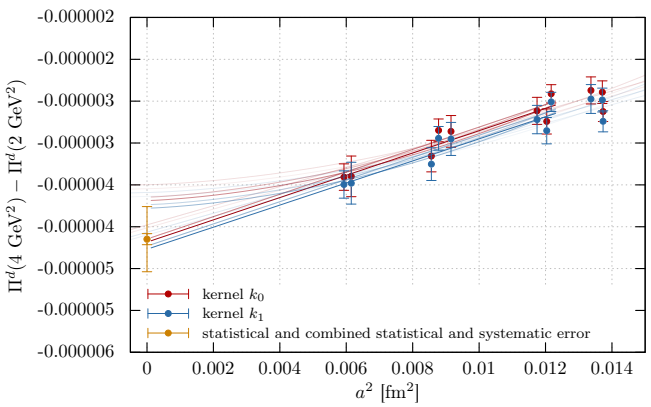


FIGURE E.17: Same as Fig. 4.14, but for $\Pi^d(4\,\mathrm{GeV}^2) - \Pi^d(2\,\mathrm{GeV}^2)$

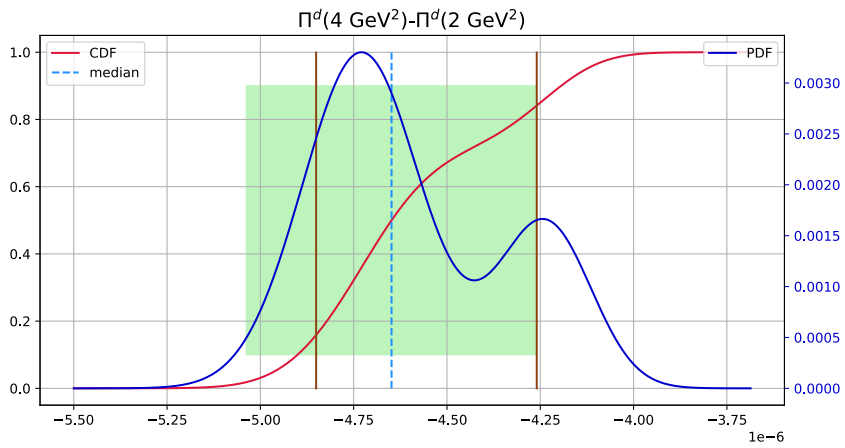


FIGURE E.18: Same as Fig. 4.17, but for $\Pi^d(4\,\mathrm{GeV}^2) - \Pi^d(2\,\mathrm{GeV}^2)$

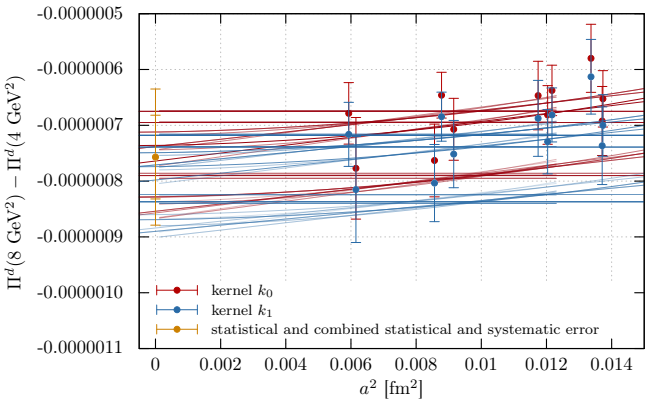


FIGURE E.19: Same as Fig. 4.14, but for $\Pi^d(8\,\mathrm{GeV}^2) - \Pi^d(4\,\mathrm{GeV}^2)$

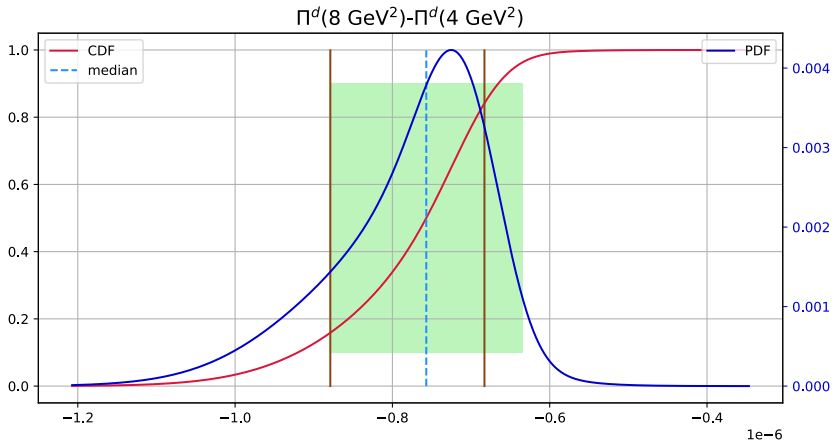


FIGURE E.20: Same as Fig. 4.17, but for $\Pi^d(8\,\mathrm{GeV}^2) - \Pi^d(4\,\mathrm{GeV}^2)$

Appendices: ALP mediated DM

F Effective ALP-DM Lagrangian from a two Higgs doublet model

The model we are considering contains an axion-like particle a and a fermionic dark matter candidate χ. We assume that a is the pseudo-Goldstone boson of an approximate global symmetry $\mathrm{U}(1)_{\mathrm{PQ}}$ which is spontaneously broken at some high scale f_a (and anomalous with respect to the standard model gauge group). A second crucial assumption is that any new fermions with standard model charges should be vector-like under both $\mathrm{U}(1)_{\mathrm{PQ}}$ and the standard model, such that any effective $aF\tilde{F}$ couplings at low energies are induced by standard model fermions only. This class of models includes e.g. the Weinberg-Wilczek axion and the DFSZ axion (and more generally all models in which an axion-like particle arises from an extended Higgs sector with no additional fermions) but not the KSVZ axion. We moreover assume that χ is massive and charged under $\mathrm{U}(1)_{\mathrm{PQ}}$. It could obtain its mass, for instance, from a coupling to a standard model singlet scalar ϕ whose vacuum expectation value $\langle\phi\rangle = f_a$ breaks $\mathrm{U}(1)_{\mathrm{PQ}}$, as contained in the DFSZ model.

The effective Lagrangian Eq. (7.3) can be UV completed by a renormalizable model containing additional Higgs doublets and singlets, with the flavour-diagonal couplings of the ALP being a consequence of the Higgs coupling structure. Here we present a simple example for pedagogical purposes.

Our starting point is the Lagrangian of a type-I two Higgs doublet model with an extra scalar singlet,

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\mathrm{kin}} &- \left(y^d_{ij}\bar{Q}_{Li}\Phi_2 D_{Rj} + y^e_{ij}\bar{L}_{Li}\Phi_2 E_R + y^u_{ij}\bar{Q}_{Li}\widetilde{\Phi}_2 U_{Rj} + \text{h.c.}\right) - \tilde{m}_1^2\,|\Phi_1|^2 \\ &- \tilde{m}_2^2\,|\Phi_2|^2 + M^2\,|\phi|^2 - \frac{\tilde{\lambda}_1}{2}|\Phi_1|^4 - \frac{\tilde{\lambda}_2}{2}|\Phi_2|^4 - \frac{\lambda_\phi}{2}|\phi|^4 - \tilde{\lambda}_3\,|\Phi_1|^2\,|\Phi_2|^2 \\ &- \tilde{\lambda}_4\,(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) - \kappa_1\,|\Phi_1|^2\,|\phi|^2 - \kappa_2\,|\Phi_2|^2|\phi|^2 - \tilde{\lambda}_{12}\left(\Phi_1^\dagger\Phi_2\phi^2 + \text{h.c.}\right).\end{aligned} \tag{F.1}$$

Here $\mathcal{L}_{\mathrm{kin}}$ contains all the gauge-kinetic terms for the SM gauge and fermion fields as well as for the scalars Φ_1, Φ_2 and ϕ. Φ_1 and Φ_2 are two Higgs doublets with the gauge quantum numbers of the SM Higgs, and ϕ is a complex scalar which is neutral under the SM. This is the most general renormalizable Lagrangian allowed by a $\mathrm{U}(1)_{\mathrm{PQ}}$ global symmetry under which Φ_2 and the SM fermions are neutral, while Φ_1 and ϕ are charged. Note that this symmetry is not a PQ

symmetry in the stricter sense of being anomalous with respect to QCD, and therefore unsuitable for solving the strong CP problem.

For $M^2 > 0$, ϕ will take a VEV $\langle\phi\rangle = f_a/\sqrt{2}$, which can be taken real and positive. We define real scalar fields σ and a by

$$\phi(x) = \frac{1}{\sqrt{2}}\big(f_a + \sigma(x)\big)\, e^{i\frac{a(x)}{f_a}} \,. \tag{F.2}$$

We anticipate that Φ_1 and Φ_2 will also take vacuum expectation values of the order $v \ll f_a$. This parameterization is useful to extract leading-order effects in v/f_a; to leading order we have $f_a = \sqrt{\frac{2}{\lambda_\phi}}M$. The fields σ and a are approximate mass eigenstates, up to mixing with the Φ-like fields with mixing angles of order v/f_a.

At low energies, σ (whose VEV is zero to leading order) can be integrated out. The effective $\Phi_{1,2}$ mass parameters and quartic couplings at low energies will be modified according to suitable matching conditions from replacing $|\phi|^2 \to f_a^2/2$ in the above Lagrangian and from four-point interactions with σ exchange, such that $\tilde{m}_i^2 \to m_i^2$ and $\tilde{\lambda}_k \to \lambda_k$. Up to higher-dimensional operators, the effective Lagrangian is now that of a type-I two-Higgs doublet model with an additional field $a(x)$,

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{kin}} &- \Big(y_{ij}^d \bar{Q}_{Li}\Phi_2 D_{Rj} + y_{ij}^e \bar{L}_{Li}\Phi_2 E_R + y_{ij}^u \bar{Q}_{Li}\widetilde{\Phi}_2 U_{Rj} + \text{h.c.}\Big) - m_1^2\,|\Phi_1|^2 \\ &- m_2^2\,|\Phi_2|^2 - \frac{\lambda_1}{2}|\Phi_1|^4 - \frac{\lambda_2}{2}|\Phi_2|^4 - \lambda_3\,|\Phi_1|^2\,|\Phi_2|^2 \\ &- \lambda_4\,(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) - \frac{\lambda_{12}\,f_a^2}{2}\Big(\Phi_1^\dagger\Phi_2\, e^{2ia/f_a} + \text{h.c.}\Big)\,.\end{aligned} \tag{F.3}$$

With the a-dependent field redefinition $\Phi_1(x) \to e^{2ia(x)/f_a}\Phi_1(x)$, a disappears from the potential. However, the Φ_1 kinetic term is not invariant,

$$\big(D_\mu\Phi_1\big)^\dagger D^\mu\Phi_1 \to \big(D_\mu\Phi_1\big)^\dagger D^\mu\Phi_1 - 2i\frac{\partial_\mu a}{f_a}\,\Phi_1^\dagger \overleftrightarrow{D}^\mu\Phi_1 + \mathcal{O}\left(\frac{a^2}{f_a^2}\right).$$

The "Higgs basis" in the space of Φ_1 and Φ_2 is defined by rotating $(\Phi_1, \Phi_2) \to (H, \Phi)$ by an angle β such that the Higgs VEV is contained in one doublet H only, $\tan\beta = \langle\Phi_2\rangle/\langle\Phi_1\rangle$. In general, β is independent of the mixing angle α which parameterizes the mixing between the mass eigenstates. However, in the decoupling limit, these two angles are aligned and the fields H and Φ contain the mass eigenstates as

$$H(x) = \begin{pmatrix} G^+(x) \\ \frac{v+h(x)+iG^0(x)}{\sqrt{2}} \end{pmatrix}, \qquad \Phi(x) = \begin{pmatrix} H^+(x) \\ \frac{H^0(x)+iA^0(x)}{\sqrt{2}} \end{pmatrix}. \tag{F.4}$$

We assume that the Φ-like states are also heavy, so they can be integrated out as well. (This step can in principle be interchanged with integrating out σ or performed simultaneously, depending on the mass hierarchies.) The effective Lagrangian becomes

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{kin}} &- \Big(Y_{ij}^d \bar{Q}_{Li} H D_{Rj} + Y_{ij}^e \bar{L}_{Li} H E_{Rj} + Y_{ij}^u \bar{Q}_{Li}\widetilde{H} U_{Rj} + \text{h.c.}\Big) \\ &+ m^2\,|H|^2 - \frac{\lambda}{2}|H|^4 - 2i\frac{\partial_\mu a}{f_a}\,\cos^2\beta\, H^\dagger \overleftrightarrow{D}^\mu H\end{aligned} \tag{F.5}$$

where $Y^{u,d,e} = y^{u,d,e} \sin\beta$.

Commonly, one then transforms to the "Georgi-Kaplan-Randall basis" in which a only couples derivatively, up to $aF\widetilde{F}$ terms. However, as discussed in the note on anomalies 6.3 these $aF\widetilde{F}$ terms are spurious and, in the end, need to be transformed away again. For completeness we explicit what one would do. By performing another a-dependent field rephasing of the fermions fields according to their PQ charges as

$$\begin{aligned} &Q_L(x) \to e^{ia(x)/(2f_a)} Q_L(x), \quad L_L(x) \to e^{ia(x)/(2f_a)} L_L(x), \\ &D_R(x) \to D_R(x) e^{-ia(x)/(2f_a)}, \quad E_R(x) \to E_R(x) e^{-ia(x)/(2f_a)}, \quad U_R(x) \to U_R(x) e^{-ia(x)/(2f_a)}\,, \end{aligned}$$

the axion disappears from the Yukawa couplings. The fermionic kinetic terms are not invariant, hence the transformation induces the operators

$$\mathcal{L} \supset -\frac{1}{2}\frac{\partial_\mu a}{f_a}\left(\bar{Q}_L\gamma^\mu Q_L + \bar{L}_L\gamma^\mu L_L - \bar{D}_R\gamma^\mu D_R - \bar{E}_R\gamma^\mu E_R - \bar{U}_R\gamma^\mu U_R\right)$$

Since the transformation is anomalous the path integral measure is not invariant and the Jacobian has to be included, giving rise to $a/f_a F\widetilde{F}$ terms in the Lagrangian. The advantage of this basis is that the Goldstone nature of a is manifest: it couples only derivatively and is the only field transforming under the non-linearly realized PQ symmetry, $a \to a + \theta$.

Instead of transforming to the GKR basis, we keep the Lagrangian of Eq. (F.5). The $(\partial a)\, H^\dagger \overleftrightarrow{D} H$ term would cause a to mix with the Z boson after electroweak symmetry breaking, which is avoided by another field redefinition:

$$H(x) \;\to\; H(x) e^{-2i\cos^2\beta\, a(x)/f_a}\,. \tag{F.6}$$

The Yukawa terms will shift accordingly, and one obtains for the final axion couplings in this field basis at leading order in $1/f_a$

$$\begin{aligned} \mathcal{L} = \mathcal{L}_{\text{kin}} + m^2|H|^2 - \frac{\lambda}{2}|H|^4 - \left(Y^d_{ij}\,\bar{Q}_{Li} H D_{Rj} + Y^e_{ij}\,\bar{L}_{Li} H E_{Rj} + Y^u_{ij}\,\bar{Q}_{Li}\widetilde{H} U_{Rj} + \text{h.c.}\right) \\ - 2\cos^2\beta\,\frac{a}{f_a}\left(i\,Y^d_{ij}\,\bar{Q}_{Li} H D_{Rj} + i\,Y^e_{ij}\,\bar{L}_{Li} H E_{Rj} + i\,Y^u_{ij}\,\bar{Q}_{Li}\widetilde{H} U_{Rj} + \text{h.c.}\right). \end{aligned} \tag{F.7}$$

The second line corresponds to the ALP-fermion couplings in Eq. (7.3) with $C_f = \cos^2\beta$, after switching to four-spinor notation.

Other models where an ALP is obtained by more general extensions of the Higgs sector will induce different axion-fermion couplings. For example, choosing a type-II two-Higgs doublet model (as in the standard DFSZ model for QCD axions [351]) would yield different couplings for up-type and down-type quarks and leptons, unless $\tan\beta = 1$.

As we saw above, it is possible to transform the dimension-5 couplings in Eq. (F.7) into $\partial_\mu a\, j^\mu_{\text{PQ}}$ terms, with $aF\widetilde{F}$ couplings appearing due to the anomaly. These are usually eliminated again after electroweak symmetry breaking, using the equations of motion to transform the $\partial_\mu a\, j^\mu_{\text{PQ}}$ terms into the Yukawa-like couplings of Eq. (7.3), which involves precisely the same anomaly term. We emphasize that additional $aF\widetilde{F}$ terms are never induced at high scales, unless there

are heavy fermions with charges under both PQ and the standard model (such as in the KSVZ model for a QCD axion).

To obtain our dark matter model, we now add two more features. The first is a mass term for a which explicitly breaks the PQ symmetry; we assume $m_a \ll f_a$ and write

$$\Delta\mathcal{L}_{\text{mass}} = \frac{m_a^2}{4}\left(\phi - \phi^*\right)^2 . \tag{F.8}$$

The second is the dark matter candidate χ, for which there are several possibilities. In this paper we focus on a Dirac dark matter candidate $\chi = \chi_L + \chi_R$ with PQ charges allowing for a coupling

$$\Delta\mathcal{L}_\chi = -\left(y_\chi\, \phi \overline{\chi}_L \chi_R + \text{h.c.}\right) \tag{F.9}$$

which, after PQ breaking, gives rise to a Dirac mass $m_\chi = y_\chi f_a/\sqrt{2}$ and an axionic coupling:

$$\Delta\mathcal{L}_{\text{eff}} = -m_\chi\, \overline{\chi}\chi - i\,\frac{m_\chi}{f_a}\, a\, \overline{\chi}\gamma^5\chi\,. \tag{F.10}$$

In this simple model, the reduced ALP-fermion couplings g_{aff} tend to be of the same order of magnitude as the reduced ALP-DM couplings $g_{a\chi\chi}$, unless $\cos\beta \ll 1$. As we show in the main text, hierarchically different couplings can lead to interesting effects. They are well-motivated in UV models with multiple axions and a clockwork-like structure for the couplings, leading to exponentially different effective decay constants for different sectors of the theory [442]. One might speculate, for instance, that the DM sector and the SM are localized on different lattice sites in theory space, giving rise to a large hierarchy $g_{aff} \ll g_{a\chi\chi}$.

Moreover, this model predicts $C_\chi = 1$ and therefore $g_{a\chi\chi} = \frac{1}{f_a}$. The DFO phase as discussed in sections 7.3.3 and 7.6.4 will therefore not be realized in this simple scenario, since DFO requires significant connector couplings $g_{a\chi\chi} \sim (100\text{ GeV})^{-1}$. A mass scale $f_a \sim 100$ GeV is of course ruled out by observation. We emphasize nevertheless that there is no obstacle in principle to raising C_χ, although doing so with the particle content of the present minimal model would require fine-tuning (by adding an additional explicitly PQ breaking term, namely a bare Dirac mass term for χ which could partially cancel the Dirac mass from spontaneous PQ breaking).

We finally remark that another possibility for the dark matter candidate would be a single Weyl dark matter candidate $\chi = \chi_L$, coupled to ϕ according to

$$\Delta\mathcal{L}_\chi = -\frac{1}{2}\left(y_\chi\, \phi \chi_L^T C \chi_L + \text{h.c.}\right) . \tag{F.11}$$

This coupling translates into a Majorana mass after PQ breaking. A third possibility would be a model with two Weyl fermions and a see-saw-like structure where both a Dirac mass and a Majorana mass are allowed, with the dark matter phenomenology depending on the hierarchy between the two.

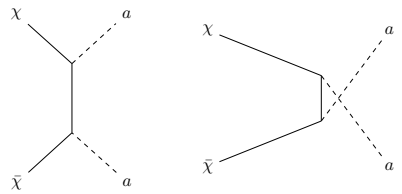
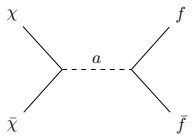
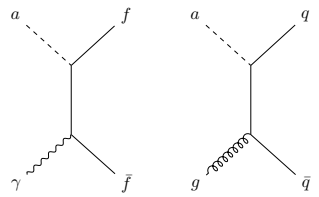
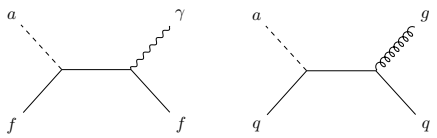
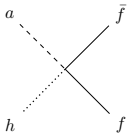
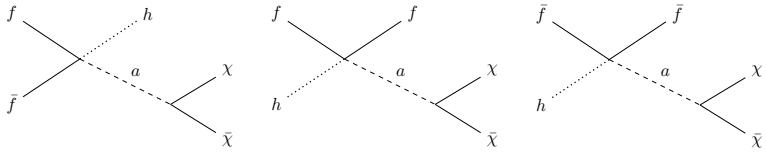
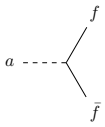
Process	Feynman-diagram
$\chi\bar{\chi} \to aa$ Eq. (G.1)	
$\chi\bar{\chi} \to f\bar{f}$ Eq. (G.2)	
$a\gamma \to f\bar{f}$, $ag \to q\bar{q}$ eqs. (G.4), (G.5)	
$af \to \gamma f$, $aq \to gq$ eqs. (G.6), (G.7)	
$ah \to f\bar{f}$ Eq. (G.8)	
$f\bar{f} \to h\chi\bar{\chi}$, $fh \to f\chi\bar{\chi}$, $\bar{f}h \to \bar{f}\chi\bar{\chi}$ Eq. (G.3)	
$a \to f\bar{f}$ Eq. (G.9)	

Table 7.4: Feynman-diagrams of the annihilation of the DM via the hidden sector and connector processes, as well as the co-annihilation of ALPs at tree level.

G Cross sections and input for Boltzmann equations

In Tab. 7.4 we collect the different annihilation processes for the DM and the ALPs which govern the Boltzmann equations in Eq. (7.7). In this section we provide expressions for the cross sections for the required $2 \to 2$ processes and the collision term for the relevant $2 \to 3$ processes.

G.1 Hidden sector process

For the annihilation of the DM particles to ALPs, i.e. annihilation in the hidden sector (for which the Feynman diagrams at tree level are shown in Tab. 7.4), the total unpolarized cross section in the centre of mass frame is given by

$$\begin{aligned}\sigma(\chi\bar{\chi}\to aa) \;=\; & \frac{(g_{a\chi\chi}m_\chi)^4}{16\pi s}\frac{\sqrt{s-4m_a^2}}{\sqrt{s-4m_\chi^2}}\Bigg(-2-\frac{m_a^4}{m_a^4-4m_a^2m_\chi^2+m_\chi^2 s} \\ & \frac{6m_a^4-4m_a^2 s+s^2}{(s-2m_a^2)\sqrt{(s-4m_a^2)(s-4m_\chi^2)}}\ln\frac{s-2m_a^2+\sqrt{(s-4m_a^2)(s-4m_\chi^2)}}{s-2m_a^2-\sqrt{(s-4m_a^2)(s-4m_\chi^2)}}\Bigg)\end{aligned} \tag{G.1}$$

G.2 Connector processes

The Feynman diagram of the $\chi\bar{\chi}\to f\bar{f}$ connector process is depicted in Tab. 7.4. For the total cross section in the centre of mass frame we find at tree level

$$\sigma(\chi\bar{\chi}\to f\bar{f}) = \frac{(g_{a\chi\chi}m_\chi)^2(g_{aff}m_f)^2 n_f^c s\sqrt{s-4m_f^2}}{16\pi\left(m_a^2-s\right)^2\sqrt{s-4m_\chi^2}}\,, \tag{G.2}$$

with n_f^c the number of colour degrees of freedom of the fermion. For reheating temperatures above a few hundred GeV contributions from the ultraviolet-dominant $2\to 3$ processes $f\bar{f}\to h\chi\bar{\chi}$, $fh\to f\chi\bar{\chi}$ and $\bar{f}h\to\bar{f}\chi\bar{\chi}$ become important. Following Ref. [341], but calculating the matrix element explicitly and including the final state masses one finds (the fact that not all involved particles are scalars does not change the integral because we neglect Fermi/Bose factors)

$$\begin{aligned}\dot{n}_\chi+3Hn_\chi = & \frac{(g_{a\chi\chi}m_\chi)^2(g_{aff}m_f)^2}{v^2}\frac{n_f^c\,T}{(4\pi)^7}\times \\ & \int_{s_{\min}}^{\infty}\mathrm{d}s\int_0^1\mathrm{d}x_2\int_{x_2}^1\mathrm{d}x_1\,s^{3/2}\,K_1\left(\frac{\sqrt{s}}{T}\right)\frac{4p_a^2 s}{\left(p_a^2-m_a^2\right)^2}\end{aligned} \tag{G.3}$$

with $p_a^2=(x_1-x_2)s+m_h^2$ and $s_{\min}=4m_f^2(T)$ for $f\bar{f}\to h\chi\bar{\chi}$ and $p_a^2=(x_1-x_2)s+m_f^2$ and $s_{\min}=(m_f(T)+m_h)^2$ for $fh\to f\chi\bar{\chi}$, $\bar{f}h\to\bar{f}\chi\bar{\chi}$.

G.3 SM-ALP processes

For the co-annihilation of ALPs to SM particles, the Feynman diagrams at tree level are shown in Tab. 7.4. Here we collect the results for the total unpolarized cross section in the centre of

mass frame:

$$\sigma(a\gamma \to f\bar{f}) = \frac{(g_{aff}m_f)^2\ \alpha_{\rm em}\ q_f^2\ n_f^c}{m_a^2 + m_\gamma^2 - s}\sqrt{\frac{\lambda(s, m_f^2, m_f^2)}{\lambda(s, m_a^2, m_\gamma^2)}}$$
$$\left[\frac{m_a^4 - 4m_a^2 m_f^2 + \left(m_\gamma^2 - s\right)^2}{\sqrt{\lambda(s, m_a^2, m_\gamma^2)(s(s-4m_f^2)}}\ln\left(\frac{\sqrt{s}\left(m_a^2+m_\gamma^2-s\right)+\sqrt{\lambda(s,m_a^2,m_\gamma^2)}\sqrt{s-4m_f^2}}{\sqrt{s}\left(m_a^2+m_\gamma^2-s\right)-\sqrt{\lambda(s,m_a^2,m_\gamma^2)}\sqrt{s-4m_f^2}}\right)\right.$$
$$\left. - \frac{m_a^2\left(m_\gamma^2 + 2m_f^2\right)\left(m_a^2 + m_\gamma^2 - s\right)}{(m_a^2 - m_\gamma^2)^2 m_f^2 + m_a^2 m_\gamma^2 s - 2(m_a^2 + m_\gamma^2)m_f^2 s + m_f^2 s^2}\right], \tag{G.4}$$

$$\sigma(ag \to q\bar{q}) = \frac{1}{6}\ \sigma(a\gamma \to f\bar{f})(\alpha_{\rm em} \to \alpha_s,\ q_f^2 \to 1,\ m_\gamma \to m_g)\ , \tag{G.5}$$

$$\sigma(af \to \gamma f) = \frac{(g_{aff}m_f)^2\ \alpha_{\rm em}\ q_f^2}{4s}\sqrt{\frac{\lambda(s, m_f^2, m_\gamma^2)}{\lambda(s, m_f^2, m_a^2)}}\int_{-1}^{1}\mathrm{d}\cos(\theta)\frac{1}{\left(s - m_f^2\right)^2\left(t - m_f^2\right)^2}$$
$$\left(m_a^2\left(m_\gamma^2\left(2m_f^2(s+t) - 2m_f^4 - s^2 - t^2\right) - 2\left(s + t - 2m_f^2\right)\left(st - m_f^4\right)\right)\right.$$
$$\left. + \left(s - m_f^2\right)\left(t - m_f^2\right)\left((s + t - 2m_f^2)^2 + 2m_a^4\right)\right), \tag{G.6}$$

$$\sigma(aq \to gq) = \frac{4}{3}\ \sigma(af \to \gamma f)(\alpha_{\rm em} \to \alpha_s,\ q_f^2 \to 1,\ m_\gamma \to m_g)\ , \tag{G.7}$$

with n_f^c the number of colour degrees of freedom of the fermion and electric charge $Q_f = eq_f$. Note that expressions (G.6) and (G.7) are equivalent to the ones for $\bar{f}a \to \bar{f}\gamma$ and $\bar{q}a \to \bar{q}g$.

For reheating temperatures above a few hundred GeV contributions from the ultraviolet-dominant process $ah \to f\bar{f}$ become important. In this case we find for the cross section in the centre of mass frame

$$\sigma(ah \to f\bar{f}) = \frac{(g_{aff}m_f)^2 n_f^c}{8\pi v^2}\sqrt{\frac{\lambda(s, m_f^2, mf^2)}{\lambda(s, m_a^2, m_h^2)}}\ . \tag{G.8}$$

G.4 Decay width of the ALP

The total decay width of the ALP plays an important role at various points in our work, particularly for the collider and cosmological constraints. For example, for the collider constraints it is important because the ALP lifetime determines whether it can be triggered on, or if it decays within the detector or outside (leading to invisible decay). At tree level, the ALP decays into leptons or quarks, however at loop level it may decay into photons or gluons. Our calculations for the leading partial decay widths are summarised below. In the perturbative regime the decay width of an ALP decaying to fermions is

$$\Gamma(a \to f\bar{f}) = \frac{(g_{aff}m_f)^2 n_f^c m_a}{8\pi}\sqrt{1 - \frac{4m_f^2}{m_a^2}}\ , \tag{G.9}$$

with n_f^c the number of colour degrees of freedom of the fermion. This expression is used for leptons and for decays into heavy quarks $c\bar{c}$ and $b\bar{b}$, provided m_a is above the threshold for the decay into a pair of D or B mesons respectively. For the contribution of the light quarks, we take into account the decays of ALPs to hadrons via

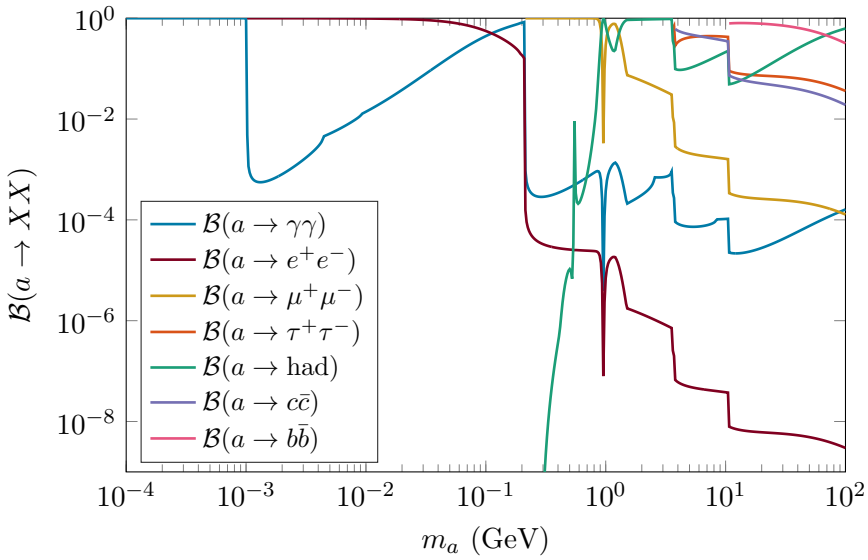


FIGURE G.1: Branching fractions of the ALP into photons, leptons and hadrons as indicated, as a function of the ALP mass.

- If the mass of the ALP is less than 1.2 GeV, the decay width is calculated using code obtained by private communication, created for Ref. [443]. This code was written for the decay of an NMSSM CP-odd Higgs, but with an appropriate choice of parameters ($P_{11} = \sqrt{2} v g_{aff}$, $\tan\beta = 1$ and decoupled neutralinos, charginos and heavy Higgs bosons) this particle can be identified with our ALP.

- Above 1.5 GeV, the decay width into light quarks is given by [57]

$$\Gamma(a \to \text{hadrons}) = g_{aff}^2 \frac{9 m_a^3 \alpha_s^2}{32\pi} \left(1 + \frac{83\alpha_s}{4\pi}\right), \tag{G.10}$$

  as in this regime the strong coupling is taken to be perturbative.

- Between 1.2 and 1.5 GeV we interpolate between the above determinations of the decay widths, in order to obtain a smooth result for the total decay width of the ALP, important particularly for the cosmological constraints.

This concludes the discussion of the tree-level decays of the ALP. In order to take into account experimental constraints on ALP decays to photons, we further require the decay width into two photons via a fermion loop. This is given by

$$\Gamma(a \to \gamma\gamma)_{\text{loop}} = \frac{\alpha^2 |T(m_a^2)|^2}{4\pi^3 m_a}, \tag{G.11}$$

where the triangle loop function is

$$T(s) = \sum_f g_{aff}\, m_f^2\, n_f^c\, q_f^2 \arcsin^2\left(\sqrt{s}/(2m_f)\right), \tag{G.12}$$

where $Q_f = e q_f$ is the electric charge of the fermion. In Fig. G.1, we provide the branching ratios for all the above mentioned decays.

G.5 DM self-interaction cross sections

For the momentum transfer cross section for particle particle scattering via a t- or u-channel exchange of an ALP we obtain

$$\sigma_T^{\chi\chi\to\chi\chi} = \frac{g_{a\chi\chi}^4 m_\chi^2 \left(5m_a^2 + 12m_\chi^2 v^2\right)}{512\pi m_\chi^2 (1+v^2)} \left(\frac{4m_\chi^2}{m_a^2 + 4m_\chi^2 v^2} + \frac{m_a^2 \log\left(\frac{m_a^2}{m_a^2+4m_\chi^2 v^2}\right)}{v^2\left(m_a^2 + 2m_\chi^2 v^2\right)} \right) , \tag{G.13}$$

while for particle-antiparticle scattering with an s- or t- channel ALP exchange we obtain for the momentum transfer cross section

$$\begin{aligned}\sigma_T^{\chi\bar{\chi}\to\chi\bar{\chi}} = & \frac{g_{a\chi\chi}^4}{512\pi E_*^2 v^4 \left(m_a^2 - 4E_*^2\right)^2} \Big(4m_\chi^4 v^4 \left(48E_*^4 - 12E_*^2 m_a^2 + m_a^4\right) \\ & + \Big(\left(-48E_*^2 m_\chi^2 v^2 + 6m_a^4 - 8m_a^2 m_\chi^2 \left(3v^2+4\right)\right) \\ & \times \log\left(m_a^2 + 2m_\chi^2 v^2\right) m_a^2 \left(m_a^2 - 4E_*^2\right) \\ & + \left(48E_*^2 m_\chi^2 v^2 - 3m_a^4 + 8m_a^2 m_\chi^2 \left(v^2+2\right)\right) \\ & \times \log\left(m_a^2 + 4m_\chi^2 v^2\right) + \left(32E_*^2 m_a^2 - 6m_a^4\right) \log(m_a) \Big) \Big) ,\end{aligned} \tag{G.14}$$

with $E_*^2 = m_\chi^2 (1+v^2)$.

H Axion production cross sections in SLAC E137 experiment

H.1 Primakoff production

The differential cross section for the production of ALPs via the Primakoff mechanism (see Fig. 7.1a) is given by Eq. (A1) in Ref. [377],

$$\frac{\mathrm{d}\sigma_Z^{\gamma\to a}}{\mathrm{d}\Omega} = 8\alpha_{\mathrm{em}} \frac{\Gamma(a\to\gamma\gamma)}{m_a^3} \mid F(t) \mid^2 \frac{\theta^2}{\left(\theta^2 + \frac{1}{4}\left(\frac{m_a}{E_\gamma}\right)^4\right)^2} , \tag{H.1}$$

where $\Gamma(a\to\gamma\gamma)$ is the loop induced ALP decay width to two photons (see app. J.1 for details), $t = (p_a - p_\gamma)^2$ is the momentum exchange with p_γ the four-momentum of the incoming electron and p_a the four-momentum of the outgoing ALP. Θ is the scattering polar angle in the laboratory frame. Z and $F(t)$ are the target atomic number and target form factor, respectively. For $t \leq 3\times 10^{-6}$ GeV^2 we use the atomic form factor, given by eqs. (A5) – (A6) in [377],

$$|F_{\mathrm{atomic}}(t)|^2 = Z^2 \left(\frac{a^2|t|}{a^2|t|+1}\right) + Z\left(\frac{a'^2|t|}{a'^2|t|+1}\right)^2 \tag{H.2}$$

with the parameters

$$a = \begin{cases} 122.8/m_e & \text{for hydrogen} \\ 111\ Z^{-1/3}/m_e & \text{for oxygen} \end{cases} \tag{H.3}$$

$$a' = \begin{cases} 282.4/m_e & \text{for hydrogen} \\ 773\ Z^{-2/3}/m_e & \text{for oxygen}\ . \end{cases} \tag{H.4}$$

For $t > 3 \times 10^{-6}$ GeV^2 the elastic scattering form factor of the nuclei is used. The dipole form factor for the proton and the elastic form factor for oxygen are given by [444]

$$F_{\mathrm{di},H}(t) = \frac{1}{(1 + \frac{|t|}{q_0^2})^2} \tag{H.5}$$

$$F_{\mathrm{el},O}(t) = Z\left(1 - \frac{a_0^2\ t}{8}\right) e^{-a_0^2\ t/4}\ , \tag{H.6}$$

with $q_0^2 = 0.71$ GeV^2 and $a_0 = 8.97$ GeV^{-1}. To obtain an expression for the momentum exchange t we need to find the relation between the energy of the incoming photon and the energy of the outgoing axion. Following [360], the momentum exchange can be expressed as

$$t = (p_\gamma - p_a)^2 = -\frac{m_a^4}{4\,E_a^2} - E_a^2\,\theta^2\ . \tag{H.7}$$

under the assumption that $\theta \ll 1$, $m_a \ll E_a, m_N$ and $p_t \approx 0$. However, since both the photon track-length distribution and our expression for the Primakoff cross section are given in terms of the photon energy, we have to find an expression for the transverse momentum squared t in terms of E_γ. To first order approximation we have $E_\gamma \approx E_a$. As pointed out in [360], the term proportional to m_a^4 is negligible since we are considering small axion masses. Expanding to second order and inverting gives

$$E_a \approx E_\gamma - \frac{(E_\gamma\theta)^2}{2m_N} \tag{H.8}$$

and consequently

$$t = (p_\gamma - p_a)^2 = -2m_N(E_\gamma - E_a) \approx (E_\gamma\theta)^2\ . \tag{H.9}$$

For our largest $m_a = 1$ GeV we get a relative error of 0.1% for the photon energy which is negligible considering the fact that we extract the track-length distribution from a graph.

H.2 Bremsstrahlung

The differential cross section for axion production by bremsstrahlung (see Fig. 7.1 b) is given by

$$\begin{aligned} \frac{\mathrm{d}^2\sigma_Z^{e\to ae}}{\mathrm{d}\Omega_a \mathrm{d}E_a} = & \frac{\alpha_{\mathrm{em}}^2 g_{aff}^2 m_e^2}{4\pi^2}\frac{E_e}{U^2}\left(x^3 - \frac{2m_a^2 x^2(1-x)}{U}\right. \\ & \left. + \frac{2m_a^2}{U^2}[m_a^2 x(1-x)^2 + m_e^2 x^3(1-x)]\right)\chi\,, \end{aligned} \tag{H.10}$$

with $U = E_e^2\theta_a^2 x + m_e^2 x + m_a^2(1-x)/x$ and $x = E_a/E_e$. χ is the integrated form factor obtained from integrating $\chi = \int_{t_{\min}} \mathrm{d}t |F(t)|^2 (t - t_{\min})/t^2$ [445],

$$\begin{aligned}\chi &= \chi_{\text{elastic}} + \chi_{\text{inelastic}} \\ &= Z^2\left[\ln\left(\frac{a^2 m_e^2 (1+l)^2}{a^2 t_{\min} + 1}\right) - 1\right] + Z\left[\ln\left(\frac{a'^2 m_e^2 (1+l)^2}{a'^2 t_{\min} + 1}\right) - 1\right],\end{aligned} \tag{H.11}$$

where $t_{\min} = \left(U/(2E_e(1-x))\right)^2$ is the minimal momentum transfer squared and $l = E_e^2\theta_a^2/m_e^2$ and a and a' are to a reasonable approximation given by eqs. (H.3). m_e is the mass of the electron, E_e is the energy of the incoming electron and positron and E_a the energy of the ALP in the laboratory frame.

H.3 Positron annihilation

Non-resonant case The non-resonant cross section for positron annihilation in the centre of mass frame is given by (see Fig. 7.1c)

$$\frac{\mathrm{d}\sigma^{e^+\to a\gamma}}{\mathrm{d}\cos(\theta^*)} = \frac{g_{aff}^2 m_f^2 \alpha_{\text{em}}}{2s}\left(\frac{m_a^2}{s - m_a^2}\right)\frac{1}{1-\beta^2\cos(\theta^*)^2}, \tag{H.12}$$

where θ^* is the scattering polar angle of the ALP in the centre of mass frame and $\beta = \sqrt{1 - 4m_e^2/s}$ is the speed of the electron target [377]. $s = 2E_+ m_e + 2m_e^2$ is the centre of mass energy squared. The maximal opening angle for the ALP to be detected is given in the laboratory frame (see appendix I). We therefore have to relate the ALP's scattering angle in the centre of mass frame, θ^*, with $\theta_{\max}^{\text{lab}}$, the angle in the laboratory frame. This can be achieved by expressing the Mandelstam variable t which is Lorentz invariant in the laboratory frame and in the centre of mass frame. We find

$$\cos(\theta^{\text{lab}}) = \frac{E_+^{\text{lab}} E_a^{\text{lab}} - E_+^* E_a^* + \cos(\theta^*)\sqrt{E_a^{*2} - m_a^2}\sqrt{E_+^{*2} - m_e^2}}{\sqrt{E_+^{\text{lab}2} - m_e^2}\sqrt{E_a^{\text{lab}2} - m_a^2}}, \tag{H.13}$$

where the asterix denotes quantities in the centre of mass frame. The ALP's energy in the the laboratory frame is

$$E_a^{\text{lab}} = \frac{s + m_a^2 + (s - m_a^2)\beta\cos(\theta^*)}{4m_e}. \tag{H.14}$$

In the centre of mass frame the energies of the ALP and of the positron are

$$E_a^* = \frac{4E_+^{*2} + m_a^2}{4E_+^*} \qquad E_+^* = \frac{1}{2}\sqrt{s} = \frac{1}{2}\sqrt{2(m_e^2 + m_e E_+^{\text{lab}})}. \tag{H.15}$$

We have now all the ingredients to express the differential cross section in terms of the scattering angle and the positron energy in the laboratory frame over which we integrate. To deal with the infrared divergence in the cross section for $s \to m_a^2$, i.e. small photon energies, we apply a cut on the centre of mass energy around $\sqrt{s} = 3$ MeV as proposed in Ref. [377].

Resonant production The resonant positron annihilation production cross section (see Fig. 7.1d) can be obtained from the Breit-Wigner formula and is in the centre of mass frame given by [446]

$$\sigma^{e^+\to a}(E_a^*) = \frac{\pi^2}{2E_a^{*2}} \frac{\Gamma(a\to e^+e^-)\Gamma(a\to XX)}{(E_a^*-m_a)^2+\Gamma_a^2/4}\,, \tag{H.16}$$

with Γ_a the total decay width of the ALP and

$$\Gamma(a\to XX) = \begin{cases} \Gamma_{\text{loop}}(a\to\gamma\gamma) & \text{for the ALP decaying into two photons} \\ \Gamma(a\to e^+e^-) & \text{for the ALP decaying into } e^+e^-. \end{cases} \tag{H.17}$$

Since the Breit-Wigner formula contains the decay width of the ALP which is strongly peaked compared to the step size in the positron energy the cross section is approximated by a delta function, i.e.

$$\sigma^{e^+\to a}(E_a^*) = \sigma_{\text{tot}}^{e^+\to a}\delta(E_a^*-m_a) \tag{H.18}$$

where

$$\sigma_{\text{tot}}^{e^+\to a}(a\to XX) = \int_{\infty}^{\infty} \mathrm{d}E_a^*\sigma^{e^+\to a}(E_a^*) = \frac{\pi^2}{2m_a^2}\Gamma(a\to e^+e^-)\mathcal{B}(a\to XX) \tag{H.19}$$

and hence independent of the positron energy. The number of events in a specific decay channel can therefore be approximated by

$$\begin{aligned}\mathcal{N}(a\to XX) &= \mathcal{N}_{e,\text{ inc}}\ \sigma_{\text{tot}}^{e^+\to a}(a\to XX)\int \mathrm{d}E_a^*\ T(E_+)\ p(E_+)\ \frac{\mathrm{d}E_+}{\mathrm{d}E_a^*}\,\delta(E_a^*-m_a)\sum_Z P_Z \\ &= \mathcal{N}_{e,\text{ inc}}\ \sigma_{\text{tot}}^{e^+\to a}(a\to XX)\ T(E_+)\ p(E_+)\ \frac{m_a}{m_e}\ \sum_Z P_Z\,,\end{aligned} \tag{H.20}$$

where $T_+(E)$ is the track-length distribution of positrons. The energy of the positron and of the ALP are in the laboratory frame given by

$$E_+^{\text{lab}} = \frac{m_a^2-m_e^2}{2m_e}\,, \qquad E_a^{\text{lab}} = E_+^{\text{lab}} + m_e \tag{H.21}$$

respectively.

I Details on the ALP's detection probability

In this appendix we will give details about the probability for the ALP in the SLAC electron beam dump experiment to decay invisibly, i.e. we consider the cases where the ALP could have been produced but not been detected. In these cases the non-observation of axions in the experiment does not allow us to set bounds on the ALP's coupling and mass. To determine whether the ALP decays inside or outside the decay volume we need its decay length. The decay length of the ALP in the laboratory frame is

$$l_a = \beta\gamma_a\tau_a = \frac{\beta\gamma_a}{\Gamma_a} \approx \frac{E_a}{m_a\Gamma_a} \tag{I.1}$$

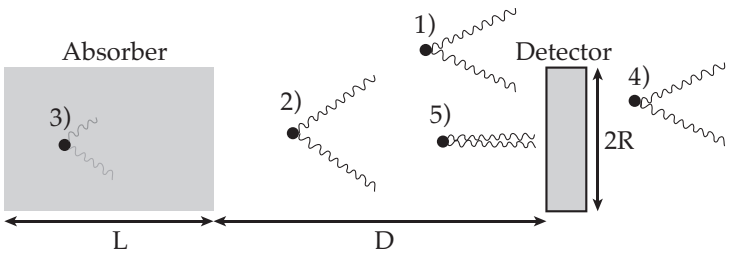


FIGURE I.1: Different scenarios for the ALPs to decay invisibly (here schematically shown for photons).

where we assume $E_a \gg m_a$ [360]. In the following discussion we consider the case where the axion decays into two photons.[13] Taking the experimental layout into account, the ALP cannot be detected if (the scenarios are summarised in Fig. I.1):

- the opening angle of the ALP is too big, $p(\theta_a > \theta_{\max}) = 0$. The detector which is approximated as cylindrical with radius $R = 1.5$ m can detect photons with a maximal separation angle of $\sin(\theta_{\max}) = R/(D+L) \approx \theta_{\max}$ (see definition in Fig. I.1), which for the dimensions of E137 corresponds to $\theta_{\text{E137, max}} = 0.22$ deg.

- the ALP decays too early and/or the opening angle between the photons is too large, so that both photons would miss the detector. The opening angle between the two photons is determined by the boost of the ALP. The minimal opening angle for high ALP boosts is given by $\theta_{\gamma,\,\min} \approx 2/\gamma_a$ [360, 369]. The distribution of the number of photons $\mathrm{d}N/\mathrm{d}\theta_\gamma$ is peaked at this minimal $\theta_{\gamma,\,\min}$ and the typical separation between the two photons arriving at the detector is therefore given by

$$d_{\gamma\gamma} = \sin\left[\theta_{\gamma,\,\min}(D+L-l_a)\right] \approx \theta_{\gamma,\,\min}(D+L-l_a) = \frac{2(D+L-l_a)}{\gamma_a}\,, \tag{I.2}$$

  where $D+L-l_a$ is the distance from the detector at which the ALP decays. $d_{\gamma\gamma}$ should not exceed the dimension of the detector, $p(d_{\gamma\gamma} > 2R) = p(D+L-l_a > RE_a/m_a) = 0$.

- the ALP decays inside the absorber and the photons get absorbed or the ALP decays behind the detector. The probability that the ALP decays inside the decay volume is given by

$$p(l_a) = p(l_a > D) - p(l_a > D+L) = e^{-D/l_a} - e^{-(D+L)/l_a}\,. \tag{I.3}$$

  Indeed, we find that this significantly reduces the number of detectable particles.

- the ALP decays in front of the detector but the separation between the two photons is too small. The typical separation between the photons has to be larger than the minimal resolution of the detector, $R_{\text{E137, min}} \approx 3$ mm, which would lead to an indistinguishable signal, $p(d_{\gamma\gamma} < R_{\min}) = 0$.

[13] In principle, the same arguments can be applied for the case where the ALP decays to an electron-positron pair. Some approximations may not be reasonable as the decay products are massive (concerning boost, typical opening angle). To derive the bounds, we have used the same expressions and obtain very good agreement with the results presented in [377] (based on a Monte-Carlo simulation) and [57].

To sum up, the probability to detect both photons is given by [360]

$$p(l_a) = \begin{cases} e^{-D/l_a} - e^{-(D+L)/l_a} & \text{if } \; E_a < 2m_a \cdot L/R_{\min}, \\ & E_a > (L+D)\cdot m_a/\left(R + 1/\Gamma_a\right) \text{ and } \theta < \theta_{\max} \\ 0 & \text{otherwise}\,. \end{cases} \tag{I.4}$$

J Details on constraints from exotic Higgs decays

To understand the sensitivity of the LHC exotic Higgs decay searches, it is first important to calculate the decay table for the axion. A detailed calculation is available in the appendix and the branching fractions are plotted in Fig. G.1 for a range of ALP masses from 10 MeV to 100 GeV. What is relevant for our discussion here is that the branching fraction into two photons remains consistently smaller than 10^{-3} as soon as the ALP mass is large enough to allow $a \to \mu^+\mu^-$. Next, as expected, once each new massive fermion mode turns on, it quickly dominates as the width in that channel is proportional to m_f^2. We also assume that as soon as the mass threshold is above 1.5 GeV ($\simeq 2m_\rho$), decays into hadrons via loop-mediated $a \to gg$ open up. Therefore, once hadronic decays become allowed, any branching into muons becomes again vanishingly small. Our calculation (assuming $a \to gg$ open only above 1 GeV) gives $\mathcal{B}(\mu^+\mu^-) < 10^{-3}$ as soon as $\tau^+\tau^-$ and $c\bar{c}$ modes open (i.e. $m_a \gtrsim 1.8$ GeV). Since all current exotic Higgs decays look for $m_a \gtrsim 10$ GeV only, we can safely assume that none of the muonic searches will be able to see the signatures of our model. We therefore need only think about the $h \to aa \to 4b$ channel. The limits on $h \to aa \to 4b$ are only able to exclude branching ratios less than one in the small mass range between 18-22.5 GeV. However, this would still require a branching fraction of 0.75 of the SM Higgs into this one channel alone. Given that in our model, these channels are dependent on 1-loop contributions only the decay width of the Higgs into these modes cannot be of comparable size to the usual tree-level 2-body SM decays. We therefore conclude that none of these searches unfortunately have any sensitivity to our model.

The current limit [382] in $h \to Z(a \to gg)$ is in principle sensitive to our model in the very narrow range 0.5 GeV $< m_a \lesssim 2.7$ GeV. However, this is precisely the mass range where theoretical calculations are wildly unpredictable due to hadronic contributions and we choose not to apply these.

Lastly, the $h \to Z\gamma$ searches can be used in certain phase space regions where the two photons from a-decay are collimated, i.e. $h \to Z(a \to \gamma\gamma)$. The $h \to ZZ^* \to 4\ell$ measurements can also be used to place a limit on new physics contributions as the off-shell Z decaying into two leptons can also be interpreted as a new particle that decays into two leptons. A study of the collimated di-gamma decay was done in [447]. However, interpreted in our model, it corresponds only to coupling values $C_{aff} \gtrsim 10^2$/GeV. Since this lies wildly outside the self-consistent EFT regime, we assume that the $h \to Z\gamma$ measurements are currently not sensitive enough to provide a useful constraint. Similarly, the 4ℓ decay mode of the Higgs currently carries the signal strength accuracy $\mu = 1.44 \pm 0.4$ [448] which translates to an upper limit on the decay width $\Gamma(h \to Za) \times \mathcal{B}(a \to 2\ell) \lesssim 0.44 \times \Gamma(h \to ZZ^* \to 4\ell)$. Again, due to the loop suppression in our decay, we end up with exclusions on $C_{aff} \gg 1$ only and there is no usable limit from this measurement.

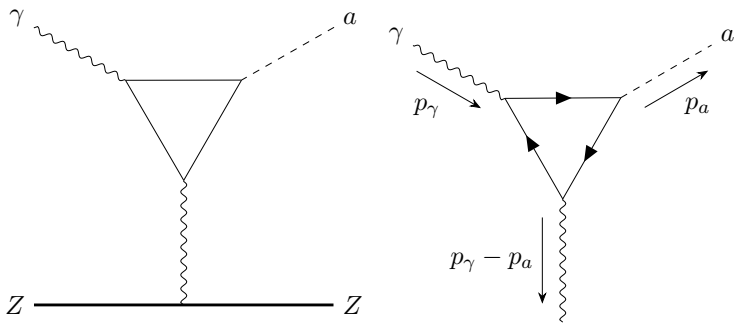


FIGURE J.1: *Left:* Feynman diagram of the Primakoff production process via a fermion loop. *Right:* Feynman diagram of the amplitude we need to calculate to replace $g_{a\gamma\gamma}$ by $g_{a\gamma\gamma}^{\text{loop}}$ in the Primakoff production process.

J.1 Loop diagrams relevant for constraint calculations

The ALP in our model can only decay into photons via a fermion loop. Applying the constraints from the effective axion-photon coupling, $g_{a\gamma\gamma}$, to the axion-fermion coupling, g_{aff}, demands more theoretical groundwork. The main task is to calculate the Primakoff cross section for this coupling. The resulting cross section can then be inserted into the expression for the ALP yield for the beam dump constraint and the expressions for the total energy outflow and the ALP opacity for the HB stars constraints. The approach we adopt is to replace the expression for the axion-photon coupling in the "decay width" in Eqs. (H.1) by an expression for the axion-photon coupling induced by a fermion loop. This coupling will now depend on the momentum transfer squared, $t = (p_\gamma - p_a)^2$. The Feynman diagram of the Primakoff process induced by a fermion loop is shown in Fig. J.1 on the left. It should be noticed that the decay width in Eq. (H.1) has to be regarded as a definition of the axion-photon coupling rather than an actual physical decay width,

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} &= 8\alpha_{\text{em}}\frac{\Gamma}{m_a^3}\mid F(t)\mid^2 \frac{\theta^2}{\left(\theta^2+\frac{1}{4}\left(\frac{m_a}{E_\gamma}\right)^4\right)^2} \\ &= \alpha_{\text{em}}\frac{g_{a\gamma\gamma}^2}{8\pi}\mid F(t)\mid^2 \frac{\theta^2}{\left(\theta^2+\frac{1}{4}\left(\frac{m_a}{E_\gamma}\right)^4\right)^2}\,. \end{aligned} \tag{J.1}$$

By comparing the amplitude of the loop diagram and the tree-level effective diagram we will deduce what $g_{a\gamma\gamma}$ should be replaced by. The Feynman diagram of the amplitude we have to consider is depicted in Fig. J.1 on the right. We proceed as in Ref. [17] and find that the amplitude of the triangle graph is given by

$$\begin{aligned}\mathcal{T}^{\alpha\beta} &= \int \frac{d^4k}{(2\pi)^4}(-1)\,\mathrm{Tr}\left[\frac{i}{\not{k}+\not{p}_\gamma - m_f}\gamma^\alpha \frac{i}{\not{k}-m_f}\gamma^\beta \frac{i}{\not{k}+\not{p}_\gamma-\not{p}_a-m_f}\gamma_5\right] \\ &= -i\frac{1}{2\pi^2}\varepsilon^{\alpha\beta\rho\sigma}p_{a,\rho}p_{\gamma,\sigma}m_f C_0(p_a^2,p_\gamma^2,(p_a-p_\gamma)^2,m_f^2,m_f^2,m_f^2)\,, \end{aligned} \tag{J.2}$$

with the three-point scalar function C_0 defined as in Ref. [449]. The matrix element squared of the diagram is given by (with $t = (p_a - p_\gamma)^2$ in Eq. (H.1))

$$|\mathcal{M}|^2_{\text{loop}} = \frac{\alpha^2_{\text{em}}(m_a^2 - t)^2 \left[\sum_f g_{aff} m_f^2 q_f^2 n_f^c |C_0(m_a^2, 0, t, m_f^2, m_f^2, m_f^2)|\right]^2}{2\pi^2} \,. \tag{J.3}$$

By calculating the same diagram but with an effective ALP-photon coupling $g_{a\gamma\gamma}$ at tree level, i.e.

$$\begin{aligned} |\mathcal{M}|^2 &= g^2_{a\gamma\gamma} \epsilon_{\mu\nu\rho\sigma} \epsilon_{\alpha\beta\kappa\lambda} g^{\alpha\mu}\, g^{\beta\nu}\, p_\gamma^\rho\, p_\gamma^\kappa\, (p_\gamma - p_a)^\sigma\, (p_\gamma - p_a)^\lambda \\ &= 2\, g^2_{a\gamma\gamma} \left((p_a \cdot p_\gamma)^2 - (p_a \cdot p_a)(p_\gamma \cdot p_\gamma)\right) = \frac{g^2_{a\gamma\gamma}(m_a^2 - t)^2}{2} \end{aligned} \tag{J.4}$$

we obtain the replacement

$$g^2_{a\gamma\gamma} \to \frac{\alpha^2_{\text{em}} \left[\sum_f g_{aff} m_f^2 q_f^2 n_f^c |C_0(m_a^2, 0, t, m_f^2, m_f^2, m_f^2)|\right]^2}{\pi^2} \tag{J.5}$$

in the expression for the Primakoff production in the E137 SLAC experiment, Eq. (H.1), and in the expression for the Primakoff production in horizontal branch stars. For the constraints from horizontal branch stars we make the approximation of zero momentum transfer, i.e. $t \approx 0$.

Bibliography

[1] M. Baak, J. Cúth, J. Haller, A. Hoecker, R. Kogler, K. Mönig, M. Schott, J. Stelzer, "The global electroweak fit at NNLO and prospects for the LHC and ILC", Eur. Phys. J. **C74**, p. 3046, (2014), arXiv: 1407.3792.

[2] M. Tanabashi *et al.*, "Review of Particle Physics", Phys. Rev. D **98**, p. 030001, (Aug 2018), `https://link.aps.org/doi/10.1103/PhysRevD.98.030001`.

[3] F. Jegerlehner, "$\alpha_{QED,eff}$(s) for precision physics at the FCC-ee/ILC", CERN Yellow Reports: Monographs **3**, p. 9, (2020).

[4] M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, "A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to $\boldsymbol{\alpha(m_Z^2)}$", arXiv: 1908.00921.

[5] A. Keshavarzi, D. Nomura, T. Teubner, "g−2 of charged leptons, $\alpha(M_Z^2)$, and the hyperfine splitting of muonium", Phys. Rev. D **101**, p. 014029, (2020), arXiv: 1911.00367.

[6] S. Borsanyi *et al.*, "Hadronic vacuum polarization contribution to the anomalous magnetic moments of leptons from first principles", Phys. Rev. Lett. **121**, p. 022002, (2018), arXiv: 1711.04980.

[7] S. Borsanyi *et al.*, "Leading hadronic contribution to the muon magnetic moment from lattice QCD", Nature **593**, p. 51, (2021), arXiv: 2002.12347.

[8] M. Cè, A. Gérardin, G. von Hippel, H. B. Meyer, K. Miura, K. Ottnad, A. Risch, T. San José, J. Wilhelm, H. Wittig, "The hadronic running of the electromagnetic coupling and the electroweak mixing angle from lattice QCD", JHEP **08**, p. 220, (2022), arXiv: 2203.08676.

[9] D. Bernecker, H. B. Meyer, "Vector Correlators in Lattice QCD: Methods and applications", Eur. Phys. J. A **47**, p. 148, (2011), arXiv: 1107.4388.

[10] T. Blum, P. A. Boyle, V. Gülpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli, J. T. Tsang, "Calculation of the hadronic vacuum polarization contribution to the muon anomalous magnetic moment", Phys. Rev. Lett. **121**, p. 022003, (2018), arXiv: 1801.07224.

[11] R. Sommer, L. Chimirri, N. Husung, "Log-enhanced discretization errors in integrated correlation functions", PoS **LATTICE2022**, p. 358, (2023), arXiv: 2211.15750.

[12] L. Parato, *A full calculation of the leading hadronic contribution to the muon anomalous magnetic moment in lattice quantum chromodynamics*. PhD thesis, Aix-Marseille U. (2022).

[13] B. Colquhoun, R. J. Dowdall, C. T. H. Davies, K. Hornbostel, G. P. Lepage, "Υ and Υ' Leptonic Widths, a_μ^b and m_b from full lattice QCD", Phys. Rev. **D91**, p. 074514, (2015), arXiv: 1408.5768.

[14] A. Keshavarzi, D. Nomura, T. Teubner, "Muon $g-2$ and $\alpha(M_Z^2)$: a new data-based analysis", Phys. Rev. **D97**, p. 114025, (2018), arXiv: 1802.02995.

[15] R. V. Harlander, M. Steinhauser, "rhad: A Program for the evaluation of the hadronic R ratio in the perturbative regime of QCD", Comput. Phys. Commun. **153**, p. 244, (2003), arXiv: hep-ph/0212294.

[16] M. Davier, Z. Fodor, A. Gerardin, L. Lellouch, B. Malaescu, F. M. Stokes, K. K. Szabo, B. C. Toth, L. Varnhorst, Z. Zhang, "Hadronic vacuum polarization: comparing lattice QCD and data-driven results in systematically improvable ways", arXiv: 2308.04221.

[17] J. Quevillon, C. Smith, "Axions are blind to anomalies", Eur. Phys. J. C **79**, p. 822, (2019), arXiv: 1903.12559.

[18] L. Di Luzio, M. Giannotti, E. Nardi, L. Visinelli, "The landscape of QCD axion models", Phys. Rept. **870**, p. 1, (2020), arXiv: 2003.01100.

[19] L. J. Hall, K. Jedamzik, J. March-Russell, S. M. West, "Freeze-in production of FIMP dark matter", Journal of High Energy Physics **2010**, p. 80, (Mar 2010).

[20] G. Bélanger, C. Delaunay, A. Pukhov, B. Zaldivar, "Dark matter abundance from the sequential freeze-in mechanism", Phys. Rev. D **102**, p. 035017, (2020), arXiv: 2005.06294.

[21] X. Chu, T. Hambye, M. H. G. Tytgat, "The Four Basic Ways of Creating Dark Matter Through a Portal", JCAP **05**, p. 034, (2012), arXiv: 1112.0493.

[22] T. Hambye, M. H. G. Tytgat, J. Vandecasteele, L. Vanderheyden, "Dark matter from dark photons: a taxonomy of dark matter production", Phys. Rev. D **100**, p. 095018, (2019), arXiv: 1908.09864.

[23] P. Gondolo, G. Gelmini, "Cosmic abundances of stable particles: Improved analysis", Nuclear Physics B **360**, p. 145, (1991), `https://www.sciencedirect.com/science/article/pii/0550321391904384`.

[24] J. H. Chang, R. Essig, S. D. McDermott, "Supernova 1987A Constraints on Sub-GeV Dark Sectors, Millicharged Particles, the QCD Axion, and an Axion-like Particle", JHEP **09**, p. 051, (2018), arXiv: 1803.00993.

[25] G. G. Raffelt, G. D. Starkman, "Stellar energy transfer by keV-mass scalars", Phys. Rev. D **40**, p. 942, (Aug 1989), `https://link.aps.org/doi/10.1103/PhysRevD.40.942`.

[26] S. D. Drell, M. Weinstein, S. Yankielowicz, "Quantum field theories on a lattice: Variational methods for arbitrary coupling strengths and the Ising model in a transverse magnetic field", Physical Review D **16**, p. 1769 (1977).

[27] A. Bharucha, F. Brümmer, N. Desai, S. Mutzel, "Axion-like particles as mediators for dark matter: beyond freeze-out", JHEP **02**, p. 141, (2023), arXiv: 2209.03932.

[28] M. Kawasaki, K. Kohri, T. Moroi, K. Murai, H. Murayama, "Big-bang nucleosynthesis with sub-GeV massive decaying particles", JCAP **12**, p. 048, (2020), arXiv: 2006.14803.

[29] F. Abe *et al.*, "Observation of top quark production in $\bar{p}p$ collisions", Phys. Rev. Lett. **74**, p. 2626, (1995), arXiv: hep-ex/9503002.

[30] S. Abachi *et al.*, "Observation of the top quark", Phys. Rev. Lett. **74**, p. 2632, (1995), arXiv: hep-ex/9503003.

[31] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Physics Letters B **716**, p. 1 , (2012), `http://www.sciencedirect.com/science/article/pii/S037026931200857X`.

[32] T. Aoyama, T. Kinoshita, M. Nio, "Theory of the Anomalous Magnetic Moment of the Electron", Atoms **7**, p. 28, (2019).

[33] G. W. Bennett *et al.*, "Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL", Phys. Rev. D **73**, p. 072003, (2006), arXiv: hep-ex/0602035.

[34] B. Abi *et al.*, "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm", Phys. Rev. Lett. **126**, p. 141801, (2021), arXiv: 2104.03281.

[35] T. Aoyama *et al.*, "The anomalous magnetic moment of the muon in the Standard Model", Phys. Rept. **887**, p. 1, (2020), arXiv: 2006.04822.

[36] D. P. Aguillard *et al.*, "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm", Phys. Rev. Lett. **131**, p. 161802, (2023), arXiv: 2308.06230.

[37] M. Cè *et al.*, "Window observable for the hadronic vacuum polarization contribution to the muon g-2 from lattice QCD", Phys. Rev. D **106**, p. 114502, (2022), arXiv: 2206.06582.

[38] C. Alexandrou *et al.*, "Lattice calculation of the short and intermediate time-distance hadronic vacuum polarization contributions to the muon magnetic moment using twisted-mass fermions", Phys. Rev. D **107**, p. 074506, (2023), arXiv: 2206.15084.

[39] T. Blum *et al.*, "Update of Euclidean windows of the hadronic vacuum polarization", Phys. Rev. D **108**, p. 054507, (2023), arXiv: 2301.08696.

[40] A. Bazavov *et al.*, "Light-quark connected intermediate-window contributions to the muon g-2 hadronic vacuum polarization from lattice QCD", Phys. Rev. D **107**, p. 114514, (2023), arXiv: 2301.08274.

[41] F. V. Ignatov *et al.*, "Measurement of the $e^+e^- \to \pi^+\pi^-$ cross section from threshold to 1.2 GeV with the CMD-3 detector", arXiv: 2302.08834.

[42] J. Erler, M. Schott, "Electroweak Precision Tests of the Standard Model after the Discovery of the Higgs Boson", Prog. Part. Nucl. Phys. **106**, p. 68, (2019), arXiv: 1902.05142.

[43] J. Erler, "Global fits to electroweak data using GAPP", in *Physics at Run II: QCD and Weak Boson Physics Workshop: 2nd General Meeting*, 6 1999.

[44] R. L. Workman, Others, "Review of Particle Physics", PTEP **2022**, p. 083C01, (2022).

[45] M. Passera, W. J. Marciano, A. Sirlin, "The Muon g-2 and the bounds on the Higgs boson mass", Phys. Rev. D **78**, p. 013009, (2008), arXiv: 0804.1142.

[46] A. Crivellin, M. Hoferichter, C. A. Manzari, M. Montull, "Hadronic Vacuum Polarization: $(g-2)_\mu$ versus Global Electroweak Fits", Phys. Rev. Lett. **125**, p. 091801, (2020), arXiv: 2003.04886.

[47] A. Keshavarzi, W. J. Marciano, M. Passera, A. Sirlin, "Muon $g-2$ and $\Delta\alpha$ connection", Phys. Rev. D **102**, p. 033002, (2020), arXiv: 2006.12666.

[48] E. de Rafael, "Constraints between $\Delta\alpha_{\rm had}(M_Z^2)$ and $(g_\mu - 2)_{\rm HVP}$", Phys. Rev. D **102**, p. 056025, (2020), arXiv: 2006.13880.

[49] B. Malaescu, M. Schott, "Impact of correlations between a_μ and $\alpha_{\rm QED}$ on the EW fit", Eur. Phys. J. C **81**, p. 46, (2021), arXiv: 2008.08107.

[50] E. Aprile *et al.*, "Constraining the Spin-Dependent WIMP-Nucleon Cross Sections with XENON1T", Phys. Rev. Lett. **122**, p. 141301, (Apr 2019), `https://link.aps.org/doi/10.1103/PhysRevLett.122.141301`.

[51] A. Boveia, C. Doglioni, "Dark Matter Searches at Colliders", Annual Review of Nuclear and Particle Science **68**, p. 429, (2018), `https://doi.org/10.1146/annurev-nucl-101917-021008`.

[52] N. Bernal, M. Heikinheimo, T. Tenkanen, K. Tuominen, V. Vaskonen, "The dawn of FIMP Dark Matter: A review of models and constraints", International Journal of Modern Physics A **32**, p. 1730023, (2017).

[53] R. D. Peccei, H. R. Quinn, "CP Conservation in the Presence of Instantons", Phys. Rev. Lett. **38**, p. 1440, (1977). [,328(1977)].

[54] R. D. Peccei, H. R. Quinn, "CP Conservation in the Presence of Pseudoparticles", Phys. Rev. Lett. **38**, p. 1440, (Jun 1977), `https://link.aps.org/doi/10.1103/PhysRevLett.38.1440`.

[55] F. Wilczek, "Problem of Strong P and T Invariance in the Presence of Instantons", Phys. Rev. Lett. **40**, p. 279, (Jan 1978), `https://link.aps.org/doi/10.1103/PhysRevLett.40.279`.

[56] S. Weinberg, "A New Light Boson?", Phys. Rev. Lett. **40**, p. 223, (Jan 1978), `https://link.aps.org/doi/10.1103/PhysRevLett.40.223`.

[57] M. Bauer, M. Neubert, A. Thamm, Journal of High Energy Physics **2017**, "Collider probes of axion-like particles",, (dec 2017).

[58] M. Chala, G. Guedes, M. Ramos, J. Santiago, "Running in the ALPs", Eur. Phys. J. C **81**, p. 181, (2021), arXiv: 2012.09017.

[59] C. Cornella, P. Paradisi, O. Sumensari, "Hunting for ALPs with Lepton Flavor Violation", JHEP **01**, p. 158, (2020), arXiv: 1911.06279.

[60] M. Bauer, M. Neubert, S. Renner, M. Schnubel, A. Thamm, "Axionlike Particles, Lepton-Flavor Violation, and a New Explanation of a_μ and a_e", Phys. Rev. Lett. **124**, p. 211803, (2020), arXiv: 1908.00008.

[61] D. Chang, W.-F. Chang, C.-H. Chou, W.-Y. Keung, "Large two loop contributions to g-2 from a generic pseudoscalar boson", Phys. Rev. D **63**, p. 091301, (2001), arXiv: hep-ph/0009292.

[62] M. A. Buen-Abad, J. Fan, M. Reece, C. Sun, "Challenges for an axion explanation of the muon $g-2$ measurement", JHEP **09**, p. 101, (2021), arXiv: 2104.03267.

[63] L. Darmé, G. Grilli di Cortona, E. Nardi, "The muon g − 2 anomaly confronts new physics in e and μ final states scattering", JHEP **06**, p. 122, (2022), arXiv: 2112.09139.

[64] O. Philipsen, *Quantenfeldtheorie und das Standardmodell der Teilchenphysik: Eine Einführung.* Springer Berlin Heidelberg, 2018.

[65] T. Osborne, "Lectures on advanced quantum field theory", 2017.

[66] M. Schwartz, *Quantum Field Theory and the Standard Model.* Quantum Field Theory and the Standard Model, Cambridge University Press, 2014.

[67] F. Mandl, G. Shaw, *Quantum Field Theory.* A Wiley-Interscience publication, Wiley, 1984.

[68] A. Zee, *Quantum Field Theory in a Nutshell: (Second Edition).* In a Nutshell, Princeton University Press, 2010.

[69] M. Neubert, "Renormalization Theory and Effective Field Theories", arXiv: 1901.06573.

[70] D. J. Gross, F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", Phys. Rev. Lett. **30**, p. 1343, (06 1973), `https://link.aps.org/doi/10.1103/PhysRevLett.30.1343`.

[71] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", Phys. Rev. Lett. **30**, p. 1346, (06 1973), `https://link.aps.org/doi/10.1103/PhysRevLett.30.1346`.

[72] M. Luscher, "Lattice QCD: From quark confinement to asymptotic freedom", Annales Henri Poincare **4**, p. S197, (2003), arXiv: hep-ph/0211220.

[73] G. Raffelt, *Stars as Laboratories for Fundamental Physics: The Astrophysics of Neutrinos, Axions, and Other Weakly Interacting Particles.* Theoretical Astrophysics, University of Chicago Press, 1996.

[74] M. D. Goodsell, "Lecture on the Strong CP puzzle", May 2018.

[75] G. 't Hooft, "Computation of the quantum effects due to a four-dimensional pseudoparticle", Phys. Rev. D **14**, p. 3432 (1976), `http://cds.cern.ch/record/428649`.

[76] S. Vandoren, P. van Nieuwenhuizen, "Lectures on instantons", arXiv: 0802.1862.

[77] "Homotopy groups of spheres." `https://en.wikipedia.org/wiki/Homotopy_groups_of_spheres`. Accessed: 2023-10-19.

[78] R. Bott, "The Stable Homotopy of the Classical Groups", Annals of Mathematics **70**, p. 313 (1959), `http://www.jstor.org/stable/1970106`.

[79] E. B. Bogomolny, "Stability of Classical Solutions", Sov. J. Nucl. Phys. **24**, p. 449 (1976).

[80] A. Belavin, A. Polyakov, A. Schwartz, Y. Tyupkin, "Pseudoparticle solutions of the Yang-Mills equations", Physics Letters B **59**, p. 85, (1975), `https://www.sciencedirect.com/science/article/pii/037026937590163X`.

[81] F. Gelis, *Quantum Field Theory: From Basics to Modern Topics*. Cambridge University Press, 2019.

[82] K. Huang, *Quarks, Leptons & Gauge Fields*. G - Reference,Information and Interdisciplinary Subjects Series, World Scientific, 1992.

[83] S. R. Coleman, "The Uses of Instantons", Subnucl. Ser. **15**, p. 805 (1979).

[84] D. Tong, "Gauge theory", Lecture notes, DAMTP Cambridge **10** (2018).

[85] G. 't Hooft, "Symmetry Breaking Through Bell-Jackiw Anomalies", Phys. Rev. Lett. **37**, p. 8, (1976).

[86] K. Fujikawa, "Path-Integral Measure for Gauge-Invariant Fermion Theories", Phys. Rev. Lett. **42**, p. 1195, (Apr 1979), `https://link.aps.org/doi/10.1103/PhysRevLett.42.1195`.

[87] K. Fujikawa, "Erratum: Path integral for gauge theories with fermions", Phys. Rev. D **22**, p. 1499, (Sep 1980), `https://link.aps.org/doi/10.1103/PhysRevD.22.1499`.

[88] C. A. Baker *et al.*, "An Improved experimental limit on the electric dipole moment of the neutron", Phys. Rev. Lett. **97**, p. 131801, (2006), arXiv: hep-ex/0602020.

[89] C. Abel *et al.*, "Measurement of the Permanent Electric Dipole Moment of the Neutron", Phys. Rev. Lett. **124**, p. 081803, (2020), arXiv: 2001.11966.

[90] D. J. Marsh, "Axion cosmology", Physics Reports **643**, p. 1 (2016). Axion cosmology.

[91] S. Profumo, *An Introduction to Particle Dark Matter*. Advanced Textbooks in Physics, World Scientific Publishing Company, 2017.

[92] M. F. Atiyah, I. M. Singer, "The index of elliptic operators on compact manifolds", Bull. Am. Math. Soc. **69**, p. 422, (1969).

[93] Z. Fodor, C. Hoelbling, S. Krieg, L. Lellouch, T. Lippert, A. Portelli, A. Sastre, K. K. Szabo, L. Varnhorst, "Up and Down Quark Masses and Corrections to Dashen's Theorem from Lattice QCD and Quenched QED", Phys. Rev. Lett. **117**, p. 082001, (Aug 2016), `https://link.aps.org/doi/10.1103/PhysRevLett.117.082001`.

[94] J. R. Ellis, M. K. Gaillard, "Strong and Weak CP Violation", Nucl. Phys. B **150**, p. 141, (1979).

[95] K. G. Wilson, "Confinement of Quarks", Phys. Rev. D **10**, p. 2445, (1974).

[96] M. Creutz, "Monte Carlo Study of Quantized SU(2) Gauge Theory", Phys. Rev. D **21**, p. 2308, (1980).

[97] G. 't Hooft, "Renormalization of Massless Yang-Mills Fields", Nucl. Phys. B **33**, p. 173, (1971).

[98] G. 't Hooft, "Renormalizable Lagrangians for Massive Yang-Mills Fields", Nucl. Phys. B **35**, p. 167, (1971).

[99] C. Becchi, A. Rouet, R. Stora, "Renormalization of Gauge Theories", Annals Phys. **98**, p. 287, (1976).

[100] I. V. Tyutin, "Gauge Invariance in Field Theory and Statistical Physics in Operator Formalism", arXiv: 0812.0580.

[101] R. F. Streater, A. S. Wightman, *PCT, Spin and Statistics, and All That.* Princeton University Press, 1989.

[102] K. Osterwalder, R. Schrader, "AXIOMS FOR EUCLIDEAN GREEN'S FUNCTIONS", Commun. Math. Phys. **31**, p. 83, (1973).

[103] K. Osterwalder, R. Schrader, "Axioms for Euclidean Green's Functions. 2.", Commun. Math. Phys. **42**, p. 281, (1975).

[104] R. P. Feynman, "Space-Time Approach to Non-Relativistic Quantum Mechanics", Rev. Mod. Phys. **20**, p. 367, (Apr 1948), `https://link.aps.org/doi/10.1103/RevModPhys.20.367`.

[105] L. Lellouch, R. Sommer, B. Svetitsky, A. Vladikas, L. F. Cugliandolo, eds., *Modern perspectives in lattice QCD: Quantum field theory and high performance computing. Proceedings, International School, 93rd Session, Les Houches, France, August 3-28, 2009*, 2011.

[106] A. Wipf, *Statistical Approach to Quantum Field Theory: An Introduction.* Lecture Notes in Physics, Springer International Publishing, 2021.

[107] A. S. Kronfeld, "Lattice QCD", in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 91): Perspectives in the Standard Model*, 2 1992.

[108] E. Follana, Q. Mason, C. Davies, K. Hornbostel, G. P. Lepage, J. Shigemitsu, H. Trottier, K. Wong, "Highly improved staggered quarks on the lattice, with applications to charm physics", Phys. Rev. **D75**, p. 054502, (2007), arXiv: hep-lat/0610092.

[109] A. S. Kronfeld, "Lattice Gauge Theory with Staggered Fermions: How, Where, and Why (Not)", PoS **LATTICE2007**, p. 016, (2007), arXiv: 0711.0699.

[110] H. Kluberg-Stern, A. Morel, O. Napoly, B. Petersson, "Spontaneous Chiral Symmetry Breaking for a U(N) Gauge Theory on a Lattice", Nucl. Phys. B **190**, p. 504, (1981).

[111] N. Kawamoto, J. Smit, "Effective Lagrangian and Dynamical Symmetry Breaking in Strongly Coupled Lattice QCD", Nucl. Phys. B **192**, p. 100, (1981).

[112] T. Kimura, S. Komatsu, T. Misumi, T. Noumi, S. Torii, S. Aoki, "Revisiting symmetries of lattice fermions via spin-flavor representation", JHEP **01**, p. 048, (2012), arXiv: 1111.0402.

[113] L. H. Karsten, J. Smit, "Lattice Fermions: Species Doubling, Chiral Invariance, and the Triangle Anomaly", Nucl. Phys. B **183**, p. 103, (1981).

[114] H. B. Nielsen, M. Ninomiya, "Absence of Neutrinos on a Lattice. 1. Proof by Homotopy Theory", Nucl. Phys. B **185**, p. 20, (1981). [Erratum: Nucl.Phys.B 195, 541 (1982)].

[115] H. B. Nielsen, M. Ninomiya, "Absence of Neutrinos on a Lattice. 2. Intuitive Topological Proof", Nucl. Phys. B **193**, p. 173, (1981).

[116] S. D. Drell, M. Weinstein, S. Yankielowicz, "Strong-coupling field theories. II. Fermions and gauge fields on a lattice", Phys. Rev. D **14**, p. 1627, (Sep 1976), `https://link.aps.org/doi/10.1103/PhysRevD.14.1627`.

[117] J. P. Costella, "A New proposal for the fermion doubling problem", arXiv: hep-lat/0207008.

[118] P. Hasenfratz, "Prospects for perfect actions", Nucl. Phys. B Proc. Suppl. **63**, p. 53, (1998), arXiv: hep-lat/9709110.

[119] A. Hasenfratz, P. Hasenfratz, F. Niedermayer, "Simulating full QCD with the fixed point action", Phys. Rev. D **72**, p. 114508, (2005), arXiv: hep-lat/0506024.

[120] P. H. Ginsparg, K. G. Wilson, "A remnant of chiral symmetry on the lattice", Phys. Rev. D **25**, p. 2649, (May 1982), `https://link.aps.org/doi/10.1103/PhysRevD.25.2649`.

[121] M. Luscher, "Exact chiral symmetry on the lattice and the Ginsparg-Wilson relation", Phys. Lett. B **428**, p. 342, (1998), arXiv: hep-lat/9802011.

[122] D. B. Kaplan, "A Method for simulating chiral fermions on the lattice", Phys. Lett. B **288**, p. 342, (1992), arXiv: hep-lat/9206013.

[123] Y. Shamir, "Chiral fermions from lattice boundaries", Nucl. Phys. B **406**, p. 90, (1993), arXiv: hep-lat/9303005.

[124] H. Neuberger, "Exactly massless quarks on the lattice", Phys. Lett. B **417**, p. 141, (1998), arXiv: hep-lat/9707022.

[125] H. Neuberger, "More about exactly massless quarks on the lattice", Phys. Lett. B **427**, p. 353, (1998), arXiv: hep-lat/9801031.

[126] M. F. Golterman, J. Smit, "Selfenergy and Flavor Interpretation of Staggered Fermions", Nucl.Phys. **B245**, p. 61, (Jan 1984), `http://dx.doi.org/10.1016/0550-3213(84)90424-3`.

[127] H. Kluberg-Stern, A. Morel, O. Napoly, B. Petersson, "Flavors of Lagrangian Susskind Fermions", Nucl. Phys. **B220**, p. 447, (1983).

[128] G. P. Lepage, "On the Absence of O(a) Errors in Staggered-Quark Discretizations", arXiv: 1111.2955.

[129] S. R. Sharpe, "Rooted staggered fermions: Good, bad or ugly?", PoS **LAT2006**, p. 022, (2006), arXiv: hep-lat/0610094.

[130] M. F. L. Golterman, "STAGGERED MESONS", Nucl. Phys. B **273**, p. 663, (1986).

[131] M. F. L. Golterman, "Irreducible Representations of the Staggered Fermion Symmetry Group", Nucl. Phys. B **278**, p. 417, (1986).

[132] R. Altmeyer, K. D. Born, M. Gockeler, R. Horsley, E. Laermann, G. Schierholz, "The Hadron spectrum in QCD with dynamical staggered fermions", Nucl. Phys. B **389**, p. 445, (1993).

[133] W.-J. Lee, S. R. Sharpe, "Partial flavor symmetry restoration for chiral staggered fermions", Phys. Rev. D **60**, p. 114503, (1999), arXiv: hep-lat/9905023.

[134] J. Gasser, H. Leutwyler, "Chiral Perturbation Theory to One Loop", Annals Phys. **158**, p. 142, (1984).

[135] R. Balian, J. M. Drouffe, C. Itzykson, "Gauge fields on a lattice. I. General outlook", Phys. Rev. D **10**, p. 3376, (Nov 1974), `https://link.aps.org/doi/10.1103/PhysRevD.10.3376`.

[136] P. Hasenfratz, "Fascinating field theory: Quantum field theory, renormalization group, and lattice regularization",

[137] P. Hasenfratz, "Lattice QCD", Acta Phys. Austriaca Suppl. **25**, p. 283, (1983).

[138] P. Hasenfratz, "Perfect actions", in *NATO Advanced Study Institute on Confinement, Duality and Nonperturbative Aspects of QCD*, p. 179, 6 1997.

[139] H. Nastase, *Introduction to Quantum Field Theory.* Cambridge University Press, 2019.

[140] K. Kanaya, "An Introduction to finite temperature quantum chromodynamics on the lattice", Prog. Theor. Phys. Suppl. **131**, p. 73, (1998), arXiv: hep-lat/9804006.

[141] T. J. Hollowood, *Renormalization Group and Fixed Points: In Quantum Field Theory*, vol48. Springer, 2013.

[142] H. Wittig, *QCD on the Lattice*, p. 137. Cham: Springer International Publishing, 2020.

[143] M. Lüscher, "Properties and uses of the Wilson flow in lattice QCD", JHEP **08**, p. 071, (2010), arXiv: 1006.4518. [Erratum: JHEP 03, 092 (2014)].

[144] S. Borsanyi *et al.*, "High-precision scale setting in lattice QCD", JHEP **09**, p. 010, (2012), arXiv: 1203.4469.

[145] R. Sommer, "A New way to set the energy scale in lattice gauge theories and its applications to the static force and α_s in SU(2) Yang-Mills theory", Nucl. Phys. B **411**, p. 839, (1994), arXiv: hep-lat/9310022.

[146] P. Weisz, "Continuum Limit Improved Lattice Action for Pure Yang-Mills Theory. 1.", Nucl. Phys. B **212**, p. 1, (1983).

[147] P. Weisz, R. Wohlert, "Continuum Limit Improved Lattice Action for Pure Yang-Mills Theory. 2.", Nucl. Phys. B **236**, p. 397, (1984). [Erratum: Nucl.Phys.B 247, 544 (1984)].

[148] M. Lüscher, P. Weisz, "Computation of the action for on-shell improved lattice gauge theories at weak coupling", Physics Letters B **158**, p. 250 (1985).

[149] B. Beinlich, F. Karsch, E. Laermann, "Improved actions for QCD thermodynamics on the lattice", Nuclear Physics B **462**, p. 415 (1996).

[150] M. Alford, W. Dimm, G. Lepage, G. Hockney, P. Mackenzie, "Lattice QCD on small computers", Physics Letters B **361**, p. 87 (1995).

[151] M. Luscher, P. Weisz, "On-Shell Improved Lattice Gauge Theories", Commun. Math. Phys. **97**, p. 59, (1985). [Erratum: Commun.Math.Phys. 98, 433 (1985)].

[152] P. Hasenfratz, F. Niedermayer, "Perfect lattice action for asymptotically free theories", Nucl. Phys. B **414**, p. 785, (1994), arXiv: hep-lat/9308004.

[153] S. Borsanyi *et al.*, "Ab initio calculation of the neutron-proton mass difference", Science **347**, p. 1452, (2015), arXiv: 1406.4088.

[154] A. Portelli, "Inclusion of isospin breaking effects in lattice simulations", PoS **LATTICE2014**, p. 013, (2015), arXiv: 1505.07057.

[155] A. Portelli, *Nonpertubative quantum chromodynamics and isospin symmetry breaking*. PhD thesis (2012). `http://www.theses.fr/2012AIXM4110`, Thèse de doctorat dirigée par Lellouch, Laurent Physique Théorique et Mathématique Aix-Marseille 2012.

[156] L. Lellouch, "Including isospin breaking in lattice QCD calculations: QED (+ QCD) in a finite volume", 2015. `https://www.benasque.org/2015qcd/talks_contr/063_Lellouch.pdf`.

[157] A. Patella, "QED Corrections to Hadronic Observables", PoS **LATTICE2016**, p. 020, (2017), arXiv: 1702.03857.

[158] A. Duncan, E. Eichten, H. Thacker, "Electromagnetic splittings and light quark masses in lattice QCD", Phys. Rev. Lett. **76**, p. 3894, (1996), arXiv: hep-lat/9602005.

[159] Z. Davoudi, J. Harrison, A. Jüttner, A. Portelli, M. J. Savage, "Theoretical aspects of quantum electrodynamics in a finite volume with periodic boundary conditions", Phys. Rev. D **99**, p. 034510, (2019), arXiv: 1810.05923.

[160] M. Hayakawa, S. Uno, "QED in finite volume and finite size scaling effect on electromagnetic properties of hadrons", Prog. Theor. Phys. **120**, p. 413, (2008), arXiv: 0804.2044.

[161] B. Lucini, A. Patella, A. Ramos, N. Tantalo, "Charged hadrons in local finite-volume QED+QCD with C* boundary conditions", JHEP **02**, p. 076, (2016), arXiv: 1509.01636.

[162] M. Luscher, "Computational strategies in lattice QCD", Les Houches Summer School: Session **93**, p. 331 (2010).

[163] Z. Fodor, C. Hoelbling, "Light hadron masses from lattice QCD", Reviews of Modern Physics **84**, p. 449 (2012).

[164] L. Giusti, C. Hoelbling, M. Lüscher, H. Wittig, "Numerical techniques for lattice QCD in the ϵ-regime", Computer Physics Communications **153**, p. 31, (2003), `https://www.sciencedirect.com/science/article/pii/S0010465502008743`.

[165] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, "Equation of state calculations by fast computing machines", J. Chem. Phys. **21**, p. 1087, (1953).

[166] C. Gattringer, C. B. Lang, *Quantum chromodynamics on the lattice*, vol788. Berlin: Springer, 2010, doi: 10.1007/978-3-642-01850-3.

[167] U. Wolff, "Monte Carlo errors with less errors", Comput. Phys. Commun. **156**, p. 143, (2004), arXiv: hep-lat/0306017. [Erratum: Comput.Phys.Commun. 176, 383 (2007)].

[168] S. Duane, J. B. Kogut, "Hybrid Stochastic Differential Equations Applied to Quantum Chromodynamics", Phys. Rev. Lett. **55**, p. 2774, (1985).

[169] S. Duane, J. B. Kogut, "The Theory of Hybrid Stochastic Algorithms", Nucl. Phys. B **275**, p. 398, (1986).

[170] S. Duane, A. D. Kennedy, B. J. Pendleton, D. Roweth, "Hybrid Monte Carlo", Phys. Lett. B **195**, p. 216, (1987).

[171] D. H. Weingarten, D. N. Petcher, "Monte Carlo integration for lattice gauge theories with fermions", Physics Letters B **99**, p. 333, (1981).

[172] A. D. Kennedy, I. Horvath, S. Sint, "A New exact method for dynamical fermion computations with nonlocal actions", Nucl. Phys. B Proc. Suppl. **73**, p. 834, (1999), arXiv: hep-lat/9809092.

[173] M. A. Clark, A. D. Kennedy, "The RHMC algorithm for two flavors of dynamical staggered fermions", Nucl. Phys. B Proc. Suppl. **129**, p. 850, (2004), arXiv: hep-lat/0309084.

[174] M. A. Clark, A. D. Kennedy, "Accelerating dynamical fermion computations using the rational hybrid Monte Carlo (RHMC) algorithm with multiple pseudofermion fields", Phys. Rev. Lett. **98**, p. 051601, (2007), arXiv: hep-lat/0608015.

[175] H. Yin, R. D. Mawhinney, "Improving DWF Simulations: the Force Gradient Integrator and the Möbius Accelerated DWF Solver", PoS **LATTICE2011**, p. 051, (2011), arXiv: 1111.5059.

[176] J. C. Sexton, D. H. Weingarten, "Hamiltonian evolution for the hybrid Monte Carlo algorithm", Nucl. Phys. B **380**, p. 665, (1992).

[177] M. Hasenbusch, "Speeding up the hybrid Monte Carlo algorithm for dynamical fermions", Phys. Lett. B **519**, p. 177, (2001), arXiv: hep-lat/0107019.

[178] P. Deuflhard, A. Hohmann, *Numerische Mathematik 1: Eine algorithmisch orientierte Einführung.* De Gruyter Studium, De Gruyter, 2018.

[179] Y. Saad, *Iterative methods for sparse linear systems.* SIAM, 2003.

[180] F. R. Hestenes, E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems", Journal of Research of the National Bureau of Standards **49**, p. 409, (1952).

[181] H. A. Van der Vorst, *Iterative Krylov methods for large linear systems.* No. 13, Cambridge University Press, 2003.

[182] H. Neff, N. Eicker, T. Lippert, J. W. Negele, K. Schilling, "On the low fermionic eigenmode dominance in QCD on the lattice", Phys. Rev. **D64**, p. 114509, (2001), arXiv: hep-lat/0106016.

[183] L. Giusti, P. Hernandez, M. Laine, P. Weisz, H. Wittig, "Low-energy couplings of QCD from current correlators near the chiral limit", JHEP **04**, p. 013, (2004), arXiv: hep-lat/0402002.

[184] T. A. DeGrand, S. Schaefer, "Improving meson two point functions in lattice QCD", Comput. Phys. Commun. **159**, p. 185, (2004), arXiv: hep-lat/0401011.

[185] M. Lüscher, "Solution of the Dirac equation in lattice QCD using a domain decomposition method", Computer physics communications **156**, p. 209 (2004).

[186] M. Cè, L. Giusti, S. Schaefer, "Domain decomposition, multi-level integration and exponential noise reduction in lattice QCD", Phys. Rev. D **93**, p. 094507, (2016), arXiv: 1601.04587.

[187] J. Brannick, R. C. Brower, M. A. Clark, J. C. Osborn, C. Rebbi, "Adaptive Multigrid Algorithm for Lattice QCD", Phys. Rev. Lett. **100**, p. 041601, (2008), arXiv: 0707.4018.

[188] M. Cè, L. Giusti, S. Schaefer, "A local factorization of the fermion determinant in lattice QCD", Phys. Rev. D **95**, p. 034503, (2017), arXiv: 1609.02419.

[189] C. Morningstar, M. J. Peardon, "Analytic smearing of SU(3) link variables in lattice QCD", Phys. Rev. D **69**, p. 054501, (2004), arXiv: hep-lat/0311018.

[190] S. Durr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, K. K. Szabo, G. Vulvert, "Lattice QCD at the physical point: Simulation and analysis details", JHEP **08**, p. 148, (2011), arXiv: 1011.2711.

[191] S. Capitani, S. Durr, C. Hoelbling, "Rationale for UV-filtered clover fermions", JHEP **11**, p. 028, (2006), arXiv: hep-lat/0607006.

[192] S. Durr, Z. Fodor, C. Hoelbling, T. Kurth, "Precision study of the SU(3) topological susceptibility in the continuum", JHEP **04**, p. 055, (2007), arXiv: hep-lat/0612021.

[193] S. Durr, C. Hoelbling, U. Wenger, "Staggered eigenvalue mimicry", Phys. Rev. D **70**, p. 094502, (2004), arXiv: hep-lat/0406027.

[194] R. Abbott *et al.*, "Sampling QCD field configurations with gauge-equivariant flow models", PoS **LATTICE2022**, p. 036, (2023), arXiv: 2208.03832.

[195] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, P. E. Shanahan, "Equivariant flow-based sampling for lattice gauge theory", Phys. Rev. Lett. **125**, p. 121601, (2020), arXiv: 2003.06413.

[196] U. Wolff, "High precision simulations with fast algorithms", in *Computational Methods in Field Theory: Proceedings of the 31. Internationale Universitätswochen für Kern-und Teilchenphysik Schladming, Austria, February 1992*, p. 127, Springer, 2005.

[197] L. Varnhorst, *Aspects of quark mass dependence in lattice QCD.* PhD thesis, Wuppertal U. (2020).

[198] H. Akaike, "A new look at the statistical model identification", IEEE transactions on automatic control **19**, p. 716 (1974).

[199] E. T. Neil, J. W. Sitison, "Improved information criteria for Bayesian model averaging in lattice field theory", arXiv: 2208.14983.

[200] L. Morel, Z. Yao, P. Cladé, S. Guellati-Khélifa, "Determination of the fine-structure constant with an accuracy of 81 parts per trillion", Nature **588**, p. 61 (2020).

[201] R. H. Parker, C. Yu, W. Zhong, B. Estey, H. Müller, "Measurement of the fine-structure constant as a test of the Standard Model", Science **360**, p. 191 (2018).

[202] D. Hanneke, S. Fogwell, G. Gabrielse, Physical Review Letters **100**, "New Measurement of the Electron Magnetic Moment and the Fine Structure Constant",, (mar 2008), https://doi.org/10.1103%2Fphysrevlett.100.120801.

[203] F. Jegerlehner, "The Running fine structure constant alpha(E) via the Adler function", Nucl. Phys. Proc. Suppl. **181-182**, p. 135, (2008), arXiv: 0807.4206.

[204] J. M. Cornwall, "Dynamical Mass Generation in Continuum QCD", Phys. Rev. D **26**, p. 1453, (1982).

[205] J. Erler, "Calculation of the QED coupling $\alpha(M_Z)$ in the modified minimal-subtraction scheme", Phys. Rev. D **59**, p. 054008, (Feb 1999), https://link.aps.org/doi/10.1103/PhysRevD.59.054008.

[206] S. Fanchiotti, B. A. Kniehl, A. Sirlin, "Incorporation of QCD effects in basic corrections of the electroweak theory", Phys. Rev. D **48**, p. 307, (1993), arXiv: hep-ph/9212285.

[207] A. Sirlin, "Radiative Corrections in the SU(2)-L x U(1) Theory: A Simple Renormalization Framework", Phys. Rev. D **22**, p. 971, (1980).

[208] C. Sturm, "Leptonic contributions to the effective electromagnetic coupling at four-loop order in QED", Nuclear Physics B **874**, p. 698, (sep 2013), https://doi.org/10.1016%2Fj.nuclphysb.2013.06.009.

[209] M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, "Reevaluation of the hadronic vacuum polarisation contributions to the Standard Model predictions of the muon $g-2$ and $\alpha(m_Z^2)$ using newest hadronic cross-section data", Eur. Phys. J. C **77**, p. 827, (2017), arXiv: 1706.09436.

[210] M. Steinhauser, "Leptonic contribution to the effective electromagnetic coupling constant up to three loops", Phys. Lett. B **429**, p. 158, (1998), arXiv: hep-ph/9803313.

[211] J. Kühn, M. Steinhauser, "A perturbation theory driven analysis of the effective QED coupling at MZ", Physics Letters B **437**, p. 425 (1998).

[212] K. Chetyrkin, J. Kühn, M. Steinhauser, "Heavy quark vacuum polarisation to three loops", Physics Letters B **371**, p. 93 (1996).

[213] K. Chetyrkin, J. Kühn, M. Steinhauser, "Heavy quark current correlators to O (αs2", Nuclear Physics B **505**, p. 40 (1997).

[214] K. Chetyrkin, J. Kühn, M. Steinhauser, "Three-loop polarization function and O (αs2) corrections to the production of heavy quarks", Nuclear Physics B **482**, p. 213 (1996).

[215] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, A. Strumia, "Higgs mass and vacuum stability in the Standard Model at NNLO", JHEP **08**, p. 098, (2012), arXiv: 1205.6497.

[216] J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, A. Riotto, A. Strumia, "Higgs mass implications on the stability of the electroweak vacuum", Phys. Lett. B **709**, p. 222, (2012), arXiv: 1112.3022.

[217] H. Flacher, M. Goebel, J. Haller, A. Hocker, K. Monig, J. Stelzer, "Revisiting the Global Electroweak Fit of the Standard Model and Beyond with Gfitter", Eur. Phys. J. C **60**, p. 543, (2009), arXiv: 0811.0009. [Erratum: Eur.Phys.J.C 71, 1718 (2011)].

[218] J. Haller, A. Hoecker, R. Kogler, K. Mönig, T. Peiffer, J. Stelzer, "Update of the global electroweak fit and constraints on two-Higgs-doublet models", Eur. Phys. J. C **78**, p. 675, (2018), arXiv: 1803.01853.

[219] C.-T. Lu, L. Wu, Y. Wu, B. Zhu, "Electroweak precision fit and new physics in light of the W boson mass", Phys. Rev. D **106**, p. 035034, (2022), arXiv: 2204.03796.

[220] J. de Blas, M. Pierini, L. Reina, L. Silvestrini, "Impact of the Recent Measurements of the Top-Quark and W-Boson Masses on Electroweak Precision Fits", Phys. Rev. Lett. **129**, p. 271801, (2022), arXiv: 2204.04204.

[221] *Theory report on the 11th FCC-ee workshop*, 2019.

[222] P. Janot, "Direct measurement of $\alpha_{QED}(m_Z^2)$ at the FCC-ee", JHEP **02**, p. 053, (2016), arXiv: 1512.05544. [Erratum: JHEP 11, 164 (2017)].

[223] N. Chen, T. Han, S. Su, W. Su, Y. Wu, "Type-II 2HDM under the Precision Measurements at the Z-pole and a Higgs Factory", JHEP **03**, p. 023, (2019), arXiv: 1808.02037.

[224] M. Baak *et al.*, "Working Group Report: Precision Study of Electroweak Interactions", in *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013.

[225] S. Borsanyi *et al.*, "Calculation of the axion mass based on high-temperature lattice quantum chromodynamics", Nature **539**, p. 69, (2016), arXiv: 1606.07494.

[226] F. Burger, K. Jansen, M. Petschlies, G. Pientka, "Leading hadronic contributions to the running of the electroweak coupling constants from lattice QCD", JHEP **11**, p. 215, (2015), arXiv: 1505.03283.

[227] F. Jegerlehner, *The Anomalous Magnetic Moment of the Muon*, vol274. Cham: Springer, 2017, doi: 10.1007/978-3-319-63577-4.

[228] M. Peskin, D. Schroeder, *An Introduction To Quantum Field Theory*. Frontiers in Physics, Avalon Publishing, 1995.

[229] M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, "Hadron Contribution to Vacuum Polarisation", `https://cds.cern.ch/record/2217138`.

[230] S. L. Adler, "Some simple vacuum-polarization phenomenology: $e^+e^- \to$ hadrons; the muonic-atom x-ray discrepancy and $g_\mu - 2$", Phys. Rev. D **10**, p. 3714, (Dec 1974), `https://link.aps.org/doi/10.1103/PhysRevD.10.3714`.

[231] S. Eidelman, F. Jegerlehner, A. Kataev, O. Veretin, "Testing non-perturbative strong interaction effects via the Adler function", Physics Letters B **454**, p. 369, (1999), `https://www.sciencedirect.com/science/article/pii/S0370269399003895`.

[232] M. Cè, T. Harris, H. B. Meyer, A. Toniato, C. Török, "Vacuum correlators at short distances from lattice QCD", JHEP **12**, p. 215, (2021), arXiv: 2106.15293.

[233] A. Francis, V. Guelpers, G. Herdoíza, H. Horch, B. Jäger, H. B. Meyer, H. Wittig, "Study of the hadronic contributions to the running of the QED coupling and the weak mixing angle", PoS **LATTICE2015**, p. 110 (2015), arXiv: 1511.04751.

[234] C. McNeile, C. T. H. Davies, E. Follana, K. Hornbostel, G. P. Lepage, "High-Precision c and b Masses, and QCD Coupling from Current-Current Correlators in Lattice and Continuum QCD", Phys. Rev. D **82**, p. 034512, (2010), arXiv: 1004.4285.

[235] Y. Aoki *et al.*, "FLAG Review 2021", Eur. Phys. J. C **82**, p. 869, (2022), arXiv: 2111.09849.

[236] T. Takaishi, "Heavy quark potential and effective actions on blocked configurations", Phys. Rev. D **54**, p. 1050, (1996).

[237] N. Ishizuka, M. Fukugita, H. Mino, M. Okawa, A. Ukawa, "Operator dependence of hadron masses for Kogut-Susskind quarks on the lattice", Nucl.Phys. **B411**, p. 875, (Jan 1994), `http://dx.doi.org/10.1016/0550-3213(94)90475-8`.

[238] A. Francis, B. Jaeger, H. B. Meyer, H. Wittig, "A new representation of the Adler function for lattice QCD", Phys. Rev. **D88**, p. 054502, (2013), arXiv: 1306.2532.

[239] N. Tantalo, "Isospin Breaking Effects on the Lattice", PoS **LATTICE2013**, p. 007, (2014), arXiv: 1311.2797.

[240] G. de Divitiis, P. Dimopoulos, R. Frezzotti, V. Lubicz, G. Martinelli *et al.*, "Isospin breaking effects due to the up-down mass difference in Lattice QCD", JHEP **1204**, p. 124, (2012), arXiv: 1110.6294.

[241] G. M. de Divitiis, R. Frezzotti, V. Lubicz, G. Martinelli, R. Petronzio, G. C. Rossi, F. Sanfilippo, S. Simula, N. Tantalo, "Leading isospin breaking effects on the lattice", Phys.Rev. **D87**, p. 114505, (jun 2013), arXiv: 1303.4896.

[242] W. A. Bardeen, A. Duncan, E. Eichten, H. Thacker, "2-dimensional QED as a testbed for interpolations between quenched and full QCD", Nucl. Phys. B Proc. Suppl. **63**, p. 811, (1998), arXiv: hep-lat/9710020.

[243] T. Blum, T. Doi, M. Hayakawa, T. Izubuchi, N. Yamada, "Determination of light quark masses from the electromagnetic splitting of pseudoscalar meson masses computed with two flavors of domain wall fermions", Phys. Rev. D **76**, p. 114508, (2007), arXiv: 0708.0484.

[244] T. Blum, R. Zhou, T. Doi, M. Hayakawa, T. Izubuchi, S. Uno, N. Yamada, "Electromagnetic mass splittings of the low lying hadrons and quark masses from 2+1 flavor lattice QCD+QED", Phys. Rev. D **82**, p. 094508, (2010), arXiv: 1006.1311.

[245] S. Borsanyi *et al.*, "Isospin splittings in the light baryon octet from lattice QCD and QED", Phys. Rev. Lett. **111**, p. 252001, (2013), arXiv: 1306.2287.

[246] A. Portelli *et al.*, "Electromagnetic corrections to light hadron masses", PoS **LATTICE2010**, p. 121, (2010), arXiv: 1011.4189.

[247] R. Zhou, S. Gottlieb, "Dynamical QCD+QED simulation with staggered quarks", PoS **LATTICE2014**, p. 024, (2014), arXiv: 1411.4115.

[248] L. Parato *et al.*, "QED and strong isospin corrections in the hadronic vacuum polarization contribution to the anomalous magnetic moment of the muon", PoS **LATTICE2021**, p. 358, (2022), arXiv: 2202.05807.

[249] G. T. Fleming, S. D. Cohen, H.-W. Lin, V. Pereyra, "Excited-State Effective Masses in Lattice QCD", Phys. Rev. D **80**, p. 074506, (2009), arXiv: 0903.2314.

[250] Z. Fodor, C. Hoelbling, S. Krieg, L. Lellouch, T. Lippert, A. Portelli, A. Sastre, K. K. Szabo, L. Varnhorst, "Up and down quark masses and corrections to Dashen's theorem from lattice QCD and quenched QED", Phys. Rev. Lett. **117**, p. 082001, (2016), arXiv: 1604.07112.

[251] N. Husung, P. Marquard, R. Sommer, "Asymptotic behavior of cutoff effects in Yang–Mills theory and in Wilson's lattice QCD", Eur. Phys. J. C **80**, p. 200, (2020), arXiv: 1912.08498.

[252] S. Mutzel, "Hadronic contribution to the running of the electromagnetic coupling constant from lattice QCD: challenges at short distances", PoS **LATTICE2022**, p. 324, (2023).

[253] B. Chakraborty, C. T. H. Davies, P. G. de Oliviera, J. Koponen, G. P. Lepage, R. S. Van de Water, "The hadronic vacuum polarization contribution to a_μ from full lattice QCD", Phys. Rev. D **96**, p. 034516, (2017), arXiv: 1601.03071.

[254] J. J. Sakurai, "Theory of strong interactions", Annals Phys. **11**, p. 1, (1960).

[255] F. Jegerlehner, R. Szafron, "$\rho^0 - \gamma$ mixing in the neutral channel pion form factor F_π^e and its role in comparing e^+e^- with τ spectral functions", Eur. Phys. J. C **71**, p. 1632, (2011), arXiv: 1101.2872.

[256] G. Colangelo, S. Durr, C. Haefeli, "Finite volume effects for meson masses and decay constants", Nucl. Phys. B **721**, p. 136, (2005), arXiv: hep-lat/0503014.

[257] Z. Davoudi, M. J. Savage, "Finite-Volume Electromagnetic Corrections to the Masses of Mesons, Baryons and Nuclei", Phys. Rev. D **90**, p. 054503, (2014), arXiv: 1402.6741.

[258] S. Borsanyi *et al.*, "Ab initio calculation of the neutron-proton mass difference", Science **347**, p. 1452, (2015), arXiv: 1406.4088.

[259] J. Bijnens, J. Harrison, N. Hermansson-Truedsson, T. Janowski, A. Jüttner, A. Portelli, "Electromagnetic finite-size effects to the hadronic vacuum polarization", Phys. Rev. D **100**, p. 014508, (2019), arXiv: 1903.10591.

[260] J. Bijnens, N. Danielsson, "Electromagnetic Corrections in Partially Quenched Chiral Perturbation Theory", Phys. Rev. **D75**, p. 014505, (2007), arXiv: hep-lat/0610127.

[261] M. F. L. Golterman, J. Smit, "Lattice Baryons With Staggered Fermions", Nucl. Phys. **B255**, p. 328, (1985).

[262] N. Husung, "Logarithmic corrections to O(a) and O(a^2) effects in lattice QCD with Wilson or Ginsparg–Wilson quarks", Eur. Phys. J. C **83**, p. 142, (2023), arXiv: 2206.03536.

[263] N. Husung, P. Marquard, R. Sommer, "The asymptotic approach to the continuum of lattice QCD spectral observables", Phys. Lett. B **829**, p. 137069, (2022), arXiv: 2111.02347.

[264] N. A. Husung, *Logarithmic corrections in Symanzik's effective theory of lattice QCD*. PhD thesis, Humboldt U., Berlin, Humboldt U., Berlin (8 2021).

[265] N. Husung, "Logarithmic corrections to a^n behaviour", 2020. `https://indico.cern.ch/event/956699/contributions/4108682/attachments/2145724/3616679/SymanzikCompressed.pdf`.

[266] N. Tantalo, "Matching lattice QC+ED to Nature", PoS **LATTICE2022**, p. 249, (2023), arXiv: 2301.02097.

[267] S. Durr *et al.*, "Ab-Initio Determination of Light Hadron Masses", Science **322**, p. 1224, (2008), arXiv: 0906.3599.

[268] F. James, "MINUIT: function minimization and error analysis reference manual", tech. rep., Cern, 1998.

[269] P. Peter, J. Uzan, *Primordial Cosmology*. Oxford Graduate Texts, OUP Oxford, 2013.

[270] Planck Collaboration, "Planck 2018 results - VI. Cosmological parameters", A&A **641**, p. A6, (2020), `https://doi.org/10.1051/0004-6361/201833910`.

[271] G. Bertone, D. Hooper, "History of dark matter", Rev. Mod. Phys. **90**, p. 045002, (2018), arXiv: 1605.04909.

[272] A. Bosma, "The dark matter problem", Celestial Mech. **72**, p. 69, (1998), arXiv: astro-ph/9812015.

[273] K. Freese, "Status of Dark Matter in the Universe", Int. J. Mod. Phys. **1**, p. 325, (2017), arXiv: 1701.01840.

[274] M. Bauer, T. Plehn, *Yet Another Introduction to Dark Matter: The Particle Physics Approach*, vol959 of *Lecture Notes in Physics*. Springer, 2019, doi: 10.1007/978-3-030-16234-4.

[275] G. Bertone, D. Hooper, J. Silk, "Particle dark matter: Evidence, candidates and constraints", Phys. Rept. **405**, p. 279, (2005), arXiv: hep-ph/0404175.

[276] W. Kapferer, *Das Rätsel Dunkle Materie: Auf der Suche nach dem Unsichtbaren*. Springer Berlin Heidelberg, 2017.

[277] A. Goudelis, *Detection of WIMP-like dark matter in some extensions of the Standard Model*. Phd thesis, Universite Paris Sud 11 (2011).

[278] A. Lenoci, *Lower mass bounds on FIMP Dark Matter*. Master thesis, Università degli studi di Padova (2021).

[279] A. M. Green, "Dark matter in astrophysics/cosmology", SciPost Phys. Lect. Notes, p. 37, (2022), `https://scipost.org/10.21468/SciPostPhysLectNotes.37`.

[280] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln", Helvetica Physica Acta **6**, p. 110 (1933).

[281] P. F. de Salas, M. Lattanzi, G. Mangano, G. Miele, S. Pastor, O. Pisanti, "Bounds on very low reheating scenarios after Planck", Phys. Rev. D **92**, p. 123534, (2015), arXiv: 1511.00672.

[282] A. D. Linde, *Particle physics and inflationary cosmology*, vol5. 1990.

[283] A. D. Sakharov, "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe", Soviet Physics Uspekhi **34**, p. 392, (may 1991), `https://dx.doi.org/10.1070/PU1991v034n05ABEH002497`.

[284] D. Toussaint, S. B. Treiman, F. Wilczek, A. Zee, "Matter-antimatter accounting, thermodynamics, and black-hole radiation", Phys. Rev. D **19**, p. 1036, (Feb 1979), `https://link.aps.org/doi/10.1103/PhysRevD.19.1036`.

[285] J. M. Cline, "Baryogenesis", in *Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime*, 9 2006.

[286] M. Laine, K. Rummukainen, "What's new with the electroweak phase transition?", Nucl. Phys. B Proc. Suppl. **73**, p. 180, (1999), arXiv: hep-lat/9809045.

[287] G. Aarts *et al.*, "Phase Transitions in Particle Physics - Results and Perspectives from Lattice Quantum Chromo-Dynamics", in *Phase Transitions in Particle Physics: Results and Perspectives from Lattice Quantum Chromo-Dynamics*, 1 2023.

[288] J. N. Guenther, "An overview of the QCD phase diagram at finite T and μ", PoS **LATTICE2021**, p. 013, (2022), arXiv: 2201.02072.

[289] B. D. Fields, P. Molaro, S. Sarkar, "Big-Bang Nucleosynthesis", Chin. Phys. C **38**, p. 339 (2014), arXiv: 1412.1408.

[290] D. Tong, "Lectures on cosmology", Lecture notes, DAMTP Cambridge, `http://www.damtp.cam.ac.uk/user/tong/cosmo.html`.

[291] J. B. Muñoz, A. Loeb, "A small amount of mini-charged dark matter could cool the baryons in the early Universe", Nature **557**, p. 684 (2018).

[292] T. R. Slatyer, "Les Houches Lectures on Indirect Detection of Dark Matter", SciPost Phys. Lect. Notes **53**, p. 1, (2022), arXiv: 2109.02696.

[293] E. A. Baltz, "Dark matter candidates", arXiv preprint astro-ph/0412170 (2004).

[294] S. Tremaine, J. E. Gunn, "Dynamical Role of Light Neutral Leptons in Cosmology", Phys. Rev. Lett. **42**, p. 407, (Feb 1979), `https://link.aps.org/doi/10.1103/PhysRevLett.42.407`.

[295] J. Alvey, N. Sabti, V. Tiki, D. Blas, K. Bondarenko, A. Boyarsky, M. Escudero, M. Fairbairn, M. Orkney, J. I. Read, "New constraints on the mass of fermionic dark matter from dwarf spheroidal galaxies", Mon. Not. Roy. Astron. Soc. **501**, p. 1188, (2021), arXiv: 2010.03572.

[296] D. S. Akerib *et al.*, "Snowmass2021 Cosmic Frontier Dark Matter Direct Detection to the Neutrino Fog", in *Snowmass 2021*, 3 2022.

[297] T. Lin, "Sub-GeV dark matter models and direct detection", SciPost Phys. Lect. Notes, p. 43, (2022), https://scipost.org/10.21468/SciPostPhysLectNotes.43.

[298] A. Mitridate, T. Trickle, Z. Zhang, K. M. Zurek, "Snowmass White Paper: Light Dark Matter Direct Detection at the Interface With Condensed Matter Physics", 2022.

[299] G. B. Gelmini, V. Takhistov, S. J. Witte, "Casting a Wide Signal Net with Future Direct Dark Matter Detection Experiments", JCAP **07**, p. 009, (2018), arXiv: 1804.01638. [Erratum: JCAP 02, E02 (2019)].

[300] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, Y. Xu, "The effective field theory of dark matter direct detection", Journal of Cosmology and Astroparticle Physics **2013**, p. 004, (feb 2013), https://doi.org/10.1088%2F1475-7516%2F2013%2F02%2F004.

[301] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, Y. Xu, "Model Independent Direct Detection Analyses", arXiv: 1211.2818.

[302] J. Brod, A. Gootjes-Dreesbach, M. Tammaro, J. Zupan, "Effective Field Theory for Dark Matter Direct Detection up to Dimension Seven", JHEP **10**, p. 065, (2018), arXiv: 1710.10218.

[303] F. Bishara, J. Brod, B. Grinstein, J. Zupan, "Chiral Effective Theory of Dark Matter Direct Detection", JCAP **02**, p. 009, (2017), arXiv: 1611.00368.

[304] F. Bishara, J. Brod, B. Grinstein, J. Zupan, "DirectDM: a tool for dark matter direct detection", arXiv: 1708.02678.

[305] E. Aprile *et al.*, "Excess electronic recoil events in XENON1T", Phys. Rev. D **102**, p. 072004, (2020), arXiv: 2006.09721.

[306] E. Aprile *et al.*, "Search for New Physics in Electronic Recoil Data from XENONnT", Phys. Rev. Lett. **129**, p. 161805, (2022), arXiv: 2207.11330.

[307] R. Bernabei, P. Belli, F. Cappella, R. Cerulli, C. J. Dai, A. d'Angelo, H. L. He, A. Incicchitti, H. H. Kuang, J. M. Ma, F. Montecchia, F. Nozzoli, D. Prosperi, X. D. Sheng, Z. P. Ye, "First results from DAMA/LIBRA and the combined results with DAMA/NaI", The European Physical Journal C **56**, p. 333, (aug 2008), https://doi.org/10.1140%2Fepjc%2Fs10052-008-0662-y.

[308] J. Amaré, S. Cebrián, D. Cintas, I. Coarasa, E. García, M. Martínez, M. A. Oliván, Y. Ortigoza, A. O. de Solórzano, J. Puimedón, A. Salinas, M. L. Sarsa, P. Villar, "Annual modulation results from three-year exposure of ANAIS-112", Phys. Rev. D **103**, p. 102005, (May 2021), https://link.aps.org/doi/10.1103/PhysRevD.103.102005.

[309] G. Adhikari *et al.*, "Three-year annual modulation search with COSINE-100", Phys. Rev. D **106**, p. 052005, (Sep 2022), https://link.aps.org/doi/10.1103/PhysRevD.106.052005.

[310] G. Adhikari *et al.*, "Strong constraints from COSINE-100 on the DAMA dark matter results using the same sodium iodide target", Sci. Adv. **7**, p. abk2699, (2021), arXiv: 2104.03537.

[311] G. Adhikari *et al.*, "An induced annual modulation signature in COSINE-100 data by DAMA/LIBRA's analysis method", Sci. Rep. **13**, p. 4676, (2023), arXiv: 2208.05158.

[312] R. Agnese *et al.*, Physical Review D **99**, "Search for low-mass dark matter with CDMSlite using a profile likelihood fit",, (Mar 2019), http://dx.doi.org/10.1103/PhysRevD.99.062001.

[313] Z. Z. Liu *et al.*, "Studies of the Earth shielding effect to direct dark matter searches at the China Jinping Underground Laboratory", Phys. Rev. D **105**, p. 052005, (Mar 2022), https://link.aps.org/doi/10.1103/PhysRevD.105.052005.

[314] J. Aalbers *et al.*, "First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment", 2022. https://arxiv.org/abs/2207.03764.

[315] I. Alkhatib, D. W. P. Amaral, T. Aralis, T. Aramaki, I. J. Arnquist, I. Ataee Langroudy, E. Azadbakht, S. Banik, D. Barker, C. Bathurst *et al.*, Physical Review Letters **127**, "Light Dark Matter Search with a High-Resolution Athermal Phonon Detector Operated above Ground",, (Aug 2021), http://dx.doi.org/10.1103/PhysRevLett.127.061801.

[316] A. Aguilar-Arevalo *et al.*, "Results on low-mass weakly interacting massive particles from a 11 kg-day target exposure of DAMIC at SNOLAB", 2020.

[317] H. B. Li *et al.*, "Limits on Spin-Independent Couplings of WIMP Dark Matter with a p -Type Point-Contact Germanium Detector", Phys. Rev. Lett. **110**, p. 261301, (jun 2013), http://link.aps.org/doi/10.1103/PhysRevLett.110.261301.

[318] P. Agnes *et al.*, "Search for Dark-Matter–Nucleon Interactions via Migdal Effect with DarkSide-50", Phys. Rev. Lett. **130**, p. 101001, (Mar 2023), https://link.aps.org/doi/10.1103/PhysRevLett.130.101001.

[319] "Supercdms dark matter limit plotter." https://supercdms.slac.stanford.edu/dark-matter-limit-plotter. Accessed: 2023-07-30.

[320] T. Hambye, M. H. Tytgat, J. Vandecasteele, L. Vanderheyden, "Dark matter direct detection is testing freeze-in", Physical Review D **98**, p. 075017 (2018).

[321] S. Heeba, F. Kahlhoefer, "Probing the freeze-in mechanism in dark matter models with $U(1)'$ gauge extensions", Physical Review D **101**, p. 035043 (2020).

[322] "Dark matter summary plots for s-channel, 2HDM+a, Higgs portal and Dark Higgs models",

[323] R. Essig, J. Mardon, M. Papucci, T. Volansky, Y.-M. Zhong, "Constraining light dark matter with low-energy e+e- colliders", Journal of High Energy Physics **2013**, p. 167 (Nov 2013).

[324] G. Bélanger *et al.*, "LHC-friendly minimal freeze-in models", JHEP **02**, p. 186, (2019), arXiv: 1811.05478.

[325] C. Pérez de los Heros, "Status, Challenges and Directions in Indirect Dark Matter Searches", Symmetry **12**, p. 1648, (2020), arXiv: 2008.11561.

[326] T. Aramaki *et al.*, "Snowmass2021 Cosmic Frontier: The landscape of cosmic-ray and high-energy photon probes of particle dark matter", arXiv: 2203.06894.

[327] R. K. Leane *et al.*, "Snowmass2021 Cosmic Frontier White Paper: Puzzling Excesses in Dark Matter Searches and How to Resolve Them", arXiv: 2203.06859.

[328] M. Heikinheimo, T. Tenkanen, K. Tuominen, "Prospects for indirect detection of frozen-in dark matter", Phys. Rev. D **97**, p. 063002, (2018), arXiv: 1801.03089.

[329] M. R. Buckley, P. J. Fox, Physical Review D **81**, "Dark matter self-interactions and light force carriers",, (Apr 2010), `http://dx.doi.org/10.1103/PhysRevD.81.083522`.

[330] S. Tulin, H.-B. Yu, K. M. Zurek, "Beyond Collisionless Dark Matter: Particle Physics Dynamics for Dark Matter Halo Structure", Phys. Rev. D **87**, p. 115007, (2013), arXiv: 1302.3898.

[331] S. Tulin, H.-B. Yu, "Dark matter self-interactions and small scale structure", Physics Reports **730**, p. 1–57, (Feb 2018), `http://dx.doi.org/10.1016/j.physrep.2017.11.004`.

[332] A. Robertson, D. Harvey, R. Massey, V. Eke, I. G. McCarthy, M. Jauzac, B. Li, J. Schaye, "Observable tests of self-interacting dark matter in galaxy clusters: cosmological simulations with SIDM and baryons", Mon. Not. Roy. Astron. Soc. **488**, p. 3646, (2019), arXiv: 1810.05649.

[333] J. J. Binney, N. W. Evans, "Cuspy dark-matter haloes and the Galaxy", Mon. Not. Roy. Astron. Soc. **327**, p. L27, (2001), arXiv: astro-ph/0108505.

[334] N. C. Relatores, A. B. Newman, J. D. Simon, R. S. Ellis, P. Truong, L. Blitz, A. Bolatto, C. Martin, M. Matuszewski, P. Morrissey, J. D. Neill, "The Dark Matter Distributions in Low-mass Disk Galaxies. II. The Inner Density Profiles", The Astrophysical Journal **887**, p. 94, (dec 2019), `https://doi.org/10.3847%2F1538-4357%2Fab5305`.

[335] J. T. Ruderman, "Dark Matter Production without Prejudice", July 2021.

[336] F. Schwabl, *Statistische Mechanik*. Springer-Lehrbuch, Springer Berlin Heidelberg, 2006.

[337] H. Stark, "Statistical Mechanics, Lecture Notes",

[338] E. W. Kolb, M. S. Turner, *The Early Universe*, vol69. 1990, doi: 10.1201/9780429492860.

[339] Y. Hochberg, E. Kuflik, R. McGehee, H. Murayama, K. Schutz, Physical Review D **98**, "Strongly interacting massive particles through the axion portal",, (dec 2018).

[340] L. Roszkowski, E. M. Sessolo, S. Trojanowski, "WIMP dark matter candidates and searches—current status and future prospects", Reports on Progress in Physics **81**, p. 066201, (may 2018).

[341] F. Elahi, C. Kolda, J. Unwin, "UltraViolet Freeze-in", JHEP **03**, p. 048, (2015), arXiv: 1410.6157.

[342] G. Villadoro, "Lecture on axions", January 2015.

[343] M. Dine, "TASI Lectures on The Strong CP Problem", Nov. 2000. arXiv:hep-ph/0011376, `http://arxiv.org/abs/hep-ph/0011376`.

[344] A. Hook, "TASI Lectures on the Strong CP Problem and Axions", July 2021. arXiv:1812.02669 [hep-ph], `http://arxiv.org/abs/1812.02669`.

[345] C. Vafa, E. Witten, "Parity Conservation in Quantum Chromodynamics", Phys. Rev. Lett. **53**, p. 535, (Aug 1984), https://link.aps.org/doi/10.1103/PhysRevLett.53.535.

[346] J. Quevillon, "Axions are blind to anomalies", 2019. https://indico.cern.ch/event/851337/contributions/3579543/attachments/1925206/3186269/quevillon_Warsaw_19.pdf.

[347] R. D. Peccei, "A short review of axions", in *High energy physics. Proceedings, 19th International Conference, ICHEP 1978, Tokyo, Japan, August 23-August 30, 1978*, p. 385, 1979.

[348] H. Murayama, G. G. Raffelt, C. Hagmann, K. van Bibber, L. J. Rosenberg, "Axions and other very light bosons: in Review of Particle Physics (RPP 1998)", Eur. Phys. J. **C3**, p. 264 (1998).

[349] J. E. Kim, "Weak-interaction singlet and strong CP invariance", Physical Review Letters **43**, p. 103 (1979).

[350] M. A. Shifman, A. Vainshtein, V. I. Zakharov, "Can confinement ensure natural CP invariance of strong interactions?", Nuclear Physics B **166**, p. 493 (1980).

[351] M. Dine, W. Fischler, M. Srednicki, "A simple solution to the strong CP problem with a harmless axion", Physics Letters B **104**, p. 199 (1981).

[352] A. Zhitnitskii, "Possible suppression of axion-hadron interactions", Sov. J. Nucl. Phys. (Engl. Transl.); (United States) **31:2** (2 1980).

[353] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, J. P. Silva, "Theory and phenomenology of two-Higgs-doublet models", Phys. Rept. **516**, p. 1, (2012), arXiv: 1106.0034.

[354] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Physics Letters B **716**, p. 30 (2012).

[355] H. Georgi, D. B. Kaplan, L. Randall, "Manifesting the Invisible Axion at Low-energies", Phys. Lett. B **169**, p. 73, (1986).

[356] M. Golterman, "Applications of chiral perturbation theory to lattice QCD", in *Les Houches Summer School: Session 93: Modern perspectives in lattice QCD: Quantum field theory and high performance computing*, p. 423, 12 2009.

[357] S. Scherer, "Introduction to chiral perturbation theory", Adv. Nucl. Phys. **27**, p. 277 (2003), arXiv: hep-ph/0210398.

[358] H. Leutwyler, "On the foundations of chiral perturbation theory", Annals Phys. **235**, p. 165, (1994), arXiv: hep-ph/9311274.

[359] I. G. Irastorza, "An introduction to axions and their detection", SciPost Phys. Lect. Notes **45**, p. 1, (2022), arXiv: 2109.07376.

[360] B. Döbrich, J. Jaeckel, F. Kahlhoefer, A. Ringwald, K. Schmidt-Hoberg, "ALPtraum: ALP production in proton beam dump experiments", Journal of High Energy Physics **2016**, p. 18, (Feb 2016).

[361] K. Saikawa, S. Shirai, "Primordial gravitational waves, precisely: The role of thermodynamics in the Standard Model", JCAP **05**, p. 035, (2018), arXiv: 1803.01038.

[362] A. Salvio, A. Strumia, W. Xue, "Thermal axion production", JCAP **01**, p. 011, (2014), arXiv: 1310.6982.

[363] A. Biswas, S. Ganguly, S. Roy, "Fermionic dark matter via UV and IR freeze-in and its possible X-ray signature", JCAP **03**, p. 043, (2020), arXiv: 1907.07973.

[364] X. Chu, T. Hambye, M. H. Tytgat, "The four basic ways of creating dark matter through a portal", Journal of Cosmology and Astroparticle Physics **2012**, p. 034, (may 2012), `https://dx.doi.org/10.1088/1475-7516/2012/05/034`.

[365] J. Edsjo, P. Gondolo, "Neutralino relic density including coannihilations", Phys. Rev. D **56**, p. 1879, (1997), arXiv: hep-ph/9704361.

[366] T. Bhattacharya, V. Cirigliano, S. Cohen, R. Gupta, H.-W. Lin, B. Yoon, "Axial, Scalar and Tensor Charges of the Nucleon from 2+1+1-flavor Lattice QCD", Phys. Rev. D **94**, p. 054508, (2016), arXiv: 1606.07049.

[367] J. Bijnens, G. Colangelo, P. Talavera, "The Vector and scalar form-factors of the pion to two loops", JHEP **05**, p. 014, (1998), arXiv: hep-ph/9805389.

[368] H. A. Weldon, "Effective fermion masses of order gT in high-temperature gauge theories with exact chiral invariance", Phys. Rev. D **26**, p. 2789, (Nov 1982), `https://link.aps.org/doi/10.1103/PhysRevD.26.2789`.

[369] M. J. Dolan, T. Ferber, C. Hearty, F. Kahlhoefer, K. Schmidt-Hoberg, "Revised constraints and Belle II sensitivity for visible and invisible axion-like particles", Journal of High Energy Physics **2017**, p. 94, (Dec 2017).

[370] M. Pospelov, A. Ritz, M. Voloshin, Physical Review D **78**, "Bosonic super-WIMPs as keV-scale dark matter",, (Dec 2008), `http://dx.doi.org/10.1103/PhysRevD.78.115012`.

[371] G. G. Raffelt, "Astrophysical axion bounds diminished by screening effects", Phys. Rev. D **33**, p. 897, (Feb 1986), `https://link.aps.org/doi/10.1103/PhysRevD.33.897`.

[372] D. Cadamuro, J. Redondo, "Cosmological bounds on pseudo Nambu-Goldstone bosons", Journal of Cosmology and Astroparticle Physics **2012**, p. 032, (feb 2012), `https://doi.org/10.1088/1475-7516/2012/02/032`.

[373] J. H. Chang, R. Essig, S. D. McDermott, "Revisiting Supernova 1987A Constraints on Dark Photons", JHEP **01**, p. 107, (2017), arXiv: 1611.03864.

[374] G. Grilli di Cortona, E. Hardy, J. Pardo Vega, G. Villadoro, "The QCD axion, precisely", JHEP **01**, p. 034, (2016), arXiv: 1511.02867.

[375] S. Hannestad, G. Raffelt, "Supernova neutrino opacity from nucleon-nucleon Bremsstrahlung and related processes", Astrophys. J. **507**, p. 339, (1998), arXiv: astro-ph/9711132.

[376] A. Bartl, R. Bollig, H.-T. Janka, A. Schwenk, "Impact of Nucleon-Nucleon Bremsstrahlung Rates Beyond One-Pion Exchange", Phys. Rev. D **94**, p. 083009, (2016), arXiv: 1608.05037.

[377] J. D. Bjorken, S. Ecklund, W. R. Nelson, A. Abashian, C. Church, B. Lu, L. W. Mo, T. A. Nunamaker, P. Rassmann, "Search for neutral metastable penetrating particles produced in the SLAC beam dump", Phys. Rev. D **38**, p. 3375, (Dec 1988), `https://link.aps.org/doi/10.1103/PhysRevD.38.3375`.

[378] V. Budnev, I. Ginzburg, G. Meledin, V. Serbo, "The two-photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation", Physics Reports **15**, p. 181, (1975), `https://www.sciencedirect.com/science/article/pii/0370157375900095`.

[379] W. R. Nelson, H. Hirayama, D. Rogers, "SLAC Report 265",

[380] L. Darmé, F. Giacchino, E. Nardi, M. Raggi, Journal of High Energy Physics **2021**, "Invisible decays of axion-like particles: constraints and prospects",, (Jun 2021), `http://dx.doi.org/10.1007/JHEP06(2021)009`.

[381] L. Marsicano, M. Battaglieri, M. Bondi', C. D. R. Carvajal, A. Celentano, M. De Napoli, R. De Vita, E. Nardi, M. Raggi, P. Valente, "Dark photon production through positron annihilation in beam-dump experiments", Phys. Rev. D **98**, p. 015031, (2018), arXiv: 1802.03794.

[382] G. Aad *et al.*, "Search for Higgs Boson Decays into a Z Boson and a Light Hadronically Decaying Resonance Using 13 TeV pp Collision Data from the ATLAS Detector", Phys. Rev. Lett. **125**, p. 221802, (2020), arXiv: 2004.01678.

[383] G. Aad *et al.*, "Search for Higgs boson decays into two new low-mass spin-0 particles in the $4b$ channel with the ATLAS detector using pp collisions at $\sqrt{s}$ = 13 TeV", Phys. Rev. D **102**, p. 112006, (2020), arXiv: 2005.12236.

[384] G. Aad *et al.*, "Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV", arXiv: 2110.13673.

[385] M. Aaboud *et al.*, "Search for Higgs boson decays into a pair of light bosons in the $bb\mu\mu$ final state in pp collision at $\sqrt{s}$ =13 TeV with the ATLAS detector", Phys. Lett. B **790**, p. 1, (2019), arXiv: 1807.00539.

[386] R. Aaij *et al.*, "Search for long-lived scalar particles in $B^+ \to K^+\chi(\mu^+\mu^-)$ decays", Phys. Rev. D **95**, p. 071101, (2017), arXiv: 1612.07818.

[387] R. Aaij *et al.*, "Search for hidden-sector bosons in $B^0 \to K^{*0}\mu^+\mu^-$ decays", Phys. Rev. Lett. **115**, p. 161802, (2015), arXiv: 1508.04094.

[388] B. Döbrich, F. Ertas, F. Kahlhoefer, T. Spadaro, "Model-independent bounds on light pseudoscalars from rare B-meson decays", Phys. Lett. B **790**, p. 537, (2019), arXiv: 1810.11336.

[389] E. Cortina Gil *et al.*, "Measurement of the very rare K$^+ \to \pi^+\nu\overline{\nu}$ decay", JHEP **06**, p. 093, (2021), arXiv: 2103.15389.

[390] J. R. Batley *et al.*, "Precise measurement of the K+- —> pi+-e+e- decay", Phys. Lett. B **677**, p. 246, (2009), arXiv: 0903.3130.

[391] J. R. Batley *et al.*, "Searches for lepton number violation and resonances in $K^{\pm} \to \pi\mu\mu$ decays", Phys. Lett. B **769**, p. 67, (2017), arXiv: 1612.04723.

[392] J. P. Lees *et al.*, "Search for $B \to K^{(*)}\nu\overline{\nu}$ and invisible quarkonium decays", Phys. Rev. D **87**, p. 112005, (2013), arXiv: 1303.7465.

[393] M. B. Gavela, R. Houtz, P. Quilez, R. Del Rey, O. Sumensari, "Flavor constraints on electroweak ALP couplings", Eur. Phys. J. C **79**, p. 369, (2019), arXiv: 1901.02031.

[394] J. Martin Camalich, M. Pospelov, P. N. H. Vuong, R. Ziegler, J. Zupan, "Quark Flavor Phenomenology of the QCD Axion", Phys. Rev. D **102**, p. 015023, (2020), arXiv: 2002.04623.

[395] A. Bharucha, T. Feldmann, M. Wick, "Theoretical and Phenomenological Constraints on Form Factors for Radiative and Semi-Leptonic B-Meson Decays", JHEP **09**, p. 090, (2010), arXiv: 1004.3249.

[396] A. Bharucha, D. M. Straub, R. Zwicky, "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules", JHEP **08**, p. 098, (2016), arXiv: 1503.05534.

[397] D. S. M. Alves, N. Weiner, "A viable QCD axion in the MeV mass range", JHEP **07**, p. 092, (2018), arXiv: 1710.03764.

[398] S. Friot, D. Greynat, E. De Rafael, "Rare kaon decays revisited", Phys. Lett. B **595**, p. 301, (2004), arXiv: hep-ph/0404136.

[399] F. Bergsma *et al.*, "Search for Axion Like Particle Production in 400-GeV Proton - Copper Interactions", Phys. Lett. B **157**, p. 458, (1985).

[400] F. Bergsma *et al.*, "A Search for Decays of Heavy Neutrinos", Phys. Lett. B **128**, p. 361, (1983).

[401] M. Millea, L. Knox, B. Fields, "New Bounds for Axions and Axion-Like Particles with keV-GeV Masses", Phys. Rev. D **92**, p. 023010, (2015), arXiv: 1501.04097.

[402] P. F. Depta, M. Hufnagel, K. Schmidt-Hoberg, "Robust cosmological constraints on axion-like particles", JCAP **05**, p. 009, (2020), arXiv: 2002.08370.

[403] B. D. Fields, K. A. Olive, T.-H. Yeh, C. Young, "Big-Bang Nucleosynthesis after Planck", JCAP **03**, p. 010, (2020), arXiv: 1912.01132. [Erratum: JCAP 11, E02 (2020)].

[404] P. F. Depta, M. Hufnagel, K. Schmidt-Hoberg, "Updated BBN constraints on electromagnetic decays of MeV-scale particles", JCAP **04**, p. 011, (2021), arXiv: 2011.06519.

[405] M. H. Reno, D. Seckel, "Primordial Nucleosynthesis: The Effects of Injecting Hadrons", Phys. Rev. D **37**, p. 3441, (1988).

[406] M. Kawasaki, K. Kohri, T. Moroi, Y. Takaesu, "Revisiting Big-Bang Nucleosynthesis Constraints on Long-Lived Decaying Particles", Phys. Rev. D **97**, p. 023502, (2018), arXiv: 1709.01211.

[407] L. Bergstrom, T. Bringmann, I. Cholis, D. Hooper, C. Weniger, "New Limits on Dark Matter Annihilation from AMS Cosmic Ray Positron Data", Phys. Rev. Lett. **111**, p. 171101, (2013), arXiv: 1306.3983.

[408] A. Cuoco, J. Heisig, M. Korsmeier, M. Krämer, "Constraining heavy dark matter with cosmic-ray antiprotons", JCAP **04**, p. 004, (2018), arXiv: 1711.05274.

[409] D. N. Spergel, P. J. Steinhardt, "Observational Evidence for Self-Interacting Cold Dark Matter", Physical Review Letters **84**, p. 3760–3763, (Apr 2000), http://dx.doi.org/10.1103/PhysRevLett.84.3760.

[410] J. S. Bullock, M. Boylan-Kolchin, "Small-Scale Challenges to the ΛCDM Paradigm", Annual Review of Astronomy and Astrophysics **55**, p. 343–387, (Aug 2017), http://dx.doi.org/10.1146/annurev-astro-091916-055313.

[411] F. Kahlhoefer, K. Schmidt-Hoberg, M. T. Frandsen, S. Sarkar, "Colliding clusters and dark matter self-interactions", Monthly Notices of the Royal Astronomical Society **437**, p. 2865–2881, (Nov 2013), http://dx.doi.org/10.1093/mnras/stt2097.

[412] P. F. Bedaque, M. I. Buchoff, R. K. Mishra, "Sommerfeld enhancement from Goldstone pseudo-scalar exchange", Journal of High Energy Physics **2009**, p. 046–046, (Nov 2009), http://dx.doi.org/10.1088/1126-6708/2009/11/046.

[413] Z.-P. Liu, Y.-L. Wu, Y.-F. Zhou, Physical Review D **88**, "Sommerfeld enhancements with vector, scalar, and pseudoscalar force carriers",, (Nov 2013), http://dx.doi.org/10.1103/PhysRevD.88.096008.

[414] M. J. Dolan, F. Kahlhoefer, C. McCabe, K. Schmidt-Hoberg, "A taste of dark matter: flavour constraints on pseudoscalar mediators", Journal of High Energy Physics **2015**, p. 171, (Mar 2015).

[415] P. Agrawal, A. Parikh, M. Reece, Journal of High Energy Physics **2020**, "Systematizing the effective theory of self-interacting dark matter",, (Oct 2020), http://dx.doi.org/10.1007/JHEP10(2020)191.

[416] F. Kahlhoefer, K. Schmidt-Hoberg, S. Wild, "Dark matter self-interactions from a general spin-0 mediator", Journal of Cosmology and Astroparticle Physics **2017**, p. 003–003, (Aug 2017), http://dx.doi.org/10.1088/1475-7516/2017/08/003.

[417] P. Carenza, O. Straniero, B. Döbrich, M. Giannotti, G. Lucente, A. Mirizzi, "Constraints on the coupling with photons of heavy axion-like-particles from Globular Clusters", Physics Letters B **809**, p. 135709, (2020), https://www.sciencedirect.com/science/article/pii/S0370269320305128.

[418] L. Di Luzio, M. Fedele, M. Giannotti, F. Mescia, E. Nardi, "Solar Axions Cannot Explain the XENON1T Excess", Phys. Rev. Lett. **125**, p. 131804, (Sep 2020), https://link.aps.org/doi/10.1103/PhysRevLett.125.131804.

[419] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi, O. Straniero, "Revisiting the Bound on Axion-Photon Coupling from Globular Clusters", Phys. Rev. Lett. **113**, p. 191302, (Nov 2014), https://link.aps.org/doi/10.1103/PhysRevLett.113.191302.

[420] M. Giannotti, I. G. Irastorza, J. Redondo, A. Ringwald, K. Saikawa, "Stellar recipes for axion hunters", Journal of Cosmology and Astroparticle Physics **2017**, p. 010, (oct 2017), https://doi.org/10.1088/1475-7516/2017/10/010.

[421] Straniero, O., Pallanca, C., Dalessandro, E., Domínguez, I., Ferraro, F. R., Giannotti, M., Mirizzi, A., Piersanti, L., "The RGB tip of galactic globular clusters and the revision of the axion-electron coupling bound", A&A **644**, p. A166, (2020), https://doi.org/10.1051/0004-6361/202038775.

[422] M. T. Dennis, J. Sakstein, "Tip of the Red Giant Branch Bounds on the Axion-Electron Coupling Revisited", arXiv: 2305.03113.

[423] P. N. Brown, G. D. Byrne, A. C. Hindmarsh, "VODE: A Variable-Coefficient ODE Solver", SIAM Journal on Scientific and Statistical Computing **10**, p. 1038, (1989), https://doi.org/10.1137/0910062.

[424] M. Lüscher, P. Weisz, U. Wolff, "A numerical method to compute the running coupling in asymptotically free theories", Nucl. Phys. B **359**, p. 221, (1991), https://cds.cern.ch/record/216786.

[425] F. Frech, "Hadronic vacuum polarization from step scaling in the Schwinger model", PoS **LATTICE2021**, p. 218, (2022).

[426] G. Abbiendi *et al.*, "Measuring the leading hadronic contribution to the muon g-2 via μe scattering", Eur. Phys. J. C **77**, p. 139, (2017), arXiv: 1609.08987.

[427] G. Abbiendi, "Status of the MUonE experiment", Phys. Scripta **97**, p. 054007, (2022), arXiv: 2201.13177.

[428] K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition", arXiv: 1610.02743.

[429] B. Geshkenbein, V. Morgunov, "On the hadronic contribution to the muon $(g-2)$-factor and to the $\alpha(M_Z)$", Physics Letters B **352**, p. 456 , (1995), http://www.sciencedirect.com/science/article/pii/0370269395004979.

[430] M. L. Swartz, "Reevaluation of the hadronic contribution to $\alpha(M_Z)$", Phys. Rev. D **53**, p. 5268, (May 1996), https://link.aps.org/doi/10.1103/PhysRevD.53.5268.

[431] N. V. Krasnikov, R. Rodenberg, "Update of the electromagnetic effective coupling constant alpha (M(Z)**2)", Nuovo Cim. **A111**, p. 217 (1998), arXiv: hep-ph/9711367.

[432] S. Groote, J. Körner, K. Schilcher, N. Nasrallah, "QCD sum rule determination of $\alpha(M_Z)$ with minimal data input", Physics Letters B **440**, p. 375 , (1998), http://www.sciencedirect.com/science/article/pii/S0370269398011071.

[433] A. Martin, J. Outhwaite, M. Ryskin, "A new determination of the QED coupling $\alpha(M_Z)$ lets the Higgs off the hook", Physics Letters B **492**, p. 69 , (2000), http://www.sciencedirect.com/science/article/pii/S0370269300010832.

[434] J. F. de Trocóniz, F. J. Ynduráin, "Calculation of $\overline{\alpha}_{\mathrm{QED}}$ on the Z", Phys. Rev. D **65**, p. 093002, (Apr 2002), https://link.aps.org/doi/10.1103/PhysRevD.65.093002.

[435] H. Burkhardt, B. Pietrzyk, "Recent BES measurements and the hadronic contribution to the QED vacuum polarization", Phys. Rev. D **84**, p. 037502, (Aug 2011), https://link.aps.org/doi/10.1103/PhysRevD.84.037502.

[436] K. Hagiwara, A. D. Martin, D. Nomura, T. Teubner, "Improved predictions for g-2 of the muon and alpha(QED) (M**2(Z))", Phys. Lett. B **649**, p. 173, (2007), arXiv: hep-ph/0611102.

[437] M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, "Reevaluation of the Hadronic Contributions to the Muon g-2 and to alpha(MZ)", Eur. Phys. J. C **71**, p. 1515, (2011), arXiv: 1010.4180. [Erratum: Eur.Phys.J.C 72, 1874 (2012)].

[438] F. Jegerlehner, "Electroweak effective couplings for future precision experiments", Nuovo Cim. **C034S1**, p. 31, (2011), arXiv: 1107.4683.

[439] K. Hagiwara, R. Liao, A. D. Martin, D. Nomura, T. Teubner, "$(g-2)_\mu$ and $\alpha(M_Z^2)$ re-evaluated using new precise data", J. Phys. **G38**, p. 085003, (2011), arXiv: 1105.3149.

[440] S. Bodenstein, C. A. Dominguez, K. Schilcher, H. Spiesberger, "Hadronic contribution to the QED running coupling $\alpha(M_Z^2)$", Phys. Rev. **D86**, p. 093013, (2012), arXiv: 1209.4802.

[441] F. Jegerlehner, "Variations on Photon Vacuum Polarization", EPJ Web Conf. **218**, p. 01003, (2019), arXiv: 1711.06089.

[442] D. E. Kaplan, R. Rattazzi, "Large field excursions and approximate discrete symmetries from a clockwork axion", Phys. Rev. D **93**, p. 085007, (2016), arXiv: 1511.01827.

[443] F. Domingo, Journal of High Energy Physics **2017**, "Decays of a NMSSM CP-odd Higgs in the low-mass region",, (mar 2017).

[444] R. Hofstadter, "Nuclear and Nucleon Scattering of High-Energy Electrons", Annual Review of Nuclear Science **7**, p. 231, (1957), `https://doi.org/10.1146/annurev.ns.07.120157.001311`.

[445] Y. S. Tsai, "Axion bremsstrahlung by an electron beam", Phys. Rev. D **34**, p. 1326, (Sep 1986), `https://link.aps.org/doi/10.1103/PhysRevD.34.1326`.

[446] B. Lu, *Search for Axion-Like Particles in the SLAC Beam Dump.* PhD thesis, VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY. (Sept. 1987).

[447] G. Brooijmans *et al.*, "Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report", in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 2 2020.

[448] G. Aad *et al.*, "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$ = 7 and 8 TeV in the ATLAS experiment", Eur. Phys. J. C **76**, p. 6, (2016), arXiv: 1507.04548.

[449] V. Shtabovenko, R. Mertig, F. Orellana, "New developments in FeynCalc 9.0", Computer Physics Communications **207**, p. 432–444, (Oct 2016), `http://dx.doi.org/10.1016/j.cpc.2016.06.008`.