

# A gravitationally induced decoherence model using Ashtekar variables

Max Joseph Fahn, Kristina Giesel\*  and Michael Kobler

Institute for Quantum Gravity, Theoretical Physics III, Department of Physics,  
Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstr. 7, 91058 Erlangen,  
Germany

E-mail: [kristina.giesel@gravity.fau.de](mailto:kristina.giesel@gravity.fau.de)

Received 15 October 2022; revised 7 February 2023

Accepted for publication 20 March 2023

Published 12 April 2023



CrossMark

## Abstract

We consider the coupling of a scalar field to linearised gravity and derive a relativistic gravitationally induced decoherence model using Ashtekar variables. The model is formulated at the gauge invariant level using suitable geometrical clocks in the relational formalism, broadening existing gauge invariant formulations of decoherence models. For the construction of the Dirac observables we extend the known observable map by a kind of dual map where the role of clocks and constraints is interchanged. We also discuss a second choice of geometrical clocks existing in the ADM literature. Then we apply a reduced phase space quantisation on Fock space and derive the final master equation choosing a Gibbs state for the gravitational environment and using the projection operator technique. The resulting master equation is not automatically of Lindblad type, a starting point sometimes assumed for phenomenological models, but still involves a residual time dependence at the level of the effective operators in the master equation due to the form of the correlation functions that we express in terms of thermal Wightman functions. Furthermore, we discuss why in the model analysed here the application of a second Markov approximation in order to obtain a set of time independent effective system operators is less straightforward than in some of the quantum mechanical models.

Keywords: decoherence, linearised gravity, master equation, Dirac observables, geometrical clocks

\* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## Contents

1. Introduction	3
1.1. Theoretical background	3
1.2. Main topics and questions addressed in this article	3
1.3. Structure and main results of this article	5
2. Classical setup for the decoherence model	6
2.1. The gravity-matter system formulated in Ashtekar variables	7
2.2. Brief review of linearised gravity in Ashtekar variables	9
2.3. Post-Minkowski approximation scheme	10
2.4. Dirac observables in the linearised model using geometrical clocks	14
3. Fock quantisation of the model	29
4. Deriving the master equation for the gravitationally induced decoherence model	33
4.1. Review of the projection operator technique	33
4.2. Review of the influence phase functional approach	37
4.3. Derivation of the master equation	42
5. Conclusions	49
Data availability statement	52
Acknowledgments	52
Appendix A. More details on the construction of Dirac observables	52
A.1. More detailed discussion on the choice of reference fields for the linearised constraints	52
A.1.1. Choice of the reference field for the linearised Gauß constraint	53
A.1.2. Choice of the reference field for the linearised Hamiltonian constraint	54
A.1.3. Choice of the reference field for the linearised spatial diffeomorphism constraint	54
A.2. Choice of basis for the calculations in Fourier space	56
A.3. Further details on computing the algebra of Dirac observables	57
A.3.1. Original matter and geometrical variables commute with clocks	57
A.3.2. Computing the algebra of the Dirac observables by means of using the Dirac observable of the corresponding Dirac bracket	58
A.3.3. Poisson algebra of linearised Dirac observables	62
A.3.4. Comparison with the Dirac observables used by Anastopoulos and Hu	64
A.4. Canonical transformation on the original phase space	65
A.5. Canonical Hamiltonian in terms of the transformed variables	68
Appendix B. Two possible choices for a gauge fixing	71
B.1. Gauge fixing following from the clocks introduced in section 2.4	71
B.2. A further gauge fixing often used for ADM variables introduced in [74]	72
Appendix C. Local invertibility of $(\mathbb{1} - \Sigma(\mathbf{t}, \mathbf{t}_0))$ in the TCL master equation	74
Appendix D. Evaluation of the correlation functions	75
References	79

## 1. Introduction

### 1.1. Theoretical background

Decoherence models play an important role in the semiclassical sector of quantum models in various areas of physics from condensed matter, quantum optics and quantum information [1–5] to cosmology [6–10], for a mathematical perspective see [11]. These models are formulated in the framework of open quantum systems [12–14] and allow for instance to address the question of the quantum-to-classical transition [15–17]. Because such models involve an interaction between the quantum system and its environment the choice of the latter is a crucial ingredient for any decoherence model. A very common choice is to couple the system to a bath of harmonic oscillators [12, 18, 19]. In this work we want to analyse gravitationally induced decoherence for which the environment is taken to be gravity. This provides a platform to formulate and analyse decoherence models based on existing models for quantum gravity. Compared to the situation in closed quantum models where the semiclassical sector is usually described by means of semiclassical states, see for instance [20–23], the semiclassical sector in the context of quantum gravity can be investigated from a different angle within decoherence models with a gravitational environment.

There already exist several models for gravitationally induced decoherence in the literature [24–38], see also the reviews in [39, 40] and the review [41] for further models in which quantum matter is coupled to gravity. Some of them have for instance been used to study the influence of decoherence on standard model matter such as neutrinos and their oscillations, see for instance [27–34]. All models have in common that the system’s dynamics is encoded in a so-called master equation that is an effective first order differential equation for the density matrix of the system, where this density matrix is obtained once the environmental degrees of freedom in the total dynamics have been traced out. One can group the existing models in two classes, that we will denote as bottom-up [25, 26, 28, 30, 33] and top-down approaches [35–38, 42–44], respectively. In the first approach one takes as a starting point a given master equation, as for instance the well known Lindblad equation [45, 46], and then constrains existing free parameters phenomenologically for a given model under consideration. In the top-down approach one starts with a given classical model, performs a quantisation of the total system and afterwards derives the master equation within that quantum model. If one aims at formulating a decoherence model based on a specific quantum gravity approach then the top-down approach is advantageous, since an interesting question in this context is whether there exist characteristic features in the decoherence model that will distinguish one quantum gravity model from another. Furthermore, in the derivation of the Lindblad equation, generally several assumptions such as the first and second Markov approximation are applied and one needs to check carefully under which circumstances these can be applied in gravitationally induced decoherence models. Compared to decoherence models derived in a quantum mechanical context, choosing quantum gravity as the environment complicates the situation in two aspects. On the one hand we have to deal with a quantum field theory where also the coupling between the system and the environment is more complex and on the other hand we consider a gauge theory and both facts need to be addressed appropriately in these models.

### 1.2. Main topics and questions addressed in this article

In this article we are interested in deriving a gravitationally induced decoherence model inspired by loop quantum gravity, see [47–49] for textbooks. Most works on the semiclassical physics of loop quantum gravity rather focus on closed quantum systems with semiclassical

states [50–53] or coarse graining in spin foam models [54, 55]. There exist a few exceptions see for instance [56] where a decoherence toy model for surface states in the context of quantum black holes is introduced or the work in [57] for a coupling of a spin foam model to an environment using the adiabatic elimination method [58] to obtain an effective Lindblad equation.

As a first step in this direction we consider a top-down approach in which we couple a scalar field to linearised gravity as it has been done in [35–37]. In contrast to [35–37] where ADM variables have been used we construct the model using Ashtekar variables [59, 60]. The latter allow in principle a quantisation using techniques from loop quantum gravity along the lines of [61, 62]. However, as a first step due to the complexity of the model and to be able to better compare our results to the existing literature, we apply a Fock quantisation to the linearised gravitational sector.

Already at the classical level the model considered in this work requires further generalisations to the already existing results. The models in [35–37] perturb the gravitational degrees of freedom only and we consider the coupling of the scalar field in the framework of a Post-Minkowski approximation scheme that allows to consider a setup where only the gravitational degrees of freedom are perturbed and still obtain a consistent formulation of the involved constraints in perturbation theory. Because we work with Ashtekar variables the model involves an additional Gauß constraint that needs to be taken into account for the gauge invariance. The route we follow here to deal with the gauge freedom is that we consider a reduced phase space quantisation. For this purpose we first construct a canonical transformation that separates the gauge and physical degrees of freedom into two sets. The latter are obtained by constructing Dirac observables in the relational formalism applying techniques from [63–68] where suitable reference fields need to be chosen. In our case, these will be so-called geometrical clocks that only depend on the gravitational degrees of freedom. These clocks differ from the ones chosen in [69], where complex Ashtekar variables have been used with the corresponding gauge fixing, whereas the Dirac observables used here have a natural relation to the gauge-fixed quantities different than the one presented in [61], which considers the vacuum case only. Geometrical clocks have also been applied in different contexts, see for instance [69–72]. Working in the relational formalism (physical) time can be defined in a relational way and circumvents some of the issues related to fluctuating light cones mentioned in [73]. In order to construct the above mentioned canonical transformation we need to generalise the observable map and introduce a kind of dual observable map where the role of constraints and reference fields is interchanged. When applied to the constraints themselves it reproduces the results for the perturbative formulation of weak Abelianisation in [69] but being more general in the sense that it can be applied to any phase space function, a property crucial for our application. As we work in the linearised theory when computing the algebra of Dirac observables that provides the starting point for the reduced phase space quantisation, it is necessary to consider Dirac observables up to second order in general to formulate the model at the gauge invariant level. In former works [36, 38] often specific gauge fixings were chosen or Dirac observables were constructed for the matter sector only [35], where to our understanding these Dirac observables are not left invariant under the spatial diffeomorphism constraint. In [37] Dirac quantisation was applied and the obtained results are consistent with the reduced phase space quantisation procedure we apply here. We also consider a second choice of geometrical clocks introduced in [74] which was carried over to Ashtekar variables in [69] and discuss its differences and similarities to the geometrical clocks chosen in this work. There exist also former work on decoherence effects caused by quantum clocks [75–77], whereas we use classical (perfect) clocks in the language of [77]. Such decoherence effects would be present in addition to the decoherence effects that we will discuss in this work and it will be interesting to compare the size of the two types of decoherence effects. A detailed analysis, requires to work with fundamental quantum clocks

and perform the construction of Dirac observables in the quantum theory for instance along the lines of [78]. However, to the knowledge of the authors this has so far only been considered in the context of quantum mechanical systems but not a full quantum field theory setup that we will need in the case considered in this work.

The physical Hamiltonian that is itself a Dirac observable and generates the dynamics of the linearised Dirac observables of the total system is then quantised using a Fock quantisation. The strategy we follow in the quantisation is that we choose to normal order the entire physical Hamiltonian operator. This differs from [35, 37, 38] where either individual operators in products of operators are normal ordered separately or no normal ordering is applied at all and such differences appear in the operator ordering in the self-interaction part of the Hamiltonian operator. For deriving a master equation in a top-down approach two prominent techniques exist, the influence functional approach [79] and the projection operator technique [80–83], where we use the latter to obtain a time convolution-less (TCL) master equation for the model considered in this work. This allows us to understand in detail how certain assumptions and approximations enter into the derivation of the master equations which will be useful also for later applications and generalisations of the model. We choose a Gibbs state as the initial state of the quantum gravitational environment and use a regularisation by means of a box of finite volume in order to implement this on Fock space, similar to the case in [84]. After computing the environmental trace we remove the regulator by taking the infinite volume limit. In order to circumvent to work with regulators one can use KMS states [85, 86], which is beyond the scope of this work. We express the final master equation in terms of thermal Wightman functions, which on the one hand is a general formalism other models can also be formulated in and reflects the structure of the involved correlation functions. On the other hand this gives an easy access to the special case where the environment is chosen to be in the vacuum state. For a better comparison with the already existing models in [35, 37, 38] we present three equivalent but different forms of the final master equation and discuss in detail the similarities and differences to [35, 37, 38].

### 1.3. Structure and main results of this article

The paper is structured as follows: After the introduction in section 1 in section 2 we present the classical formulation of the decoherence model considered in this article. Section 2.2 briefly reviews the existing results on linearised gravity in terms of Ashtekar variables and we consider the coupling of a scalar field to linearised gravity in a Post-Minkowski approximation scheme in section 2.3. Section 2.4 presents the construction of Dirac observables for the linearised model using geometrical clocks. Furthermore, we introduce a canonical transformation on the full phase space in which the geometrical clocks as well as the constraints are chosen as some of the new canonical coordinates, allowing to clearly separate physical and gauge degrees of freedom in the model. This decomposition, often also called Kuchař decomposition [87, 88] that was recently considered in [89] to discuss different choices of gauge in the context of linear cosmological perturbation theory. In the generic situation where a chosen set of set of clocks is not necessary abelian *a priori*, as it was for instance be assumed in [69] and no comments on the case of non-abelian gauge fixing conditions can be found in [89], this requires the application of the already existing observable map as well as its dual version introduced in this section. This opens the possibility for constructing such canonical transformations and hence performing a Kuchar decomposition in principle also in other models and beyond linear order. For each choice of geometrical clocks there exists a corresponding gauge in which the gauge invariant variables have a natural and simple interpretation. The advantage of working at the gauge invariant level is that even if we choose a different gauge we still have a gauge

invariant formulation of the model, with an in general different interpretation of the gauge invariant variables but the equations are valid in any chosen gauge. If one is interested in always choosing the set of geometrical clocks for which the Dirac observables have a natural interpretation, then this is also possible in this framework. Since in the work presented here the choice of geometrical clocks is implemented by a canonical transformation, it can be inverted and thus afterwards one can choose a different gauge and hence also switch between different choices of geometrical clocks, which can again be carried over to choosing different gauges. We have discussed to examples for different set of geometrical clocks in appendix A.3.4.

Given the reduced phase space derived in section 2.4 and the physical Hamiltonian, we briefly discuss the Fock quantisation of the model in section 3 with a focus on the individual contributions in the physical Hamiltonian operator. The derivation of the final master equation is presented in section 4. We include a brief review of the projection operator technique as well as the influence functional approach in sections 4.1 and 4.2, respectively. This serves as a preparation for the application of the first technique to the model in this work, presented in section 4.3 where we obtain the explicit form of the final master equation. In order to be able to compare our results better with the already existing ones in the literature we present the final master equation in three equivalent forms. The first version can be found in (4.52) and is written in terms of a dissipation and noise kernel. Next in (4.58) an equivalent form of the final master equation is presented which is very close to the notation used in [35, 37, 38] and allows a direct comparison to their work. Furthermore, the third equivalent version of the master equation, found in (4.60), clearly separates the contributions from the zero and finite temperature cases. These equivalent version of the final master equation are further compared to already existing master equations for similar models in the literature. The main differences to the existing results are: First the model presented here is formulated in terms of Dirac observables with respect to chosen geometrical clocks or also known as dynamical reference systems. The formalism allows different choices of such sets of geometrical clocks and thus also an analysis how such choices affect a given model in future work. Furthermore, the final master equation we obtain is not automatically of Lindblad type. Often approximations such as the second Markov approximation and the rotating wave approximation are applied to arrive at a Lindbladian master equation. At least to the authors it is not clear from the outset that such techniques, which are mostly used in the quantum mechanical context, can be directly transferred to the context of quantum field theories and, according to the authors, require further careful and detailed analysis, as also shown, for example, by the work in [90]. The resulting final master equation in this work provides an good starting point for such an investigation. Finally, in section 5 we summarise our results and conclude. We present more detailed computations in the appendix that we refer to in the main text of this article.

## 2. Classical setup for the decoherence model

Following the usual approach for gravitationally induced decoherence models, in the model presented in this work here, we will choose as the environmental degrees of freedom the gravitational ones and consider as the system's ones a coupled Klein–Gordon scalar field and discuss the corresponding action and necessary boundary contributions in terms of Ashtekar variables in section 2.1. As a simpler setup than full GR we will specialise this action to a coupling to linearised gravity only. For this purpose we consider the linearisation of vacuum gravity in Ashtekar variables in section 2.2 and combine this in the framework of a Post-Minkowski approximation to couple the scalar field as discussed in section 2.3. Because we are working with a gauge theory before later quantising the system and tracing out the environmental

degrees of freedom to obtain an effective reduced model for the system, we need to take care of the gauge freedom involved in the theory. Here, one can either consider a specific gauge fixing or work at the gauge invariant level by formulating the model in terms of gauge invariant variables also known as Dirac observables in the context of general relativity, where we will choose the latter strategy in this work as discussed in section 2.4. Once the dynamics of the model is written by means of Dirac observables we can split the total physical Hamiltonian that generates the dynamics of the Dirac observables into a system and environmental part and we will take this then as our starting point for the quantisation in section 3.

### 2.1. The gravity-matter system formulated in Ashtekar variables

As the starting point for deriving the decoherence model in this work we choose the following action involving gravity coupled to a Klein–Gordon scalar field in the ADM decomposition [91] expressed in terms of Ashtekar variables [59, 60, 92, 93]:

$$S = \int_{\mathbb{R}} dt \int_{\sigma} d^3x \left( \frac{1}{\kappa\beta} \dot{A}_a^i(\vec{x}, t) E^a_i(\vec{x}, t) + \dot{\phi}(\vec{x}, t) \pi(\vec{x}, t) - \left[ \Lambda^i(\vec{x}, t) G_i(\vec{x}, t) + N^a(\vec{x}, t) C_a(\vec{x}, t) + N(\vec{x}, t) C(\vec{x}, t) \right] \right), \quad (2.1)$$

where  $\beta$  is the Barbero-Immirzi parameter,  $\kappa = \frac{8\pi G_N}{c^4}$  with Newton's gravitational constant  $G_N$ ,  $c$  denotes the speed of light and we work with the mostly-plus signature of the metric. From here on, we will set  $c = \hbar = 1$ . The gravitational degrees of freedom are encoded in the Ashtekar variables consisting of an  $SU(2)$ -connection  $A_a^i(\vec{x}, t)$  and the canonically conjugate densitised triads  $E^a_i(\vec{x}, t)$ . The matter sector includes the scalar field  $\phi(\vec{x}, t)$  as well as its canonically conjugate momentum  $\pi(\vec{x}, t)$ , an overdot denotes a derivative with respect to the temporal coordinate. That general relativity is a fully constrained theory is reflected by the fact that apart from the symplectic potential in the action, there appears a sum of constraints  $G_i(\vec{x}, t)$ ,  $C_a(\vec{x}, t)$  and  $C(\vec{x}, t)$  only multiplied by different Lagrange multipliers  $\Lambda^i(\vec{x}, t)$ ,  $N^a(\vec{x}, t)$  and  $N(\vec{x}, t)$ , where the latter two are usually referred to as the shift vector and the lapse function, respectively. In Ashtekar variables the spatial diffeomorphism constraint reads

$$C_a(\vec{x}, t) = \frac{1}{\kappa\beta} F_{ab}^i(\vec{x}, t) E^b_i(\vec{x}, t) + \pi(\vec{x}, t) \partial_a \phi(\vec{x}, t), \quad (2.2)$$

where  $F_{ab}^i(\vec{x}, t) := \partial_a A_b^i(\vec{x}, t) - \partial_b A_a^i(\vec{x}, t) + \epsilon_{ijk} A_a^j(\vec{x}, t) A_b^k(\vec{x}, t)$  denotes the curvature related to the connection variable  $A$ . The Gauß constraint, which is absent if working in ADM variables, is given by

$$G_i(\vec{x}, t) = \frac{1}{2\kappa\beta} (\partial_a E^a_i(\vec{x}, t) + \epsilon_{ij}^k A_a^j(\vec{x}, t) E^a_k(\vec{x}, t)) \quad (2.3)$$

with the completely antisymmetric tensor  $\epsilon$ . Finally, the Hamiltonian constraint takes the form

$$C(\vec{x}, t) = \frac{1}{2\kappa} \left( F_{ab}^i(\vec{x}, t) - \frac{\beta^2 + 1}{\beta^2} \epsilon_{ilm} (A_a^l(\vec{x}, t) - \Gamma_a^l(\vec{x}, t)) (A_b^m(\vec{x}, t) - \Gamma_b^m(\vec{x}, t)) \right) \times \frac{\epsilon^{ijk} E^a_j(\vec{x}, t) E^b_k(\vec{x}, t)}{\sqrt{\det(E^c_n)}} + \frac{\pi^2(\vec{x}, t)}{2\sqrt{\det(E^c_n)}} + \frac{1}{2\sqrt{\det(E^c_n)}} \times \delta^{ij} E^a_i(\vec{x}, t) E^b_j(\vec{x}, t) (\partial_a \phi(\vec{x}, t)) (\partial_b \phi(\vec{x}, t)) + \frac{\sqrt{\det(E^c_n)}}{2} m^2 \phi^2(\vec{x}, t). \quad (2.4)$$

Here,  $m$  denotes the mass of the scalar field and  $\Gamma_a^i$  is the spin connection, considered a functional of the densitised triads, i.e.

$$\Gamma_a^i = -\frac{1}{2}\epsilon^{ijk}E^b{}_k \left( E_a^j{}_{,b} - E_b^j{}_{,a} + E^c{}_j E_a^l E_c^l{}_{,b} + E_a^j \frac{\det(E^c{}_l)_{,b}}{\det(E^c{}_l)} - E_b^j \frac{\det(E^c{}_l)_{,a}}{2\det(E^c{}_l)} \right), \quad (2.5)$$

where we abbreviated partial derivatives with a comma, that is  $E_a^j{}_{,b} := \partial_b E_a^j$  and suppressed the spatial and temporal arguments of the involved field variables. The elementary phase space variables satisfy the following Poisson algebra on the classical phase space:

$$\{E^a{}_i(\vec{x}, t), E^b{}_j(\vec{y}, t)\} = \{A_a^i(\vec{x}, t), A_b^j(\vec{y}, t)\} = 0, \quad \{A_a^i(\vec{x}, t), E^b{}_j(\vec{y}, t)\} = \beta\kappa\delta_a^b\delta_j^i\delta^3(\vec{x} - \vec{y}) \quad (2.6)$$

$$\{\phi(\vec{x}, t), \phi(\vec{y}, t)\} = \{\pi(\vec{x}, t), \pi(\vec{y}, t)\} = 0, \quad \{\phi(\vec{x}, t), \pi(\vec{y}, t)\} = \delta^{(3)}(\vec{x} - \vec{y}), \quad (2.7)$$

where all the remaining ones vanish. As the decoherence model presented in this article requires to linearise the above action around a flat Minkowski background in the course of the upcoming section, we choose asymptotically flat boundary conditions [94] also addressed in [95, 96] for the case of real Ashtekar variables. With the fall-off behaviour of the Ashtekar variables discussed in section 2.2 and the linearisation of the Lagrange multipliers chosen such that both are consistent with an asymptotically flat Universe, it turns out that the Gauß constraint and the gravitational part of the spatial diffeomorphism constraint do not cause any issues and need no further modification in accordance with [69]. However, the gravitational part of the Hamiltonian constraint as well as its variation is not well-defined in the asymptotically flat limit and contains divergences. To circumvent these, we introduce a suitable boundary term given by

$$T[N] = -2 \int_{\partial\sigma} dS_a N A_b^k \epsilon^{ijk} \frac{E^a{}_i E^b{}_j}{\sqrt{\det E}} \quad (2.8)$$

to the smeared Hamiltonian constraint, which makes it well-defined and functionally differentiable. Note that on the constraint hypersurface where the Gauß constraint vanishes  $T[1]$  is the so-called ADM-energy [97]. Further we assume, as is usually done, that the initial data of the matter variables have compact support, yielding matter contributions that are well-defined in the asymptotically flat case. Given this, the canonical Hamiltonian that we will work with in the following and that is consistent with our chosen boundary conditions reads

$$\begin{aligned} \mathbf{H}_{\text{can}} &= C(N) + \vec{C}(\vec{N}) + \vec{G}(\vec{\Lambda}) + T[N] \\ &= \int_{\sigma} d^3x \left[ \frac{1}{\kappa\beta} N^a F_{ab}{}^j E^b{}_j + N^a \pi \nabla_a \phi + \frac{1}{2\kappa\beta} \Lambda^i \partial_a E^a{}_i + \frac{1}{2\kappa\beta} \Lambda^i \epsilon_{ikl} A_a^k E^a{}_l \right. \\ &\quad - N e^{jkl} \frac{1}{\sqrt{\det E}} E^a{}_k E^b{}_l \frac{1}{2\kappa\beta^2} \epsilon_{jmn} (A_a^m A_b^n + (\beta^2 + 1)\Gamma_a^m \Gamma_b^n - 2\Gamma_a^m A_b^n) \\ &\quad + \frac{1}{\sqrt{\det E}} \left( -\frac{1}{\kappa} e^{jkl} E^a{}_k E^b{}_l A_b^j \partial_a N + N \frac{\pi^2}{2} + \frac{N}{2} E^a{}_i E^b{}_i (\partial_a \phi) (\partial_b \phi) \right) \\ &\quad \left. + N \frac{\sqrt{\det E}}{2} m^2 \phi^2 \right], \quad (2.9) \end{aligned}$$

where  $C(N) := \int_{\sigma} d^3x C(\vec{x}, t) N(\vec{x}, t)$ ,  $\vec{C}(\vec{N}) := \int_{\sigma} d^3x C_a(\vec{x}, t) N^a(\vec{x}, t)$  and  $\vec{G}(\vec{\Lambda}) := \int_{\sigma} d^3x G_i(\vec{x}, t) \Lambda^i(\vec{x}, t)$ .

In the following two subsections we will consider a linearisation of the gravity-matter system. For this purpose we first briefly review the linearisation of the vacuum case and afterwards

discuss how the scalar field can be coupled to linearised gravity in the framework of a Post-Minkowski approximation scheme.

## 2.2. Brief review of linearised gravity in Ashtekar variables

A quantisation of the full theory of general relativity is a highly complex task. As outlined in the introduction, one of the motivations for the work presented in this article comes from scenarios that involve weak couplings between the matter system and the gravitational field. Therefore, we will consider a linearisation of general relativity around a Minkowski spacetime and then consider a perturbation for the gravitational degrees of freedom using  $\kappa$  as an expansion parameter. Here we will focus on vacuum gravity using Ashtekar variables first, analogous to the work in for instance [61] or [98]. We denote the linear perturbations with a prefix  $\delta$ . The Ashtekar variables as well as the Lagrange multipliers then become

$$E^a_i = \bar{E}^a_i + \kappa \delta E^a_i = \delta^a_i + \kappa \delta E^a_i \quad \delta E : O(r^{-1}) \text{ even} \quad (2.10)$$

$$A_a^i = \bar{A}_a^i + \kappa \delta A_a^i = 0 + \kappa \delta A_a^i \quad \delta A : O(r^{-2}) \text{ odd} \quad (2.11)$$

$$N = \bar{N} + \kappa \delta N = 1 + \kappa \delta N \quad \delta N : O(r^{-1}) \quad (2.12)$$

$$N^a = \bar{N}^a + \kappa \delta N^a = 0 + \kappa \delta N^a \quad \delta N^a : O(r^{-1}) \quad (2.13)$$

$$\Lambda^i = \bar{\Lambda}^i + \kappa \delta \Lambda^i = 0 + \kappa \delta \Lambda^i \quad \delta \Lambda : O(r^{-2}) \text{ even}, \quad (2.14)$$

where the fall-off behaviour in the limit of spatial infinity of the first order perturbations has been chosen in accordance with the asymptotically flat boundary condition [94]. The split into background and perturbed quantities is chosen such that the background corresponds to a Minkowski spacetime that remains unchanged by the action of the canonical Hamiltonian of the system, i.e.

$$\overline{\{E^a_i, \mathbf{H}_{\text{can}}\}} = 0 \quad (2.15)$$

$$\overline{\{A_a^i, \mathbf{H}_{\text{can}}\}} = 0, \quad (2.16)$$

where the overline stands for evaluation with respect to the background Minkowski spacetime. In the case of vacuum gravity (after computation of the Poisson brackets) this amounts to  $\bar{E}^a_i = \delta^a_i$ ,  $\bar{A}_a^i = 0$ ,  $\bar{N} = 1$  and  $\bar{N}^a = 0$ . As one can easily compute, the spin connection vanishes in the background:

$$\begin{aligned} \Gamma_a^i &= \bar{\Gamma}_a^i + \kappa \delta \Gamma_a^i \approx 0 + \kappa \delta \Gamma_a^i \\ &= -\frac{\kappa}{2} \epsilon^{ijk} \delta_k^b [-\delta_a^l \delta_c^j \partial_b (\delta E^c_l) + \delta_c^j \delta_b^l \partial_a (\delta E^c_l) - \delta_{ac} \partial_b (\delta E^c_j) + \delta_a^j \delta_c^l \partial_b (\delta E^c_l)]. \end{aligned} \quad (2.17)$$

The Poisson algebra of the linearised gravitational variables can be inherited from (2.6):

$$\begin{aligned} \{\delta E^a_i(\vec{x}, t), \delta E^b_j(\vec{y}, t)\} &= \{\delta A_a^i(\vec{x}, t), \delta A_b^j(\vec{y}, t)\} = 0, \\ \{\delta A_a^i(\vec{x}, t), \delta E^b_j(\vec{y}, t)\} &= \frac{\beta}{\kappa} \delta_a^b \delta_j^i \delta^3(\vec{x} - \vec{y}). \end{aligned} \quad (2.18)$$

In the next subsection we will address how the matter contributions can be included consistently in the linearised framework.

### 2.3. Post-Minkowski approximation scheme

In the context of general relativistic perturbation theory, one usually chooses a background solution and then considers perturbations of the gravitational and matter sector around it assuming that the perturbations are small compared to the chosen background quantities. In this work we linearise around a flat Minkowski spacetime, which is a vacuum solution of general relativity. Hence, any considered perturbation in the matter sector will not be small compared to vanishing matter degrees of freedom in the vacuum case. However, we can formulate a model that involves the coupling between matter and linearised gravity in a Post-Minkowski approximation scheme. For the model presented in this article we will need to consider the zeroth and first order in the Post-Minkowski formalism only and this will be the underlying guiding principle of how the matter sector will be included into the Hamiltonian formulation and in particular into the constraints. We will briefly sketch the main steps that lead to the linearised action that we take as our starting point. The Post-Minkowski formalism is based on the Landau–Lifshitz formulation of general relativity, see for instance the book [99] for an introduction to the subject, and one starts by introducing the so-called gothic metric defined as

$$\mathfrak{g}^{\mu\nu} := \sqrt{-\det(g_{\rho\sigma})} g^{\mu\nu}, \quad (2.19)$$

by means of which we can introduce the following tensor density

$$H^{\mu\nu\rho\sigma} := \mathfrak{g}^{\mu\rho} \mathfrak{g}^{\nu\sigma} - \mathfrak{g}^{\mu\sigma} \mathfrak{g}^{\nu\rho}, \quad (2.20)$$

where  $H^{\mu\nu\rho\sigma}$  carries the same symmetries as the Riemann tensor. Following the presentation in [99] one can use (2.20) to rewrite Einstein's equations  $G_{\mu\nu} = \kappa T_{\mu\nu}$  in the following form:

$$\partial_\nu \partial_\sigma H^{\mu\nu\rho\sigma} = 2\kappa(-\det(g)) [T^{\mu\rho} + t_{LL}^{\mu\rho}]. \quad (2.21)$$

Here,  $T^{\mu\rho}$  denotes the energy-momentum tensor of the matter degrees of freedom and  $t_{LL}^{\mu\rho}$  is the so-called Landau–Lifshitz pseudotensor which consists of a sum of (contractions of) terms  $\partial\mathfrak{g}\partial\mathfrak{g}$  and does not transform as a tensor under coordinate transformations. This quantity corresponds to the distribution of energy of the gravitational field in spacetime. If one aims at formulating the corresponding linearised theory, as a first step, one applies a partial gauge fixing on the gothic metric by the harmonic coordinate condition

$$\partial_\mu \mathfrak{g}^{\mu\nu} = 0. \quad (2.22)$$

For later convenience it is useful to introduce the following quantity also known as the metric potentials  $\mathfrak{h}^{\mu\nu}$  that have the following relation to the gothic metric  $\mathfrak{g}^{\mu\nu}$ :

$$\mathfrak{g}^{\mu\nu} = \eta^{\mu\nu} - \mathfrak{h}^{\mu\nu}. \quad (2.23)$$

The harmonic coordinate condition then carries over to  $\partial_\mu \mathfrak{h}^{\mu\nu} = 0$ . In this gauge, Einstein's equations in (2.21) have the form, see e.g. equations (6.51) and (6.52) in [99]:

$$\square \mathfrak{h}^{\mu\nu} = -2\kappa(-\det(g)) [T^{\mu\nu}[\Phi, g] + t_{LL}^{\mu\nu}[\mathfrak{h}] + t_H^{\mu\nu}[\mathfrak{h}]], \quad (2.24)$$

with the d'Alembertian  $\square$  in flat spacetime and

$$t_H^{\mu\nu}[\mathfrak{h}] := \frac{1}{2\kappa(-\det(g))} (\partial_\rho \mathfrak{h}^{\mu\sigma} \partial_\sigma \mathfrak{h}^{\nu\rho} - \mathfrak{h}^{\rho\sigma} \partial_\rho \partial_\sigma \mathfrak{h}^{\mu\nu}). \quad (2.25)$$

Among the individual contributions in (2.24) the energy-momentum tensor  $T^{\mu\nu}$  depends on the matter variables here denoted by  $\Phi$  and on the metric  $g$ , while the other two contributions

depend on the modified gothic metric  $\mathfrak{h}$  only<sup>1</sup>. The idea of the Post-Minkowski approximation scheme is to construct an iterative solution for (2.24). For this purpose one chooses that in lowest order the gothic metric agrees with the Minkowski metric and for higher orders considers an expansion of  $\mathfrak{h}^{\mu\nu}$  in terms of powers of  $\kappa$  according to

$$\mathfrak{g}^{\mu\nu} = \eta^{\mu\nu} + \sum_{n=1}^{\infty} \kappa^n \mathfrak{h}_{(n)}^{\mu\nu}. \quad (2.26)$$

As in this work we linearise our entire system around a flat Minkowski background and are therefore only interested in weak gravity, for the model in this article we can consider truncations of this perturbative expansion<sup>2</sup>.

An iterative solution of Einstein's equations is constructed by the following procedure. Let us assume one has a solution for the potentials  $\mathfrak{h}$  of order  $\kappa^k$ . Then one is able to compute the contributions of the quantities on the right hand side of (2.24) with it. Due to the prefactor  $\kappa$ , the right hand side will then be of order  $\kappa^{k+1}$ , so the entire equation will yield a solution for  $\mathfrak{h}_{(k+1)}$ . As can be seen from (2.26), the perturbative ansatz in the Post-Minkowski approximation scheme is constructed in a way that in the zeroth order where  $\mathfrak{h} = 0$  one has  $\mathfrak{g} = \eta$ , from which one directly obtains  $g = \eta$ . With this, one can construct the right hand side of (2.24) using this zeroth order solution. As  $t_{LL}$  and  $t_H$  both contain  $\mathfrak{h}$  in each summand, they vanish for  $\mathfrak{h} = 0$  and the only non-trivial contribution comes from the energy-momentum tensor  $T^{\mu\nu}$  on the right hand side. As in zeroth order  $g = \eta$ , it simplifies to the energy momentum tensor  $T^{\mu\nu}$  on a flat Minkowski spacetime. Given the right hand side of (2.24), the left hand side amounts to the first order modified gothic metric, that is

$$\kappa \square \mathfrak{h}_{(1)}^{\mu\nu} = -2\kappa T^{\mu\nu}[\Phi, \eta]. \quad (2.27)$$

Finally, one needs to relate  $\mathfrak{h}_{(1)}^{\mu\nu}$  to the (inverse) metric perturbations  $\delta h^{\mu\nu}$ . Since in this work we are interested in the linearised theory only, we can use  $g^{\mu\nu} = \eta^{\mu\nu} - \kappa \delta h^{\mu\nu} + O(\kappa^2)$  and obtain

$$\mathfrak{h}_{(1)}^{\mu\nu} = \delta h^{\mu\nu} - \frac{1}{2} \eta^{\mu\nu} \eta_{\rho\sigma} \delta h^{\rho\sigma}. \quad (2.28)$$

The last equations allows us to rewrite the linearised equations in a first order Minkowski approximation as

$$-\frac{\kappa}{2} \left( \square(\delta h^{\mu\nu}) - \frac{1}{2} \eta^{\mu\nu} \eta_{\rho\sigma} \square(\delta h^{\rho\sigma}) \right) = \kappa T^{\mu\nu}[\Phi, \eta]. \quad (2.29)$$

From (2.29) we can directly read off how the interaction term in the action needs to look like, namely

$$S_I = \frac{\kappa}{2} \int_M d^4x \delta h_{\mu\nu} T^{\mu\nu}[\Phi, \eta]. \quad (2.30)$$

The overall factor  $\frac{\kappa}{2}$  takes into account that we start with an overall factor of  $\frac{1}{2\kappa}$  and then consider linear perturbations of the metric  $g_{\mu\nu} = \eta_{\mu\nu} + \kappa \delta h_{\mu\nu}$  yielding a factor  $\kappa^2$  for second-order perturbations in the vacuum case. Note that, as expected, the same interaction contribution to the action can be obtained by minimally coupling matter to gravity and then linearising

<sup>1</sup> Later we consider a Klein–Gordon scalar field for the matter sector but here  $\Phi$  is understood symbolically for all possible matter choices that one would like to couple in a given model.

<sup>2</sup> Note also the formal role of the expansion parameter  $\kappa$  that is further explained in box 6.4 in [99]: Depending on the chosen system of units,  $\kappa$  could also be equal to one.

the metric degrees of freedom around a Minkowski spacetime [35–37, 100]. Let us denote the matter Lagrangian density by  $\mathcal{L}_\Phi$ , then we have

$$\begin{aligned} \int_M d^4x \sqrt{-\det(g)} \mathcal{L}_\Phi &\approx \int_M d^4x \mathcal{L}_\Phi|_{g=\eta} + \int_M d^4x (-\kappa \delta h^{\mu\nu}) \left( \frac{\delta(\sqrt{-\det(g)} \mathcal{L}_\Phi)}{\delta g^{\mu\nu}} \right) \Big|_{g=\eta} \\ &= S_\Phi^\eta + \frac{\kappa}{2} \int_M d^4x \delta h^{\mu\nu} T_{\mu\nu}[\Phi, \eta], \end{aligned} \quad (2.31)$$

where  $S_\Phi^\eta$  denotes the matter action on flat spacetime. For the model discussed in this work we choose the matter Lagrangian of a Klein–Gordon scalar field

$$\mathcal{L}_\phi = -\frac{1}{2} g^{\mu\nu} D_\mu \phi D_\nu \phi - \frac{m^2}{2} \phi^2, \quad (2.32)$$

where  $D_\mu$  denotes the torsion-free, covariant metric-compatible derivative with respect to  $g$ . Performing an ADM decomposition as we did for the vacuum case we end up with

$$S_\phi = \int dt \int_\sigma d^3x \left( \dot{\phi} \pi - [N^a C_a^\phi + N C^\phi] \right) \quad (2.33)$$

with

$$C_a^\phi := \pi \nabla_a \phi, \quad (2.34)$$

$$C^\phi := \frac{\pi^2}{2\sqrt{\det q}} + \frac{\sqrt{\det q}}{2} q^{ab} (\nabla_a \phi)(\nabla_b \phi) + \sqrt{\det q} \frac{m^2}{2} \phi^2, \quad (2.35)$$

where  $\pi$  denotes the canonically conjugate momentum associated to  $\phi$ . If we linearise the model including the scalar field as discussed above and take into account all contributions up to first order in  $\kappa$ , we obtain the following constraints in the linearised theory:

$$\delta G_i = \frac{\kappa}{2\beta} (\partial_a (\delta E^a_i) + \epsilon_{ik}^l \delta_l^a \delta A_a^k), \quad (2.36)$$

$$\delta C_a = \frac{\kappa}{\beta} \delta_j^b (\partial_a (\delta A_b^j) - \partial_b (\delta A_a^j)) + \kappa \underbrace{\pi \partial_a \phi}_{=: p_a(\phi, \pi)}, \quad (2.37)$$

$$\delta C = \kappa \epsilon_j^{kl} \delta_k^a \delta_l^b \partial_a (\delta A_b^j) + \kappa \underbrace{\frac{1}{2} [\pi^2 + \partial_a \phi \partial^a \phi + m^2 \phi^2]}_{=: \epsilon(\phi, \pi)}. \quad (2.38)$$

These constraints are also consistent with the 00 and 0a components of the linearised Einstein's equations that one obtains in the Post-Minkowski approximation scheme. Now these include gravitational- as well as matter contributions, where we introduced the momentum density  $p_a(\phi, \pi)$  and the energy density  $\epsilon(\phi, \pi)$  of the scalar field. Note that these constraints are abelian up to first order in  $\kappa$ . The corresponding background constraints of the zero order Post-Minkowski approximation scheme all vanish trivially, since  $\bar{C} = \bar{C}^{\text{geo}} = 0$ ,  $\bar{C}_a = \bar{C}_a^{\text{geo}} = 0$  and  $\bar{G}_j = 0$ . The action of the linearised theory is given by

$$\begin{aligned} S_{\text{lin}} = \int dt \left( \int_\sigma d^3x \left[ \frac{\kappa}{\beta} \delta \dot{A}_a^i \delta E^a_i + \pi \dot{\phi} \right] \right. \\ \left. - \underbrace{\int_\sigma d^3x \left[ \epsilon(\phi, \pi) + \delta N \delta C + \delta N^a \delta C_a + \delta \Lambda^i \delta G_i + \delta \mathcal{H}_I + \frac{1}{\kappa} \delta^2 C^{\text{geo}} \right]}_{=: \delta \mathbf{H}_{\text{can}}} \right) \end{aligned} \quad (2.39)$$

with the interaction part  $\delta\mathcal{H}_I := \frac{\kappa}{2}\delta h_{ab}T^{ab}$  with  $\delta h_{ab}$  understood as a function of the densitised triads and the second order of the geometrical Hamiltonian constraint  $\delta^2 C^{geo}$ . Note that the overall  $\frac{1}{\kappa}$  in front of the geometrical term has been partly cancelled by the  $\kappa$  involved in the linearised quantities for  $N, N^a$  and  $\Lambda^i$ , while the constraints and the interaction term are still linear in  $\kappa$ . In the above form of the linearised action it was already used that in the linearised framework with asymptotically flat boundary conditions the boundary term precisely cancels the first order of the (geometrical) Hamiltonian constraint in accordance with [69], i.e.  $T = -\bar{N}\delta C^{geo} = -\delta C^{geo}$  where we used that  $\bar{N} = 1$ . Hence, neither the linearised Hamiltonian constraint with background lapse nor the boundary term do appear in the action and canonical Hamiltonian anymore.

The linearised total canonical Hamiltonian, which follows from (2.9), is given by:

$$\begin{aligned} \delta\mathbf{H}_{\text{can}} = \int_{\sigma} d^3x \left[ \epsilon(\phi, \pi) + \kappa \delta N^a \pi \partial_a \phi + \kappa \delta N \epsilon(\phi, \pi) + \kappa (\partial_a \phi) (\partial^b \phi) \delta E^a_i \delta^i_b \right. \\ + \frac{\kappa}{2} m^2 \phi^2 \delta E^a_i \delta^i_a - \frac{\kappa}{2} \delta E^a_i \delta^i_a \epsilon(\phi, \pi) + \frac{\kappa}{\beta} \delta N^a \delta^b_j (\partial_a (\delta A_b^j) - \partial_b (\delta A_a^j)) \\ + \frac{\kappa}{2\beta} \delta \Lambda^i \partial_a (\delta E^a_i) + \frac{\kappa}{2\beta} \delta \Lambda^i \epsilon_{ik}^l \delta_l^a \delta A_a^k - \kappa \epsilon_j^{kl} \delta_k^a \delta_l^b \delta A_b^j \partial_a (\delta N) \\ \left. - \frac{\kappa}{2\beta^2} \epsilon^{jkl} \epsilon_{jmn} \delta_k^a \delta_l^b (\delta A_a^m \delta A_b^n + (\beta^2 + 1) \delta \Gamma_a^m \delta \Gamma_b^n - 2 \delta \Gamma_a^m \delta A_b^n) \right]. \quad (2.40) \end{aligned}$$

Let us briefly comment on the individual contributions in different orders of  $\kappa$ . The  $\kappa^0$ -order encodes the equations of motion for the uncoupled system and environment. In our case these are the Klein–Gordon equation and the equations of motion for the background connection and densitised triad variables. Since the connection vanishes in the background and the triad is just given by  $\bar{E}_i^a = \delta_i^a$  the corresponding equations of motion trivially vanish and this is again consistent with the vanishing of the background constraints that generate this trivial dynamics for the gravitational degrees of freedom. Note that the energy density of the scalar field  $\epsilon(\phi, \pi)$  in  $\kappa^0$ -order is not part of the background constraints but just contributes to the non-vanishing part of the Hamiltonian in this order. Without the boundary term a further term that contributes to the  $\kappa^0$  is the linearised Hamiltonian given by  $\bar{N}\delta C^{geo} = \bar{N}(\delta C - \kappa\epsilon)$ . On the linearised phase space this term generates the background equations. If we compute the Poisson brackets of  $\delta A_a^i, \delta E^a_i$  with  $\bar{N}\delta C^{geo}$  they both vanish for  $\bar{N} = 1$  demonstrating again that the background equations are trivially fulfilled. In linear order in  $\kappa$  in the covariant case we obtain a part that is quadratic in the perturbations of all metric components. Here, this corresponds to the terms that are either quadratic in the linearised Ashtekar variables or involve them linearly together with the linearised lapse and shift variables. These terms together in  $\delta\mathbf{H}_{\text{can}}$  will generate the left hand side of all linearised Einstein's equations including the 00 and 0a component. The right hand side of the linearised Einstein's equations will be obtained from all contributions that involve gravitational as well as matter variables in the  $\kappa^1$ -order. These will be  $\delta\mathcal{H}_I$  together with the parts where the matter variables occur in combination with the linearised lapse and shift.

As can easily be seen, most of the twenty phase space variables of the linearised theory,

$$(\phi, \pi) \quad (\delta A_a^i, \delta E^b_j), \quad (2.41)$$

are not observable, as they transform non-trivially under the linearised constraints. In the next section 2.4 we use the relational formalism to construct a set of independent gauge invariant quantities.

#### 2.4. Dirac observables in the linearised model using geometrical clocks

As mentioned above in a gauge theory we can either apply a specific gauge fixing or formulate the model in terms of gauge invariant quantities. In case we work at the level of perturbation theory, as done here, one often takes the approach that gauge invariance is guaranteed up to possible corrections that are of higher order in perturbation theory than one is truncating at, we will follow the same strategy here. To construct gauge invariant variables on the linearised phase space we consider the relational formalism [63–65] together with an observable map that maps the elementary variables of the linearised phase space to their corresponding gauge invariant quantities [66–68, 101]. The formalism is based on a choice of reference fields, one for each constraint in the system, and then the observable map returns for a given tensor field of certain rank on phase space its corresponding gauge invariant extension. Explicitly, it can be written as a power series in the reference fields weighted by contributions that involve nested Poisson brackets of the tensor fields and the constraints. The physical interpretation of these gauge invariant quantities, also known as Dirac observables in the framework of general relativity, is that they give the value of the tensor field at those values where the reference fields take specific values. For general relativity and its spatial diffeomorphism and Hamiltonian constraints this particularly means that we can formulate the dynamics of a given tensor field on phase space with respect to the reference fields and these provide a notion of physical spatial and temporal coordinates. The observable map introduced in [66–68] has the property that when applied to the chosen reference fields it maps them to phase space independent quantities and thus such a map can describe neither a canonical transformation on the full phase space one starts with nor a transformation that preserves the number of elementary phase space variables. For this reason we will slightly modify the map from [66–68] in two aspects. First we modify its action on reference fields and second we further introduce also the kind of dual version of this map to treat the choice of the reference fields and the constraints more on an equal footing, allowing to use the modified map to construct a canonical transformation for the variables  $(\phi, \pi), (\delta A_a^i, \delta E^b_j)$ . For the purpose of this work we can restrict our discussion to the linearised phase space. The transformation that we aim at constructing will map the set of elementary variables  $(\phi, \pi), (\delta A_a^i, \delta E^b_j)$  to a new set of canonical variables containing the Dirac observables associated with  $(\phi, \pi)$ , seven chosen reference fields that are canonically conjugate to the seven linearised constraints as well as two canonical pairs of further Dirac observables in the gravitational sector. Using this kind of canonical transformation for the elementary phase space variables, the physical phase space can be more easily accessed because the constraints as well as the reference fields are among the new canonical variables.

As we will discuss below for a modification of the observable maps in [66–68, 101] in our work here, we add a so called dual observable map and the combination of both allow us to construct the canonical transformation on the full phase space. Let us briefly introduce the observable map from [66–68, 101] as well as the dual one and then apply them to the model. We consider a system with a set of first class constraints  $\{C_I\}$  and elementary phase space variables  $(q^A, p_A)$ . Then we choose a set of reference variables  $\{T^I\}$  that satisfy  $\det(\{T^I, C_J\}) \neq 0$  where we abbreviate  $\{T^I, C_J\} := M^I_J$ . We can define an equivalent set of constraints given by

$$C'_I = \sum_J (M^{-1})^J_I C_J. \quad (2.42)$$

Then by construction we have that the constraints  $C'_I$  are weakly canonically conjugate to the  $T^I$ 's. Given this, one further chooses a set of functions  $\{\tau^I\}$  that can depend on the spatial and temporal coordinates by means of which one can construct an observable map which maps a function  $f$  on the phase space to its corresponding observable denoted by  $\mathcal{O}_{f, \{T\}}$

$$\begin{aligned} \mathcal{O}_{f,\{T\}}(\tau^I) &= [\exp(\xi^I\{C_I, \cdot\}) \cdot f] \Big|_{\xi^I:=T^I-\tau^I} \\ &= f + (T^I - \tau^I)\{C_I, f\} + \frac{1}{2!}(T^I - \tau^I)(T^J - \tau^J)\{C_J, \{C_I, f\}\} + \dots, \end{aligned} \tag{2.43}$$

where the label  $\{T\}$  refers to the chosen set of reference variables. The observable  $\mathcal{O}_{f,\{T\}}(\tau^I)$  returns the value of  $f$  at those values where the reference variables  $T^I$  take the values  $\tau^I$ . As can be shown [66, 68] we have  $\{C_I, \mathcal{O}_{f,T}\} \approx 0$  where  $\approx$  denotes weak equivalence, that is on the constraint hypersurface defined by the  $C_I$ 's and thus the  $\mathcal{O}_{f,\{T\}}$  are weak Dirac observables. In case the constraints  $C_I$  are already canonically conjugate to the  $T^I$ 's then the  $\mathcal{O}_{f,\{T\}}$  are even strong Dirac observables and weak equalities are replaced by equality signs. Let us introduce  $\mathcal{G}^I := T^I - \tau^I$  that we can also understand as a choice of coordinate gauge fixing condition if we require  $\mathcal{G}^I \approx 0$ . Then (2.43) can be understood as a power series in the  $\mathcal{G}^I$ 's with nested Poisson brackets involving the constraints  $C_I$  that are in this case equal to the  $C_I^J$ 's. In case we have  $\{T^I, C_J\} = \delta^I_J$  because the  $\tau^I$  do not depend on phase space variables we also obtain  $\{\mathcal{G}^I, C_J\} = \delta^I_J$ . Thus, the gauge fixing conditions and the constraints build canonically conjugate pairs. If this is the case, we can construct a dual version of the observable map in (2.43) where the role of the gauge fixing conditions and constraints are interchanged given by

$$\begin{aligned} \mathcal{O}_{f,\{C\}}^{\text{dual}} &= [\exp(-\xi_I\{\mathcal{G}^I, \cdot\}) \cdot f] \Big|_{\xi_I:=C_I} \\ &= f - C_I\{\mathcal{G}^I, f\} + \frac{1}{2!}C_I C_J\{\mathcal{G}^I, \{f, \mathcal{G}^J\}\} \pm \dots \end{aligned} \tag{2.44}$$

where the label  $\{C\}$  denotes that we have interchanged the role of gauge fixing conditions and the constraints for the dual map. This is also the reason for the additional factor of  $(-1)^n$  compared to the observable map in (2.43). Note that this dual map does not depend on any functions  $\tau^I$  since the constraint hypersurface is defined by  $C_I \approx 0$ . Similar to  $\mathcal{O}_{f,\{T\}}(\tau^I)$  which can be understood as a family of gauge invariant extension of  $f$  parameterised by  $\tau^I$  for  $\mathcal{G}^I \neq 0$ , the quantity  $\mathcal{O}_{f,\{C\}}^{\text{dual}}$  is an extension from the  $C_I \approx 0$  constraint hypersurface with the property that it commutes with all the  $\mathcal{G}^I$  by construction.

Our strategy is to construct a canonical transformation to a new set of elementary variables. For this purpose, we apply a combination of the two observable maps yielding quantities that commute by construction with all constraints  $C_I$  and all  $\mathcal{G}_I$ 's. Hence, we can choose the set  $(\mathcal{G}^I, C_I)$  as new canonical variables together with the number of independent Dirac observables obtained from the observable and its dual map associated with the set  $(q^I, p_I)$ . How many independent such Dirac observables exist depends on the number of constraints.

Note that the introduction of the equivalent set of constraints in (2.42) is also called weak Abelianisation [66] and ensures that the corresponding Hamiltonian vector fields of the  $C_I$ 's weakly commute. As a consequence, the order in which we apply the constraints  $C_I$  in the nested Poisson bracket is irrelevant. Now the situation relevant for us is that the constraints satisfy  $\{T^I, C_J\} = \delta^I_J$  and thus the  $C_I$ 's can be expressed linearly in the momenta conjugate to the  $T^I$ 's. Using that the set  $\{C_I\}$  is first class, this is sufficient to show that the constraints are abelian and hence their corresponding Hamiltonian vector fields commute and the order how we apply the  $C_I$  also does not matter in our case. However, if we consider the entire set  $(\mathcal{G}^I, C_I)$  then the  $\mathcal{G}^I$  and  $C_I$  do not commute, not even weakly and neither is the entire set first class and in general also the subset of the  $\mathcal{G}^I$  is not abelian. This has the effect that the final observable that we obtain by applying the observable map in combination with its dual in general depends on the order in which we apply the two maps as well as the order in which the  $\mathcal{G}^I$  occur in the nested Poisson brackets. This causes no problem for the model considered in this article, it

just means that there exist different choices of possible coordinate transformations on phase space, but important for us is rather that we can choose one among those.

In the following we will generalise the observable map to the field theoretic setup and perturbations theory which can easily be done. In the context of perturbation theory it is sufficient to require gauge invariance or some specific form of the Poisson algebra up to corrections that are higher than the order that is considered in perturbation theory. This means we can truncate the power series for the map and its dual at some order in accordance with the desired order in perturbation theory that we consider for the linearised model. For instance if we want to compute Poisson brackets of observables up to linear order then we need to perturb both observables up to second order and collect all terms that contribute up to linear order to the final result. Such a perturbative approach for constructing observables has for instance also been used in [69–71]. Alternatively, if available we can also take the result of the corresponding Poisson bracket in full general relativity and perturb it up to linear order. But since many quantities we work with will be known at the perturbative level only the latter option is often not possible.

The first step in the relational formalism consist of choosing suitable reference fields for the given set of constraints. In our case these will be the Hamiltonian, spatial diffeomorphism and Gauß constraint. If we are interested in results up to corrections of second order in the perturbations inside the Poisson brackets we need to consider perturbations up to second order of these constraints. The properties our chosen reference fields should have are:

- (i) Each chosen reference field consists of linear perturbations of the elementary gravitational variables and its derivatives only. These are also known as linearised geometrical clocks.
- (ii) Each chosen reference field is in lowest order canonically conjugated to one of the constraints.
- (iii) Each reference field commutes in lowest order with all constraints except the one that it is in lowest order canonically conjugated to.
- (iv) All reference fields mutually commute.

Let us introduce the notation

$$\delta\mathcal{C}_I(\vec{x}, t) := (\delta C(\vec{x}, t), \delta C_a(\vec{x}, t), \delta G_i(\vec{x}, t)), \quad \delta^2\mathcal{C}_I(\vec{x}, t) := (\delta^2 C(\vec{x}, t), \delta^2 C_a(\vec{x}, t), \delta^2 G_i(\vec{x}, t))$$

for the set of linearised constraints  $\delta\mathcal{C}_I(\vec{x}, t)$  and  $\delta^2\mathcal{C}_I(\vec{x}, t)$  for the set of second-order perturbations. For each of the individual constraints we have to choose one reference field and we introduce the following notation  $\mathcal{G}^I(\vec{x}, t)$  for this set with

$$\begin{aligned} \mathcal{G}^I(\vec{x}, t) &:= (\delta\mathcal{G}(\vec{x}, t), \delta\mathcal{G}^a(\vec{x}, t), \delta\mathcal{G}^j(\vec{x}, t)) \\ \delta\mathcal{G}(\vec{x}, t) &:= \delta T(\vec{x}, t) - \delta\tau(\vec{x}, t), \quad \delta\mathcal{G}^a(\vec{x}, t) := \delta T^a(\vec{x}, t) - \delta\sigma^a(\vec{x}, t), \\ \delta\mathcal{G}^j(\vec{x}, t) &:= \delta\Xi^j(\vec{x}, t) - \delta\xi^j(\vec{x}, t), \end{aligned} \tag{2.45}$$

where  $\delta T, \delta T^a, \delta\Xi^j$  denote the individual reference fields for the Hamiltonian, spatial diffeomorphism and Gauß constraint respectively. We denote background quantities with a bar and assume that in the background the gauge fixing conditions as well as the constraints are satisfied, that is

$$\begin{aligned} \bar{\mathcal{G}}(\vec{x}, t) &:= \bar{T}(\vec{x}, t) - \bar{\tau}(\vec{x}, t) = 0, \bar{C} = 0, \quad \bar{\mathcal{G}}^a(\vec{x}, t) := \bar{T}^a(\vec{x}, t) - \bar{\sigma}^a(\vec{x}, t) = 0, \bar{C}_a = 0, \\ \bar{\mathcal{G}}^j(\vec{x}, t) &:= \bar{T}^j(\vec{x}, t) - \bar{\xi}^j(\vec{x}, t) = 0, \bar{G}_j = 0. \end{aligned}$$

With the assumptions (i)–(iv) above we know that  $\delta^2 \mathcal{G}^I = 0$  and further we have

$$\begin{aligned} \{\mathcal{G}^I(x), \mathcal{C}_J(y)\} &: = \{\delta \mathcal{G}^I(x), \delta \mathcal{C}_J(y)\} + \{\delta \mathcal{G}^I(x), \delta^2 \mathcal{C}_J(y)\} + O(\delta^2, \kappa^2) \\ &= \frac{1}{\kappa} \delta_J^I \delta^{(3)}(x, y) + \{\delta \mathcal{G}^I(x), \delta^2 \mathcal{C}_J(y)\} + O(\delta^2, \kappa^2), \end{aligned} \quad (2.46)$$

where  $O(\delta^2, \kappa^2)$  means that we neglect all terms that are second-order in the perturbations and/or of order  $\kappa^2$  and the factor  $\frac{1}{\kappa}$  has been chosen because it is also involved in the Poisson bracket of the elementary gravitational variables. This means there could be terms being of order  $\kappa^n$  with  $n \leq 1$  that do not contribute because they involve second or higher orders of  $\delta$ . That those terms can be present is caused by an asymmetry in  $\delta$  and  $\kappa$  due to the fact that we only perturb the gravitational degrees of freedom but not the matter ones and further that the individual terms in the action involve different powers of  $\kappa$  from the beginning. We realise that the linearised constraints  $\delta \mathcal{C}_I$  are canonically conjugated to the reference fields  $\mathcal{G}^I$  but in general there can be non-vanishing contribution in linear order coming from the Poisson bracket  $\{\delta \mathcal{G}^I(x), \delta^2 \mathcal{C}_J(y)\}$ . To ensure that we have a vanishing contribution in linear order we will use the dual observable map as discussed below. For this purpose we need to adapt the observable map and its dual to field theory and perturbation theory and for both maps we need the observable map up to second order. Taking this into account the observable formula up to second order reads

$$\begin{aligned} \mathcal{O}_{f, \{T\}}(\delta \tau, \delta \sigma^a, \delta \xi^i) &= \delta f + \delta^2 f + \kappa \int_{\sigma} d^3 y \delta \mathcal{G}^I(y) \{\delta \mathcal{C}_I(y), \delta f\} \\ &\quad + \kappa \int_{\sigma} d^3 y \delta^2 \mathcal{G}^I(y) \{\delta \mathcal{C}_I(y), \delta f\} \\ &\quad + \kappa \int_{\sigma} d^3 y \delta \mathcal{G}^I(y) \left( \{\delta^2 \mathcal{C}_I(y), \delta f\} + \{\delta \mathcal{C}_I(y), \delta^2 f\} \right) \\ &\quad + \frac{\kappa^2}{2} \int_{\sigma} d^3 y \delta \mathcal{G}^I(y) \int_{\sigma} d^3 z \delta \mathcal{G}^J(z) \left( \{\delta \mathcal{C}_J(z), \{\delta^2 \mathcal{C}_I(y), \delta f\}\} \right. \\ &\quad \left. + \{\delta \mathcal{C}_J(z), \{\delta \mathcal{C}_I(y), \delta^2 f\}\} \right) + O(\delta^3, \kappa^3), \end{aligned} \quad (2.47)$$

where we used that  $\bar{\mathcal{G}} = \bar{\mathcal{G}}^a = \bar{\mathcal{G}}^j = 0$  and allowed possible non-vanishing second-order perturbations of  $\mathcal{G}^I$  that might be present if one wants to relax the assumptions (i)–(iv) from above. Because we consider perturbations around flat spacetime, in the background it holds that  $\bar{\mathcal{C}} = 0$ ,  $\bar{\mathcal{C}}_a = 0$  as well as  $\bar{\mathcal{G}}_i = 0$ , all trivially vanish and hence we also have  $\bar{\mathcal{G}} = 0$ ,  $\bar{\mathcal{G}}^a = 0$  and  $\bar{\mathcal{G}}^j = 0$ . Hence, for the background there is no gauge freedom we have to deal with and thus no corresponding observables to construct. Note that if we construct the observable for elementary phase space variables then  $\delta^2 f = 0$  and the observable formula above simplifies. This construction of observables order by order in  $\kappa$  plays also a role when we consider the linearised Hamiltonian  $\delta \mathbf{H}_{\text{can}}$  which, as can be seen from (2.40), has contributions in  $\kappa^0$  and linear order in  $\kappa$ . The transformation behaviour under the linearised constraints of the matter variables can be found in the appendix A.1 in (A.7) and (A.11) and the results are an expression linear in  $\kappa$ . This again demonstrates that in the limit where  $\kappa$  is sent to zero, which corresponds to the situation that we consider a scalar field on Minkowski only with no coupling to linearised gravity, the elementary variables  $\phi, \pi$  are suitable observables. Once we consider the coupling with linearised gravity perturbations, this is no longer the case and we need the gauge invariant version of these variables. For the  $\delta \mathbf{H}_{\text{can}}$  this means that in the  $\kappa^0$  term we can still work with

the original  $\phi, \pi$  whereas for the linear order in  $\kappa$  we also need to construct Dirac observables for the matter sector, see for instance also the discussion in [100] in the context of the covariant theory.

Note that as before, the linearised phase space in the Post-Minkowski approximation scheme involves the linear perturbations of the gravitational degrees of freedom as well as the variables  $(\phi, \pi)$  for the matter sector. In a similar way we obtain the linearised dual observable map given by

$$\begin{aligned} \mathcal{O}_{f, \{C\}}^{\text{dual}} &= \delta f + \delta^2 f - \kappa \int_{\sigma} d^3 y \delta \mathcal{C}_I(y) \{ \delta \mathcal{G}^I(y), \delta f \} \\ &\quad - \kappa \int_{\sigma} d^3 y \delta^2 \mathcal{C}_I(y) \{ \delta \mathcal{G}^I(y), \delta f \} \\ &\quad - \kappa \int_{\sigma} d^3 y \delta \mathcal{C}_I(y) \left( \{ \delta^2 \mathcal{G}^I(y), \delta f \} + \{ \delta \mathcal{G}^I(y), \delta^2 f \} \right) \\ &\quad + \frac{\kappa^2}{2} \int_{\sigma} d^3 y \delta \mathcal{C}_I(y) \int_{\sigma} d^3 z \delta \mathcal{C}_J(z) \left( \{ \delta \mathcal{G}^J(z), \{ \delta^2 \mathcal{G}^I(y), \delta f \} \} \right. \\ &\quad \left. + \{ \delta \mathcal{G}^J(z), \{ \delta \mathcal{G}^I(y), \delta^2 f \} \} \right) + \mathcal{O}(\delta^3, \kappa^3). \end{aligned} \quad (2.48)$$

The alternating signs compared to the observable map in (2.43) are needed since the order of how the constraints and the reference fields enter into the Poisson bracket is switched. Note that if we choose for instance  $\mathcal{G}^I$  such that  $\delta^2 \mathcal{G}^I = 0$  and not the non-linear  $\mathcal{C}_I$  but only the linearised parts as the clocks then we drop all terms that involve  $\delta^2 \mathcal{G}^I, \delta^2 \mathcal{C}_I$  in the dual observable map (2.48) and this will be exactly the application we need later.

As we will work with so-called geometrical clocks, in a first step we will choose the set of reference fields among the linearised elementary gravitational degrees of freedom. As a consequence, we will work with geometrical clocks for which  $\delta^2 \mathcal{G}^I = 0$  and hence they satisfy assumption (i). To also fulfil (ii) we further choose  $\mathcal{G}^I$  such that

$$\{ \delta \mathcal{G}^I(x), \delta \mathcal{C}_J(y) \} = \frac{1}{\kappa} \delta^I_J \delta^{(3)}(x, y),$$

where the factor  $\frac{1}{\kappa}$  is chosen because it is also involved in Poisson bracket of the elementary geometrical variables. These requirements still allow several choices for suitable  $\mathcal{G}^I$  and the specific choice for  $\mathcal{G}^I$  taken here is motivated by the fact that we can relate the set of gauge invariant variables to a gauge fixing often chosen in the context of linearised gravity as will be discussed at the end of this subsection. For the reference fields  $\delta T(\vec{x}, t), \delta \Xi^I(\vec{x}, t)$ , for the Hamiltonian constraint and the Gauß constraint respectively we choose

$$\delta T(\vec{x}, t) := -\frac{1}{\kappa \beta} \left[ \frac{1}{2} \delta_c^i \epsilon_a^{cb} \partial_b (\delta E^a_i * G^\Delta)(\vec{x}, t) + \delta_i^a (\delta A_a^i * G^\Delta)(\vec{x}, t) \right], \quad (2.49)$$

$$\delta \Xi^i(\vec{x}, t) := \frac{2}{\kappa} \partial^a (\delta A_a^i * G^\Delta)(\vec{x}, t), \quad (2.50)$$

with the three dimensional, spatial convolution

$$(\delta E^a_i * G^\Delta)(\vec{x}, t) := \int d^3 y \delta E^a_i(\vec{y}, t) G^\Delta(\vec{x} - \vec{y}), \quad (2.51)$$

the derivative of the convolution acting with respect to  $\vec{x}$ , hence only on  $G^\Delta(\vec{x} - \vec{y})$ , and the Green's function of the Laplacian

$$G^\Delta(\vec{x} - \vec{y}) := \int \frac{d^3k}{(2\pi)^3} \frac{1}{\|\vec{k}\|^2} e^{i\vec{k}(\vec{x} - \vec{y})}, \quad (2.52)$$

such that  $\Delta G^\Delta(\vec{x} - \vec{y}) = -\delta(\vec{x} - \vec{y})$ . Thus the reference fields are, as in general expected, non-local quantities because they are local in momentum space. As discussed in the appendix A.1 these reference fields satisfy

$$\{\delta T(\vec{x}, t), \delta C(\vec{y}, t)\} = \frac{1}{\kappa} \delta(\vec{x} - \vec{y}), \quad \{\delta \Xi^i(\vec{x}, t), \delta G_j(\vec{y}, t)\} = \frac{1}{\kappa} \delta_j^i \delta(\vec{x} - \vec{y}), \quad \{\delta T(\vec{x}, t), \delta \Xi^i(\vec{y}, t)\} = 0.$$

Furthermore, the reference fields  $\delta T, \delta \Xi^i$  commute with all remaining linearised constraints and hence they satisfy (iii) and (iv) as well. A more detailed discussion can be found in the appendix A.1. Reference fields that are canonically conjugate to the components of the spatial diffeomorphism constraint and commute with the linearised Hamiltonian as well as the linearised Gauß constraint are given by

$$\delta \tilde{T}^a(\vec{x}, t) := \frac{1}{\kappa} \left( \delta_b^a \delta_c^i \partial^c - \frac{1}{2} \delta_b^i \partial^a + \delta^{ac} \delta_c^i \partial_b \right) (\delta E^b{}_i * G^\Delta)(\vec{x}, t). \quad (2.53)$$

Although the reference fields  $\delta \tilde{T}^a(\vec{x}, t)$  mutually commute, they do not have vanishing Poisson brackets with all the remaining reference fields and violate assumption (iv). To obtain reference fields for the spatial diffeomorphism constraint we can employ the dual observables map for  $\delta \tilde{T}^a(\vec{x}, t)$  yielding a quantity that by construction commutes with  $\delta \mathcal{G}, \delta \mathcal{G}^a, \delta \mathcal{G}^j$  and hence with all reference fields. As the clock for the dual observable map we choose the geometric part of the linearised Hamiltonian constraint, that is  $\delta C - \kappa \epsilon$  as well as the linearised Gauß constraint. Note that  $\delta C - \kappa \epsilon \not\approx 0$  and as a consequence  $\delta \tilde{T}^a \not\approx \delta T^a$  but this causes no problems because two different choices of set of clocks need not necessarily to be weakly equivalent. In order that the observable formula can be applied we just need  $\{\delta T^I, \delta \mathcal{C}_J\} = \{\delta \mathcal{G}^I, \delta \mathcal{C}_J\} = \delta_J^I$  which is satisfied also if we use the geometric contribution of  $\delta C$  only. For  $\delta \tilde{T}^a(\vec{x}, t)$  the part corresponding to the spatial diffeomorphism constraint will not contribute in the dual observable map. Since we use the linearised constraint as the clock and not  $\mathcal{C}_I$  we can drop all terms in involving  $\delta^2 \mathcal{C}_I$ . Using this together with  $\delta^2 \mathcal{G}^I = 0$  the dual observable map in (2.44) simplifies in this case to

$$\mathcal{O}_{\delta \tilde{T}^a, \{\delta C\}}^{\text{dual}} =: \delta T^a = \delta \tilde{T}^a - \kappa \int_{\sigma} d^3y (\delta C - \kappa \epsilon)(y) \{\delta T(y), \delta \tilde{T}^a\} - \kappa \int_{\sigma} d^3y \delta G_j(y) \{\delta \Xi^j(y), \delta \tilde{T}^a\}.$$

The explicit result we obtain has the form

$$\begin{aligned} \delta T^a(\vec{x}, t) &:= \frac{1}{\kappa} \left( \delta_b^a \delta_c^i \partial^c - \frac{1}{2} \delta_b^i \partial^a + \delta^{ac} \delta_c^i \partial_b \right) (\delta E^b{}_i * G^\Delta)(\vec{x}, t) \\ &\quad + \frac{2\beta}{\kappa^2} \left[ \frac{1}{2} \delta_b^i \partial^a \partial^b (\delta G_i * G^{\Delta\Delta})(\vec{x}, t) - \delta^{ab} \delta_b^i (\delta G_i * G^\Delta)(\vec{x}, t) \right] \\ &\quad + \frac{1}{2\kappa^2} \partial^a [(\delta C - \kappa \epsilon) * G^{\Delta\Delta}](\vec{x}, t), \end{aligned} \quad (2.54)$$

where  $G^{\Delta\Delta}$  denotes the Green's function of the squared Laplacian, that is

$$G^{\Delta\Delta}(\vec{x} - \vec{y}) = \int \frac{d^3k}{(2\pi)^3} \frac{1}{\|\vec{k}\|^4} e^{i\vec{k}(\vec{x} - \vec{y})} = \int d^3z G^\Delta(\vec{x} - \vec{z}) G^\Delta(\vec{y} - \vec{z}). \quad (2.55)$$

Note that in [69] also geometrical clocks were used based on the ADM clocks introduced in [74, 102]. This set of clocks is not equivalent to ours and when requiring that all clocks vanish this corresponds to a different gauge fixing, as is discussed at the end of appendix B.1.

Now we are in the situation, that the seven linearised constraints as well as the seven reference fields form in lowest order two abelian subalgebras and further obey the following Poisson algebra:

$$\{\delta T(\vec{x}, t), \delta C(\vec{y}, t)\} = \frac{1}{\kappa} \delta(\vec{x} - \vec{y}) \quad \{\delta T(\vec{x}, t), \delta C_a(\vec{y}, t)\} = 0 \quad \{\delta T(\vec{x}, t), \delta G_i(\vec{y}, t)\} = 0 \quad (2.56)$$

$$\{\delta T^a(\vec{x}, t), \delta C(\vec{y}, t)\} = 0 \quad \{\delta T^a(\vec{x}, t), \delta C_b(\vec{y}, t)\} = \frac{1}{\kappa} \delta_b^a \delta(\vec{x} - \vec{y}) \quad \{\delta T^a(\vec{x}, t), \delta G_i(\vec{y}, t)\} = 0 \quad (2.57)$$

$$\{\delta \Xi^i(\vec{x}, t), \delta C(\vec{y}, t)\} = 0 \quad \{\delta \Xi^i(\vec{x}, t), \delta C_a(\vec{y}, t)\} = 0 \quad \{\delta \Xi^i(\vec{x}, t), \delta G_j(\vec{y}, t)\} = \frac{1}{\kappa} \delta_j^i \delta(\vec{x} - \vec{y}). \quad (2.58)$$

Since  $\delta \mathcal{G}$ ,  $\delta \mathcal{G}^c$ ,  $\delta \mathcal{G}^j$  differ from  $\delta T$ ,  $\delta T^c$ ,  $\delta \Xi^j$  by some phase space independent quantity only, we can replace the reference fields by the corresponding  $\delta \mathcal{G}$ 's and the Poisson algebra above will not change. We have defined the Poisson algebra with a factor  $\frac{1}{\kappa}$  because the gravitational degrees of freedom involve a similar factor. The algebra reference fields have been defined such that the algebra of them and the linearised constraints no longer involves the Barbero-Immirzi parameter on the right hand side for the reason that  $\kappa$  is the parameter labelling the order of perturbations. Going back to the Poisson algebra of  $\mathcal{G}^I$  and  $\mathcal{C}_I$  in (2.46) so far we have chosen reference fields that satisfy this algebra. However, the in general non-vanishing contribution in linear order will prevent us from chosen  $\mathcal{G}^I$  and  $\mathcal{C}_I$  as new canonical coordinates. To achieve this we will apply the dual observable map to  $\mathcal{C}_I$  and obtain an equivalent set of constraints  $\mathcal{C}'_I$  satisfying

$$\{\mathcal{G}^I(x), \mathcal{C}'_J(y)\} = \{\delta \mathcal{G}^I(x), \delta \mathcal{C}'_J(y) + \delta^2 \mathcal{C}'_J(y)\} = \frac{1}{\kappa} \delta_J^I \delta^{(3)}(x, y) + O(\delta^2, \kappa^2). \quad (2.59)$$

With this algebra given we can then choose  $\delta \mathcal{G}^I$  as new configuration variables and  $\delta \mathcal{C}'_I$  as new momentum variables. Because in lowest order we have already the correct algebra relations we only need a modification in the linear order. To accomplish this we apply the dual observable map to all  $\mathcal{C}_I$  and drop the linear order terms. The latter would just abelianise the lowest order and cancel the  $\mathcal{C}_I$  contribution in the observable something we do not want. The second order that we keep will then modify the linear order in a way that we obtain the algebra shown in (2.59). Since we want an equivalent set of constraints here we are forced to choose as clocks the linearised constraints  $\delta \mathcal{C}_I$  involving also the matter contributions in  $\delta C$  and  $\delta C_a$ . Taking all this into account the equivalent constraints  $\mathcal{C}'_I$  are given by

$$\begin{aligned} \mathcal{C}'_I &:= \mathcal{O}_{\mathcal{C}_I, \{\delta C\}}^{\text{dual}} = \delta \mathcal{C}_I + \delta^2 \mathcal{C}_I - \kappa \int_{\sigma} d^3 y \delta \mathcal{C}_I(y) \{\delta \mathcal{G}^I(y), \delta^2 \mathcal{C}_I\} \\ &+ \frac{\kappa^2}{2} \int_{\sigma} d^3 y \delta \mathcal{C}_I(y) \int_{\sigma} d^3 z \delta \mathcal{C}_K(z) \{\delta \mathcal{G}_J(z), \{\delta \mathcal{G}_K(y), \delta^2 \mathcal{C}_I\}\} + O(\delta^3, \kappa^3), \end{aligned} \quad (2.60)$$

where we used again that  $\delta^2 \mathcal{G}^I = 0$  in our case. Up to linear order we have  $\delta \mathcal{C}'_I = \delta \mathcal{C}_I$  showing that as expected we have modified the constraint only in second order. This modifications become not relevant when we work with the constraints directly in the linearised theory since they are of second order but need to be considered when we compute Poisson brackets and want

to consider the result up to linear order. For a set of mutually commuting reference fields, that we consider here, in a more general context it was proven in [69] that the constraints  $C'_I$  are abelian up to corrections of second order and applying this result here we have

$$\{C'_I(x), C'_J(y)\} = \{\delta C'_I(x) + \delta^2 C'_I(x), \delta C'_J(y) + \delta^2 C'_J(y)\} = 0 + \mathcal{O}(\delta^2, \kappa^2)$$

and thus in the order of perturbation theory we consider here, we can treat them as abelian constraints. Because we can just consider higher orders in the dual observable map we can always ensure that the constraints are abelian up to a chosen order in perturbation theory and if the entire sum of the dual observable map converges even in the full unperturbed theory. Note that a similar result has already been discussed in [69] in the context of applying weak abelianisation order by order in perturbation theory and afterwards modifying the constraints by terms that involve higher order powers of the linearised constraints. Here we can rediscover that case in the framework of the dual observable map which seems to us to be slightly more general because it cannot only be applied to the constraints as this corresponds to the case discussed in [69] but to any phase space function as we did for instance for the clock of the spatial diffeomorphism constraint before. Further, the strategy discussed in [69] separates weak Abelianisation and the addition of terms higher order in the linearised constraints where this happens all at once here using the dual observable map from second order on.

Given the abelian constraints  $C'_I(x)$  now we can use them in the observable map in (2.43) and construct gauge invariant quantities that commute by construction with the constraints  $C'_I(x)$  up to  $\mathcal{O}(\delta^2, \kappa^2)$  corrections. We want to apply the observable map to the elementary phase space variables of the matter and gravitational degrees of freedom and use  $\delta\mathcal{G}^I$  as our clocks. Therefore, we can drop all terms involving  $\delta^2 f$  and  $\delta^2 \mathcal{G}^I$  and the observable map reduces to

$$\begin{aligned} (\delta f)^{GI}(\delta\tau, \delta\sigma^a, \delta\xi^j) &:= \mathcal{O}_{\delta f, \{\delta T\}}^{GI}(\delta\tau, \delta\sigma^a, \delta\xi^j) \\ &= \mathcal{O}_{\delta f, \{\delta T\}}^{(1)}(\delta\tau, \delta\sigma^a, \delta\xi^j) + \mathcal{O}_{\delta f, \{\delta T\}}^{(2)}(\delta\tau, \delta\sigma^a, \delta\xi^j) + \mathcal{O}(\delta^3, \kappa^3), \end{aligned} \quad (2.61)$$

where the label  $GI$  means gauge invariant with

$$\begin{aligned} \mathcal{O}_{\delta f, \{\delta T\}}^{(1)}(\delta\tau, \delta\sigma^a, \delta\xi^j) &:= \delta f + \kappa \int_{\sigma} d^3 y \delta\mathcal{G}^I(y) \{\delta C'_I(y), \delta f\} \\ \mathcal{O}_{\delta f, \{\delta T\}}^{(2)}(\delta\tau, \delta\sigma^a, \delta\xi^j) &:= \kappa \int_{\sigma} d^3 y \delta\mathcal{G}^I(y) \{\delta^2 C'_I(y), \delta f\} \\ &\quad + \frac{\kappa^2}{2} \int_{\sigma} d^3 y \delta\mathcal{G}^I(y) \int_{\sigma} d^3 z \delta\mathcal{G}^J(z) \left( \{\delta C'_J(z), \{\delta^2 C'_I(y), \delta f\}\} \right) \end{aligned} \quad (2.62)$$

where  $\mathcal{O}_{\delta f, \{\delta T\}}^{(1)}$  and  $\mathcal{O}_{\delta f, \{\delta T\}}^{(2)}$  denote all contributions to the observable in linear and second order respectively. Since in this work we perturb the geometrical degrees of freedom only for the matter sector we just replace  $\delta f$  by  $f = \phi, \pi$  in the formula above. That this works so easily in the matter sector might not be obvious in the first place for the following reason: If we consider in  $\mathcal{O}_{\delta f, \{\delta T\}}^{(1)}$  the second term then for  $\delta f$  chosen from geometric degrees of freedom this will be of zeroth order and hence trivially commute with the linearised or higher order constraints. In contrast, for the matter variables the corresponding term still involves the matter variables linearly and hence does not commute with the linearised constraints or higher order ones and thus yields an additional contribution compared to the geometric case. However, since the Poisson brackets of the gravitational degrees of freedom involve a factor  $\frac{1}{\kappa}$ , whereas the Poisson brackets of the matter variables do not, these additional contributions come with

an extra factor of  $\kappa$ . In the linearised theory we require for Dirac observables to commute with the constraints up to order  $\delta$  and  $\kappa$  and hence, these kind of contributions will only occur in higher orders in  $\kappa$ . A similar argument applies to these kind of additional contributions in  $\mathcal{O}_{\delta f, \{\delta T\}}^{(2)}$  and therefore we can indeed apply the observable formula also to the matter variables. To obtain the explicit form of these Dirac observables we apply the observable map in (2.62) to  $(\phi, \pi, \delta A_a^i, \delta E^a_i)$  and obtain

$$\begin{aligned} \phi^{GI}(\delta\sigma^c, \delta\tau, \delta\xi^j) &= \phi(\vec{x}, t) - \kappa^2(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t))\pi(\vec{x}, t) \\ &\quad - \kappa^2(\delta T^c(\vec{x}, t) - \delta\sigma^c(\vec{x}, t))\partial_c\phi(\vec{x}, t) \\ &\quad + \mathcal{O}_{\phi, \{\delta T\}}^{(2)}(\delta\tau, \delta\sigma^a, \delta\xi^j) + \mathcal{O}(\delta^3, \kappa^3), \end{aligned} \quad (2.63)$$

$$\begin{aligned} \pi^{GI}(\delta\sigma^c, \delta\tau, \delta\xi^j) &= \pi(\vec{x}, t) - \kappa^2\partial_a[(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t))\partial^a\phi(\vec{x}, t)] \\ &\quad - \kappa^2\partial_c[(\delta T^c(\vec{x}, t) - \delta\sigma^c(\vec{x}, t))\pi(\vec{x}, t)] + \kappa^2(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t))m^2\phi(\vec{x}, t) \\ &\quad + \mathcal{O}_{\pi, \{\delta T\}}^{(2)}(\delta\tau, \delta\sigma^a, \delta\xi^j) + \mathcal{O}(\delta^3, \kappa^3), \end{aligned} \quad (2.64)$$

$$\begin{aligned} (\delta E^a_i)^{GI}(\delta\sigma^a, \delta\tau, \delta\xi^j) &= \delta E^a_i(\vec{x}, t) - \kappa(\delta_i^a\partial_c - \delta_c^a\delta_i^b\partial_b)(\delta T^c(\vec{x}, t) - \delta\sigma^c(\vec{x}, t)) \\ &\quad + \kappa\beta\epsilon_c^{ab}\delta_i^c\partial_b(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t)) - \frac{\kappa}{2}\epsilon_{ij}^k\delta_k^a(\delta\Xi^j(\vec{x}, t) - \delta\xi^j(\vec{x}, t)) \\ &\quad + \mathcal{O}_{\delta E^a_i, \{\delta T\}}^{(2)}(\delta\tau, \delta\sigma^a, \delta\xi^j) + \mathcal{O}(\delta^3, \kappa^2), \end{aligned} \quad (2.65)$$

$$\begin{aligned} (\delta A_a^i)^{GI}(\delta\sigma^a, \delta\tau, \delta\xi^j) &= \delta A_a^i(\vec{x}, t) + \frac{\kappa}{2}\partial_a(\delta\Xi^i(\vec{x}, t) - \delta\xi^i(\vec{x}, t)) \\ &\quad + \mathcal{O}_{\delta A_a^i, \{\delta T\}}^{(2)}(\delta\tau, \delta\sigma^a, \delta\xi^j) + \mathcal{O}(\delta^3, \kappa^2), \end{aligned} \quad (2.66)$$

where we displayed only the linear order in explicit form and  $(\delta\tau(\vec{x}, t), \delta\sigma^c(\vec{x}, t), \delta\xi^j(\vec{x}, t))$  denote the spacetime functions corresponding to the reference fields  $(\delta T, \delta T^c, \delta\Xi^j)$ , see  $\delta\mathcal{G}, \delta\mathcal{G}^c$  and  $\delta\mathcal{G}^j$  in (2.45). The so constructed observables are Dirac observables for all values of the functions  $(\delta\tau(\vec{x}, t), \delta\sigma^c(\vec{x}, t), \delta\xi^j(\vec{x}, t))$  and thus each can be understood as a family of Dirac observables parameterised by  $\delta\tau(\vec{x}, t), \delta\sigma^c(\vec{x}, t)$  and  $\delta\xi^j(\vec{x}, t)$  respectively. Now let us choose specific spacetime functions for  $\delta\tau(\vec{x}, t), \delta\sigma^c(\vec{x}, t)$  and  $\delta\xi^j(\vec{x}, t)$ . A specific choice for  $\delta\tau(\vec{x}, t), \delta\sigma^c(\vec{x}, t)$  and  $\delta\xi^j(\vec{x}, t)$  can be chosen as follows: For the Gauß constraint we choose vanishing parameters  $\delta\xi^j(\vec{x}, t)$ . The remaining parameters we associate with the temporal and spatial coordinates respectively. Taking into account that the full non-linear parameters  $\tau$  and  $\sigma^c$  with  $\tau = \bar{\tau} + \kappa\delta\tau + \mathcal{O}(\kappa^2)$  and  $\sigma^c = \bar{\sigma}^c + \kappa\delta\sigma^c + \mathcal{O}(\kappa^2)$  should be set to  $t$  and  $x_\sigma^c$  respectively and that the clocks vanish in the background as well as for higher orders than the linear one, we only need to choose the linearised parameters  $\delta\tau$  and  $\delta\sigma^c$  yielding

$$\delta\tau(\vec{x}, t) := \frac{t}{\kappa}, \quad \delta\sigma^c(\vec{x}, t) := \frac{x_\sigma^c}{\kappa}, \quad \delta\xi^j(\vec{x}, t) := 0, \quad (2.67)$$

with  $x_\sigma^c$  being the unique solution where  $\delta T^c(\vec{x}, t) = \frac{1}{\kappa}\sigma^c$  with  $\sigma^c = \text{const}$ . We reinsert these choices into  $\delta\mathcal{G}, \delta\mathcal{G}^c$  and  $\delta\mathcal{G}^j$ , respectively, in (2.45). Then we can understand the Dirac observables as functions of  $(\vec{x}_\sigma, t)$ .

We have that  $(\phi^{GI}, \pi^{GI})$  are already the independent Dirac observables in the matter sector. However, the 18 observables  $(\delta A_a^i)^{GI}, (\delta E^a_i)^{GI}$  are not all independent, but only four of them are in total. For completeness we have presented the explicit forms of  $(\delta A_a^i)^{GI}$  and  $(\delta E^a_i)^{GI}$  in Fourier space in (A.45) and (A.46) respectively. We realise that  $\delta\tau$  and  $\delta\sigma^c$  enter explicitly into these formulae. However, for the choice discussed here where  $\delta\tau$  is linearly in  $t$  and  $\delta\sigma^c$  is constant the contributions both vanish because they come with spatial derivatives acting on  $\delta t$  and  $\delta\sigma^c$ . In case we choose  $\delta\sigma^c$  linearly in  $x^c$  then  $(\delta E^a_i)^{GI}$  has an additional term involving

a Kronecker delta that also survives when we express  $(\delta E^a_i)^{GI}$  in terms of the independent gauge invariant degrees of freedom similar to what was observed in [74]. As a consequence, considering  $(\delta E^a_i)^{GI}$  as an isolated quantity its fall-off behaviour gets modified and does no longer satisfy the requirement in (2.10). This, however causes no issue because we need this quantity as an intermediate step only to reinsert it into the physical Hamiltonian. The final result of the physical Hamiltonian has a suitable fall-off behaviour because these critical terms are combined with scalar field contributions for which we assumed, as usually done, that the initial data has compact support, see also our discussion below (2.76). The four independent degrees of freedom in the gravitational sector are encoded in the symmetric<sup>3</sup> transverse-traceless part of the the variables  $(\delta A_a^i)^{GI}$ ,  $(\delta E^a_i)^{GI}$  and therefore the 4 independent Dirac observables in the gravitational sector are given by

$$\delta \mathcal{A}_a^i(\vec{x}_\sigma, t) := P_a^{ij}(\delta A_b^j)^{GI}(\vec{x}_\sigma, t) \quad \delta \mathcal{E}^a_i(\vec{x}_\sigma, t) := P_a^{ij}(\delta E^b_j)^{GI}(\vec{x}_\sigma, t), \quad (2.68)$$

where  $P_a^{ij}$  and  $P_a^{ij}$  denote the projector on the transverse-traceless part in position space. Its explicit form is given by

$$\begin{aligned} P_a^{ij} X(\vec{x}_\sigma, t) = \frac{1}{2} \left\{ \left[ \delta_{aj} \delta^{bi} + \delta_a^b \delta_j^i - \delta_a^i \delta_j^b \right] X(\vec{x}_\sigma, t) + \partial_a \partial^i \partial_j \partial^b \left( X * G^{\Delta\Delta} \right) (\vec{x}_\sigma, t) \right. \\ \left. + \left[ \delta_{aj} \partial^i \partial^b + \delta^{ib} \partial_a \partial_j + \delta_a^b \partial^i \partial_j + \delta_j^i \partial_a \partial^b - \delta_a^i \partial_j \partial^b - \delta_j^b \partial_a \partial^i \right] \left( X * G^\Delta \right) (\vec{x}_\sigma, t) \right\}, \end{aligned} \quad (2.69)$$

where  $\partial_i := \delta_i^c \partial_c$ ,  $\partial^i := \delta_c^i \partial^c$ ,  $\delta_{aj} := \delta_{ac} \delta_j^c$ , and

$$P_a^{ij} X(\vec{x}_\sigma, t) = \delta^{ac} \delta_{ik} \delta^{jm} \delta_{bd} P_c^k P_m^d X(\vec{x}_\sigma, t). \quad (2.70)$$

As expected these projectors are non-local in position space but local in momentum space, see the appendix A.2 for more details. A comparison between the local projector in position space also often used in the literature in the context of the transverse-traceless gauge has been analysed in [103].

From now on we will drop the label  $\sigma$  at  $\vec{x}_\sigma$  and by an abuse of notation denote it by  $\vec{x}$  again and always keep in mind that these  $x^a$  coordinates are related to values that the reference field  $T^a$  takes. By construction the Dirac observables Poisson commute with all constraints  $\mathcal{C}'_I$ . Since the set  $\mathcal{C}'_I$  is weakly equivalent to the set  $\mathcal{C}_I$  the Dirac observables are weak Dirac observables with respect to  $\mathcal{C}_I$  where the weak equivalence refers to the constraint hypersurface defined by the linearised constraints  $\delta \mathcal{C}_I$  that, as discussed above, agree with  $\delta \mathcal{C}'_I$ . Furthermore, as discussed in the appendix A all Dirac observables truncated at linear order, that is neglecting contributions from  $\mathcal{O}_{\delta f, \{\delta T\}}^{(2)}$ , Poisson commute with all linearised constraints. What still remains to be discussed is the algebra of the independent Dirac observables. For this purpose we will consider the relation between the algebra of Dirac observables and the observable associated to the corresponding Dirac bracket. As proven in theorem 2.2 in [101], see also [66] for an alternative proof in general we have

$$\{O_{f, \{T\}}, O_{g, \{T\}}\} \approx O_{\{f, g\}^*, \{T\}}$$

where  $\{f, g\}^*$  denotes the Dirac bracket associated with the total set  $(\mathcal{G}^I, \mathcal{C}_J)$ , that, when requiring  $\mathcal{G}^I$  to be gauge fixing conditions is a second class system of constraints for which a Dirac

<sup>3</sup> If one considers the soldered version of the quantities, i.e.  $\delta \mathcal{A}_{ab} := \delta \mathcal{A}_a^i \delta_i^c \delta_{cb}$  and analogously for  $\delta \mathcal{E}^{ab} := \delta \mathcal{E}^a_i \delta_i^c \delta^{cb}$ .

bracket can be constructed. If we have such a relation also at the perturbative level then it provides an efficient way to compute the observable algebra because if we are interested in the algebra up to linear order then we can just expand the observable on the right hand side up to linear order whereas we need to consider the observables up to second order on the left hand side. In the following we will not consider the most general case here but just discuss the necessary result we need for the model derived in this work.

As proven in the appendix under the assumption that we have linearised clocks, that is  $\mathcal{G}^I = \delta\mathcal{G}^I$  with  $\delta^k\mathcal{G}^I = 0$  for all  $k \geq 2$  and we consider observables of quantities  $f = \delta f$  for which  $\delta^k f = 0$  with  $k \geq 2$  then one can show that

$$\left\{ \mathcal{O}_{\delta f, \{\delta T\}}^{(1)} + \mathcal{O}_{\delta f, \{\delta T\}}^{(2)}, \mathcal{O}_{\delta g, \{\delta T\}}^{(1)} + \mathcal{O}_{\delta g, \{\delta T\}}^{(2)} \right\} = \{\delta f, \delta g\}^* + \mathcal{O}_{\{\delta f, \delta g\}^*, \{\delta T\}}^{(1)} + \mathcal{O}(\delta^2, \kappa^2), \quad (2.71)$$

where the Dirac bracket  $\{\delta f, \delta g\}^*$  corresponds to the set  $(\mathcal{G}^I = \delta\mathcal{G}^I, \mathcal{C}'_I = \mathcal{C}'_I^{(1)} + \mathcal{C}'_I^{(2)})$ . Again as above because we do not perturb the matter degrees of freedom we can apply the same result here if we replace  $\delta f$  by  $f$  again with  $f = \phi, \pi$ . As discussed in appendix A.3.2 the case interesting for us is when  $f, g$  are the elementary phase space variables of the linearised theory. Given this and using that for the reference fields we have  $\mathcal{G}^I = \delta\mathcal{G}^I$  as well as  $\{\mathcal{G}^I(x), \mathcal{C}'_J(y)\} = \frac{1}{\kappa} \delta^I_J \delta^{(3)}(x, y) + \mathcal{O}(\delta^2, \kappa^2)$  the explicit form of the Dirac bracket up to linear order reads

$$\begin{aligned} \{\delta f(x), \delta g(y)\}^* &:= \{\delta f(x), \delta g(y)\} - \kappa \int_{\sigma} d^3 z \{\delta f(x), \delta \mathcal{C}'_L(z)\} \delta_M^L \{\delta \mathcal{G}^M(z), \delta g(y)\} \\ &\quad - \kappa \int_{\sigma} d^3 z \{\delta f(x), \delta^2 \mathcal{C}'_L(z)\} \delta_M^L \{\delta \mathcal{G}^M(z), \delta g(y)\} \\ &\quad + \kappa \int_{\sigma} d^3 z \{\delta g(x), \delta \mathcal{C}'_L(z)\} \delta_M^L \{\delta \mathcal{G}^M(z), \delta f(y)\} \\ &\quad + \kappa \int_{\sigma} d^3 z \{\delta g(x), \delta^2 \mathcal{C}'_L(z)\} \delta_M^L \{\delta \mathcal{G}^M(z), \delta f(y)\} + \mathcal{O}(\delta^2, \kappa^2). \end{aligned}$$

Now for  $\delta f$  and  $\delta g$  that have the property that they commute with all  $\delta\mathcal{G}^I$  the Dirac bracket reduces to the ordinary Poisson bracket. If  $\delta f, \delta g$  are further elementary phase space variables that satisfy standard canonical Poisson brackets then their Poisson bracket is phase space independent and as a consequence, the Dirac observable of the Dirac bracket has only a zeroth order contributions that agrees with the original Poisson bracket. As shown in appendix A.3.1 the variables  $(\phi, \pi), (\delta\mathcal{A}_a^i, \delta\mathcal{E}^a_i)$  all commute with all  $\delta T^I$  which is equivalent that they commute with all  $\delta\mathcal{G}^I$ . Thus, for these set of variables the Dirac bracket agrees with the original Poisson bracket. Consequently, we can immediately conclude that the non-vanishing Poisson brackets of the set of Dirac observables  $\phi^{GI}, \pi^{GI}, \delta\mathcal{A}_a^i, \delta\mathcal{E}^a_i$  read

$$\{\phi^{GI}(\vec{x}, t), \pi^{GI}(\vec{y}, t)\} = \delta(\vec{x} - \vec{y}) + \mathcal{O}(\delta^2, \kappa^2), \quad (2.72)$$

$$\{\delta\mathcal{A}_a^i(\vec{x}, t), \delta\mathcal{E}^b_j(\vec{y}, t)\} = \frac{\beta}{\kappa} P_a^{i,b} \delta(\vec{x} - \vec{y}) + \mathcal{O}(\delta^2, \kappa^2). \quad (2.73)$$

All remaining ones vanish up to corrections of order  $\mathcal{O}(\delta^2, \kappa^2)$ . The matter variables satisfy the same Poisson algebra as their gauge variant counter parts and for the geometric gauge invariant degrees of freedom the Poisson bracket involves on the right hand side as well the expected projector on the transverse and traceless part. Considering this together with the Poisson algebra of the  $\mathcal{G}^I$  and  $\mathcal{C}'_I$  in (2.59) we realise we can perform a canonical transformation from the set of variables

$$(\phi, \pi), \quad (\delta A_a^i, \delta E^a_i)$$

to the new set of variables

$$(\delta \mathcal{G}, C'), \quad (\delta \mathcal{G}^a, C'_a), \quad (\delta \mathcal{G}^j, G'_j), \quad (\phi^{GI}, \pi^{GI}), \quad (\delta \mathcal{A}_a^i, \delta \mathcal{E}^a_i)$$

with

$$\delta \mathcal{G} = \delta T - t, \quad \delta \mathcal{G}^a = \delta T^a - x^a_\sigma, \quad \delta \mathcal{G}^j = \delta \Xi^j. \tag{2.74}$$

As discussed in the seminal work in [74], one can reformulate the corresponding forms of the geometric clocks so that they no longer explicitly involve the coordinates. This is possible here because the way the geometric clocks are defined is non-local due to the Green's function involved in the position space representation. The equivalent local conditions are obtained by applying the Laplacian, which has the consequence that neither  $t$  nor  $x^a$  are explicitly involved any more. It is also these local quantities that are involved in the perturbed quantities, and therefore the use of such geometric clocks does not contradict the requirement that the perturbations be small. The reason is that these coordinate conditions only set certain components of the metric perturbations and their conjugate moments to zero, see also the more detailed discussion in [74] and references therein.

Because the definition of  $C'_I$  as can be seen from (2.60) involves second-order contributions to implement this canonical transformation on the linearised phase space we restrict  $C'_I$  to its linear part and using that in our case  $\delta C'_I = \delta C_I$  the new set of canonical variables on the linearised phase space is given by

$$(\delta \mathcal{G}, \delta C), \quad (\delta \mathcal{G}^a, \delta C_a), \quad (\delta \mathcal{G}^j, \delta G_j), \quad (\phi^{GI}, \pi^{GI}), \quad (\delta \mathcal{A}_a^i, \delta \mathcal{E}^a_i),$$

the explicit transformation can be found in appendix A.4. An advantage of this new set of elementary phase space variables is that it allows to clearly separate the physical degrees of freedom from the remaining gauge degrees of freedom at the level of the full phase space. The physical phase space corresponds to the subspace involving the subalgebra of all Dirac observables  $(\phi^{GI}, \pi^{GI}), (\delta \mathcal{A}_a^i, \delta \mathcal{E}^a_i)$  including the expected six physical degrees of freedom. The price to pay in this context is that, as far as the position space is considered, we need to work with highly non-local clocks that become local in momentum space. This is not unexpected here considering the fact that the projector to the transverse-traceless part of the gravitational perturbations is local in momentum but non-local in position space<sup>4</sup>.

Finally, let us briefly sketch the strategy we followed here that can also be applied to a different choice of reference fields. Suppose we are interested in perturbation theory up to linear order.

- First choose a set of linearised reference fields, that is  $\delta^2 \mathcal{G}^I = 0$  such that they build canonical pairs with the linearised constraints, that is  $\{\delta \tilde{\mathcal{G}}^I(x), \delta C_J(y)\} = \delta^I_J \delta^{(3)}(x, y)$ .
- In case not all reference fields and hence  $\delta \tilde{\mathcal{G}}^I(x)$  mutually commute then apply the dual observable map with the linearised constraints as the clocks up to second order<sup>5</sup> to those that do not to obtain  $\mathcal{G}^I$ .
- Next apply the dual observable map up to second order neglecting the linear order to all constraints  $C_I$  to get the constraints  $C'_I$  that are abelian and canonically conjugate to the reference fields  $\mathcal{G}^I$  up to corrections of second order.

<sup>4</sup> In [103] it is discussed that there are situations particularly in the case of sources where a replacement of the non-local projector in position space by the also widely used local projector can yield different physical effects.

<sup>5</sup> Note for this choice of clocks and since we apply it to the linearised clocks only the linear terms in the power series will contribute.

- Define the observable map by means of the constraints  $C'_I$  using as reference fields  $\mathcal{G}^I$ .
- Choose next to  $(\delta\mathcal{G}^I, \delta C'_I)$  further independent phase space variables on the linearised phase space denoted by  $\delta q^I, \delta p_I$ .
- Apply the observable map to  $(\delta q^I, \delta p_I)$  to obtain the physical gauge invariant degrees of freedom.
- Compute the algebra of the Dirac observables of  $(\delta q^I, \delta p_I)$  using (2.71).
- If we further apply the dual observable map to these Dirac observables and they still satisfy the same Poisson algebra as their gauge variant counterparts we can also say that given the set of variables  $(\delta\mathcal{G}^I, \delta C'_I, \delta q^I, \delta p_I)$  then applying the observable map and its dual to the entire set with the modification that we exclude the linear order in the observable map and its dual for the set  $(\delta\mathcal{G}^I, \delta C'_I)$  defines a canonical transformation on the entire phase space.

In our case the final application of the dual observable map on the variables  $(\delta q^I, \delta p_I)$  corresponding to  $\phi^{\text{GI}}, \pi^{\text{GI}}, \delta A_a^i, \delta E_i^a$  acts trivially since the original gauge variant quantities already commute with our chosen geometrical clocks. We expect that one can extend this strategy without any problems to higher orders in perturbation theory and for the aspect of constructing abelian constraints this has been done in [69]. How and if also the further steps can be generalised to higher order goes beyond the scope of this work and will be discussed more in detail elsewhere.

The work in [35] also aims at constructing Dirac observables in the framework of the ADM formalism. In appendix A.3.4 we compare their construction to the one presented here. To our understanding the explicit form of the observable for the scalar field momentum presented in [35] is not invariant under transformations generated by the linearised spatial diffeomorphism constraint.

We will close this subsection with discussing the dynamics of our constructed Dirac observables. This is the last ingredient missing in order to take the gauge invariant formulation as the starting point for the quantisation in the next section. As we know from full general relativity, the dynamics of Dirac observables cannot be generated by the canonical Hamiltonian because by construction, all Dirac observables commute with the constraints and  $\mathbf{H}_{\text{can}}$  is just a linear combination of the smeared constraints. The generator of the dynamics of the Dirac observables is the so-called physical Hamiltonian [101]. If we rewrite the Hamiltonian constraint linearly in the clock momentum of the temporal reference field, that is  $C = P_T + h$ , then the physical Hamiltonian corresponds to the Dirac observable associated with the phase space function  $h$ . Here we need to adapt this to the framework of perturbations theory and consider the fact that in addition we have an interaction Hamiltonian as well as a non-vanishing Hamiltonian in  $\kappa^0$  order. The perturbed Hamiltonian constraint up to second order is given by  $C = \delta C + \delta^2 C^{\text{geo}} + O(\delta^2, \kappa^2)$ . The reference field  $\delta T$  is canonically conjugate to the linearised Hamiltonian constraint  $\delta C$  and thus we can identify  $\delta C$  with the momentum variable conjugate to  $\delta T$  that we denote by  $\delta P_T$ , then we have  $C = \delta P_T + \delta^2 C^{\text{geo}} + O(\delta^2, \kappa^2)$ .<sup>6</sup> The Dirac observable associated with  $\delta^2 C^{\text{geo}}$  corresponds to that part of the physical Hamiltonian that is generating the dynamics of the purely gravitational physical degrees of freedom. The Dirac observable corresponding to the interaction Hamiltonian  $\mathcal{H}_I = \frac{\kappa}{2} \delta h_{ab} T^{ab}$  with  $\delta h_{ab} = -\delta_b^i \delta_{ac} \delta E^c_i - \delta_a^i \delta_{bc} \delta E^c_i + \delta_{ab} \delta_c^i \delta E^c_i$  encodes the interaction between the matter and gravitational physical degrees of freedom. As we will see below when expressed in terms of the independent Dirac observables it decomposes into a term where the physical triad variables

<sup>6</sup> Note that the boundary term  $T[N]$  in  $\delta H_{\text{can}}$  cancels exactly the contribution  $\bar{N} \delta C = \delta C$  in  $\delta H_{\text{can}}$  but since the clock momentum  $\delta P_T = \delta C$  is anyway not part of the physical Hamiltonian this causes no problems.

$\delta\mathcal{E}^a_i$  couple to the physical matter variables as well as a self-interaction term that involves a coupling between the physical matter variables only. Finally, in  $\kappa^0$ -order the physical Hamiltonian involves the matter Hamiltonian for the scalar field on Minkowski spacetime. Summarising the above discussion we denote the physical Hamiltonian of the model considered in this work by  $\delta\mathbf{H}$  that takes the following form

$$\delta\mathbf{H} = \mathcal{H}_\phi + O_{\mathcal{H}_{\text{geo}},\{T\}} + O_{\mathcal{H}_I,\{T\}} \quad \text{with} \quad \mathcal{H}_{\text{geo}} := \int_{\sigma} d^3x \delta^2 C^{\text{geo}}(x), \quad \mathcal{H}_\phi := \int_{\sigma} d^3x \epsilon(\vec{x}, t). \quad (2.75)$$

As can be seen in (2.76) it can be written entirely in terms of the Dirac observables  $(\phi^{GI}, \pi^{GI})$ ,  $(\delta\mathcal{A}_a^i, \delta\mathcal{E}^a_i)$  and therefore is by construction gauge invariant up to corrections of order  $O(\delta^2, \kappa^2)$ . Here, as discussed above, we insert the Dirac observables in all terms being linearly in  $\kappa$  but keep  $\phi, \pi$  in the  $\kappa^0$  contribution because at zeroth order  $\phi, \pi$  are suitable observables. Because in second-order perturbation theory the constraints  $\mathcal{C}'_I$  are only weakly equivalent to the original constraints  $\mathcal{C}_I$  and the observable of the interaction Hamiltonian  $\mathcal{H}_I$  is linearly in the perturbations we need to consider the Dirac observable of  $\delta\mathcal{E}^a_i$  up to second order in  $\mathcal{H}_I$  when computing Poisson brackets with the constraints. Hence, as far as the set of original constraints  $\mathcal{C}_I$  are considered  $\delta\mathbf{H}$  is a weak Dirac observable.

Explicitly we obtain, see appendix A.5:

$$\begin{aligned} \delta\mathbf{H} = & \int_{\mathbb{R}^3} d^3x \epsilon(\vec{x}, t) + \kappa \int_{\mathbb{R}^3} d^3k \frac{1}{2\beta^2} \sum_{r \in \{\pm\}} \left( \delta\mathcal{A}^r(\vec{k}, t) \delta\mathcal{A}^r(-\vec{k}, t) + 2r \|\vec{k}\| \delta\mathcal{E}^r(\vec{k}, t) \delta\mathcal{A}^r(-\vec{k}, t) \right. \\ & \left. + (\beta^2 + 1) \|\vec{k}\|^2 \delta\mathcal{E}^r(\vec{k}, t) \delta\mathcal{E}^r(-\vec{k}, t) \right) + \kappa \int_{\mathbb{R}^3} d^3k \delta_b^i \varphi_a^b(\vec{k}, t) \delta\mathcal{E}^a_i(-\vec{k}, t) \\ & + \kappa \int d^3k ik_b \kappa \delta\sigma^{ia} \left[ \varphi_a^b - \delta_a^b \left( \varphi_c^c + m^2 \tilde{V} - \epsilon \right) \right] + \kappa \int d^3k \left[ i\kappa \|\vec{k}\| \delta\tau \hat{k}^b p'_b \right] \\ & - \kappa \int_{\mathbb{R}^3} \frac{d^3k}{\|\vec{k}\|^2} \left\{ - p_c(\vec{k}, t) p_d(-\vec{k}, t) \left[ \delta^{cd} - \frac{1}{4} \hat{k}^c \hat{k}^d \right] + \epsilon(\vec{k}, t) \epsilon(-\vec{k}, t) \left[ \frac{1}{4} + \frac{3}{2} \right] \right. \\ & \left. - \frac{3}{2} m^2 \tilde{V}(\vec{k}, t) \epsilon(-\vec{k}, t) - \varphi_a^a(\vec{k}, t) \epsilon(-\vec{k}, t) \right\}, \quad (2.76) \end{aligned}$$

with

$$\epsilon(\vec{x}, t) := \epsilon(\phi^{GI}(\vec{x}, t), \pi^{GI}(\vec{x}, t)), \quad p_a(\vec{x}, t) := p_a(\phi^{GI}(\vec{x}, t), \pi^{GI}(\vec{x}, t)), \quad (2.77)$$

$$\tilde{V}(\vec{x}, t) := \phi^{GI}(\vec{x}, t)^2, \quad \varphi_a^b(\vec{x}, t) := (\partial_a \phi^{GI}(\vec{x}, t)) (\partial^b \phi^{GI}(\vec{x}, t)), \quad (2.78)$$

where as before we considered terms up to linear order in  $\kappa$  only. Note that an underbow denotes the spatial Fourier transform of the corresponding quantity, i.e.

$$\underset{\sim}{f}(\vec{k}, t) := \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{\mathbb{R}^3} d^3x e^{-i\vec{k}\vec{x}} f(\vec{x}, t), \quad \hat{k}^a := \frac{1}{\|\vec{k}\|} k^a.$$

Note that for the  $\kappa^0$  order we also use the operator valued distributions of the physical matter variables because the difference to the gauge variant variables is of higher order than  $\kappa^0$  and is neglected to obtain a model that is consistent with the Post-Minkowski approximation order by order in  $\kappa$ . As can be seen from the explicit form of the physical Hamiltonian in (2.76) for our choice of  $\delta\tau$  and  $\delta\sigma^c$  in (2.67),  $\delta\tau$  being linear in  $t$  does not contribute to the physical Hamiltonian and  $\delta\sigma^c$  is also absent if we chose it to be constant and enters via Kronecker delta

in case we choose it to be linearly in  $x^c$  in exact agreement to what was found in [74] for the ADM variables. This additional Kronecker delta contribution for the second choice combines with terms including the scalar field and its derivatives respectively and has thus a suitable fall-off behaviour.

The different contributions of the Hamiltonian in the individual lines can be interpreted as follows: The first line corresponds to the Hamiltonian of a scalar field on a Minkowski background. The next two lines encode the Hamiltonian of a gravitational field in vacuum, the third integral denotes the interaction between the gravitational degrees of freedom and the matter field and finally the remaining contributions encode the gravitational self-interaction of the scalar field. Even though the latter does not contain any gravitational degrees of freedom, it only appears due to the coupling between matter and gravity and results from the coupling of the gauge degrees of freedom in the gravitational sector to the scalar field. Once these gauge degrees of freedom are expressed in terms of the independent physical degrees of freedom we end up with this result. The latter contribution was neglected in [69] by using the argument that only the transverse-traceless components need to be considered. While in the vacuum case one can use a gauge fixing that sets all but the transverse-traceless gravitational degrees of freedom to zero, this is no longer a valid choice for a gauge-fixing if we couple the scalar field, because the constraints involve additional contributions from the scalar field and suitable gauge fixing will result in such a contribution of gravitational self-interaction of the scalar field. Hence, to our understanding, it is not justified to neglect this contributions in the order of perturbation theory considered in [69]. This finalises the discussion on the classical model that will be the starting point for the Fock quantisation in the next section. Having identified the physical degrees of freedom in the model, now we are in the situation to separate the total system into a system and an environmental part, as usually done in the framework of decoherence models. We choose the following separation:

$$\text{system: } \phi^{GI}, \pi^{GI} \quad \text{environment: } \delta\mathcal{A}_a^i, \delta\mathcal{E}_i^a$$

and hence the gauge invariant matter sector becomes the system and the physical degrees freedom in the gravitational sector provide the environment.

Finally let us comment on the relation of the fully gauge invariant framework to the gauge fixing discussed in the appendix B.1. As discussed for instance in [104] for certain choices of the  $\mathcal{G}^I$  one can relate a model formulated in terms of fully gauge invariant quantities to a corresponding gauge-fixed model. The observables constructed here also fall into this class of models, which means that the gauge-fixed Hamiltonian and the gauge invariant physical Hamiltonian formally agree if we replace all gauge invariant quantities by their gauge-fixed counterparts. Practically, this can be achieved by setting  $\mathcal{G} = \mathcal{G}^a = \mathcal{G}^j = 0$ , that is strongly equal to zero and then we can formally identify the Dirac observables with the gauge-fixed quantities involved in the gauge fixing discussed in the appendix B.1. For this specific gauge fixing, the constructed observables have a standard interpretation. The Hamiltonian of the gauge fixed theory can be obtained by inserting all constraints, gauge fixing conditions as well as the Lagrange multipliers that ensure the stability of the gauge fixing into  $\delta\mathbf{H}_{\text{can}}$ . As can be seen from the explicit form of  $\delta\mathbf{H}_{\text{can}}$  in (A.106) the Hamiltonian of the gauge fixed theory agrees formally with  $\delta\mathbf{H}$  under the identification of the Dirac observables with their corresponding gauge-fixed quantities. In order to compare our results with the existing literature we discuss in appendix B.2 the geometrical clocks used in [69] that are denoted as ADM clocks there because they were introduced in the seminal paper [74]. Also this set of geometrical clocks encodes the physical gravitational degrees of freedom in the symmetric transverse-traceless linearised connection and triad variables. The main difference we see among the two

sets of geometrical clocks is how the reference field associated with the linearised Gauß constraint is chosen. In our work we choose a Lorentz-like condition if we set the corresponding gauge fixing condition  $\delta\mathcal{G}^j = 0$ , whereas in [69] the resulting  $\delta\mathcal{G}^j = 0$  is only equivalent to a Lorentz-like condition in the vacuum case but not if we couple a scalar field. Then the resulting reference fields for the linearised spatial diffeomorphism and Hamiltonian constraint also differ because for both sets of clocks we require that the all geometrical clocks mutually commute. The physical Hamiltonian for the ADM clocks is also presented in (B.2) and it turns out that compared to the physical Hamiltonian in (2.76) it only differs in the term describing the self-interaction of the scalar field. There we find that it differs only by numerical factors in all but one term and it involves one additional term where in our case only a trace enters, but for the ADM clocks an additional contraction with the Fourier basis is present, see (B.18) in appendix B.2.

However, working at the gauge invariant level and not choosing one specific gauge fixing allows us also to choose a different gauge fixing than the one discussed in the appendix B.1 and then the observables and their dynamics discussed here are still valid because everything was formulated in a gauge invariant manner. The only difference for other gauge fixing choices is that the interpretation of the gauge invariant Dirac observables and their relation to the gauge-fixed quantities is modified because then, the gauge fixing does not necessarily correspond to setting all  $\mathcal{G}^I$  equal to zero. Hence in this sense, if we consider a class of models where a relation to one gauge fixing is possible, then there is always a convenient choice for the reference fields in order to make such a relation as simple as possible. This was exactly our motivation for choosing the reference fields for this model the way we did, apart from the additional requirements (i) to (iii) listed below (2.44).

### 3. Fock quantisation of the model

In this section we present the results of a Fock quantisation of the model under consideration. We work with units where  $\hbar = c = 1$ . For convenience we choose to quantise the following scalar and tensor fields:

$$\phi^{GI}(\vec{x}, t) \quad \pi^{GI}(\vec{x}, t) \quad \delta\mathcal{E}^a_i(\vec{x}, t) \quad \delta\mathcal{C}_a^i(\vec{x}, t), \quad (3.1)$$

where the latter is the Fourier transform of  $\delta\mathcal{C}_a^i(\vec{k}, t) = \delta\mathcal{C}^+(\vec{k}, t)m_a(\vec{k})m^i(\vec{k}) + \delta\mathcal{C}^-(\vec{k}, t)\bar{m}_a(\vec{k})\bar{m}^i(\vec{k})$  with  $\delta\mathcal{C}^\pm(\vec{k}, t) = -\frac{1}{\beta} \left( \delta\mathcal{A}^\pm(\vec{k}, t) \pm \|\vec{k}\| \delta\mathcal{E}^\pm(\vec{k}, t) \right)$ ; see appendix A.2 for the definition of the basis  $(\hat{k}_a, m_a(\vec{k}), \bar{m}_a(\vec{k}))$  in Fourier space. The reason for this choice is that by using  $\delta\mathcal{C}$  instead of  $\delta\mathcal{A}$  the terms in the linearised Hamiltonian containing gravitational degrees of freedom only in (2.76) can be rewritten in the following way:

$$\begin{aligned} & \kappa \int_{\mathbb{R}^3} d^3x \frac{1}{2} \left[ \delta^{ab} \delta_{ij} \delta\mathcal{C}_a^i(\vec{x}, t) \delta\mathcal{C}_b^j(\vec{x}, t) + \delta_{bc} \delta^{ij} (\partial_a \delta\mathcal{E}^b_i(\vec{x}, t)) (\partial^a \delta\mathcal{E}^c_j(\vec{x}, t)) \right] \\ & = \kappa \int_{\mathbb{R}^3} d^3x \sum_{r \in \{+, -\}} \frac{1}{2} \left[ \delta\mathcal{C}^r(\vec{x}, t) \delta\mathcal{C}^r(\vec{x}, t) + (\partial_a \delta\mathcal{E}^r(\vec{x}, t)) (\partial^a \delta\mathcal{E}^r(\vec{x}, t)) \right], \quad (3.2) \end{aligned}$$

which has the same form as the energy density of two massless scalar fields  $\delta E^\pm(\vec{x}, t)$  since

$$\{\delta E^\pm(\vec{x}, t), \delta\mathcal{C}^\pm(\vec{y}, t)\} = -\frac{1}{\beta} \{\delta E^\pm(\vec{x}, t), \delta\mathcal{A}^\pm(\vec{y}, t)\} = \frac{1}{\kappa} \delta^3(\vec{x} - \vec{y}). \quad (3.3)$$

A mode expansion of the fields and the Fock quantisation yield the following operator-valued distributions for the physical degrees of freedom:

$$\phi^{GI}(\vec{x}, t) = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}} \frac{1}{\sqrt{2\omega_k}} \left[ a_k e^{-i\omega_k t + i\vec{k}\vec{x}} + a_k^\dagger e^{i\omega_k t - i\vec{k}\vec{x}} \right] \quad (3.4)$$

$$\pi^{GI}(\vec{x}, t) = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}} (-i) \sqrt{\frac{\omega_k}{2}} \left[ a_k e^{-i\omega_k t + i\vec{k}\vec{x}} - a_k^\dagger e^{i\omega_k t - i\vec{k}\vec{x}} \right] \quad (3.5)$$

$$\delta\mathcal{E}^a_i(\vec{x}, t) = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}} \sqrt{\frac{1}{2\kappa\Omega_k}} \sum_{r \in \{\pm\}} \left[ [P^r(-\vec{k})]_i^a b_k^r e^{-i\Omega_k t + i\vec{k}\vec{x}} + [P^r(\vec{k})]_i^a (b_k^r)^\dagger e^{i\Omega_k t - i\vec{k}\vec{x}} \right] \quad (3.6)$$

$$\delta\mathcal{C}_a^i(\vec{x}, t) = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}} (-i) \sqrt{\frac{\Omega_k}{2\kappa}} \sum_{r \in \{\pm\}} \left[ [P^r(-\vec{k})]_a^i b_k^\pm e^{-i\Omega_k t + i\vec{k}\vec{x}} - [P^r(\vec{k})]_a^i (b_k^\pm)^\dagger e^{i\Omega_k t - i\vec{k}\vec{x}} \right], \quad (3.7)$$

where the  $a_k^{(\dagger)}$  denote annihilation (creation) operator-valued distributions for scalar particles while the  $(b_k^\pm)^{(\dagger)}$  denote annihilation (creation) operator-valued distributions for gravitons with polarisation label  $\pm$ . We have also introduced  $\omega_k := \sqrt{\vec{k}^2 + m^2}$  and  $\Omega_k := \sqrt{\vec{k}^2}$  as well as the transverse-traceless projectors  $[P^r(\vec{k})]_a^i$  defined in (B.14), which play the role of the polarisation tensors of the quantised fields. From their definitions follows that  $[P^r(\vec{k})]_i^a = [P^r(-\vec{k})]_i^a$  and hence  $\delta\mathcal{E}^a_i(\vec{x}, t)^\dagger = \delta\mathcal{E}^a_i(\vec{x}, t)$  and  $\delta\mathcal{C}_a^i(\vec{x}, t)^\dagger = \delta\mathcal{C}_a^i(\vec{x}, t)$ . Note that in contrast to [35] the polarisation tensors are different for the positive and negative frequency modes. The creation and annihilation operator-valued distributions satisfy the following commutation relations:

$$[a_k, a_l^\dagger] = \delta^3(\vec{k} - \vec{l}) \quad [a_k, a_l] = [a_k^\dagger, a_l^\dagger] = 0 \quad (3.8)$$

$$[(b_k^\pm), (b_l^\pm)^\dagger] = \delta^3(\vec{k} - \vec{l}) \quad [(b_k^\pm), b_l^\pm] = [(b_k^\pm)^\dagger, (b_l^\pm)^\dagger] = 0, \quad (3.9)$$

here we omit the vector arrow on the mode labels in all index positions in order to keep our notation more compact. The total Fock space consists of a tensor product of three individual bosonic Fock spaces, one for the scalar particles, one for the + polarised gravitons and one for the - polarised ones. As usual the annihilation and creation operators for different fields or polarisations mutually commute. The Hamiltonian operator corresponding to (2.76) can then be implemented in the Schrödinger picture as

$$H = \int_{\mathbb{R}^3} d^3k \left\{ \omega_k a_k^\dagger a_k + \Omega_k [(b_k^+)^\dagger b_k^+ + (b_k^-)^\dagger b_k^-] \right\} + \sqrt{\frac{\kappa}{2}} \int d^3k \frac{1}{\sqrt{\Omega_k}} \sum_{r \in \{\pm\}} \left[ b_k^r J_r^\dagger(\vec{k}) + (b_k^r)^\dagger J_r(\vec{k}) \right] + \kappa U \otimes \mathbb{1}_\mathcal{E}, \quad (3.10)$$

where we have introduced some kind of normal-ordered current operator for the scalar field that couples to the gravitational environment and which is quadratic in the scalar field and its derivatives:

$$J_r(\vec{k}) := \hat{\varphi}_a^b(\vec{k}, 0) [P^r(\vec{k})]_b^a = \int \frac{d^3p}{(2\pi)^{\frac{3}{2}}} \frac{1}{2\sqrt{\omega_p\omega_{k+p}}} \left[ p_a p^b [P^r(\vec{k})]_b^a \right] \left( 2a_p^\dagger a_{k+p} + a_{-p} a_{k+p} + a_p^\dagger a_{-p-k}^\dagger \right). \quad (3.11)$$

Furthermore,  $U$  denotes a self-interaction operator that is present only due to coupling of the scalar field to linearised gravity, as its contribution to the Hamiltonian operator will vanish if the coupling constant  $\kappa$  is set to zero. It involves fourth powers of the annihilation and creation operators of the scalar field and its momentum and its contributions can be understood as additional self-interaction vertices of the scalar field that are not present in the corresponding free theory. We chose to implement this operator in a completely normal ordered form, that is  $:U:$ , where  $:\cdot:$  denotes normal ordering. This is in contrast to the quantisation procedure of similar systems in [35, 37, 38], where either no specific ordering is mentioned [37], or the individual operators corresponding to the operators  $\epsilon, p_a, \tilde{V}, \varphi$  and  $J_r$  introduced above are normal ordered [35, 38], but no normal ordering is applied to the entire self-interaction operator  $U$ . For the normal ordered self-interaction operator we obtain

$$\begin{aligned}
U := & \int d^3k \frac{ik_b}{2} \kappa \delta\sigma^a(-\vec{k}, 0) \left[ \varphi_a^b(\vec{k}, 0) - \delta_a^b \left( \varphi_c^c(\vec{k}, 0) + m^2 \tilde{V}(\vec{k}, 0) - \epsilon(\vec{k}, 0) \right) \right] \\
& - \int d^3k \left[ 2i\kappa |\vec{k}| |\delta\tau(-\vec{k}, 0) \hat{k}^b p_b(\vec{k}, 0) \right] \\
& - \int_{\mathbb{R}^3} \frac{d^3k}{|\vec{k}|^2} \left\{ - : \hat{p}_c(\vec{k}, 0) \hat{p}_a(-\vec{k}, 0) : \left[ \delta^{cd} - \frac{1}{4} \hat{k}^c \hat{k}^d \right] + : \hat{\epsilon}(\vec{k}, 0) \hat{\epsilon}(-\vec{k}, 0) : \left[ \frac{1}{4} + \frac{3}{2} \right] \right. \\
& \left. - \frac{3}{2} m^2 : \hat{V}(\vec{k}, 0) \hat{\epsilon}(-\vec{k}, 0) : - : \hat{\varphi}_a^a(\vec{k}, 0) \hat{\epsilon}(-\vec{k}, 0) : \right\}. \tag{3.12}
\end{aligned}$$

We realise that the parameters  $\delta\tau$  and  $\delta\sigma^c$  enter into the self-interaction operator  $U$  of the scalar field only. For our choice of  $\delta\tau$  in (2.67) where  $\delta\tau$  is linear in  $t$  the term including  $\delta\tau$  drops out completely. The same happens for all contributions involving  $\delta\sigma^c$  if we choose it to be constant. In case we choose it linearly in  $x^c$  then we have  $\delta\sigma_{,a}^c = \delta_c^a$  and hence the corresponding contributions in Fourier space will not vanish but  $x^c$  will not enter explicitly into the self-interaction operator  $U$ .

Here we introduced several new operators involved in the physical Hamiltonian operator of the model. First of all a (normal-ordered) operator corresponding to the scalar field's momentum density:

$$\begin{aligned}
\hat{p}_a(\vec{k}, t) := & \frac{1}{2} \int_{\mathbb{R}^3} \frac{d^3q}{(2\pi)^{\frac{3}{2}}} q_a \sqrt{\frac{\omega_{k-q}}{\omega_q}} \left[ a_q a_{k-q} e^{-it(\omega_q + \omega_{k-q})} + a_{-q}^\dagger a_{k-q} e^{-it(\omega_{k-q} - \omega_q)} \right. \\
& \left. - a_{q-k}^\dagger a_q e^{it(\omega_{k-q} - \omega_q)} - a_{-q}^\dagger a_{q-k}^\dagger e^{it(\omega_q + \omega_{k-q})} \right]. \tag{3.13}
\end{aligned}$$

Additionally, an operator corresponding to the scalar field's energy density:

$$\begin{aligned}
\hat{\epsilon}(\vec{k}, t) := & \frac{1}{4} \int_{\mathbb{R}^3} \frac{d^3q}{(2\pi)^{\frac{3}{2}}} \left\{ - \sqrt{\omega_q \omega_{k-q}} \left[ a_q a_{k-q} e^{-it(\omega_q + \omega_{k-q})} - a_{-q}^\dagger a_{k-q} e^{-it(\omega_{k-q} - \omega_q)} \right] \right. \\
& \left. - a_{q-k}^\dagger a_q e^{it(\omega_{k-q} - \omega_q)} + a_{-q}^\dagger a_{q-k}^\dagger e^{it(\omega_q + \omega_{k-q})} \right] \\
& - \frac{q_a(k^a - q^a) - m^2}{\sqrt{\omega_q \omega_{k-q}}} \left[ a_q a_{k-q} e^{-it(\omega_q + \omega_{k-q})} + a_{-q}^\dagger a_{k-q} e^{-it(\omega_{k-q} - \omega_q)} \right. \\
& \left. + a_{q-k}^\dagger a_q e^{it(\omega_{k-q} - \omega_q)} + a_{-q}^\dagger a_{q-k}^\dagger e^{it(\omega_q + \omega_{k-q})} \right] \left. \right\}. \tag{3.14}
\end{aligned}$$

Finally, two more operators that correspond to certain different terms of the scalar field's energy momentum tensor that appear in the classical Hamiltonian, namely to  $\phi^2$  and  $\partial_a \phi \partial^b \phi$ :

$$\begin{aligned} \widehat{V}(\vec{k}, t) := & \frac{1}{2} \int_{\mathbb{R}^3} \frac{d^3 q}{(2\pi)^{\frac{3}{2}}} \frac{1}{\sqrt{\omega_q \omega_{k-q}}} \left[ a_q a_{k-q} e^{-it(\omega_q + \omega_{k-q})} + a_{-q}^\dagger a_{k-q} e^{-it(\omega_{k-q} - \omega_q)} \right. \\ & \left. + a_{q-k}^\dagger a_q e^{it(\omega_{k-q} - \omega_q)} + a_{-q}^\dagger a_{q-k}^\dagger e^{it(\omega_q + \omega_{k-q})} \right] \end{aligned} \quad (3.15)$$

$$\begin{aligned} \widehat{\phi}_a^b(\vec{k}, t) := & -\frac{1}{2} \int_{\mathbb{R}^3} \frac{d^3 q}{(2\pi)^{\frac{3}{2}}} \frac{q_a (k-q)^b}{\sqrt{\omega_q \omega_{k-q}}} \left[ a_q a_{k-q} e^{-it(\omega_q + \omega_{k-q})} + a_{-q}^\dagger a_{k-q} e^{-it(\omega_{k-q} - \omega_q)} \right. \\ & \left. + a_{q-k}^\dagger a_q e^{it(\omega_{k-q} - \omega_q)} + a_{-q}^\dagger a_{q-k}^\dagger e^{it(\omega_q + \omega_{k-q})} \right]. \end{aligned} \quad (3.16)$$

Note that all these constituents are symmetric in the sense that  $\underline{\epsilon}^\dagger(-\vec{k}, t) = \underline{\epsilon}(\vec{k}, t)$  which implies  $\epsilon^\dagger(\vec{x}, t) = \epsilon(\vec{x}, t)$ , and hence also  $U^\dagger = U$ . In the interaction picture, which we denote by a tilde, the Hamiltonian operator that involves the gravity-matter interaction is given by

$$\begin{aligned} \widetilde{H}(t) = \sqrt{\kappa} \underbrace{\frac{1}{\sqrt{2}} \int d^3 k \frac{1}{\sqrt{\Omega_k}} \sum_r \left[ b_k^r e^{-i\Omega_k t} J_r^\dagger(\vec{k}, t) + (b_k^r)^\dagger e^{i\Omega_k t} J_r(\vec{k}, t) \right]}_{=: \widetilde{H}_{\mathcal{T}}(t)} + \kappa \widetilde{U}(t) \otimes \mathbb{1}_{\mathcal{E}} \end{aligned} \quad (3.17)$$

with an appropriate current operator obtained directly from the one in the Schrödinger picture by using the time-dependent constituents,

$$\begin{aligned} J_r(\vec{k}, t) := & \widehat{\varphi}_a^b(\vec{k}, t) [P^r(\vec{k})]_b^a \\ = & \int \frac{d^3 p}{(2\pi)^{\frac{3}{2}}} \frac{1}{2\sqrt{\omega_p \omega_{k+p}}} \left[ p_a p^b [P^r(\vec{k})]_b^a \left( a_p^\dagger a_{k+p} e^{it(\omega_p - \omega_{k+p})} + a_{-p} a_{k+p} e^{-it(\omega_p + \omega_{k+p})} \right) \right. \\ & \left. + a_{-p-k}^\dagger a_{-p} e^{it(\omega_{k+p} - \omega_p)} + a_p^\dagger a_{-p-k}^\dagger e^{it(\omega_p + \omega_{k+p})} \right], \end{aligned} \quad (3.18)$$

and the total normal-ordered self-interaction operator

$$\begin{aligned} \widetilde{U}(t) := & \int d^3 k \frac{ik_b}{2} \kappa \delta \sigma^a(-\vec{k}, t) \left[ \varphi_a^b(\vec{k}, t) - \delta_a^b \left( \varphi_c^c(\vec{k}, t) + m^2 \widehat{V}(\vec{k}, t) - \underline{\epsilon}(\vec{k}, t) \right) \right] \\ & - \int d^3 k \left[ 2i\kappa \|\vec{k}\| \delta \tau(-\vec{k}, t) \hat{k}^b \underline{p}_b(\vec{k}, t) \right] \\ & - \int_{\mathbb{R}^3} \frac{d^3 k}{\|\vec{k}\|^2} \left\{ - : \hat{p}_c(\vec{k}, t) \hat{p}_d(-\vec{k}, t) : \left[ \delta^{cd} - \frac{1}{4} \hat{k}^c \hat{k}^d \right] + : \hat{\epsilon}(\vec{k}, t) \hat{\epsilon}(-\vec{k}, t) : \left[ \frac{1}{4} + \frac{3}{2} \right] \right. \\ & \left. - \frac{3}{2} m^2 : \widehat{V}(\vec{k}, t) \hat{\epsilon}(-\vec{k}, t) : - : \widehat{\varphi}_a^a(\vec{k}, t) \hat{\epsilon}(-\vec{k}, t) : \right\}. \end{aligned} \quad (3.19)$$

With this, the classical model has been carried over to the quantum field theoretic framework. In the next section, we will derive the so called master equation that governs the effective time evolution of the matter system without the need of tracking all the details of the gravitational degrees of freedom.

#### 4. Deriving the master equation for the gravitationally induced decoherence model

In this section we derive the master equation for the gravitationally induced decoherence model considered in this article from first principle. For this purpose we need to trace out the environmental and thus the gravitational degrees of freedom. Before we apply this to the model in section 4.3 we briefly review two methods to derive master equations in the context of decoherence models. The first method, namely the projection operator technique, will be discussed in section 4.1 and a second method based on an influence phase functional is presented in section 4.2.

##### 4.1. Review of the projection operator technique

The starting point in this top-down approach of the derivation of every quantum master equation is the microscopic dynamics in the form of the Liouville-von Neumann equation, regardless of the exact approach that is taken thereafter. With an interaction Hamiltonian operator  $\tilde{H}(t)$  and the density operator of the total system in the interaction picture given by  $\tilde{\rho}(t)$ , where we denote operators in the interaction picture with a tilde, the evolution equation initially amounts to:

$$\frac{\partial}{\partial t} \tilde{\rho}(t) = -i\alpha[\tilde{H}(t), \tilde{\rho}(t)] =: \alpha\mathcal{L}(t)\tilde{\rho}(t), \quad (4.1)$$

where  $\alpha$  denotes the dimensionless<sup>7</sup> coupling strength between the system and the environment and  $\mathcal{L}(t)$  is the so-called Liouville superoperator. The goal is to eliminate the explicit evolution of the many environmental degrees of freedom in order to arrive at an effective equation for the relevant system's degrees of freedom, generally of the following form:

$$\frac{\partial}{\partial t} \tilde{\rho}_S(t) = -i\alpha[\tilde{H}_{LS}(t), \tilde{\rho}_S(t)] + \tilde{\mathcal{D}}(\tilde{\rho}_S(t)) \quad (\text{interaction picture}) \quad (4.2)$$

$$\frac{\partial}{\partial t} \rho_S(t) = -i\alpha[H_S(t) + H_{LS}(t), \rho_S(t)] + \mathcal{D}(\rho_S(t)) \quad (\text{Schrödinger picture}). \quad (4.3)$$

In general, there can be a Lamb shift-type contribution  $H_{LS}(t)$  in addition to the unitary dynamics generated by  $H_S(t)$  in the Schrödinger picture equation, alongside the dissipator  $\mathcal{D}(\rho_S(t))$ , the latter encodes both decoherence and dissipation. Note that the dissipator can in general retain some time dependence even after the transition to the Schrödinger picture, this critically depends on the model and the applied approximations. At this point there are various possibilities to proceed, the easiest and most frequently used path is towards the so-called Lindblad equation. We would like to mention that the Lindblad equation is a very specific form of a master equation in the sense that it originates from a bounded generator of a quantum-dynamical semigroup [45, 46]. Naturally, the Lindblad equation is completely positive and trace-preserving, a feature not all master equations exhibit [13] without further adjustments as we will see at the end of this section. We will start with a fully analytical and in principle exact approach towards the Nakajima-Zwanzig equation, which will provide the basis for the TCL master equation we will later use for the model considered in this work. This relies fundamentally on the results in [80–83]. Hereby we will closely follow the pedagogical outline given in

<sup>7</sup> Due to the involvement of gravity, the coupling will not be dimensionless in the application of this formalism, see chapter (2.3) for further details.

[13] and [12]. The density operator  $\tilde{\rho}(t)$  in (4.1) is trace class by construction, hence is contained in the set  $\mathcal{B}(\mathcal{H})$  of linear, bounded operators. We proceed to define a pair of projection superoperators called  $\mathcal{P}$  and  $\mathcal{Q}$ :

$$\mathcal{P} : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{H}), \quad \mathcal{P}(\rho) := \text{tr}_{\mathcal{E}}(\rho) \otimes \rho_{\mathcal{E}}, \quad (4.4)$$

$$\mathcal{Q} : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{H}), \quad \mathcal{Q}(\rho) := \rho - \mathcal{P}(\rho), \quad (4.5)$$

where  $\rho_{\mathcal{E}}$  is a stationary environmental state that we will specify later and  $\text{tr}_{\mathcal{E}}$  defines the environmental trace, i.e. the trace over all degrees of freedom that do not belong to the system of interest. The fact that these operators satisfy  $\mathcal{P}^2 = \mathcal{P}$ ,  $\mathcal{Q}^2 = \mathcal{Q}$ ,  $\mathcal{P} + \mathcal{Q} = \mathbb{1}$  and  $\mathcal{P}\mathcal{Q} = \mathcal{Q}\mathcal{P} = 0$  and thus are indeed projectors is directly evident from the definitions above. The operators  $\mathcal{P}$  and  $\mathcal{Q}$  signify the relevant and irrelevant part of the combined dynamics, respectively, in the sense that everything projected with  $\mathcal{P}$  exclusively contains the dynamics of the system degrees of freedom, with only an effective influence coming from the traced-out environment. These operators do furthermore commute with the partial time-derivative and hence can be directly applied to the Liouville-von Neumann equation (4.1) to obtain:

$$\mathcal{P} \frac{\partial}{\partial t} \tilde{\rho}(t) = \frac{\partial}{\partial t} \mathcal{P} \tilde{\rho}(t) = \alpha \mathcal{P} \mathcal{L}(t) \tilde{\rho}(t) = \alpha (\mathcal{P} \mathcal{L}(t) \mathcal{P} \tilde{\rho}(t) + \mathcal{P} \mathcal{L}(t) \mathcal{Q} \tilde{\rho}(t)), \quad (4.6)$$

$$\mathcal{Q} \frac{\partial}{\partial t} \tilde{\rho}(t) = \frac{\partial}{\partial t} \mathcal{Q} \tilde{\rho}(t) = \alpha \mathcal{Q} \mathcal{L}(t) \tilde{\rho}(t) = \alpha (\mathcal{Q} \mathcal{L}(t) \mathcal{P} \tilde{\rho}(t) + \mathcal{Q} \mathcal{L}(t) \mathcal{Q} \tilde{\rho}(t)). \quad (4.7)$$

Hereby we used that the projection operators are linear. In the last step of each line we used that the two projectors combine to unity. In the end we would like to find a solution for the relevant part  $\mathcal{P} \tilde{\rho}(t)$  of the density operator, the projectors do however not commute with the Liouvillian  $\mathcal{L}(t)$ . As a first step to eliminate the appearance of  $\mathcal{Q} \tilde{\rho}(t)$  in the first and hence for us interesting and relevant equation (4.6), we write down a formal solution of equation (4.7):

$$\mathcal{Q} \tilde{\rho}(t) = \mathcal{G}(t, t_0) \mathcal{Q} \tilde{\rho}(t_0) + \alpha \int_{t_0}^t ds \mathcal{G}(t, s) \mathcal{Q} \mathcal{L}(s) \mathcal{P} \tilde{\rho}(s), \quad \mathcal{G}(t, s) := \mathcal{T}_{\leftarrow} \exp \left[ \int_s^t d\tau \mathcal{Q} \mathcal{L}(\tau) \right], \quad (4.8)$$

where  $\mathcal{T}_{\leftarrow}$  stands for time-ordering with the largest argument to the very left. Applying a partial derivative with respect to  $t$  yields equation (4.7) while the initial condition for  $t = t_0$  is satisfied due to  $\mathcal{G}(t_0, t_0) = \mathbb{1}$  because of a vanishing integral, so the stated solution is unique. Inserting (4.8) into the relevant equation (4.6) gives:

$$\frac{\partial}{\partial t} \mathcal{P} \tilde{\rho}(t) = \alpha \mathcal{P} \mathcal{L}(t) \mathcal{P} \tilde{\rho}(t) + \alpha \mathcal{P} \mathcal{L}(t) \mathcal{G}(t, t_0) \mathcal{Q} \tilde{\rho}(t_0) + \alpha^2 \int_{t_0}^t ds \mathcal{P} \mathcal{L}(t) \mathcal{G}(t, s) \mathcal{Q} \mathcal{L}(s) \mathcal{P} \tilde{\rho}(s). \quad (4.9)$$

This is the *Nakajima-Zwanzig* equation, albeit admittedly not less complicated than the attempt of solving the initial problem, mostly due to the appearance of the integral kernel

$$\mathcal{K}(t, s) := \mathcal{P} \mathcal{L}(t) \mathcal{G}(t, s) \mathcal{Q} \mathcal{L}(s) \mathcal{P}. \quad (4.10)$$

It is worth noting that this manifestly non-Markovian equation (4.9) is in principle exact, there was no need for approximations whatsoever until this point. If we assume that the interaction Hamiltonian density is of the general form  $\sum_{\alpha} S_{\alpha} \otimes E_{\alpha}$  with a discrete summation index  $\alpha$  (omitting model-specific time- or mode dependencies for the sake of brevity) and  $E_{\alpha}$  consisting of linear contributions of annihilation- and creation-type variables, the first contribution in equation (4.9) vanishes. This is simply due to the fact that odd numbers of the environmental interaction term do not contribute in the environmental trace, given a suitable initial state, specifically a Gibbs state. This observation is contained in the very definition (4.4)

of the relevant projector  $\mathcal{P}$ , which turns out to be useful even beyond the mere simplification of equation (4.9) as we will see later in this section. For this broad class of models, a method called *cumulant expansion* turns out to be a powerful tool for a step-by-step approximation of a time-convolutionless approach to non-Markovian dynamics, first developed in [105] and [106], respectively. Up to now the main complexity of the Nakajima-Zwanzig equation remains, the convolution integral makes it nearly impossible to be solved analytically. To this extent, we introduce the two propagators  $F(t, s)$  and  $G(t, s)$  such that  $G(t, s)\tilde{\rho}(s) = \tilde{\rho}(t)$  and  $F(t, s) \circ G(t, s) = \mathbb{1}$  with  $s \leq t \in \mathbb{R}$ . Note the distinction  $\mathcal{G}(t, s) \neq G(t, s)$  in the previously used notation, the latter being the propagator of the *total* system and not merely the one of the irrelevant projection. Hence:

$$\frac{\partial}{\partial t} \left( F(t, s) \circ G(t, s) \tilde{\rho}(s) \right) = \left( \frac{\partial}{\partial t} F(t, s) \right) \tilde{\rho}(t) + \alpha F(t, s) \mathcal{L}(t) \tilde{\rho}(t) = 0. \quad (4.11)$$

This can be easily solved formally for the inverse propagator:

$$\frac{\partial}{\partial t} F(t, s) = -\alpha F(t, s) \mathcal{L}(t) \iff F(t, s) = \mathcal{T}_{\rightarrow} \exp \left[ -\alpha \int_s^t d\tau \mathcal{L}(\tau) \right], \quad (4.12)$$

where  $\mathcal{T}_{\rightarrow}$  denotes anti-time-ordering, that is with the largest argument to the very right. As the next step, we insert  $\tilde{\rho}(s) = F(t, s)(\mathcal{P} + \mathcal{Q})\tilde{\rho}(t)$  into the solution of equation (4.7) and obtain:

$$\begin{aligned} \mathcal{Q}\tilde{\rho}(t) &= \mathcal{G}(t, t_0) \mathcal{Q}\tilde{\rho}(t_0) + \alpha \int_{t_0}^t ds \mathcal{G}(t, s) \mathcal{Q}\mathcal{L}(s) \mathcal{P}F(t, s)(\mathcal{P} + \mathcal{Q})\tilde{\rho}(t) \\ &=: \mathcal{G}(t, t_0) \mathcal{Q}\tilde{\rho}(t_0) + \Sigma(t, t_0)(\mathcal{P} + \mathcal{Q})\tilde{\rho}(t), \end{aligned} \quad (4.13)$$

where we collected the convolution part into a separate, now central entity  $\Sigma(t, t_0)$ :

$$\Sigma(t, t_0) := \alpha \int_{t_0}^t ds \mathcal{G}(t, s) \mathcal{Q}\mathcal{L}(s) \mathcal{P}F(t, s). \quad (4.14)$$

Now our aim is to solve (4.13) for  $\mathcal{Q}\tilde{\rho}(t)$  by collecting all terms on one side:

$$(\mathbb{1} - \Sigma(t, t_0)) \mathcal{Q}\tilde{\rho}(t) = \mathcal{G}(t, t_0) \mathcal{Q}\tilde{\rho}(t_0) + \Sigma(t, t_0) \mathcal{P}\tilde{\rho}(t). \quad (4.15)$$

The key point is the treatment of  $(\mathbb{1} - \Sigma(t, t_0))$  or rather the inverse of it. Evidently, the initial conditions  $\Sigma(t_0, t_0) = 0$  and  $\Sigma(t, t_0)|_{\alpha=0} = 0$  are easy to check, our approach involves the representation of  $(\mathbb{1} - \Sigma(t, t_0))^{-1}$  as the limit of a geometric series of operators, that is:

$$(\mathbb{1} - \Sigma(t, t_0))^{-1} = \sum_{n=0}^{\infty} (\Sigma(t, t_0))^n. \quad (4.16)$$

It is indeed the case that there exists some neighbourhood around  $t_0$  where the expression in (4.16) is invertible. These steps require some additional mathematical justification, for a more detailed discussion on this point we refer the reader to appendix C in the appendix. Hence, under the assumption that we restrict to an interval  $[t_0, t]$  where invertibility is guaranteed:

$$\mathcal{Q}\tilde{\rho}(t) = (\mathbb{1} - \Sigma(t, t_0))^{-1} \mathcal{G}(t, t_0) \mathcal{Q}\tilde{\rho}(t_0) + (\mathbb{1} - \Sigma(t, t_0))^{-1} \Sigma(t, t_0) \mathcal{P}\tilde{\rho}(t). \quad (4.17)$$

Finally inserting this into the Nakajima-Zwanzig equation (4.9) for  $\mathcal{P}\tilde{\rho}(t)$  yields:

$$\begin{aligned} \frac{\partial}{\partial t} \mathcal{P}\tilde{\rho}(t) &= \alpha \mathcal{P}\mathcal{L}(t) \left( \mathbb{1} + (\mathbb{1} - \Sigma(t, t_0))^{-1} \Sigma(t, t_0) \right) \mathcal{P}\tilde{\rho}(t) + \alpha \mathcal{P}\mathcal{L}(t) (\mathbb{1} - \Sigma(t, t_0))^{-1} \mathcal{G}(t, t_0) \mathcal{Q}\tilde{\rho}(t_0) \\ &=: K(t, t_0) \mathcal{P}\tilde{\rho}(t) + \mathcal{I}(t, t_0) \mathcal{Q}\tilde{\rho}(t_0). \end{aligned} \quad (4.18)$$

This is the so-called TCL master equation. Note that this master equation, although time-local, still conveys non-Markovian processes in contributions with higher than second order in  $\alpha$ . In principle it is even exact in the neighbourhood where  $(\mathbb{1} - \Sigma(t, t_0))^{-1}$  is invertible, considering terms of all orders. We will not further investigate the inhomogeneity  $\mathcal{I}(t, t_0)\mathcal{Q}\tilde{\rho}(t_0)$  at this point since we are focused on separable initial conditions for now, for which this expression vanishes. To this extent, consider the explicit form of the TCL generator  $K(t, t_0)$  in terms of the geometric series expansion of the inverse of  $(\mathbb{1} - \Sigma(t, t_0))$ , neglecting  $\mathcal{P}\mathcal{L}(t)\mathcal{P}$  based on the assumption that the environmental part of the interaction Hamiltonian density is linearly comprised of creation- and annihilation-type operators:

$$\begin{aligned} K(t, t_0) &= \alpha\mathcal{P}\mathcal{L}(t)(\mathbb{1} - \Sigma(t, t_0))^{-1}\Sigma(t, t_0)\mathcal{P} = \alpha\mathcal{P}\mathcal{L}(t)\sum_{n=1}^{\infty}(\Sigma(t, t_0))^n\mathcal{P} \\ &= \alpha\mathcal{P}\mathcal{L}(t)\sum_{n=1}^{\infty}\left(\sum_{m=1}^{\infty}\alpha^m\Sigma_m(t, t_0)\right)^n\mathcal{P}, \end{aligned} \quad (4.19)$$

where we noticed that it is apparent that  $\Sigma(t, t_0)$ , given its explicit form in equation (4.14), can be expanded in powers of  $\alpha$  as well. Since we are interested in a weak coupling scenario, it is sufficient to consider contributions up to second order in  $\alpha$ , that is only the lowest-order terms that occur in the series expansion of  $\Sigma(t, t_0)$ . In our particular case this entails *only* the second-order contributions since  $\mathcal{P}\mathcal{L}(t)\mathcal{P} = 0$ . This fact amounts to the Born approximation on the one hand, by virtue of the assumption of separable initial conditions with a stationary thermal environment and to the (first) Markov approximation on the other hand. The latter fact becomes evident once we realise that by eliminating the convolution integral with the density operator altogether without correcting for it with higher-order terms, we end up with an equation that has simply replaced  $\tilde{\rho}(s)$  with  $\tilde{\rho}(t)$ . The second Markov approximation is commonly depicted as the limit  $t_0 \rightarrow -\infty$ , based on the assumption that the characteristic time scales of system and environment are vastly different, inducing a peakedness of the correlation functions. The latter arise naturally as functions of the form  $\text{tr}_\varepsilon(E_\alpha(t)E_\beta(s)\rho_\varepsilon)$  in the explicit computation of the environmental trace with  $E_\alpha(t)$  meaning the environmental part of the interaction Hamiltonian. This peakedness property means that an extension of the integration interval has a negligible effect on the effective dynamics of the system, which however has to be checked explicitly for any given model. In the specific decoherence model we consider, we do not take this limit yet, all of the previous assumptions are nevertheless justified based on the weak coupling of matter to gravity. The only relevant contribution is the one where  $\mathcal{G}(t, s)$  and  $F(t, s)$  do not contribute, since everything beyond the identity contains the coupling constant.

$$K(t, t_0) = \alpha^2\mathcal{P}\mathcal{L}(t)\Sigma_1(t, t_0)\mathcal{P} = \alpha^2\mathcal{P}\mathcal{L}(t)\int_{t_0}^t ds\mathcal{Q}\mathcal{L}(s)\mathcal{P} = \alpha^2\mathcal{P}\int_{t_0}^t ds\mathcal{L}(t)\mathcal{L}(s)\mathcal{P}. \quad (4.20)$$

Note that higher-order contributions in the coupling constant get increasingly difficult to evaluate explicitly, based on the complexity of (4.19), see for example [107] for the fourth-order TCL generator. The cumulant expansion procedure [105, 106] provides a systematic framework in which the individual contributions in  $K(t, t_0)$  can be evaluated, especially for models in which all terms involving  $\mathcal{P}\mathcal{L}(t)\mathcal{P}$  vanish due to the explicit form of the interaction. Now we can explicitly write down the desired master equation we will use for the model considered in this article later:

$$\begin{aligned}\frac{\partial}{\partial t}\mathcal{P}\tilde{\rho}(t) &= \alpha^2\mathcal{P}\int_{t_0}^t ds\mathcal{L}(t)\mathcal{L}(s)\tilde{\rho}(t) \\ &= -\alpha^2\int_{t_0}^t ds\text{tr}_{\mathcal{E}}\left\{\left[\tilde{H}(t),\left[\tilde{H}(s),\text{tr}_{\mathcal{E}}\{\tilde{\rho}(t)\}\otimes\tilde{\rho}_{\mathcal{E}}\right]\right]\right\}\otimes\tilde{\rho}_{\mathcal{E}}.\end{aligned}\quad (4.21)$$

This second-order TCL master equation could have been obtained in a straightforward fashion: First integrate the von Neumann equation (4.1) and subsequently insert the formal solution into the original equation, akin to the second-order truncation of the Dyson series. Secondly, assume separable initial conditions and replace  $\tilde{\rho}_S(s) \rightarrow \tilde{\rho}_S(t)$ , eliminating the convolution integral. It is however not clear from this perspective how non-Markovian corrections or initial conditions with nonvanishing entanglement entropy can be dealt with. The presented projection operator technique in conjunction with the TCL formalism answers these questions order by order and simultaneously provides a differential (as opposed to an integro-differential) equation for  $\tilde{\rho}(t)$ . An alternative but largely equivalent route towards a time-local master equation is presented in the next section. Lastly we would like to add a few sentences regarding the relation between the TCL master equation (4.21) and the frequently used Lindblad equation [45, 46]. Trace preservation and the notion of positivity are of paramount importance when it comes to solutions of the master equation. However, not every master equation is guaranteed to admit solutions that adhere to these principles. The Lindblad form ensures that these criteria are automatically met:

$$\frac{\partial}{\partial t}\rho_S(t) = -i\alpha[H_S(t) + H_{LS}(t), \rho_S(t)] + \sum_{\alpha\beta}\lambda_{\alpha\beta}\left(L_{\alpha}\rho_S(t)L_{\beta}^{\dagger} - \frac{1}{2}\{L_{\beta}^{\dagger}L_{\alpha}, \rho_S(t)\}\right), \quad (4.22)$$

with a finite set of traceless Lindblad operators  $L_{\alpha}$  and positive, semi-definite coefficients  $\lambda_{\alpha\beta}$ . This equation of so-called *first* standard form can be diagonalised by a unitary transformation of the operators  $L_{\alpha}$ , after which it is commonly called the *second* standard form. It is immediately clear that not every equation of the form (4.21) exhibits Lindblad form, firstly because of the remaining time dependence that has to be removed by hand based on the properties of the environmental correlation functions. Secondly, the operator structure also differs in a way that positivity is not automatically ensured, even if the derivation has been performed in a top-down manner [108]. There is no universal pathway that remedies these issues, for certain models it turns out to be feasible to expand the dissipator to obtain Lindblad form [12], whereas it can be sufficient to neglect certain contributions in the course of a rotating-wave approximation [109].

#### 4.2. Review of the influence phase functional approach

A different albeit mostly equivalent approach of formulating a master equation is the influence phase functional approach, which has the same starting point as the previously introduced projection operator formalism. During the course of this brief review we will follow the idea first presented in [79] while we also follow the pedagogical presentation in [12]. Regarding the notation in terms of propagators and thermal Wightman functions, we work with the same notation as for instance used in [110] or [10]. As before, we start at the von Neumann equation in the interaction picture:

$$\frac{\partial}{\partial t}\tilde{\rho}(t) = -i\alpha[\tilde{H}(t), \tilde{\rho}(t)] =: \alpha\mathcal{L}(t)\tilde{\rho}(t), \quad \tilde{\rho}_S(t) = \text{tr}_{\mathcal{E}}\left\{\mathcal{T}_{\leftarrow}\exp\left\{\alpha\int_{t_0}^t\int_{\mathbb{R}^3}d\tau d^3x\ell(x)\right\}\tilde{\rho}(t_0)\right\}, \quad (4.23)$$

where  $\ell(x) = \ell(\tau, \vec{x})$  is the superoperator density associated to  $\mathcal{L}(\tau)$  and  $\mathcal{T}_{\leftarrow} := \mathcal{T}_{\leftarrow}^{(S)}\mathcal{T}_{\leftarrow}^{(\varepsilon)}$  corresponds to chronological time-ordering of system and environmental operators, respectively, with larger time arguments to the left of the expression. Contrary to the Nakajima-Zwanzig formalism, the strategy is to directly eliminate the environmental time-ordering at the level of the iterated Dyson series (4.23). Note that system's operators can be considered as commuting under the time-ordering  $\mathcal{T}_{\leftarrow}^{(S)}$ . We assume an interaction Hamiltonian density of the general form  $\mathcal{H}(x) = \sum_{\alpha} S_{\alpha}(x) \otimes E_{\alpha}(x)$ , the indices  $\alpha, \beta$  can be understood as (a mixture of either co- or contravariant spacetime- or internal) multi-indices depending on the specific model. The goal of this section is to reformulate equation (4.23) into a more tractable form similar to

$$\tilde{\rho}_S(t) = \exp\{i\Phi_t[S^+, S^-]\}\tilde{\rho}_S(t_0), \quad (4.24)$$

where  $\Phi_t[S^+, S^-]$  is called the influence phase functional and the  $S^{\pm}$  are superoperators associated to the system of interest. A convenient decomposition of the time ordering in equation (4.23) can be straightforwardly shown by induction for *almost* commuting environmental operators, i.e. the case that  $[E_{\alpha}(x), E_{\beta}(x')] \propto ic_{\alpha\beta}(x, x')\mathbb{1}_{\varepsilon}$  with  $c_{\alpha\beta}(x, x') \in \mathbb{R}$ . This is done by introducing an ansatz where the time-ordering amounts to a phase factor times the ordinary, non-time-ordered exponential of the Liouvillian, leading to a differential equation which can be directly integrated. A very similar procedure has been performed at the level of the time-evolution operator in [111] for the spin-boson model. This can be carried over almost immediately by recalling the adjoint action of these operators on the density operator and the fact that the  $S_{\alpha}(t)$  can be ordered arbitrarily *inside* the time-ordering  $\mathcal{T}_{\leftarrow}^{(S)}$ . This is the setup this work will rely on, however in the more general case of strictly non-commuting environmental operators, the proof still holds but relies on the application of the finite-temperature Wick theorem [12, 112]. The result is:

$$\begin{aligned} \tilde{\rho}_S(t) &= \mathcal{T}_{\leftarrow}^{(S)} \text{tr}_{\varepsilon} \left\{ \mathcal{T}_{\leftarrow}^{(\varepsilon)} \exp \left\{ \alpha \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \ell(x) \right\} \tilde{\rho}(t_0) \right\} \\ &= \mathcal{T}_{\leftarrow}^{(S)} \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^{\tau} \int_{\mathbb{R}^3} d\tau' d^3x' [\ell(x), \ell(x')] \right\} \text{tr}_{\varepsilon} \\ &\quad \times \left\{ \exp \left\{ \alpha \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \ell(x) \right\} \tilde{\rho}(t_0) \right\} \\ &= \mathcal{T}_{\leftarrow}^{(S)} \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^{\tau} \int_{\mathbb{R}^3} d\tau' d^3x' [\ell(x), \ell(x')] \right\} \\ &\quad \times \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^{\tau} \int_{\mathbb{R}^3} d\tau' d^3x' \langle \ell(x)\ell(x') \rangle_{\varepsilon} \right\} \tilde{\rho}_S(t_0), \end{aligned} \quad (4.25)$$

with the environmental expectation value  $\langle \ell(x)\ell(x') \rangle_{\varepsilon} := \text{tr}_{\varepsilon} \{ \ell(x)\ell(x')\rho_{\varepsilon} \}$  under the assumption that the initial state  $\tilde{\rho}(t_0) = \tilde{\rho}_S(t_0) \otimes \tilde{\rho}_{\varepsilon}(t_0)$  is uncorrelated and hence separable. Note that the integrals in the first exponent are nested, i.e. the commutator of the Liouvillians is explicitly time-ordered whereas the second exponent holds integrals with identical limits. Furthermore we would like to point out that the last line of (4.25) only acts upon  $\tilde{\rho}_S(t_0)$ . This is due to the fact that for an uncorrelated thermal initial state, the action of the exponentiated Liouvillian can be directly computed in terms of a cumulant expansion with a vanishing first order, terminating after the second order, assuming that  $\tilde{\rho}_{\varepsilon}$  describes a Gaussian state with respect to the environmental interaction variables. Hence it is possible to absorb the expectation value

$\langle \ell(x)\ell(x') \rangle_\varepsilon$  into the exponential. Use of the Jacobi identity and the above definitions lets us further simplify (4.25):

$$\begin{aligned} [\ell(x), \ell(x')] \tilde{\rho} &= - [[\mathcal{H}(x), \mathcal{H}(x')], \tilde{\rho}] = - \sum_{\alpha} \sum_{\beta} [[S_{\alpha}(x) \otimes E_{\alpha}(x), S_{\beta}(x') \otimes E_{\beta}(x')], \tilde{\rho}_S \otimes \tilde{\rho}_{\varepsilon}] \\ &\cong - \sum_{\alpha} \sum_{\beta} [E_{\alpha}(x), E_{\beta}(x')] [S_{\alpha}(x) S_{\beta}(x') \otimes \mathbf{1}_{\varepsilon}, \tilde{\rho}_S \otimes \tilde{\rho}_{\varepsilon}]. \end{aligned}$$

Note that the last equality holds under consideration of the overall system time-ordering to the very left of expression (4.25) and under the assumption that the commutator  $[E_{\alpha}(x), E_{\beta}(x')]$  is purely a phase factor, as it is the case for the model in this work. We have indicated that a certain equality only holds under additional time-ordering  $\mathcal{T}_{\leftarrow}^{(S)}$  with  $\cong$ . This justifies acting upon  $\tilde{\rho}_S$  instead of  $\tilde{\rho}_S \otimes \tilde{\rho}_{\varepsilon}$  in (4.25), where the environmental part has been completely reduced to the expectation value in the second exponential. The action of these operator densities on the density operator is to be understood in a formal sense, all quantities are spatially integrated once we insert these identities back into the phase functional. In order to represent the action of the exponentiated commutators, let us introduce the following notation:

$$S_{\alpha}^{+}(x) \tilde{\rho}_S := S_{\alpha}(x) \tilde{\rho}_S, \quad S_{\alpha}^{-}(x) \tilde{\rho}_S := \tilde{\rho}_S S_{\alpha}(x),$$

where the commutator of Liouvillians  $[\ell(x), \ell(x')]$  immediately assumes the form

$$[\ell(x), \ell(x')] = - \sum_{\alpha} \sum_{\beta} \left( [E_{\alpha}(x), E_{\beta}(x')] S_{\alpha}^{+}(x) S_{\beta}^{+}(x') + [E_{\alpha}(x), E_{\beta}(x')] S_{\alpha}^{-}(x) S_{\beta}^{-}(x') \right).$$

This concludes the first part of the derivation of the influence phase functional based on (4.25):

$$\begin{aligned} \tilde{\rho}_S(t) &= \mathcal{T}_{\leftarrow}^{(S)} \exp \left\{ - \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \theta(\tau - \tau') \sum_{\alpha} \sum_{\beta} [E_{\alpha}(x), E_{\beta}(x')] S_{\alpha}^{+}(x) S_{\beta}^{+}(x') \right\} \\ &\quad \times \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \theta(\tau - \tau') \sum_{\alpha} \sum_{\beta} [E_{\alpha}(x), E_{\beta}(x')] S_{\alpha}^{-}(x) S_{\beta}^{-}(x') \right\} \\ &\quad \times \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \langle \ell(x)\ell(x') \rangle_{\varepsilon} \right\} \tilde{\rho}_S(t_0). \end{aligned} \quad (4.26)$$

Next we have a closer look at the thermal expectation value  $\langle \ell(x)\ell(x') \rangle_{\varepsilon}$  and apply its definition:

$$\begin{aligned} &\int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \langle \ell(x)\ell(x') \rangle_{\varepsilon} \tilde{\rho}_S \\ &= \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \text{tr}_{\varepsilon} \left( - [\mathcal{H}(x), [\mathcal{H}(x'), \tilde{\rho}_S \otimes \tilde{\rho}_{\varepsilon}]] \right) \\ &\cong \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \sum_{\alpha} \sum_{\beta} \left( - \langle E_{\beta}(x') E_{\alpha}(x) \rangle_{\varepsilon} S_{\alpha}^{+}(x) S_{\beta}^{+}(x') \right. \\ &\quad \left. - \langle E_{\alpha}(x) E_{\beta}(x') \rangle_{\varepsilon} S_{\alpha}^{-}(x) S_{\beta}^{-}(x') + \langle E_{\beta}(x') E_{\alpha}(x) \rangle_{\varepsilon} S_{\alpha}^{+}(x) S_{\beta}^{-}(x') + \langle E_{\alpha}(x) E_{\beta}(x') \rangle_{\varepsilon} S_{\alpha}^{-}(x) S_{\beta}^{+}(x') \right), \end{aligned}$$

where it is important that the last equality again only holds under the time-ordering still present in (4.26). This is because we exchanged the operators  $S_\alpha^+(x), S_\beta^+(x')$  and switched the integration variables and summation indices respectively. Afterwards we ended up with what seem to be switched arguments in the correlation function  $\langle E_\beta(x')E_\alpha(x) \rangle_\varepsilon$ , likewise for the contribution from  $S_\alpha^-(x)S_\beta^+(x')$ . This gives us the possibility to group the individual terms in the influence functional (4.25) such that we can express it in terms of thermal Wightman functions.

$$\begin{aligned} \tilde{\rho}_S(t) = \mathcal{T}_{\leftarrow}^{(S)} \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \sum_{\alpha} \sum_{\beta} \right. \\ \times \left[ - \left( \theta(\tau - \tau') [E_\alpha(x), E_\beta(x')] + \langle E_\beta(x')E_\alpha(x) \rangle_\varepsilon \right) S_\alpha^+(x) S_\beta^+(x') \right. \\ \left. + \left( \theta(\tau - \tau') [E_\alpha(x), E_\beta(x')] - \langle E_\alpha(x)E_\beta(x') \rangle_\varepsilon \right) S_\alpha^-(x) S_\beta^-(x') \right. \\ \left. + \langle E_\beta(x')E_\alpha(x) \rangle_\varepsilon S_\alpha^+(x) S_\beta^-(x') + \langle E_\alpha(x)E_\beta(x') \rangle_\varepsilon S_\alpha^-(x) S_\beta^+(x') \right] \left. \right\} \tilde{\rho}_S(t_0). \end{aligned} \quad (4.27)$$

Upon closer inspection, these combinations of correlation functions amount to the propagators:

$$\begin{aligned} G_{\alpha\beta}^{++}(x-x') &:= \langle \mathcal{T}_{\leftarrow}^{(\varepsilon)}(E_\alpha(x)E_\beta(x')) \rangle_\varepsilon = \theta(\tau - \tau') [E_\alpha(x), E_\beta(x')] + \langle E_\beta(x')E_\alpha(x) \rangle_\varepsilon \\ &= \langle E_\alpha(x)E_\beta(x') \rangle_\varepsilon \theta(\tau - \tau') + \langle E_\beta(x')E_\alpha(x) \rangle_\varepsilon \theta(\tau' - \tau) \\ &=: G_{\alpha\beta}^>(x-x')\theta(\tau - \tau') + G_{\alpha\beta}^<(x-x')\theta(\tau' - \tau), \end{aligned} \quad (4.28)$$

and similarly although anti-time-ordered for the other contribution:

$$\begin{aligned} -G_{\alpha\beta}^{--}(x-x') &:= -\langle \mathcal{T}_{\rightarrow}^{(\varepsilon)}(E_\alpha(x)E_\beta(x')) \rangle_\varepsilon = \theta(\tau - \tau') [E_\alpha(x), E_\beta(x')] - \langle E_\alpha(x)E_\beta(x') \rangle_\varepsilon \\ &= -\langle E_\alpha(x)E_\beta(x') \rangle_\varepsilon \theta(\tau' - \tau) - \langle E_\beta(x')E_\alpha(x) \rangle_\varepsilon \theta(\tau - \tau') \\ &=: -G_{\alpha\beta}^>(x-x')\theta(\tau' - \tau) - G_{\alpha\beta}^<(x-x')\theta(\tau - \tau'). \end{aligned} \quad (4.29)$$

Inserting  $G_{\alpha\beta}^{++}(x-x')$  and  $G_{\alpha\beta}^{--}(x-x')$  into equation (4.27) yields

$$\begin{aligned} \tilde{\rho}_S(t) = \mathcal{T}_{\leftarrow}^{(S)} \exp \left\{ \frac{\alpha^2}{2} \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^t \int_{\mathbb{R}^3} d\tau' d^3x' \sum_{\alpha} \sum_{\beta} \right. \\ \times \left[ - \left( G_{\alpha\beta}^>(x-x')\theta(\tau - \tau') + G_{\alpha\beta}^<(x-x')\theta(\tau' - \tau) \right) S_\alpha^+(x) S_\beta^+(x') \right. \\ \left. - \left( G_{\alpha\beta}^>(x-x')\theta(\tau' - \tau) + G_{\alpha\beta}^<(x-x')\theta(\tau - \tau') \right) S_\alpha^-(x) S_\beta^-(x') \right. \\ \left. + G_{\alpha\beta}^<(x-x') S_\alpha^+(x) S_\beta^-(x') + G_{\alpha\beta}^>(x-x') S_\alpha^-(x) S_\beta^+(x') \right] \left. \right\} \tilde{\rho}_S(t_0). \end{aligned}$$

Now we would like to eliminate the Heaviside functions  $\theta(\tau - \tau')$  from this expression in order to be left purely with the thermal Wightman functions  $G_{\alpha\beta}^{\lessgtr}(x-x')$ . This is achieved by expanding the mixed contributions with  $1 = \theta(\tau - \tau') + \theta(\tau' - \tau)$  and renaming

the integration variables in the individual terms while changing the integration domain into a nested integral [110]:

$$\begin{aligned} \tilde{\rho}_S(t) = & \mathcal{T}_{\leftarrow}^{(S)} \exp \left\{ \alpha^2 \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^{\tau} \int_{\mathbb{R}^3} d\tau' d^3x' \sum_{\alpha} \sum_{\beta} \right. \\ & \times \left[ -G_{\alpha\beta}^>(x-x') S_{\alpha}^+(x) S_{\beta}^+(x') - G_{\alpha\beta}^<(x-x') S_{\alpha}^-(x) S_{\beta}^-(x') \right. \\ & \left. \left. + G_{\alpha\beta}^<(x-x') S_{\alpha}^+(x) S_{\beta}^-(x') + G_{\alpha\beta}^>(x-x') S_{\alpha}^-(x) S_{\beta}^+(x') \right] \right\} \tilde{\rho}_S(t_0). \end{aligned} \quad (4.30)$$

This concludes the derivation of the influence phase functional, while strictly speaking  $\Phi_t[S^+, S^-]$  is both a functional in the  $S^{\pm}$  operators as well as a superoperator that acts on the space of density operators associated to the system. Explicitly, we obtain

$$\begin{aligned} i\Phi_t[S^+, S^-] := & \alpha^2 \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^{\tau} \int_{\mathbb{R}^3} d\tau' d^3x' \sum_{\alpha} \sum_{\beta} \\ & \times \left[ -G_{\alpha\beta}^>(x-x') S_{\alpha}^+(x) S_{\beta}^+(x') - G_{\alpha\beta}^<(x-x') S_{\alpha}^-(x) S_{\beta}^-(x') \right. \\ & \left. + G_{\alpha\beta}^<(x-x') S_{\alpha}^+(x) S_{\beta}^-(x') + G_{\alpha\beta}^>(x-x') S_{\alpha}^-(x) S_{\beta}^+(x') \right]. \end{aligned} \quad (4.31)$$

The influence phase functional is often depicted in a slightly different form, including functions called the dissipation- and noise kernel based on their interpretation in well-established decoherence models. The transition from the expression in (4.31) is rather straightforward. To this extent, let us first introduce  $D_{\alpha\beta}(x-x')$  and  $D_{\alpha\beta}^1(x-x')$  based on the thermal Wightman functions:

$$D_{\alpha\beta}(x-x') := \langle [E_{\alpha}(x), E_{\beta}(x')] \rangle_{\varepsilon} = i(G_{\alpha\beta}^>(x-x') - G_{\alpha\beta}^<(x-x')), \quad (4.32)$$

$$D_{\alpha\beta}^1(x-x') := \langle \{E_{\alpha}(x), E_{\beta}(x')\} \rangle_{\varepsilon} = G_{\alpha\beta}^>(x-x') + G_{\alpha\beta}^<(x-x'). \quad (4.33)$$

These relations can be readily inverted:

$$\begin{aligned} G_{\alpha\beta}^>(x-x') &= \frac{1}{2} (D_{\alpha\beta}^1(x-x') - iD_{\alpha\beta}(x-x')), \\ G_{\alpha\beta}^<(x-x') &= \frac{1}{2} (D_{\alpha\beta}^1(x-x') + iD_{\alpha\beta}(x-x')). \end{aligned}$$

In terms of the noise- and dissipation kernel, the influence phase functional is given by:

$$\begin{aligned} i\Phi_t[S^+, S^-] := & \alpha^2 \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^{\tau} \int_{\mathbb{R}^3} d\tau' d^3x' \sum_{\alpha} \sum_{\beta} \\ & \times \left[ \frac{i}{2} D_{\alpha\beta}(x-x') (S_{\alpha}^+(x) S_{\beta}^+(x') - S_{\alpha}^-(x) S_{\beta}^-(x') + S_{\alpha}^+(x) S_{\beta}^-(x') - S_{\alpha}^-(x) S_{\beta}^+(x')) \right. \\ & \left. - \frac{1}{2} D_{\alpha\beta}^1(x-x') (S_{\alpha}^+(x) S_{\beta}^+(x') + S_{\alpha}^-(x) S_{\beta}^-(x') - S_{\alpha}^+(x) S_{\beta}^-(x') - S_{\alpha}^-(x) S_{\beta}^+(x')) \right]. \end{aligned} \quad (4.34)$$

We realise that these particular combinations of  $S^\pm$ -superoperators can be cast into a familiar form in terms of combinations of commutators and anti-commutators

$$S_\alpha^c(x)(\cdot) := [S_\alpha(x), \cdot] = (S_\alpha^+(x) - S_\alpha^-(x))(\cdot), \quad S_\alpha^a(x)(\cdot) := \{S_\alpha(x), \cdot\} = (S_\alpha^+(x) + S_\alpha^-(x))(\cdot).$$

Finally, the most convenient (and most abundant) version of the influence phase functional reads:

$$i\Phi_I[S^+, S^-] := \alpha^2 \int_{t_0}^t \int_{\mathbb{R}^3} d\tau d^3x \int_{t_0}^\tau \int_{\mathbb{R}^3} d\tau' d^3x' \sum_\alpha \sum_\beta \times \left[ \frac{i}{2} D_{\alpha\beta}(x-x') S_\alpha^c(x) S_\beta^a(x') - \frac{1}{2} D_{\alpha\beta}^1(x-x') S_\alpha^c(x) S_\beta^c(x') \right]. \quad (4.35)$$

At this stage it is evident that the projection operator formalism and the influence phase functional approach are equivalent formulations. Strictly speaking, this is only the case for Gaussian environmental initial states [12]. Up to second order this equality can be easily verified by deriving equation (4.31) with respect to time and consequently applying all the commutators and anti-commutators to the density operator to the right. Once we recall the definition of the kernels, this precisely amounts to the Born-Markov master equation prior to the second Markov approximation regarding the upper integral limit. A similar procedure can be implemented order by order [107].

#### 4.3. Derivation of the master equation

In this section we apply the projection operator method reviewed in section 4.1 to the model in this work. Assuming factorising initial conditions and a Gibbs state for the environment, we can take the TCL master equation (4.21) to second order in the coupling strength  $\alpha = \sqrt{\kappa}$  and evaluate it for the model under our consideration described by the interaction Hamiltonian operator shown in (3.17):

$$\frac{\partial}{\partial t} \tilde{\rho}_S(t) = -i\kappa \left[ \tilde{U}(t), \tilde{\rho}_S(t) \right] - \kappa \int_0^t ds \operatorname{tr}_E \left\{ \left[ \tilde{H}_{TI}(t), \left[ \tilde{H}_{TI}(s), \tilde{\rho}_S(t) \otimes \rho_E \right] \right] \right\}. \quad (4.36)$$

Now to obtain the final master equations we have to evaluate the trace over the environmental degrees of freedom in this expression and thus to obtain the correlation functions. As a first step we take into account that the second term on the right hand side of (4.36) can be written in terms of thermal Wightman functions, using the expression of  $\tilde{H}_{TI}(t)$  in position space given by

$$\tilde{H}_{TI} = \sqrt{\kappa} \int_{\mathbb{R}^3} d^3x \delta_b^i \varphi_a^b(\vec{x}, t) \delta \mathcal{E}_i^a(\vec{x}, t), \quad (4.37)$$

where  $\varphi_a^b(\vec{x}, t)$  denotes the three dimensional Fourier transforms of  $\varphi_a^b(\vec{k}, t)$  which was defined in (3.16). Following the procedure outlined in [110], where in our case the environmental part of the interaction is linear in the environmental fields and we use a Gibbs state for the environment, hence all one-point correlation functions vanish, we can define the thermal Wightman functions:

$$G^{>a}_i{}^b{}_j(\vec{x} - \vec{y}, t - s) := \langle \delta \mathcal{E}_i^a(\vec{x}, t) \delta \mathcal{E}_j^b(\vec{y}, s) \rangle_\epsilon = P^a{}_i{}^b{}_j G^>(\vec{x} - \vec{y}, t - s) \quad (4.38)$$

$$G^{<a b}_{i j}(\vec{x} - \vec{y}, t - s) := \langle \delta \mathcal{E}^b_j(\vec{y}, s) \delta \mathcal{E}^a_i(\vec{x}, t) \rangle_\epsilon = P^{a b}_{i j} G^<(\vec{x} - \vec{y}, t - s) = G^{>a b}_{i j}(\vec{y} - \vec{x}, s - t), \quad (4.39)$$

where  $\langle \cdot \rangle_\epsilon$  denotes the expectation value with respect to a thermal Gibbs state and  $P^{a b}_{i j}$  is the transverse-traceless projector given in (2.69). The detailed derivations of the identities in (4.38) and (4.39) as well as the derivations for the results up to (4.46) can be found in appendix D. Note that this definition coincides with the one in (4.28) adapted to the model we consider here, hence the  $\alpha$  and  $\beta$  indices from (4.28) here correspond each to the indices carried by  $\delta \mathcal{E}$ , i.e. to a spatial index  $a$  as well as an internal index  $i$ . The spectral representations of the thermal Wightman functions in terms of a spectral density denoted by  $\rho(k^0, |\vec{k}|)$  read:

$$G^>(\vec{x} - \vec{y}, t - s) = \int_{\mathbb{R}^4} \frac{d^4 k}{(2\pi)^{\frac{4}{2}}} \rho^>(k^0, |\vec{k}|) e^{-ik^0(t-s) + i\vec{k}(\vec{x}-\vec{y})} \quad (4.40)$$

$$G^<(\vec{x} - \vec{y}, t - s) = \int_{\mathbb{R}^4} \frac{d^4 k}{(2\pi)^{\frac{4}{2}}} \rho^<(k^0, |\vec{k}|) e^{-ik^0(t-s) + i\vec{k}(\vec{x}-\vec{y})} \quad (4.41)$$

with

$$\rho^>(k^0, |\vec{k}|) := (1 + N(k^0))\rho(k^0, |\vec{k}|) \quad \rho^<(k^0, |\vec{k}|) := N(k^0)\rho(k^0, |\vec{k}|) \quad (4.42)$$

and the spectral density

$$\rho(k^0, |\vec{k}|) := \sqrt{\frac{\pi}{2}} \frac{1}{\Omega_k} [\delta(k^0 - \Omega_k) - \delta(k^0 + \Omega_k)]. \quad (4.43)$$

From this decomposition, it is immediately possible to see the additional effect caused by the finite temperature in the environment, evident by the presence of the Bose–Einstein distribution

$$N(k^0) := \frac{1}{e^{\beta k^0} - 1} = \frac{\coth\left(\frac{\beta k^0}{2}\right)}{2} - \frac{1}{2}, \quad (4.44)$$

which vanishes for zero temperature parameter  $\Theta = \beta^{-1} = 0$  of the Gibbs state. As usual the two thermal Wightman functions can be combined to build the thermal Feynman propagator:

$$\begin{aligned} G^{(F)a b}_{i j}(\vec{x} - \vec{y}, t - s) &= \langle \mathcal{T}_\epsilon \delta \mathcal{E}^a_i(\vec{x}, t) \delta \mathcal{E}^b_j(\vec{y}, s) \rangle_\epsilon \\ &= P^{a b}_{i j} G^>(\vec{x} - \vec{y}, t - s)\theta(t - s) + P^{a b}_{i j} G^<(\vec{x} - \vec{y}, t - s)\theta(s - t), \end{aligned} \quad (4.45)$$

where  $\theta(t)$  denotes the Heaviside step function which is one for non-negative arguments and zero otherwise. A four-dimensional Fourier transform yields

$$\underline{G}^{(F)a b}_{i j}(p) = \underline{P}^{a b}_{i j}(\vec{p}) \left[ \frac{-i}{p^\mu p_\mu - i\epsilon} + 2\pi N(\Omega_p) \delta(p^\mu p_\mu) \right], \quad (4.46)$$

where the first part is the ordinary Feynman propagator one obtains at a vanishing temperature parameter  $\Theta$ , that is when the Gibbs state merges into a vacuum state, and a second part the thermal contribution that obviously vanishes for  $\Theta = 0$ . Note that we obtain a decomposition into the ordinary Feynman propagator and a thermal correction, this is caused by the fact that we use a mode expansion that involves a splitting into positive and negative frequency modes. As a consequence, normal ordered expectation values with respect to thermal states are in general non-vanishing as for instance  $\langle b_k^\dagger b_k \rangle_\epsilon = N(\Omega_k) \neq 0$ . Vanishing expectation values of normal ordered products are usually an important property in the proof of Wick's theorem. Given that we are working with a Gaussian state for the environment and an interaction

Hamiltonian being linear in the environmental fields, we can calculate the relevant expectation values in the model we consider directly without the use of Wick's theorem. However, for different and more complicated models the application of Wick's theorem might be of advantage and for this purpose it is often convenient to consider a different splitting than in the zero temperature case such that the expectation values of normal ordered products with respect to thermal states also vanish. A detailed discussion on possible splittings in this context as well as the relation between different choices of splittings is presented in [112]. Now we can write down the second term on the right side of in the master equation (4.36) in terms of the thermal Wightman functions:

$$\begin{aligned} & \kappa \int_0^t ds \operatorname{tr}_{\mathcal{E}} \left\{ \left[ \tilde{H}_{TI}(t), \left[ \tilde{H}_{TI}(s), \tilde{\rho}_S(t) \otimes \rho_{\mathcal{E}} \right] \right] \right\} \\ &= -\kappa \int_0^t ds \int_{\mathbb{R}^3} d^3x \int_{\mathbb{R}^3} d^3y \sum_{r \in \{\pm\}} \left\{ \left[ J_r(\vec{x}, t) J_r(\vec{y}, s) \tilde{\rho}_S(t) - J_r(\vec{y}, s) \tilde{\rho}_S(t) J_r(\vec{x}, t) \right] \right. \\ & \quad \left. \times G^>(\vec{x} - \vec{y}, t - s) + \left[ \tilde{\rho}_S(t) J_r(\vec{y}, s) J_r(\vec{x}, t) - J_r(\vec{x}, t) \tilde{\rho}_S(t) J_r(\vec{y}, s) \right] G^<(\vec{x} - \vec{y}, t - s) \right\}. \end{aligned} \quad (4.47)$$

Since we consider a Gibbs state on Fock space which is not well defined, we would need to work with KMS states [85, 86] or alternatively regularise the system within a finite volume, where we will follow the latter in this work. For this purpose we put the system into a box of volume  $V = L^3$  allowing us to explicitly show the above identities and evaluate thermal two-point functions. This kind of regularisation leads to the discreteness of modes that belong to the set  $\mathbb{K}$  now and to a replacement of the operator-valued distributions by operators:

$$\int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3} \longrightarrow \frac{1}{V} \sum_{\vec{k} \in \mathbb{K}} \quad b_k^r \longrightarrow \sqrt{\frac{V}{(2\pi)^3}} b_{\vec{k}}^r, \quad (4.48)$$

where a rescaling of the operators was introduced in order to keep the exponential  $e^{-\beta H_{\mathcal{E}}}$  dimensionless in the regularised model:

$$e^{-\beta H_{\mathcal{E}}} = \exp \left\{ -\beta \sum_{r \in \{+, -\}} \sum_{\vec{k} \in \mathbb{K}} \Omega_{\vec{k}} \left[ (b_{\vec{k}}^r)^\dagger b_{\vec{k}}^r \right] \right\}. \quad (4.49)$$

The Gibbs state for the environment is then of the form

$$\rho_{\mathcal{E}} = Z_{\mathcal{E}}^{-1} e^{-\beta H_{\mathcal{E}}} \quad (4.50)$$

with the partition sum given by

$$Z_{\mathcal{E}} := \operatorname{tr}_{\mathcal{E}} \left\{ e^{-\beta H_{\mathcal{E}}} \right\} = \left[ \prod_{j \in \mathbb{N}} \frac{1}{1 - e^{-\beta \Omega_{\vec{k}_j}}} \right]^2, \quad (4.51)$$

where the  $j \in \mathbb{N}$  label the  $\vec{k}_j \in \mathbb{K}$ . Moreover, for the calculation we had to assume  $\vec{k} = 0 \notin \mathbb{K}$ , which corresponds to  $\Omega_{\vec{k}} = 0$  and is the usual infrared divergence. More details on computing this partition sum can be found in appendix D. Removing the regulator by taking the limit

$L \rightarrow \infty$  in the Schrödinger picture we obtain the following TCL master equation for the system density operator  $\rho_S(t)$ :

$$\begin{aligned} \frac{\partial}{\partial t} \rho_S(t) = & -i [H_S + \kappa U, \rho_S(t)] - \frac{\kappa}{2} \int_0^t ds \sum_r \int_{\mathbb{R}^3} d^3k \left\{ iD(\vec{k}, t-s) \right. \\ & \left. \times \left[ J_r(\vec{k}), \left\{ J_r(-\vec{k}, s-t), \rho_S(t) \right\} \right] + D_1(\vec{k}, t-s) \left[ J_r(\vec{k}), \left[ J_r(-\vec{k}, s-t), \rho_S(t) \right] \right] \right\} \end{aligned} \quad (4.52)$$

with the matter system's Hamiltonian operator

$$H_S := \int_{\mathbb{R}^3} d^3k \omega_k a_k^\dagger a_k \quad (4.53)$$

as well as the operator  $U$  for the gravitational self-interaction of the scalar field that was defined in (3.12), the operator-valued distribution  $J_r(\vec{k}, t)$  defined in (3.18) and  $J_r(\vec{k}) = J_r(\vec{k}, 0)$ . The quantities  $D(\vec{k}, t-s)$  and  $D_1(\vec{k}, t-s)$  are the three-dimensional Fourier transforms of the dissipation and noise kernel. They can be related to the commutator's and anticommutator's correlation functions and also to the thermal Wightman functions  $G^>(\vec{x}-\vec{y}, t-s)$  and  $G^<(\vec{x}-\vec{y}, t-s)$ , see also (4.32) and (4.33):

$$\begin{aligned} [\delta\mathcal{E}^a_i(\vec{x}, t), \delta\mathcal{E}^b_j(\vec{y}, s)] &= P^a_i b_j [G^>(\vec{x}-\vec{y}, t-s) - G^<(\vec{x}-\vec{y}, t-s)] \\ &= iP^a_i b_j \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}} e^{i\vec{k}(\vec{x}-\vec{y})} D(\vec{k}, t-s) \end{aligned} \quad (4.54)$$

$$\begin{aligned} \text{tr}_{\mathcal{E}} \left( \left\{ \delta\mathcal{E}^a_i(\vec{x}, t), \delta\mathcal{E}^b_j(\vec{y}, s) \right\} \right) &= P^a_i b_j [G^>(\vec{x}-\vec{y}, t-s) + G^<(\vec{x}-\vec{y}, t-s)] \\ &= P^a_i b_j \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}} e^{i\vec{k}(\vec{x}-\vec{y})} D_1(\vec{k}, t-s), \end{aligned} \quad (4.55)$$

where we used that for the model considered here the commutator in the first equation is proportional to the identity operator and thus can be pulled out of the trace. Hence, we get

$$D(\vec{k}, t-s) := -\frac{\sin(\Omega_k(t-s))}{\Omega_k} \quad (4.56)$$

$$D_1(\vec{k}, t-s) := [2N(\Omega_k) + 1] \frac{\cos(\Omega_k(t-s))}{\Omega_k}. \quad (4.57)$$

To compare our results better with those already existing in the literature like for instance the ones in [35, 37, 38] it is of advantage to rewrite the master equation in (4.52) in two equivalent ways. The first alternative and equivalent form is given by

$$\begin{aligned} \frac{\partial}{\partial t} \rho_S(t) = & -i [H_S + \kappa U, \rho_S(t)] \\ & + \frac{\kappa}{4} \int d^3k \sum_r \left\{ \frac{\coth\left(\frac{\beta\Omega_k}{2}\right)}{\Omega_k} \left[ J_r(\vec{k}), \left[ \rho_S(t), \tilde{\mathcal{J}}_r^\dagger(\vec{k}, t) \right] \right] + \text{h.c.} \right. \\ & \left. + \frac{1}{\Omega_k} \left[ J_r(\vec{k}), \left\{ \rho_S(t), \tilde{\mathcal{J}}_r^\dagger(\vec{k}, t) \right\} \right] + \text{h.c.} \right\}, \end{aligned} \quad (4.58)$$

with

$$\tilde{J}_r(\vec{k}, t) := \int_0^t ds e^{-i\Omega_k(t-s)} J_r(\vec{k}, s-t) \quad (4.59)$$

and where h.c. denotes the hermitian conjugate. This form of master equation is similar to the master equations derived in [35, 38]. While the first reference investigates also a model where a scalar field is coupled to linearised gravity, the second one replaces the scalar field by photons. Let us briefly discuss the similarities and difference of these models and the one considered here: At the classical level one of the main difference is that we formulate the model in terms of Ashtekar variables whereas the models in [35, 38] are based on ADM variables. Thus, we have to deal with an additional Gauß constraint in the model that also needs to be taken into account when gauge invariance is considered. While we presented a way to work with a gauge invariant formulation of the classical theory by means of constructing suitable Dirac observables and also showed that one can construct a canonical transformation on the original phase space that provides a natural separation of the physical and gauge sector of the phase space by the transformation in appendix A.4, in the papers [35, 38] a specific gauge fixing is used to eliminate the gauge degrees of freedom. Furthermore, we consider a standard boundary term [94] that ensures that the action in terms of Ashtekar variables as well as its variation is well defined and we expect that the corresponding ADM boundary term should be included in the formulation of the models in [35, 38] as well. Note that in our case the boundary terms cancels the term in  $\delta\mathbf{H}_{can}$  that is linear in the linearised geometric contribution to the Hamiltonian constraint and the remaining matter part combines with the  $\frac{1}{\kappa}$  in front of the action to the energy density of the scalar field in  $\kappa^0$  order. In contrast in [35, 38] no perturbations of linear order have been considered in the canonical Hamiltonian but the  $\kappa^0$  order is present. Therefore, effects of the missing boundary term cannot be seen in the final master equation when comparing the two results.

For the quantised model let us comment on the work in [35, 38] separately, starting with the latter. One difference in the Hamiltonian operators in our result and the one in [35] is the form of the self-interaction part because we choose a different normal ordering compared to [35]. We chose to normal order the entire self-interaction part of the Hamiltonian operator, whereas in [35] only the individual contributions  $\hat{p}_a$ ,  $\hat{\epsilon}$ ,  $\hat{V}$  and  $\hat{\varphi}_b^a$  were normal ordered. Albeit the current operator valued distribution  $J_r$ , defined in (3.11) is the same as in [35], it appears in their Hamiltonian (3.10) with a different factor of  $\sqrt{2}$  compared to our coupling. This difference has no physical effect because it is absorbed at the level of the master equation for the following reason: The different factor of  $\sqrt{2}$  arises because the Hamiltonian for the pure gravitational part contains an additional factor of  $\frac{1}{2}$  when expressed in Ashtekar variables and working with  $\delta\mathcal{E}$  and  $\delta\mathcal{C}$  compared to the one in ADM variables in [35]. This leads to an additional factor  $\frac{1}{\sqrt{2}}$  in our Fock quantised gravitational variables compared to the ADM variables. In addition in terms of ADM variables the interaction term in the Hamiltonian density reads  $-\frac{\kappa}{2}\delta h_{ab}^{TT}T^{ab}$  while in Ashtekar variables it is given by  $+\kappa\delta\mathcal{E}^a\delta^{bi}T_{ab}$ . This difference can be explained by the relation between ADM and Ashtekar variables. In case of ADM variables the gravitational physical degrees of freedom are given by the transverse-traceless components  $\delta h_{ab}^{TT} := P_{ab}^{cd}\delta h_{cd}$  and the interaction reads  $\delta h_{ab}^{TT}T^{ab}$  which is then quantised in [35] and where the coupling to the gravitational gauge degrees of freedom to the energy momentum tensor is contained in the self-interaction part. In contrast in our case, writing  $\delta h_{ab}$  in terms of the perturbed cotriad, that is  $\delta h_{ab}(E) = -\delta_b^i\delta_{ac}\delta E^c_i - \delta_a^i\delta_{bc}\delta E^c_i + \delta_{ab}\delta_c^i\delta E^c_i$  and then using symmetry of the energy-momentum tensor, we obtain  $\delta h_{ab}(E)T^{ab} \rightarrow [-2\delta_a^i\delta_{bc} + \delta_{ab}\delta_c^i]\delta E^c_i T^{ab}$ , where the gravitational physical degrees of freedom only enter in the first term and we obtain in the linearised theory

just  $-2\delta\mathcal{E}^a_i\delta_b^i T_a^b$ . This difference already present in the classical Hamiltonian has no effect on the final equations of motion because with our convention the Poisson bracket between  $\delta\mathcal{A}_a^i, \delta\mathcal{E}^a_i$  involves an additional factor of  $\frac{1}{2}$  compared to the Poisson brackets used in [35, 38], cancelling the additional factor of 2. The minus sign is also cancelled because in our case the momentum variables couple to the energy momentum tensor whereas in [35, 38] the coupling is via configuration variables. This kind of cancellation carries of course over to the commutator at the quantum level.

A further difference is that the final master equation presented in [35] is of Lindblad form (4.22). This is a significant difference because as we will show in the following, the master equation in (4.52) we obtain is not of Lindblad type yet and it requires further assumptions and approximations respectively to be of this form. For the purpose of discussing this we rewrite the master equation in (4.52) in another equivalent form

$$\begin{aligned} \frac{\partial}{\partial t}\rho_S(t) = & -i[H_S + \kappa U, \rho_S(t)] - \kappa \sum_r \int_{\mathbb{R}^3} \frac{d^3k}{2\Omega_k} \\ & \times \left\{ [J_r^\dagger(\vec{k}), \tilde{J}_r(\vec{k}, t)\rho_S(t)] + N(\Omega_k)[J_r^\dagger(\vec{k}), [\tilde{J}_r(\vec{k}, t), \rho_S(t)]] + \text{h.c.} \right\}. \end{aligned} \quad (4.60)$$

In this form, one can clearly separate the dissipator into a part associated with zero temperature (terms without  $N(\Omega_k)$ ) and a second part corresponding to finite temperature (terms with  $N(\Omega_k)$ ). All three forms of the master equation derived in this paper are time-convolutionless. However, they cannot (yet) be brought into Lindblad form, as there are still time-dependent quantities apart from the reduced density operator itself present in the dissipator. Nevertheless, we can bring the master equation into a form which is similar to the first standard form introduced in (4.22) and which admits the splitting of the dissipator into a Lamb-shift Hamiltonian and a remaining contribution. To this extent, we start from the formulation in (4.60). First, we note that the  $J_r$ -operators can be split in the following way into pairs of creation and / or annihilation operator-valued distributions:

$$J_r(\vec{k}) = - \int \frac{d^3p}{(2\pi)^{\frac{3}{2}}} \sum_{a=1}^4 j_r^a(\vec{k}, \vec{p}) \quad (4.61)$$

$$\tilde{J}_r(\vec{k}, t) = - \int \frac{d^3p}{(2\pi)^{\frac{3}{2}}} \sum_{a=1}^4 j_r^a(\vec{k}, \vec{p}) f(\Omega_k + \omega_a(\vec{k}, \vec{p}); t) \quad (4.62)$$

with

$$j_r^1(\vec{k}, \vec{p}) := a_p^\dagger a_{k+p} \frac{1}{2\sqrt{\omega_p\omega_{k+p}}} \left[ p_a p^b [P^{-r}(\vec{k})]_b^a \right] \quad \omega_1(\vec{k}, \vec{p}) := \omega_p - \omega_{k+p} \quad (4.63)$$

$$j_r^2(\vec{k}, \vec{p}) := a_{-p-k}^\dagger a_{-p} \frac{1}{2\sqrt{\omega_p\omega_{k+p}}} \left[ p_a p^b [P^{-r}(\vec{k})]_b^a \right] \quad \omega_2(\vec{k}, \vec{p}) := \omega_{k+p} - \omega_p \quad (4.64)$$

$$j_r^3(\vec{k}, \vec{p}) := a_{-p} a_{k+p} \frac{1}{2\sqrt{\omega_p\omega_{k+p}}} \left[ p_a p^b [P^{-r}(\vec{k})]_b^a \right] \quad \omega_3(\vec{k}, \vec{p}) := -\omega_p - \omega_{k+p} \quad (4.65)$$

$$j_r^4(\vec{k}, \vec{p}) := a_p^\dagger a_{-k-p}^\dagger \frac{1}{2\sqrt{\omega_p\omega_{k+p}}} \left[ p_a p^b [P^{-r}(\vec{k})]_b^a \right] \quad \omega_4(\vec{k}, \vec{p}) := \omega_p + \omega_{k+p} \quad (4.66)$$

and

$$f(\omega; t) := \int_0^t ds e^{-i\omega(t-s)} = \frac{i}{\omega} (e^{-i\omega t} - 1). \quad (4.67)$$

Using this expansion in (4.60), the dissipator can be rewritten as

$$\mathcal{D}[\rho_S] = \frac{\kappa}{2} \sum_{r,a,b} \int_{\mathbb{R}^3} \frac{d^3k d^3p d^3l}{(2\pi)^{\frac{9}{2}}} \left\{ \Delta_{ab}(\vec{p}, \vec{l}; \vec{k}, t) \left[ j_r^b(\vec{k}, \vec{l}) \rho_S(t), j_r^a(\vec{k}, \vec{p})^\dagger \right] + h.c. \right\} \quad (4.68)$$

with

$$\begin{aligned} \Delta_{ab}(\vec{p}, \vec{l}; \vec{k}, t) &:= \frac{1}{\Omega_k} \left[ (N(\Omega_k) + 1) f(\Omega_k + \omega_b(\vec{k}, \vec{l}); t) + N(\Omega_k) f(-\Omega_k + \omega_b(\vec{k}, \vec{l}); t) \right] \\ &= 2 \int_0^t ds G^>(\vec{k}, t-s) e^{-i\omega_b(\vec{k}, \vec{l})(t-s)}, \end{aligned} \quad (4.69)$$

where  $G^>(\vec{k}, t-s)$  denotes the three dimensional Fourier transform of the Wightman function  $G^>(\vec{x}-\vec{y}, t-s)$ . Combining the new coefficient functions in the following form

$$\begin{aligned} S_{ab}(\vec{p}, \vec{l}; \vec{k}, t) &:= \frac{1}{2i} \left( \Delta_{ab}(\vec{p}, \vec{l}; \vec{k}, t) - \Delta_{ba}^*(\vec{l}, \vec{p}; \vec{k}, t) \right) \\ &= \frac{1}{2\Omega_k} \left[ [N(\Omega_k) + 1] \left\{ \frac{e^{-i(\Omega_k + \omega_b(\vec{k}, \vec{l}))t} - 1}{\Omega_k + \omega_b(\vec{k}, \vec{l})} + \frac{e^{i(\Omega_k + \omega_a(\vec{k}, \vec{p}))t} - 1}{\Omega_k + \omega_a(\vec{k}, \vec{p})} \right\} \right. \\ &\quad \left. - N(\Omega_k) \left\{ \frac{e^{i(\Omega_k - \omega_b(\vec{k}, \vec{l}))t} - 1}{\Omega_k - \omega_b(\vec{k}, \vec{l})} + \frac{e^{-i(\Omega_k - \omega_a(\vec{k}, \vec{p}))t} - 1}{\Omega_k - \omega_a(\vec{k}, \vec{p})} \right\} \right] \end{aligned} \quad (4.70)$$

$$\begin{aligned} R_{ab}(\vec{p}, \vec{l}; \vec{k}, t) &:= \Delta_{ab}(\vec{p}, \vec{l}; \vec{k}, t) + \Delta_{ba}^*(\vec{l}, \vec{p}; \vec{k}, t) \\ &= \frac{i}{\Omega_k} \left[ [N(\Omega_k) + 1] \left\{ \frac{e^{-i(\Omega_k + \omega_b(\vec{k}, \vec{l}))t} - 1}{\Omega_k + \omega_b(\vec{k}, \vec{l})} - \frac{e^{i(\Omega_k + \omega_a(\vec{k}, \vec{p}))t} - 1}{\Omega_k + \omega_a(\vec{k}, \vec{p})} \right\} \right. \\ &\quad \left. - N(\Omega_k) \left\{ \frac{e^{i(\Omega_k - \omega_b(\vec{k}, \vec{l}))t} - 1}{\Omega_k - \omega_b(\vec{k}, \vec{l})} - \frac{e^{-i(\Omega_k - \omega_a(\vec{k}, \vec{p}))t} - 1}{\Omega_k - \omega_a(\vec{k}, \vec{p})} \right\} \right], \end{aligned} \quad (4.71)$$

we can split the dissipator into two parts:

$$\mathcal{D}[\rho_S] = -i\kappa[H_{LS}, \rho_S(t)] + \mathcal{D}_{\text{first}}[\rho_S] \quad (4.72)$$

with the Lamb-shift Hamiltonian operator

$$H_{LS} := \frac{1}{2} \int \frac{d^3k d^3p d^3l}{(2\pi)^{\frac{9}{2}}} \sum_{r,a,b} S_{ab}(\vec{p}, \vec{l}; \vec{k}, t) j_r^a(\vec{k}, \vec{p})^\dagger j_r^b(\vec{k}, \vec{l}) \quad (4.73)$$

and a new dissipator term being in a form similar to the first standard form:

$$\begin{aligned} \mathcal{D}_{\text{first}}[\rho_S] &:= \frac{\kappa}{2} \int \frac{d^3k d^3p d^3l}{(2\pi)^{\frac{9}{2}}} \sum_{r,a,b} R_{ab}(\vec{p}, \vec{l}; \vec{k}, t) \\ &\quad \times \left( j_r^b(\vec{k}, \vec{l}) \rho_S(t) j_r^a(\vec{k}, \vec{p})^\dagger - \frac{1}{2} \left\{ j_r^a(\vec{k}, \vec{p})^\dagger j_r^b(\vec{k}, \vec{l}), \rho_S(t) \right\} \right), \end{aligned} \quad (4.74)$$

where the label ‘first’ refers to the dissipator in first standard form. The difference to the first standard form shown in (4.22) is that the coefficient function  $R_{ab}$  is still time-dependent and thus the corresponding master equation is not of Lindblad type. Often a Lindblad form can be obtained for a given master equation by applying the second Markov approximation, that is by formally sending the upper limit of the integral involved in (4.69) or directly  $t$  in (4.70) and (4.71) to infinity leading to time-independent coefficients. Note that this is *a priori* problematic here as the complex exponentials in (4.70) and (4.71) do not have a well-defined limit for  $t \rightarrow \infty$ . This can be considered in the context of distributions along the lines of the Fourier transform of the step functions, see for instance [113] and this will be analysed together with the remaining integration over the modes in detail in future work in [114]. If we ignore such subtleties for the moment and simply send  $t$  formally to infinity then the master equation from [35] coincides structurally with the one in (4.58) apart from the explicit form of their last term, which is given by  $\{J_r, [\rho_S, \tilde{J}_r^\dagger]\} + h.c.$ . In contrast in our derivation we actually end up with  $[J_r, \{\rho_S, \tilde{J}_r^\dagger\}] + h.c.$ , which yields different terms and from our calculation we do not see a way how their result can be reproduced. In [35] no detailed derivation of this result is presented but the authors cite [12], where however also such terms are not involved in the master equations.

As already mentioned, the master equation in [38] involves a photon field instead of a scalar field and thus the projectors involved are those that project on the physical degrees of freedom of the photon field. Most of the differences have already been pointed out above because the procedure in [38] is quite similar to the one followed in [35]. Compared to [35], more details on the derivation of the Lindblad equation are presented in [38] and it is derived by means of a Born-Markov and a (weak) rotating wave approximation. The form of their master equations after the Born-Markov approximation is structurally the same as the one in (4.58), also the commutator-anticommutator structure agrees.

Note that the master equation in (4.60) looks, apart from the expected differences due to their usage of ADM variables, similar to the one derived in [37], where a Dirac quantisation was carried out and the physical degrees of freedom were identified by imposing the gauge conditions in the quantum theory and specialising to the transverse-traceless degrees of freedom. In a second step, they then derived a master equation using the influence functional approach for a general bosonic field. Their final master equation is a TCL master equation that, similar to the result derived in this paper, does also not exhibit Lindblad form.

The physical investigation of the final master equation is complex, as the equation is in general not completely positive, in contrast to an equation of Lindblad-type. Apart from applying a second Markov approximation in order to arrive at second standard form, it is required to diagonalise the coefficient functions with respect to the labels  $(a, \vec{p})$  and  $(b, \vec{l})$ . In this context an application of a rotating wave approximation is beneficial as has for instance been used in [38]. For the model considered in our work this needs a further detailed analysis because applying the above-proposed additional approximations to cast it into a completely positive form in general one loses some features of the dynamics of the system. Hence, we postpone this kind of detailed analysis into our companion paper [114] where we also will discuss the special case of a one-particle master equations.

## 5. Conclusions

In this article we considered a scalar field coupled to linearised gravity in a Post Minkowski approximation scheme as the classical starting point for a gravitationally induced decoherence model. To enable a later generalisation of the model to the case that the gravitational sector is quantised using techniques from loop quantum gravity we used Ashtekar variables to construct

the model. In contrast to [35–38] we include a suitable boundary term in our action ensuring that the action as well as its variation is well defined. As a preparation for applying a reduced phase space quantisation, the model is formulated at the gauge invariant level by means of suitable Dirac observables. The latter have been constructed using geometrical clocks in the relational formalism. Moreover, we constructed a canonical transformation on the full phase space that separates the gauge and physical degrees of freedom into two sets by introducing a dual observable map in which the role of constraints and clocks is interchanged. That the relevance of physical degrees of freedom and the implementation of gauge invariance plays a pivotal role in decoherence models has for instance been pointed out in [115, 116], showing that if this is not taken care of it can lead to a misleading interpretation of results [36]. The results for the reduced phase space quantisation obtained here extend the results in the literature where so far ADM variables were used and the construction of Dirac observables [35] was discussed only briefly and to the understanding of the authors not in a complete manner. We further considered a second choice of a set of geometrical clocks used in [69] based on the ADM clocks in [74] and showed that the difference in the corresponding physical Hamiltonians only lies in the term encoding the self-interaction of the scalar field. That term, however, was not discussed in detail in [69] where only the coupling of the transverse-traceless variables to matter was considered which seems to our understanding problematic in the non-vacuum case. Compared to former models in the literature we further analysed two different choices of geometrical clocks and discussed differences and similarities for these two choices. In particular for these two choices we obtain that the physical Hamiltonian differs in the self-interaction part of the scalar field only. In addition we also briefly investigated how different choices of the parameters for a given set of geometrical clocks associated with physical spatial and temporal coordinates affects the form of the physical Hamiltonian and its required fall-off behaviour in the linearised theory. We showed that for the choices taken in this work no issues arise as far as the fall-off behaviour of the physical Hamiltonian is concerned and we can confirm the results in [74] using ADM variables that for the choices considered here the final physical Hamiltonian does not depend on the physical spatial and temporal coordinates. The results obtained in [37] using a Dirac quantisation approach in the ADM framework are consistent with our results. The environment is chosen to be the physical sector of the quantum gravitational sector corresponding to the symmetric and transverse-traceless projections of the Dirac observables associated with the linearised Ashtekar variables.

In this model we quantised the physical Hamiltonian using a Fock quantisation which should be understood as a first step towards the aim of formulating gravitationally induced decoherence models inspired by loop quantum gravity. In the quantisation we choose to normal order the entire physical Hamiltonian which differs from the operator orderings chosen in [35, 37, 38]. However, what kind of possible effects these different orderings might have at the level of the master equations is complex to analyse at the current level of the model due to its complexity. We will postpone such an analysis to future work where we will apply further approximations and consider also the special case of one particle [114]. The final master equation was derived in the full relativistic framework by assuming a Gibbs state for the quantum gravitational environment and applying the the projection operator technique, no restriction to the non-relativistic sector was considered, as for instance done in [44, 117–119] during the derivation. We further discuss that for the assumptions the discussed model is based upon, the influence functional approach yields the same result.

The final master equation we end up with is not of Lindblad type and not of first standard form since it still carries time-dependent coefficient functions. Applying the the second Markov approximation for the model considered here is less trivial than in for instance some

of the standard decoherence models in quantum mechanics because in the limit in which we send the upper limit of the temporal integral to infinity, we can no longer work with ordinary functions but need to consider distributions along the lines of [113]. Combined with the fact that the result obtained from the remaining integration over all modes might not have the properties ordinary test functions have, it requires a careful analysis how this limit can be taken in general. Moreover, since properties like the positivity of the density matrix might only be valid for a restricted temporal interval, the second Markov approximation might be problematic from this aspect as well. An example of a model where the second Markov approximation cannot be applied can be found in [56]. The reason here is that the chosen environmental operators are non-dynamical, yielding correlation functions with no dependence on the temporal coordinate. In the model in [56] this results in a set of effective system operators in the final master equation that depend linearly on the temporal coordinate preventing them from applying the second Markov approximation. In contrast, the environmental operators in the model here depend linearly on the densitised triad operators present in the interaction Hamiltonian. This means that the correlation functions contain a combination of complex time dependent exponentials, where the application of the second Markov approximation is not excluded *a priori*. This aspect will be analysed in future work in [114].

Let us ignore those subtleties for the moment and just formally apply the second Markov approximation and compare the resulting master equation to the results in [35]. Here we obtain a slightly different result for the last term in their equation, where the roles of commutators and anticommutators are switched compared to our result and we are not able to reproduce the dissipator in the form presented in [35]. The final master equation in [38] is based on a model in which photons are coupled to linearised gravity. Compared to [35] the derivation of the master equation is presented more in detail involving a Born-Markov as well as a rotating wave approximation. Our results structurally agree with the ones in [38] considering that the projector on the physical degrees of freedom are adapted to a coupling with photons. In this case commutators and anti-commutators enter in the same way as in our result. Our results structurally agree with the master equation derived in [37] where the model is based on Dirac quantisation. In particular this model also ends up with time-dependent coefficients in the master equations that is, as in our case, not of Lindblad type unless further approximations are applied. Because the model presented in this work here uses Ashtekar variables compared to the models in [35–38, 40] it can be easier generalised to models where fermions are coupled whereas in the other cases one has to reformulate the model in terms of tetrads first. As discussed when reviewing the projection operator and influence functional technique, the formulation of the final master equation in terms of thermal Wightman functions can be done in a rather model-independent way and thus provides a good framework comparing and/or generalising a given decoherence model.

Note that in principle there are conceivable differences between the various top-down approaches depending on the nature of the environmental correlation functions. A stochastic environmental noise [42–44] does in general not capture the full complexity of the system-environment interaction due to its fundamentally classical nature [40]. Truncating the TCL master equation at a certain order does however also not guarantee an outcome with a sound physical interpretation for arbitrarily large time scales, positivity violations [108] can warrant the introduction of additional terms in order to restore positivity [12]. In some cases, additional approximations are able to ensure Lindblad form. Hence, it is indispensable to evaluate the features of the master equation for every model individually based on the assumptions and approximations performed in the course of its derivation.

The results obtained in this work build the basis for further investigations and generalisations of gravitationally induced decoherence models along the lines discussed in

this article. The formulation of the model in a gauge invariant manner at the classical level using geometrical clocks can be easily carried over to couplings of matter other than just a scalar field because the linearised constraints will be of the form that they split into a purely geometric and matter contribution which simplified the construction of the canonical transformation and Dirac observables. Because with the analysis of the results obtained here using a Fock quantisation, we understand now in detail how our results compare to those where ADM variables are used. If we consider a loop quantisation for the gravitational sector in future work, we can therefore judge then clearly what kind of differences and effects stem from the fact that we do no longer apply Fock quantisation. Using a loop quantisation we expect that the derivation of a master equation will be more difficult in two aspects. First the transition to the interaction picture will be more complex because the underlying algebra is the one of holonomies and fluxes and secondly this also has the consequence that the commutator of two environmental operators in general will no longer just be almost commuting, that is proportional to the identity operator, which was a crucial property in the derivation of the master equation. Next to understanding the effects of different choices of quantisation, we also aim at investigating more in detail the physical properties of the model. Since this is a rather complicated task due to the still involved time-dependence in the final master equation, we want to consider the one-particle sector in future work where additional approximations can be handled more easily and such a specialisation further allows to make contact to the already existing bottom-up models used for instance in the context of decoherence effects in neutrino oscillations. Following this route we will be able to evaluate whether some of the free parameters involved in such bottom-up models are determined by the model derived in this work. For neutrino oscillations for instance we expect this to be the dependence on the power with which the neutrino energy enters into the contribution relevant for decoherence effects. Since we did not consider a coupling of fermions here we understand such a one-particle model as a first toy model to work with and postpone the derivation of such a model to future work [114].

### Data availability statement

No new data were created or analysed in this study.

### Acknowledgments

M J F and M K both thank the Heinrich-Böll foundation for financial support. M J F and K G thank Thomas Eberl for illuminating discussions on decoherence models in the context of neutrino oscillations. M J F thanks Thomas Thiemann for useful comments on his master's thesis on which part of the work here is based on.

### Appendix A. More details on the construction of Dirac observables

In this appendix we present more details on the construction of the Dirac observables that we use in the main text to formulate the decoherence model at fully gauge invariant level.

#### A.1. More detailed discussion on the choice of reference fields for the linearised constraints

Defining the smeared linearised Gauß, Hamiltonian and spatial diffeomorphism constraints as

$$\delta G[\delta\Lambda](t) := \int_{\sigma} d^3y \delta\Lambda^i(\vec{x}, t) \delta G_i(\vec{x}, t) \quad (\text{A.1})$$

$$\delta C[\delta N](t) := \int_{\sigma} d^3x \delta N(\vec{x}, t) \delta C(\vec{x}, t) \quad (\text{A.2})$$

$$\overrightarrow{\delta C}[\overrightarrow{\delta N}](t) := \int_{\sigma} d^3x \delta N^a(\vec{x}, t) \delta C_a(\vec{x}, t) \quad (\text{A.3})$$

one can evaluate the gauge transformations they infer on the phase space variables. As the Gauß constraint does not contain matter degrees of freedom, it leaves the matter fields unchanged and just modifies the geometrical degrees of freedom:

$$\{\phi(\vec{x}, t), \delta G[\delta\Lambda](t)\} = \{\pi(\vec{x}, t), \delta G[\delta\Lambda](t)\} = 0 \quad (\text{A.4})$$

$$\{\delta E^a_i(\vec{x}, t), \delta G[\delta\Lambda](t)\} = \frac{1}{2} \epsilon^{ijk} \delta\Lambda_j(\vec{x}, t) \delta^a_k \quad (\text{A.5})$$

$$\{\delta A_a^i(\vec{x}, t), \delta G[\delta\Lambda](t)\} = -\frac{1}{2} \partial_a(\delta\Lambda^i(\vec{x}, t)). \quad (\text{A.6})$$

For the Hamiltonian constraint we find the following gauge transformations:

$$\{\phi(\vec{x}, t), \delta C[\delta N](t)\} = \kappa \delta N(\vec{x}, t) \pi(\vec{x}, t) \quad (\text{A.7})$$

$$\{\pi(\vec{x}, t), \delta C[\delta N](t)\} = \kappa \partial_a[\delta N(\vec{x}, t) \partial^a \phi(\vec{x}, t)] - \kappa \delta N(\vec{x}, t) m^2 \phi(\vec{x}, t) \quad (\text{A.8})$$

$$\{\delta E^a_i(\vec{x}, t), \delta C[\delta N](t)\} = -\beta \epsilon^{ijk} \delta^a_j \delta^b_k \partial_b(\delta N(\vec{x}, t)) \quad (\text{A.9})$$

$$\{\delta A_a^i(\vec{x}, t), \delta C[\delta N](t)\} = 0. \quad (\text{A.10})$$

Finally, the gauge transformations induced by the spatial diffeomorphism constraint are

$$\{\phi(\vec{x}, t), \overrightarrow{\delta C}[\overrightarrow{\delta N}](t)\} = \kappa \delta N^a(\vec{x}, t) \partial_a \phi(\vec{x}, t) \quad (\text{A.11})$$

$$\{\pi(\vec{x}, t), \overrightarrow{\delta C}[\overrightarrow{\delta N}](t)\} = \kappa \partial_a[\delta N^a(\vec{x}, t) \pi(\vec{x}, t)] \quad (\text{A.12})$$

$$\{\delta E^a_i(\vec{x}, t), \overrightarrow{\delta C}[\overrightarrow{\delta N}](t)\} = -(\delta_i^b \partial_b(\delta N^a(\vec{x}, t)) - \delta_i^a \partial_b(\delta N^b(\vec{x}, t))) \quad (\text{A.13})$$

$$\{\delta A_a^i(\vec{x}, t), \overrightarrow{\delta C}[\overrightarrow{\delta N}](t)\} = 0. \quad (\text{A.14})$$

*A.1.1. Choice of the reference field for the linearised Gauß constraint.* As the choice of reference fields can be understood as gauge fixing constraints in a corresponding gauge fixed theory, we choose a reference field for the Gauß constraint that implements a Lorentz-like gauge condition analogous to [61] for the connection perturbation, i.e.

$$\partial^a(\delta A_a^i(\vec{x}, t)) = 0. \quad (\text{A.15})$$

Evaluation of the Poisson bracket between this condition and the Gauß constraint (2.36) using the linearised Poisson bracket (2.18) yields

$$\{\partial^a(\delta A_a^i(\vec{x}, t)), \delta G_j(\vec{y}, t)\} = -\frac{1}{2} \delta_j^i \Delta_{\vec{x}} \delta^3(\vec{x} - \vec{y}), \quad (\text{A.16})$$

where  $\Delta_{\vec{x}}$  denotes the Laplacian with respect to  $\vec{x}$ . We will drop the subscript if only one coordinate is involved. In order to have the commutation relation

$$\{\delta\Xi^i(\vec{x}, t), \delta G_j(\vec{y}, t)\} = \frac{1}{\kappa} \delta_j^i \delta^3(\vec{x} - \vec{y}), \quad (\text{A.17})$$

we define the reference fields for the Gauß constraint as

$$\delta\Xi^i(\vec{x}, t) := \frac{2}{\kappa} \partial^a (\delta A_a^i * G^\Delta)(\vec{x}, t) = \frac{2}{\kappa} \int d^3y \partial_{\vec{x}}^a G^\Delta(\vec{x} - \vec{y}) \delta A_a^i(\vec{y}) \quad (\text{A.18})$$

with  $\partial_{\vec{x}}^a := \frac{\partial}{\partial x_a}$ . The abbreviation  $G^\Delta(\vec{x} - \vec{y})$  denotes the Green's function of the Laplacian,

$$G^\Delta(\vec{x} - \vec{y}) = \int \frac{d^3k}{(2\pi)^3} \frac{1}{|\vec{k}|^2} e^{i\vec{k}(\vec{x} - \vec{y})}. \quad (\text{A.19})$$

Due to the fact that the connection remains invariant under the Hamiltonian and spatial diffeomorphism constraint, this reference field also remains invariant under transformations induced by these two constraints.

**A.1.2. Choice of the reference field for the linearised Hamiltonian constraint.** The Hamiltonian constraint leaves the connection invariant and only transforms the part  $\epsilon_a^{cb} \delta_c^i \partial_b \delta E^a_i(\vec{x}, t)$  of the densitised triad in the following manner:

$$\{\epsilon_a^{cb} \delta_c^i \partial_b \delta E^a_i(\vec{x}, t), \delta C(\vec{y}, t)\} = 2\beta \Delta_{\vec{x}} \delta^3(\vec{x} - \vec{y}). \quad (\text{A.20})$$

This suggests to define a reference field

$$\delta\tilde{T}(\vec{x}, t) := -\frac{1}{2\beta\kappa} \epsilon_a^{cb} \delta_c^i \partial_b (\delta E^a_i * G^\Delta)(\vec{x}, t). \quad (\text{A.21})$$

However, in this form the reference field is not invariant under gauge transformations induced by the Gauß constraint:

$$\{\delta\tilde{T}(\vec{x}, t), \delta G_j(\vec{y}, t)\} = -\frac{1}{2\beta\kappa} \delta_j^a \partial_a^\vec{x} G^\Delta(\vec{x} - \vec{y}), \quad (\text{A.22})$$

where  $\partial_a^\vec{x}$  denotes the partial derivative with respect to  $x^a$ . To cure this, we seek to subtract some combination of geometrical phase space variables that transforms precisely the same way as  $\delta\tilde{T}$  under the Gauß constraint and remains invariant under the Hamiltonian constraint. The latter is true for any combination of the connection variables, and it turns out that the trace of the connection solves the problem. Hence a good choice for a reference field corresponding to the Hamiltonian constraint is

$$\delta T(\vec{x}, t) := -\frac{1}{\kappa\beta} \left[ \frac{1}{2} \epsilon_a^{cb} \delta_c^i \partial_b (\delta E^a_i * G^\Delta)(\vec{x}, t) + \delta_i^a (\delta A_a^i * G^\Delta)(\vec{x}, t) \right], \quad (\text{A.23})$$

which also commutes with the spatial diffeomorphism constraint. Additionally, it commutes with the reference field for the Gauß constraint:

$$\{\delta T(\vec{x}, t), \Xi^i(\vec{y}, t)\} = 0. \quad (\text{A.24})$$

**A.1.3. Choice of the reference field for the linearised spatial diffeomorphism constraint.** In a last step we have to find a suitable reference field for the spatial diffeomorphism constraint. The list of requirements for this field is motivated by the application in the main part of this work and we will construct it in several steps. In the end, the final reference field  $\delta T^a(\vec{x}, t)$  should

- consists of linearised elementary gravitational degrees of freedom only
- fulfill  $\{\delta T^a(\vec{x}, t), \delta C_b(\vec{y}, t)\} = \frac{1}{\kappa} \delta_b^a \delta^3(\vec{x} - \vec{y})$ ,
- commute with the remaining linearised constraints and
- commute with the remaining reference fields.

For the first step it is helpful to realise from (A.13) that only the trace and one of the longitudinal parts of the densitised triad will be modified by this constraint in the following manner:

$$\{\delta_c^i \partial^c \delta E^a_i(\vec{x}, t), \delta C_b(\vec{y}, t)\} = (\partial_{\vec{x}}^a \partial_{\vec{y}}^b - \delta_b^a \Delta_{\vec{x}}) \delta^3(\vec{x} - \vec{y}) \quad (\text{A.25})$$

$$\{\delta_a^i \delta E^a_i(\vec{x}, t), \delta C_b(\vec{y}, t)\} = 2 \partial_{\vec{x}}^b \delta^3(\vec{x} - \vec{y}). \quad (\text{A.26})$$

A suitable combination for a reference field that is canonically conjugated to the linearised spatial diffeomorphism constraint is therefore

$$\tilde{\delta T}^a(\vec{x}, t) := \frac{1}{\kappa} \left( \delta_b^a \delta_c^i \partial^c - \frac{1}{2} \delta_b^i \partial^a \right) (\delta E^b_i * G^\Delta)(\vec{x}, t). \quad (\text{A.27})$$

Unfortunately, this combination is not invariant under the linearised Gauß constraint:

$$\{\tilde{\delta T}^a(\vec{x}, t), \delta G_j(\vec{y}, t)\} = -\frac{1}{2\kappa} \epsilon_{cb}{}^a \delta_j^c \partial_{\vec{x}}^b G^\Delta(\vec{x} - \vec{y}). \quad (\text{A.28})$$

To cure this, we can add a suitable form of

$$\partial_a \delta E^a_i(\vec{x}, t), \quad (\text{A.29})$$

which is just a scalar density and therefore invariant under the linearised spatial diffeomorphism constraint. It turns out that

$$\tilde{\delta T}^a(\vec{x}, t) := \frac{1}{\kappa} \left( \delta_b^a \delta_c^i \partial^c - \frac{1}{2} \delta_b^i \partial^a + \delta^{ac} \delta_c^i \partial_b \right) (\delta E^b_i * G^\Delta)(\vec{x}, t) \quad (\text{A.30})$$

is invariant under the linearised Gauß constraint and still canonically conjugated to the linearised spatial diffeomorphism constraint. A quick calculation also shows that it is invariant under the gauge transformations generated by the linearised Hamiltonian constraint. The last bullet point which we required, namely vanishing Poisson bracket with the other reference fields, is not yet fulfilled. To construct an extension of (A.30) that commutes with the remaining reference fields in the lowest order and maintains the additional requirements listed in the beginning, we use the linearised dual observable map introduced in (2.48). The advantage is that we already know that the constraints are abelian up to first order in  $\kappa$ , so this extension will have exactly the same behaviour as  $\tilde{\delta T}^a(\vec{x}, t)$  in the Poisson brackets with the linearised constraints. As the clock in the dual observable map we choose the geometric contribution of the linearised Hamiltonian constraint given by  $\delta C - \kappa \epsilon$  because we want the final clock to depend on the geometrical degrees of freedom only. Since we have  $\{\delta T(x), (\delta C - \kappa \epsilon)(y)\} = \frac{1}{\kappa} \delta(x, y)$  this is a suitable choice. The result of the application of the dual linearised observable map on  $\tilde{\delta T}^a$  yields a suitable reference field corresponding to the spatial diffeomorphism constraint:

$$\begin{aligned} \delta T^a(\vec{x}, t) := & \frac{1}{\kappa} \left( \delta_b^a \delta_c^i \partial^c - \frac{1}{2} \delta_b^i \partial^a + \delta^{ac} \delta_c^i \partial_b \right) (\delta E^b_i * G^\Delta)(\vec{x}, t) \\ & + \frac{2\beta}{\kappa^2} \left[ \frac{1}{2} \delta_b^i \partial^a \partial^b (\delta G_i * G^{\Delta\Delta})(\vec{x}, t) - \delta^{ab} \delta_b^i (\delta G_i * G^\Delta)(\vec{x}, t) \right] \\ & + \frac{1}{2\kappa^2} \partial^a [(\delta C - \kappa \epsilon) * G^{\Delta\Delta}](\vec{x}, t), \end{aligned} \quad (\text{A.31})$$

where  $G^{\Delta\Delta}$  denotes the Green's function of the squared Laplacian, that is

$$G^{\Delta\Delta}(\vec{x}-\vec{y}) = \int \frac{d^3k}{(2\pi)^3} \frac{1}{||\vec{k}||^4} e^{i\vec{k}(\vec{x}-\vec{y})} = \int d^3z G^{\Delta}(\vec{x}-\vec{z}) G^{\Delta}(\vec{y}-\vec{z}). \quad (\text{A.32})$$

This reference field is defined on the entire phase space. On the constraint hypersurface, the additional constraint terms drop and we are left again with  $\delta\tilde{T}^a(\vec{x}, t)$ , which still commutes with the constraints in the desired way. This concludes the construction of the reference fields for the model under consideration.

### A.2. Choice of basis for the calculations in Fourier space

When discussing the physical degrees of freedom and also in the context of gauge fixing and the relation between the fully gauge invariant and gauge fixed model, it turns out to be helpful to work in Fourier space. For these calculations we choose a certain complex basis in Fourier space, which is the same as in [61] consisting of

$$\hat{k}^a := \frac{k^a}{||\vec{k}||} \quad m^a(\vec{k}) \quad \bar{m}^a(\vec{k}), \quad (\text{A.33})$$

where  $\vec{k}$  is the momentum and we understand the plane perpendicular to it as a complex plane spanned by the two unit vectors  $m^a(\vec{k})$  and  $\bar{m}^a(\vec{k})$ , where the bar denotes complex conjugation. This basis is orthonormal in the following manner:

$$\hat{k}^a \hat{k}_a = m^a(\vec{k}) \bar{m}_a(\vec{k}) = \bar{m}^a(\vec{k}) m_a(\vec{k}) = 1 \quad (\text{A.34})$$

$$\hat{k}^a m_a(\vec{k}) = \hat{k}^a \bar{m}_a(\vec{k}) = m^a(\vec{k}) m_a(\vec{k}) = \bar{m}^a(\vec{k}) \bar{m}_a(\vec{k}) = 0. \quad (\text{A.35})$$

The expansion of the metric in this basis is

$$q_{ab}(\vec{k}) = \hat{k}_a \hat{k}_b + m_a(\vec{k}) \bar{m}_b(\vec{k}) + \bar{m}_a(\vec{k}) m_b(\vec{k}), \quad (\text{A.36})$$

the soldering form reads

$$\delta_a^i(\vec{k}) = \hat{k}_a \hat{k}^i + m_a(\vec{k}) \bar{m}^i(\vec{k}) + \bar{m}_a(\vec{k}) m^i(\vec{k}), \quad (\text{A.37})$$

and we fix the orientation of the basis vectors in the complex plane relative to each other by imposing

$$\epsilon_{abc} \hat{k}^a m^b(\vec{k}) \bar{m}^c(\vec{k}) = -i. \quad (\text{A.38})$$

We additionally impose (see [120] section 5.3):

$$m^a(-\vec{k}) = \bar{m}^a(\vec{k}). \quad (\text{A.39})$$

We expand the Fourier transform of the gravitational fields in this basis, following the notation in [61]:

$$\begin{aligned} \delta E^a_i(\vec{k}) &= \delta E^+(\vec{k}) m^a(\vec{k}) m_i(\vec{k}) + \delta E^-(\vec{k}) \bar{m}^a(\vec{k}) \bar{m}_i(\vec{k}) + \delta E^1(\vec{k}) \hat{k}^a m_i(\vec{k}) + \delta E^{\bar{1}}(\vec{k}) \hat{k}^a \bar{m}_i(\vec{k}) \\ &+ \delta E^2(\vec{k}) m^a(\vec{k}) \hat{k}_i + \delta E^{\bar{2}}(\vec{k}) \bar{m}^a(\vec{k}) \hat{k}_i + \delta E^3(\vec{k}) \hat{k}^a \hat{k}_i \\ &+ \delta E^4(\vec{k}) m^a(\vec{k}) \bar{m}_i(\vec{k}) + \delta E^5(\vec{k}) \bar{m}^a(\vec{k}) m_i(\vec{k}) \end{aligned} \quad (\text{A.40})$$

and

$$\begin{aligned} \delta A_a^i(\vec{k}) &= \delta A^+(\vec{k}) m_a(\vec{k}) m^i(\vec{k}) + \delta A^-(\vec{k}) \bar{m}_a(\vec{k}) \bar{m}^i(\vec{k}) + \delta A^1(\vec{k}) \hat{k}_a m^i(\vec{k}) + \delta A^{\bar{1}}(\vec{k}) \hat{k}_a \bar{m}^i(\vec{k}) \\ &+ \delta A^2(\vec{k}) m_a(\vec{k}) \hat{k}^i + \delta A^{\bar{2}}(\vec{k}) \bar{m}_a(\vec{k}) \hat{k}^i + \delta A^3(\vec{k}) \hat{k}_a \hat{k}^i \\ &+ \delta A^4(\vec{k}) m_a(\vec{k}) \bar{m}^i(\vec{k}) + \delta A^5(\vec{k}) \bar{m}_a(\vec{k}) m^i(\vec{k}). \end{aligned} \quad (\text{A.41})$$

From this expansion it is evident that  $\delta A^\pm$  and  $\delta E^\pm$  encode the transverse traceless degrees of freedom with corresponding projectors

$$P_{i^j b}^a(\vec{k}) := \bar{m}^a(\vec{k}) \bar{m}_i(\vec{k}) m^j(\vec{k}) m_b(\vec{k}) + m^a(\vec{k}) m_i(\vec{k}) \bar{m}^j(\vec{k}) \bar{m}_b(\vec{k}) \quad (\text{A.42})$$

for  $\delta E$  and  $P_{a^i j}^{i^j b}(\vec{k}) := \delta_{ac} \delta^{il} \delta_{jm} \delta^{bd} P_{l^m d}^c(\vec{k})$  for  $\delta A$ . Their position space version was given in (2.69) and involves highly-nonlocal terms. Hence we set

$$\delta \mathcal{E}^a_i(\vec{k}) := \delta E^+(\vec{k}) m^a(\vec{k}) m_i(\vec{k}) + \delta E^-(\vec{k}) \bar{m}^a(\vec{k}) \bar{m}_i(\vec{k}) \quad (\text{A.43})$$

$$\delta \mathcal{A}_a^i(\vec{k}) := \delta A^+(\vec{k}) m_a(\vec{k}) m^i(\vec{k}) + \delta A^-(\vec{k}) \bar{m}_a(\vec{k}) \bar{m}^i(\vec{k}). \quad (\text{A.44})$$

The Dirac observables constructed in (2.65) and (2.66) can be expressed in terms of this basis up to linear order and read, where we drop the  $\vec{k}$  dependency for better readability and already set  $\delta \xi^j = 0$ :

$$\begin{aligned} (\delta E^a_i)^{Gl} &= \delta \mathcal{E}^a_i + \hat{k}^a m_i \left( \delta E^1 + \frac{1}{\|\vec{k}\|} \delta A^1 \right) + \hat{k}^a \bar{m}_i \left( \delta E^{\bar{1}} - \frac{1}{\|\vec{k}\|} \delta A^{\bar{1}} \right) - m^a \hat{k}_i \frac{1}{\|\vec{k}\|} \delta A^2 \\ &+ \bar{m}^a \hat{k}_i \frac{1}{\|\vec{k}\|} \delta A^{\bar{2}} + \hat{k}^a \hat{k}_i \delta E^3 - m^a \bar{m}_i \frac{2}{\|\vec{k}\|} \delta A^4 + \bar{m}^a m_i \frac{2}{\|\vec{k}\|} \delta A^5 \\ &+ i\kappa (\delta_i^a k_c - \delta_c^a \delta_i^b k_b) \delta \sigma^c - i\beta \kappa \epsilon_c^{ab} \delta_i^c k_b \delta \tau + O(\delta^2, \kappa) \end{aligned} \quad (\text{A.45})$$

$$(\delta A_a^i)^{Gl} = \delta \mathcal{A}_a^i + \delta A^2 m_a \hat{k}^i + \delta A^{\bar{2}} \bar{m}_a \hat{k}^i + \delta A^4 m_a \bar{m}^i + \delta A^5 \bar{m}_a m^i + O(\delta^2, \kappa). \quad (\text{A.46})$$

### A.3. Further details on computing the algebra of Dirac observables

In this appendix we present further details on the algebra of Dirac observables used in the main text.

**A.3.1. Original matter and geometrical variables commute with clocks.** In this appendix we will discuss that the original phase space variables in the matter sector, these are  $(\phi, \pi)$  as well as the transverse-traceless components of the gravitational sector, these are  $(P_{a^i j}^{i^j b}(\delta A_b^j), P_{i^j b}^{a^i j}(\delta E_b^j))$  will commute with all clocks  $\delta T, \delta T^a$  and  $\delta \Xi^j$  chosen for the Hamiltonian, spatial diffeomorphism and Gauß constraint respectively.

First we note that the clocks being geometrical ones do not contain any matter variables and also the additional contributions form the dual observable map involved in the diffeomorphism clock contains the geometric degrees of freedom only. Hence, all clocks trivially commute with the matter fields  $(\phi, \pi)$ . To see that they also commute with the variables  $(P_{a^i j}^{i^j b}(\delta A_b^j), P_{i^j b}^{a^i j}(\delta E_b^j))$ , it is convenient to express them in Fourier space. This can be found in section A.4 and one can see that all clocks are independent of  $\delta A^\pm$  and  $\delta E^\pm$ . As the latter

are the transverse-traceless degrees of freedom and these are elementary phase space variables they have vanishing Poisson brackets with all remaining degrees of freedom and thus also commute with all clocks.

*A.3.2. Computing the algebra of the Dirac observables by means of using the Dirac observable of the corresponding Dirac bracket.* In this appendix we will show how the algebra of the Dirac observables is related to the Dirac observable of the Dirac bracket. Here we will discuss the special case and order in perturbation theory that is relevant for the model under consideration in this work and we refer the reader for the proof in the general case to [67, 101]. In the application where we need this relation we will assume that we consider linearised geometrical clocks  $\delta T^I(x)$  with  $\delta^2 T^I(x) = 0$  that only depend on the gravitational degrees of freedom. We then introduce  $\delta \mathcal{G}^I(x) := \delta T^I - \tau^I(x)$  that mutually commute with a set of constraints  $\mathcal{C}'_J(x)$  up to corrections of order  $\delta^2$  that is

$$M^I_J(x, y) := \{\delta \mathcal{G}^I(x), \delta \mathcal{C}'_J(y)\} = \frac{1}{\kappa} \delta^I_J \delta^{(3)}(x, y) + \mathcal{O}(\delta^2).$$

Given the definition of  $M^I_J(x, y)$  above its inverse is simply given by

$$(M^{-1})^I_J(x, y) := \kappa \delta^I_J \delta^{(3)}(x, y) + \mathcal{O}(\delta^2, \kappa^2)$$

with

$$\int d^3 z M^I_J(x, z) (M^{-1})^J_K(z, y) = \delta^I_K \delta^{(3)}(x, y).$$

Note that we chose to denote the constraints with a prime here because this is the notation used in the main text. This is exactly the setup needed to analyse the algebra of Dirac observables for the model in this work. Since we only perturb the gravitational degrees of freedom but not the matter ones, it is convenient to discuss these two types separately. We will start with the geometric sector and afterwards discuss in detail how the results can be carried over to the algebra of the matter variables. In the framework of field theory the Dirac bracket  $\{\cdot, \cdot\}^*$  reads

$$\begin{aligned} \{f(x), g(y)\}^* &= \{f(x), g(y)\} - \int d^3 z' \int d^3 z'' \{f(x), \mathcal{C}'_L(z')\} (M^{-1})^L_M(z', z'') \{\mathcal{G}^M(z''), g(y)\} \\ &\quad + \int d^3 z' \int d^3 z'' \{g(y), \mathcal{C}'_L(z')\} (M^{-1})^L_M(z', z'') \{\mathcal{G}^M(z''), f(x)\} \\ &= \{f(x), g(y)\} - \int d^3 z \{f(x), \mathcal{C}'_L(z)\} \delta^L_M \{T^M(z), g(y)\} \\ &\quad + \int d^3 z \{g(y), \mathcal{C}'_L(z)\} \delta^L_M \{T^M(z), f(x)\}, \end{aligned} \tag{A.47}$$

where we used the explicit form of  $(M^{-1})^J_K(z, y)$  together with the fact that inside the Poisson bracket we can replace  $\mathcal{G}^M$  by  $T^M$  because they differ only by a phase space independent function. Quantities involving gravitational variables are perturbed in the model in this article. Let us denote their linear and second-order perturbations by  $\delta f$  and  $\delta^2 f$  respectively. Then we consider the perturbation of the Dirac bracket up to linear order that is given by

$$\begin{aligned}
 \{f(x), g(y)\}^* &= \{\delta f(x), \delta g(y)\} - \kappa \int d^3z \{\delta f(x), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\
 &\quad + \kappa \int d^3z \{\delta g(y), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\} \\
 &\quad \times \{\delta f(x), \delta^2 g(y)\} + \{\delta^2 f(x), \delta g(y)\} \\
 &\quad - \kappa \int d^3z \{\delta^2 f(x), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\
 &\quad - \kappa \int d^3z \{\delta f(x), \delta^2 C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\
 &\quad - \kappa \int d^3z \{\delta f(x), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta^2 g(y)\} \\
 &\quad + \kappa \int d^3z \{\delta^2 g(y), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\} \\
 &\quad + \kappa \int d^3z \{\delta g(y), \delta^2 C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\} \\
 &\quad + \kappa \int d^3z \{\delta g(y), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta^2 f(x)\} + O(\delta^2, \kappa^2). \tag{A.48}
 \end{aligned}$$

Now for elementary phase space variables we have  $\delta^2 f = 0$  and  $\delta^2 g = 0$  then the Dirac bracket simplifies to

$$\begin{aligned}
 \{f(x), g(y)\}^* &= \{\delta f(x), \delta g(y)\} - \kappa \int d^3z \{\delta f(x), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\
 &\quad + \kappa \int d^3z \{\delta g(y), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\} \\
 &\quad - \kappa \int d^3z \{\delta f(x), \delta^2 C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\
 &\quad + \kappa \int d^3z \{\delta g(y), \delta^2 C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\} + O(\delta^2, \kappa^2). \tag{A.50}
 \end{aligned}$$

The Dirac observable associated with  $\delta f$  is up to second order given by

$$\begin{aligned}
 O_{\delta f, \{T\}} &= \delta f - \kappa \int d^3z \delta \mathcal{G}^K(z) \{\delta f, \delta C'_K(z)\} - \kappa \int d^3z \delta \mathcal{G}^K(z) \{\delta f, \delta^2 C'_K(z)\} \\
 &\quad - \frac{\kappa^2}{2} \int d^3z \int d^3z' \delta \mathcal{G}^K(z) \delta \mathcal{G}^L(z') \{\{\delta f, \delta^2 C'_L(z)\}, \delta C'_M(z')\} + O(\delta^3, \kappa^2) \\
 &=: O_{\delta f, \{T\}}^{(1)} + O_{\delta f, \{T\}}^{(2)} + O(\delta^3, \kappa^2), \tag{A.51}
 \end{aligned}$$

where, as in the main text, we include all terms of order  $\delta$  and  $\delta^2$  in  $O_{\delta f, \{T\}}^{(1)}$  and  $O_{\delta f, \{T\}}^{(2)}$  respectively. Our aim is to show the following: For linearised quantities  $\delta f, \delta g$  with  $\delta^2 f = 0$  and  $\delta^2 g = 0$  with a set of constraints  $C'_I$  and reference fields  $\mathcal{G}^I$  that satisfy  $\{\mathcal{G}^I(x), C'_J(y)\} = \delta^I_J \delta^{(3)}(x, y) + O(\delta^2, \kappa^2)$  we have

$$\{O_{\delta f, \{T\}}, O_{\delta g, \{T\}}\} = O_{\{\delta f, \delta g\}^*, \{T\}} + O(\delta^2, \kappa^2), \quad (\text{A.52})$$

where the crucial property for our later application is that we have a strong equality here and not a weak one that is present in the general case [67, 101]. For the purpose of showing the equality in (A.53) we introduce the following notation:

$$\begin{aligned} \{f, g\}^*|_{\delta^0} &:= \{\delta f(x), \delta g(y)\} - \kappa \int d^3z \{\delta f(x), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\ &\quad + \kappa \int d^3z \{\delta g(y), \delta C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\} \\ \{f, g\}^*|_{\delta^1} &:= -\kappa \int d^3z \{\delta f(x), \delta^2 C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta g(y)\} \\ &\quad + \kappa \int d^3z \{\delta g(y), \delta^2 C'_L(z)\} \delta_M^L \{\delta T^M(z), \delta f(x)\}, \end{aligned} \quad (\text{A.53})$$

where we again used that  $\delta^2 f = \delta^2 g = 0$  and that we can replace  $\delta \mathcal{G}^M$  by  $\delta T^M$  inside the Poisson brackets. Given this notation we can write  $O_{\{\delta f, \delta g\}^*, \{T\}}$  up to linear order as

$$\begin{aligned} O_{\{\delta f, \delta g\}^*, \{T\}} &= \{f, g\}^*|_{\delta^0} + \{f, g\}^*|_{\delta^1} - \kappa \int d^3z \delta \mathcal{G}^K(z) \{\{f, g\}^*|_{\delta^1}, \delta C'_K(z)\} + O(\delta^2, \kappa^2) \\ &= \{f, g\}^*|_{\delta^0} + \{f, g\}^*|_{\delta^1} - \kappa^2 \int d^3z \int d^3z' \delta \mathcal{G}^L(z) \\ &\quad \times \left( -\delta_M^K \{\delta T^M(z'), \delta g\} \{\{\delta f, \delta^2 C'_K(z')\}, \delta C'_L(z)\} \right. \\ &\quad \left. - \delta_M^K \{\delta T^M(z'), \delta f\} \{\{\delta g, \delta^2 C'_K(z')\}, \delta C'_L(z)\} \right) + O(\delta^2, \kappa^2). \end{aligned} \quad (\text{A.54})$$

Hence, we need to show that  $\{O_{\delta f, \{T\}}^{(1)} + O_{\delta f, \{T\}}^{(2)}, O_{\delta g, \{T\}}^{(1)} + O_{\delta g, \{T\}}^{(2)}\}$  agrees with the right hand side of (A.54). To confirm this we insert the Dirac observables of  $\delta f$  and  $\delta g$  up to second order in (A.51) into the Poisson bracket and collect all terms in the individual orders. In order  $\delta^0$  we obtain

$$\begin{aligned} \delta^0 : \{ \delta f(x), \delta g(y) \} &- \kappa \int d^3z \{\delta f(x), \delta T^L(z)\} \{\delta g(y), \delta C'_L(z)\} \\ &- \kappa \int d^3z \{\delta T^L(z), \delta g(y)\} \{\delta f(x), \delta C'_L(z)\} \\ &= \{f(x), g(y)\}^*|_{\delta^0}, \end{aligned} \quad (\text{A.55})$$

where we used the antisymmetry of the Poisson bracket as well as again used that we can replace  $\delta \mathcal{G}^M$  by  $\delta T^M$  inside a given Poisson bracket. For the linear order in  $\delta$  the result is given by

$$\begin{aligned}
 \delta^1 : & \underbrace{-\kappa \int d^3 z \mathcal{G}^M(z) \{\delta f(x), \{\delta g(y), \delta^2 \mathcal{C}'_M(z)\}\}}_{=:A_1} - \kappa \underbrace{\int d^3 z \{\delta f(x), \delta T^M(z)\} \{\delta g(y), \delta^2 \mathcal{C}'_M(z)\}}_{=:B_1} \\
 & \underbrace{-\kappa \int d^3 z \mathcal{G}^K(z) \{\{\delta f(x), \delta^2 \mathcal{C}'_K(z)\}, \delta g(y)\}}_{=:A_2} - \kappa \underbrace{\int d^3 z \{\delta T^K(z), \delta g(y)\} \{\delta f(x), \delta^2 \mathcal{C}'_K(z)\}}_{=:B_2} \\
 & \underbrace{-\kappa^2 \int d^3 z \int d^3 z' \{\delta f(x), \delta T^M(z)\} \delta \mathcal{G}^N(z') \{\{\delta g(y), \delta^2 \mathcal{C}'_M(z)\}, \delta \mathcal{C}'_N(z')\}}_{=:C_1} \\
 & \underbrace{-\kappa^2 \int d^3 z \int d^3 z' \{\delta T^K(z), \delta g(y)\} \delta \mathcal{G}^L(z') \{\{\delta f(x), \delta^2 \mathcal{C}'_K(z)\}, \delta \mathcal{C}'_L(z')\}}_{=:C_2}, \tag{A.56}
 \end{aligned}$$

where we again used that we can replace  $\delta \mathcal{G}^M$  by  $\delta T^M$  inside a given Poisson bracket.

Now let us discuss the individual contributions separately. First we show that  $A_1 + A_2 = 0$ . Using the Jacobi identity for  $A_1$  we obtain

$$\begin{aligned}
 A_1 &= -\kappa \int d^3 z \mathcal{G}^M(z) \{\delta f(x), \{\delta g(y), \delta^2 \mathcal{C}'_M(z)\}\} \tag{A.57} \\
 &= \kappa \int d^3 z \mathcal{G}^M(z) \left( \{\delta g(y), \{\delta^2 \mathcal{C}'_M(z), \delta f(x)\}\} + \underbrace{\{\delta^2 \mathcal{C}'_M(z), \underbrace{\{\delta f(x), \delta g(y)\}}_{\sim \delta^0}\}}_{=0} \right) \\
 &= \kappa \int d^3 z \mathcal{G}^M(z) \{\delta g(y), \{\delta^2 \mathcal{C}'_M(z), \delta f(x)\}\} \\
 &= \kappa \int d^3 z \mathcal{G}^M(z) \{\{\delta f(x), \delta^2 \mathcal{C}'_M(z)\}, \delta g(y)\} \\
 &= -A_2. \tag{A.58}
 \end{aligned}$$

Comparing  $B_1 + B_2$  with  $\{f, g\}^*|_{\delta^1}$  in (A.53) and using the antisymmetry of the Poisson bracket we realise that we have

$$B_1 + B_2 = \{f, g\}^*|_{\delta^1}. \tag{A.59}$$

Comparing  $C_1 + C_2$  with the last two terms in (A.54) and applying a suitable relabelling of the integration variables it turns out that these two expressions are exactly identical. Thus, collecting all intermediate results we have shown

$$O_{\{\delta f, \delta g\}^*, \{T\}} = A_1 + A_2 + B_1 + B_2 + C_1 + C_2 + O(\delta^2, \kappa^2) = \{O_{\delta f, \{T\}}, O_{\delta g, \{T\}}\} + O(\delta^2, \kappa^2),$$

which is exactly the equality in (A.52) we wanted to prove.

Now let us discuss the case of elementary variables from the matter sector. Because we use geometrical clocks if we consider for  $\delta f$  and  $\delta g$  both matter variables, then since these will commute with all geometrical clocks we immediately have  $\{\delta f, \delta g\}^* = \{\delta f, \delta g\}$  so that we can work with the Poisson bracket instead of Dirac bracket in (A.52) and as discussed in the main text the observable algebra drastically simplifies. The other case is if we consider the algebra

of observables of one geometric and one matter quantity and without loss of the generality let us assume that  $\delta f$  contains geometric and  $\delta g$  matter variables. Then in general the Dirac bracket can differ from the Poisson bracket by those terms where  $\delta f$  is involved in a Poisson bracket with the geometric clocks. In this case the proof just presented carries over if we just replace  $\delta g$  by  $g$  for the following reason: The Poisson bracket  $\{g(y), \delta C'_L(z)\}$  for  $g(y)$  a function of elementary matter variables is of order  $\delta^0$  and hence also contributes in this case to the  $\delta^0$ -order of the Dirac bracket. The second relevant Poisson bracket  $\{g(y), \delta^2 C'_L(z)\}$  is as before of order  $\delta^1$  because in the order of perturbation theory we consider, the contributions to this Poisson bracket come from terms that involve the gravitational perturbation linearly and the matter variables in quadratic order so that the final result will still be of order  $\delta^1$ . Therefore, we can also apply the equality shown in (A.52) to matter variables by simply replacing  $\delta f$  by  $f$  and  $\delta g$  by  $g$  if we restrict  $f, g$  to be elementary phase space variables from the set  $\phi(x), \pi(y)$ . Further because the Poisson brackets of the matter variables do not involve a factor  $\frac{1}{\kappa}$  in contrast to the gravitational variables additional factors  $\kappa$  can arise when the Dirac bracket with matter variables is considered. The application we need in the main text is even a more special case so we will not discuss these terms more in detail here since they will not be needed in any further computation. The main motivation for presenting this proof here in detail was to understand under which assumptions regarding perturbation theory we can ensure that the Dirac bracket agrees with the Poisson bracket and how, if this is not the case, the explicit form of the Dirac observable of the Dirac bracket looks like in perturbation theory.

**A.3.3. Poisson algebra of linearised Dirac observables.** In this appendix we will compute the Poisson algebra of the linearised Dirac observables explicitly and show that the zeroth order contributions are given by the standard canonical Poisson brackets, a result that we use in the main text of this work. In the notation of the main text for all observables  $O_{f,\{T\}} = O_{f,\{T\}}^{(1)} + O_{f,\{T\}}^{(2)} + O(\delta^3, \kappa^3)$  we will neglect the contributions coming from  $O_{f,\{T\}}^{(2)} + O(\delta^3, \kappa^3)$ . This means that the explicit computations presented in this appendix here do only consider the zeroth order result of the corresponding Poisson algebra. This causes no problem since we also know this as well as the linear contribution to the Poisson algebra already from the relation to the Dirac observable of the corresponding Dirac bracket as discussed in the main text. The computation presented here should be rather understood as double checking our results for the explicit form of the Dirac observables that would yield incorrect results for the algebra in zeroth order if they were not computed correctly. For the gauge invariant geometrical degrees of freedom, one can use the following way to obtain their algebra:

$$\{\delta A_a^i(\vec{x}, t), \delta E^b_j(\vec{y}, t)\} = \frac{\beta}{\kappa} \delta_j^i \delta_a^b \delta(\vec{x} - \vec{y}) \implies \{\delta A_a^i(\vec{k}, t), \delta E^b_j(\vec{p}, t)\} = \frac{\beta}{\kappa} \delta_j^i \delta_a^b \delta(\vec{k} + \vec{p}) \quad (\text{A.60})$$

and hence for  $r, u \in \{\pm\}$ :

$$\begin{aligned} \{\delta A^r(\vec{k}, t), \delta E^u(\vec{p}, t)\} &= \bar{m}^a(r\vec{k}) \bar{m}_i(r\vec{k}) \bar{m}_b(u\vec{p}) \bar{m}^j(u\vec{p}) \{\delta A_a^i(\vec{k}, t), \delta E^b_j(\vec{p}, t)\} \\ &= \frac{\beta}{\kappa} \bar{m}^a(r\vec{k}) \bar{m}_i(r\vec{k}) m_a(u\vec{k}) m^b(u\vec{k}) \delta(\vec{k} + \vec{p}) \\ &= \delta^{ru} \frac{\beta}{\kappa} \delta(\vec{k} + \vec{p}). \end{aligned} \quad (\text{A.61})$$

Analogously one can show that  $\{\delta A^r(\vec{k}, t), \delta A^u(\vec{p}, t)\} = \{\delta E^r(\vec{k}, t), \delta E^u(\vec{p}, t)\} = 0$ . From this and the fact that  $O_{\delta A_a^i, \{\delta T\}}^{(1)}(\vec{x}, t) = O_{P\delta A_a^i, \{\delta T\}}^{(1)}(\vec{x}, t) = P O_{\delta A_a^i, \{\delta T\}}^{(1)}(\vec{x}, t) = P A_a^i(\vec{x}, t)$  and similarly for  $\delta E$ , follows that

$$\begin{aligned}
\{O_{\delta\mathcal{A}_a^i, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\delta\mathcal{E}^b, \{\delta T\}}^{(1)}(\vec{y}, t)\} &= \int \frac{d^3k d^3p}{(2\pi)^3} e^{i\vec{k}\vec{x} + i\vec{p}\vec{y}} \underline{P}_{a^i l^c}(\vec{k}) \underline{P}_{j^m d}(\vec{p}) \{\delta A_c^l(\vec{k}, t), \delta E_m^d(\vec{p}, t)\} \\
&= \frac{\beta}{\kappa} \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}(\vec{x}-\vec{y})} \underline{P}_{a^i l^c}(\vec{k}) \underline{P}_{j^l c}(-\vec{k}) \\
&= \frac{\beta}{\kappa} \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}(\vec{x}-\vec{y})} \underline{P}_{a^i j^b}(\vec{k}) \\
&= \frac{\beta}{\kappa} P_{a^i j^b} \delta(\vec{x}-\vec{y}), \tag{A.62}
\end{aligned}$$

where we used in the fourth line that  $\underline{P}_{j^l c}(-\vec{k}) = \underline{P}_{j^l c}(\vec{k}) = P_{c^b j^l}(\vec{k})$ , as well as

$$\{O_{\delta\mathcal{A}_a^i, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\delta\mathcal{A}_a^i, \{\delta T\}}^{(1)}(\vec{y}, t)\} = 0 \tag{A.63}$$

$$\{O_{\delta\mathcal{E}^a, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\delta\mathcal{E}^b, \{\delta T\}}^{(1)}(\vec{y}, t)\} = 0. \tag{A.64}$$

For the matter variables we obtain, dropping terms in  $O(\delta^2, \kappa^2)$  and higher:

$$\{O_{\phi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\phi, \{\delta T\}}^{(1)}(\vec{y}, t)\} = -\kappa^2 (\delta\mathcal{G}^c(\vec{x}, t)) \delta^3(\vec{x}-\vec{y}) + \kappa^2 (\delta\mathcal{G}^c(\vec{y}, t)) \delta^3(\vec{x}-\vec{y}) = 0 \tag{A.65}$$

$$\begin{aligned}
\{O_{\pi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\pi, \{\delta T\}}^{(1)}(\vec{y}, t)\} &= \kappa^2 ((\partial_a^y \delta\mathcal{G}(\vec{y}, t) + \partial_a^x \delta\mathcal{G}(\vec{x}, t)) \partial_y^a \delta^3(\vec{x}-\vec{y}) \\
&\quad + (\delta\mathcal{G}(\vec{y}, t) - \delta\mathcal{G}(\vec{x}, t)) \Delta \delta^3(\vec{x}-\vec{y})) \tag{A.66}
\end{aligned}$$

$$\begin{aligned}
\{O_{\phi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\pi, \{\delta T\}}^{(1)}(\vec{y}, t)\} &= \delta^3(\vec{x}-\vec{y}) - \kappa^2 \partial_a^y [\delta\mathcal{G}^a(\vec{y}, t) \delta^3(\vec{x}-\vec{y})] \\
&\quad - \kappa^2 \delta\mathcal{G}^a(\vec{x}, t) \partial_a^x \delta^3(\vec{x}-\vec{y}). \tag{A.67}
\end{aligned}$$

Due to the presence of derivatives acting on the delta distributions, it is not immediately evident that all additional terms (which are of  $O(\kappa)$  due to the presence of a factor  $\kappa^{-1}$  in the reference fields) vanish. To see this, it is convenient to consider the smeared version of the linearised observables, then it turns out that:

$$\int d^3x \int d^3y f(\vec{x}) g(\vec{y}) \{O_{\pi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\pi, \{\delta T\}}^{(1)}(\vec{y}, t)\} = 0 \tag{A.68}$$

$$\int d^3x \int d^3y f(\vec{x}) g(\vec{y}) \{O_{\phi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\pi, \{\delta T\}}^{(1)}(\vec{y}, t)\} = \int d^3x f(\vec{x}) g(\vec{x}). \tag{A.69}$$

Hence we indeed end up with the desired algebra also for the matter observables:

$$\begin{aligned}
\{O_{\phi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\phi, \{\delta T\}}^{(1)}(\vec{y}, t)\} &= 0 \\
\{O_{\pi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\pi, \{\delta T\}}^{(1)}(\vec{y}, t)\} &= 0 \\
\{O_{\phi, \{\delta T\}}^{(1)}(\vec{x}, t), O_{\pi, \{\delta T\}}^{(1)}(\vec{y}, t)\} &= \delta^3(\vec{x}-\vec{y}). \tag{A.70}
\end{aligned}$$

As  $O_{\delta\mathcal{A}_i, \{\delta T\}}^{(1)}$  and  $O_{\delta\mathcal{E}^i, \{\delta T\}}^{(1)}$  commute with the clocks and do not depend on the original matter fields, all Poisson brackets among the  $(O_{\delta\mathcal{A}, \{\delta T\}}^{(1)}, O_{\delta\mathcal{E}, \{\delta T\}}^{(1)})$  vanish up to  $O(\delta^2, \kappa^2)$  and the  $(O_{\phi, \{\delta T\}}^{(1)}, O_{\pi, \{\delta T\}}^{(1)})$  vanish up to  $O(\kappa^2)$ .

**A.3.4. Comparison with the Dirac observables used by Anastopoulos and Hu.** The model in [35] also considers gauge invariant quantities for the matter variables in the ADM framework and we would like to compare their quantities to the constructed Dirac observables in this work. There, an extension of the basic gauge variant matter variables  $\phi$  and  $\pi$  is proposed that is required to commute with the linearised constraints. These observables take the form

$$\tilde{\phi}(\vec{x}, t) = \phi(\vec{x} - \vec{q}, t - \tau) \quad \tilde{\pi}(\vec{x}, t) = \pi(\vec{x} - \vec{q}, t - \tau) - \partial_a q^a,$$

where  $\vec{q}$  and  $\tau$ , similarly to the geometrical clocks used in our work, consist of purely geometrical quantities in ADM variables. For these as well as the matter fields in [35] the following transformation behaviour under the Hamiltonian and spatial diffeomorphism constraint respectively is used<sup>8</sup>:

$$\begin{array}{ll} \text{Under the Hamilton constraint : } \phi \rightarrow \phi + \lambda \frac{\delta H_0}{\delta \pi} & \pi \rightarrow \pi - \lambda \frac{\delta H_0}{\delta \phi} \\ & \vec{q} \rightarrow \vec{q} \quad \tau \rightarrow \tau + \lambda \\ \text{Under the spatial diffeomorphism constraint : } \phi \rightarrow \phi + \kappa \lambda^a \partial_a \phi & \pi \rightarrow \pi + \kappa \partial_a (\lambda^a \pi) \\ & q^i \rightarrow q^i + \lambda^i \quad \tau \rightarrow \tau, \end{array}$$

where  $H_0 = \int d^3x \epsilon(\vec{x}, t)$ . Starting with the transformations under the Hamilton constraint, which is just  $\epsilon(\vec{x}, t)$  in the matter part, we obtain

$$\left\{ \phi(\vec{x}), \int d^3y \lambda(\vec{y}) \epsilon(\vec{y}) \right\} = \lambda(\vec{x}) \pi(\vec{x}) = \lambda(\vec{x}) \frac{\delta H_0}{\delta \pi}(\vec{x})$$

which hence indeed confirms their proposed transformation of the matter field, its momentum transforms differently reflecting the fact that it is a scalar density of weight one:

$$\begin{aligned} \left\{ \pi(\vec{x}), \int d^3y \lambda(\vec{y}) \epsilon(\vec{y}) \right\} &= \partial_a (\lambda(\vec{x}) \partial^a \phi(\vec{x})) - \lambda(\vec{x}) V'(\phi(\vec{x})) = -\lambda(\vec{x}) \frac{\delta H_0}{\delta \phi}(\vec{x}) \\ &\quad + (\partial_a \lambda(\vec{x})) \partial^a \phi(\vec{x}), \end{aligned}$$

where we assumed that the scalar field obeys a Klein–Gordon equation and that the energy density possesses an arbitrary potential  $V(\phi)$ . The transformation behaviour used in [35] under the Hamiltonian constraint only agrees with the above expression if  $\lambda$  is constant that in general cannot be assumed and hence the second term coming from the Poisson bracket needs to be included in the transformation of  $\pi(\vec{x})$  and to our understanding has been omitted in [35].

Given the correct general transformation behaviours, we want to check whether the Dirac observables used in [35]  $\tilde{\phi}$  and  $\tilde{\pi}$  are indeed gauge invariant. First it is worth noticing that due to the fact that we are working in a linearised theory, to our understanding the quantity  $\tilde{\phi}(\vec{x}, t)$  has to be understood as a Taylor expansion in the modifications truncated after linear order. In general this way of constructing observables looks like the canonical version of the strategy

<sup>8</sup> Note that in [35] the transformation for  $\pi$  is given by  $\pi \rightarrow \pi - \lambda \frac{\delta H_0}{\delta \pi}$  where we expect the  $\frac{\delta}{\delta \pi}$  to be a typo and corrected this accordingly.

followed in [100]. To show this, we will add an additional  $\kappa$  in front of the modifications and also include it into the constraints. In [35] this was done for the spatial diffeomorphism constraint only, as its action on the matter fields indeed yields a correction term of order  $\kappa$ . As shown in the discussion on the Post-Minkowski approximation scheme in 2.3, for consistency such a factor is also required in the Hamiltonian constraint. Therefore, we obtain

$$\begin{aligned}\tilde{\phi}(\vec{x}, t) &= \phi(\vec{x} - \kappa\vec{q}(\vec{x}, t), t - \kappa\tau(\vec{x}, t)) = \phi(\vec{x}, t) - \kappa q^a(\vec{x}, t)\partial_a\phi(\vec{x}, t) - \kappa\tau(\vec{x}, t)\dot{\phi}(\vec{x}, t) \\ \tilde{\pi}(\vec{x}, t) &= \pi(\vec{x} - \kappa\vec{q}(\vec{x}, t), t - \kappa\tau(\vec{x}, t)) - \partial_a q^a(\vec{x}, t) = \pi(\vec{x}, t) - \kappa q^a\partial_a\pi(\vec{x}, t) - \kappa\tau\dot{\pi}(\vec{x}, t) \\ &\quad - \partial_a q^a(\vec{x}, t).\end{aligned}$$

Now under the Hamilton constraint  $\tilde{\phi}$  transforms as:

$$\tilde{\phi}(\vec{x}, t) \rightarrow \tilde{\phi}(\vec{x}, t) + \kappa(\lambda\pi)(\vec{x}, t) - \kappa(\lambda\dot{\phi})(\vec{x}, t) = \tilde{\phi}(\vec{x}, t),$$

where we neglected terms of higher order in  $\kappa$  and see that indeed  $\tilde{\phi}$  remains invariant under the transformations induced by the Hamilton constraint. For  $\tilde{\pi}$  one obtains:

$$\tilde{\pi}(\vec{x}, t) \rightarrow \tilde{\pi}(\vec{x}, t) + \kappa(\lambda\dot{\pi})(\vec{x}, t) - \kappa(\lambda\dot{\pi})(\vec{x}, t) = \tilde{\pi}(\vec{x}, t).$$

Thus, working with the correct transformation behaviour of  $\pi$  under the Hamiltonian constraint, also  $\tilde{\pi}$  remains invariant. For the spatial diffeomorphism constraint a quick calculation shows that its action on the matter fields is indeed given by the terms stated above and in [35]. Under its action we obtain:

$$\begin{aligned}\tilde{\phi} &\rightarrow \tilde{\phi} + \kappa\lambda^a\partial_a\phi - \kappa\lambda^a\partial_a\phi = \tilde{\phi} \\ \tilde{\pi} &\rightarrow \tilde{\pi} - \kappa\lambda^a\partial_a\pi - \partial_a\lambda^a \neq \tilde{\pi}\end{aligned}$$

which demonstrates that the  $\tilde{\pi}$  proposed in [35] is no (linearised) Dirac observable with respect to the spatial diffeomorphism constraint. Thus, actual linearised observables for  $\phi(\vec{x}, t)$ ,  $\pi(\vec{x}, t)$  in [35] would be given by

$$\begin{aligned}\tilde{\phi}(\vec{x}, t) &= \phi(\vec{x}, t) - \kappa q^a(\vec{x}, t)\partial_a\phi(\vec{x}, t) - \kappa\tau(\vec{x}, t)\dot{\phi}(\vec{x}, t) \\ \tilde{\pi}(\vec{x}, t) &= \pi(\vec{x}, t) - \kappa q^a\partial_a(q^a\pi)(\vec{x}, t) - \kappa\tau\dot{\pi}(\vec{x}, t),\end{aligned}$$

where it is crucial that the partial derivative in the second term in  $\tilde{\pi}$  acts on both  $q^a(\vec{x}, t)$  and  $\pi(\vec{x}, t)$ . Compared to our framework, the basic idea is similar to extend the matter fields with geometrical variables to obtain gauge invariant observables. Hence, the quantities  $q^a$  and  $\tau$  play the role of geometrical clocks in our setup. In our approach we in addition require that  $q^a$  and  $\tau$  mutually commute which requires to use an explicit choice of geometrical clocks that satisfies this additional property.

#### A.4. Canonical transformation on the original phase space

In this appendix the aim is to find an expression of the original phase space variables in terms of the new ones which consist of the physical degrees of freedom, the constraints and the clocks.

For this, we first transform the clocks and constraints into momentum space using the basis introduced above in appendix A.2 and end up with the following expressions:

$$\delta \mathcal{G} = \frac{1}{2\beta\kappa|\vec{k}|}(\delta E^5 - \delta E^4) - \frac{1}{\beta\kappa|\vec{k}|^2}(\delta A^3 + \delta A^4 + \delta A^5) - \delta\tau \quad (\text{A.71})$$

$$\delta C = \kappa|\vec{k}|(\delta A^4 - \delta A^5) + \kappa\epsilon \quad (\text{A.72})$$

$$m_a \delta \mathcal{G}^a = \frac{i}{\kappa|\vec{k}|}(\delta E^{\bar{1}} + \delta E^{\bar{2}}) - \frac{2\beta}{\kappa^2|\vec{k}|^2}m^i \delta G_i - m_a \delta \sigma^a \quad (\text{A.73})$$

$$\bar{m}_a \delta \mathcal{G}^a = \frac{i}{\kappa|\vec{k}|}(\delta E^1 + \delta E^2) - \frac{2\beta}{\kappa^2|\vec{k}|^2}\bar{m}^i \delta G_i - \bar{m}_a \delta \sigma^a \quad (\text{A.74})$$

$$\hat{k}_a \delta \mathcal{G}^a = \frac{i}{2\kappa|\vec{k}|}(3\delta E^3 - \delta E^4 - \delta E^5) - \frac{3\beta}{\kappa^2|\vec{k}|^2}\hat{k}^i \delta G_i + \frac{i}{2\kappa|\vec{k}|^3}\left(\frac{1}{\kappa}\delta C - \epsilon\right) - \hat{k}_a \delta \sigma^a \quad (\text{A.75})$$

$$m^a \delta C_a = -\frac{i\kappa|\vec{k}|}{\beta}\delta A^{\bar{2}} + \kappa m^a p_a \quad (\text{A.76})$$

$$\bar{m}^a \delta C_a = -\frac{i\kappa|\vec{k}|}{\beta}\delta A^2 + \kappa \bar{m}^a p_a \quad (\text{A.77})$$

$$\hat{k}^a \delta C_a = \frac{i\kappa|\vec{k}|}{\beta}(\delta A^4 + \delta A^5) + \kappa \hat{k}^a p_a \quad (\text{A.78})$$

$$m^i \delta G_i = \frac{\kappa i}{2\beta}(|\vec{k}|\delta E^{\bar{1}} - \delta A^{\bar{1}} + \delta A^{\bar{2}}) \quad (\text{A.79})$$

$$\bar{m}^i \delta G_i = \frac{\kappa i}{2\beta}(|\vec{k}|\delta E^1 - \delta A^2 + \delta A^1) \quad (\text{A.80})$$

$$\hat{k}^i \delta G_i = \frac{\kappa i}{2\beta}(|\vec{k}|\delta E^3 - \delta A^5 + \delta A^4) \quad (\text{A.81})$$

$$m_i \delta \mathcal{G}^i = \frac{2i}{\kappa|\vec{k}|}\delta A^{\bar{1}} - m_i \delta \xi^i \quad (\text{A.82})$$

$$\bar{m}_i \delta \mathcal{G}^i = \frac{2i}{\kappa|\vec{k}|}\delta A^1 - \bar{m}_i \delta \xi^i \quad (\text{A.83})$$

$$\hat{k}_i \delta \mathcal{G}^i = \frac{2i}{\kappa|\vec{k}|}\delta A^3 - \hat{k}_i \delta \xi^i. \quad (\text{A.84})$$

In total, these are 14 equations that can be solved uniquely for the 14 old variables. The solution then is

$$\delta A^1 = -\frac{i\kappa|\vec{k}|}{2}\bar{m}_i \left[ \delta \mathcal{G}^i + \delta \xi^i \right] \quad (\text{A.85})$$

$$\delta A^{\bar{1}} = -\frac{i\kappa|\vec{k}|}{2}m_i \left[ \delta \mathcal{G}^i + \delta \xi^i \right] \quad (\text{A.86})$$

$$\delta A^2 = \frac{i\beta}{|\vec{k}|}\bar{m}^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) \quad (\text{A.87})$$

$$\delta A^{\bar{2}} = \frac{i\beta}{|\vec{k}|}m^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) \quad (\text{A.88})$$

$$\delta A^3 = -\frac{i\kappa|\vec{k}|}{2}\hat{k}_i \left[ \delta \mathcal{G}^i + \delta \xi^i \right] \quad (\text{A.89})$$

$$\delta A^4 = -\frac{i\beta}{2|\vec{k}|}\hat{k}^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) + \frac{1}{2|\vec{k}|} \left( \frac{1}{\kappa}\delta C - \epsilon \right) \quad (\text{A.90})$$

$$\delta A^5 = -\frac{i\beta}{2|\vec{k}|}\hat{k}^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) - \frac{1}{2|\vec{k}|} \left( \frac{1}{\kappa}\delta C - \epsilon \right) \quad (\text{A.91})$$

$$\delta E^1 = -\frac{2i\beta}{\kappa|\vec{k}|}\bar{m}^i \delta \mathcal{G}_i + \frac{i\beta}{|\vec{k}|^2}\bar{m}^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) + \frac{i\kappa}{2}\bar{m}^i \left[ \delta \mathcal{G}_i + \delta \xi_i \right] \quad (\text{A.92})$$

$$\delta E^{\bar{1}} = -\frac{2i\beta}{\kappa|\vec{k}|}m^i \delta \mathcal{G}_i - \frac{i\beta}{|\vec{k}|^2}m^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) - \frac{i\kappa}{2}m^i \left[ \delta \mathcal{G}_i + \delta \xi_i \right] \quad (\text{A.93})$$

$$\delta E^2 = -i\kappa|\vec{k}|\bar{m}_a \left[ \delta \mathcal{G}^a + \delta \sigma^a \right] - \frac{i\kappa}{2}\bar{m}_i \left[ \delta \mathcal{G}^i + \delta \xi^i \right] - \frac{\beta i}{|\vec{k}|^2}\bar{m}_a \left( \frac{1}{\kappa}\delta C^a - p^a \right) \quad (\text{A.94})$$

$$\delta E^{\bar{2}} = -i\kappa|\vec{k}|m_a \left[ \delta \mathcal{G}^a + \delta \sigma^a \right] + \frac{i\kappa}{2}m_i \left[ \delta \mathcal{G}^i + \delta \xi^i \right] + \frac{\beta i}{|\vec{k}|^2}m_a \left( \frac{1}{\kappa}\delta C^a - p^a \right) \quad (\text{A.95})$$

$$\delta E^3 = -\frac{2i\beta}{\kappa|\vec{k}|}\hat{k}^i \delta \mathcal{G}_i - \frac{1}{|\vec{k}|^2} \left( \frac{1}{\kappa}\delta C - \epsilon \right) \quad (\text{A.96})$$

$$\begin{aligned} \delta E^4 = & -\kappa\beta|\vec{k}| \left[ \delta \mathcal{G} + \delta \tau \right] + i\kappa|\vec{k}|\hat{k}^a \left[ \delta \mathcal{G}_a + \delta \sigma_a \right] - \frac{1}{|\vec{k}|^2} \left( \frac{1}{\kappa}\delta C - \epsilon \right) + \frac{i\kappa}{2}\hat{k}^i \left[ \delta \mathcal{G}_i + \delta \xi^i \right] \\ & + \frac{i\beta}{|\vec{k}|^2}\hat{k}^a \left( \frac{1}{\kappa}\delta C_a - p_a \right) \end{aligned} \quad (\text{A.97})$$

$$\begin{aligned} \delta E^5 = & \kappa\beta|\vec{k}| \left[ \delta \mathcal{G} + \delta \tau \right] + i\kappa|\vec{k}|\hat{k}^a \left[ \delta \mathcal{G}_a + \delta \sigma_a \right] - \frac{1}{|\vec{k}|^2} \left( \frac{1}{\kappa}\delta C - \epsilon \right) - \frac{i\kappa}{2}\hat{k}^i \left[ \delta \mathcal{G}_i + \delta \xi^i \right] \\ & - \frac{i\beta}{|\vec{k}|^2}\hat{k}^a \left( \frac{1}{\kappa}\delta C_a - p_a \right). \end{aligned} \quad (\text{A.98})$$

All the appearing matter degrees of freedom must be substituted by their gauge invariant extensions plus the corresponding correction terms, what can be read off from (2.63) and (2.64), from which we obtain up to first order in  $\kappa$ :

$$\phi(\vec{x}, t) = \phi^{GI}(\delta\sigma^c, \delta\tau, \delta\xi^j) + \kappa^2(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t))\pi(\vec{x}, t) + \kappa^2(\delta T^c(\vec{x}, t) - \delta\sigma^c(\vec{x}, t))\partial_c\phi(\vec{x}, t) \quad (\text{A.99})$$

$$\begin{aligned} \pi(\vec{x}, t) = & \pi^{GI}(\delta\sigma^c, \delta\tau, \delta\xi^j) + \kappa^2\partial_a[(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t))\partial^a\phi(\vec{x}, t)] \\ & + \kappa^2\partial_c[(\delta T^c(\vec{x}, t) - \delta\sigma^c(\vec{x}, t))\pi(\vec{x}, t)] - \kappa^2(\delta T(\vec{x}, t) - \delta\tau(\vec{x}, t))m^2\phi(\vec{x}, t). \end{aligned} \quad (\text{A.100})$$

Note that when considering terms where any of the gravitational variables  $\delta\mathcal{A}$  or  $\delta E$  appear with a prefactor  $\kappa$ , then we can drop all the correction terms of order  $\kappa$  in (A.99) and (A.100) as they would then lead to terms of order  $\kappa^2$ .

#### A.5. Canonical Hamiltonian in terms of the transformed variables

Using the transformation derived in the previous subsection, it is possible to express the canonical Hamiltonian  $\delta\mathbf{H}_{\text{can}}$  in (2.40) in terms of the new phase space variables consisting of the physical gauge invariant degrees of freedom, the constraints and the clocks. In a first step we rewrite the Hamiltonian in the following way:

$$\begin{aligned} \delta\mathbf{H}_{\text{can}} = & \int_{\sigma} d^3x \left[ \epsilon(\phi, \pi) + \delta N^a \delta C_a + \delta N \delta C + \delta \Lambda^i \delta G_i \right. \\ & + \kappa(\partial_a\phi)(\partial^b\phi)\delta E^a{}_i\delta_b^i + \frac{\kappa}{2}m^2\phi^2\delta E^a{}_i\delta_a^i - \frac{\kappa}{2}\delta E^a{}_i\delta_a^i\epsilon(\phi, \pi) + \\ & \left. - \frac{\kappa}{2\beta^2}e^{jkl}\epsilon_{jmn}\delta_k^a\delta_l^b(\delta A_a{}^m\delta A_b{}^n + (\beta^2 + 1)\delta\Gamma_a{}^m\delta\Gamma_b{}^n - 2\delta\Gamma_a{}^m\delta A_b{}^n) \right], \end{aligned} \quad (\text{A.101})$$

where in the first line there is the background matter Hamiltonian and the constraints, in the second line the interaction part and in the last line the second order of the Hamilton constraint. The first line contains the background Hamiltonian of the scalar field, which does not need to be transformed as there are no background constraints and hence it is already gauge invariant, see discussion above. In order to obtain the physical Hamiltonian as described in (2.75) in terms of the new variables, one first has to determine the observables (A.45) and (A.46) in terms of the new variables, where we already set  $\delta\xi^j = 0$  as explained in the main text:

$$\begin{aligned} (\delta E^a{}_i)^{GI} = & \delta\mathcal{E}^a{}_i - \frac{2i\beta}{\kappa|\vec{k}|}\hat{k}^a\delta\mathcal{G}_i + \frac{\beta}{\kappa|\vec{k}|^2}\epsilon^{ba}{}_i\left[\delta\mathcal{C}_b - \kappa p_b\right] - \frac{1}{\kappa|\vec{k}|^2}\delta_i^a\left[\delta\mathcal{C} - \kappa\epsilon\right] \\ & - i\kappa|\vec{k}|\delta\sigma^b\left(\delta_b^a\hat{k}_i - \delta_i^a\hat{k}_b\right) - \kappa|\vec{k}|\beta\delta\tau i\hat{k}_c\epsilon_i{}^{ac} + \mathcal{O}(\delta^3, \kappa^2) \end{aligned} \quad (\text{A.102})$$

$$\begin{aligned} (\delta A_a{}^i)^{GI} = & \delta\mathcal{A}_a{}^i + \frac{i\beta}{2\kappa|\vec{k}|}(2\hat{k}_i\delta^{ab} - \hat{k}^b\delta_i^a - \hat{k}^i\hat{k}_a\hat{k}^b)\left[\delta\mathcal{C}_b - \kappa p_b\right] - \frac{1}{2\kappa|\vec{k}|}i\epsilon_{ba}{}^i\hat{k}^b\left[\delta\mathcal{C} - \kappa\epsilon\right] \\ & + \mathcal{O}(\delta^3, \kappa^2). \end{aligned} \quad (\text{A.103})$$

The interaction part in (2.75),  $O_{\mathcal{H}_i, \{T\}}$ , then assumes the following form:

$$\begin{aligned}
O_{\mathcal{H}_i, \{T\}} &= \kappa \int d^3k \varphi_a^b \delta \mathcal{E}'^a \delta_b^i + \kappa \int d^3k i k_b \delta \sigma'^a \left[ \varphi_a^b - \delta_a^b \left( \varphi_c^c + m^2 \tilde{V} - \epsilon \right) \right] \\
&+ \int \frac{d^3k}{\|\vec{k}\|^2} \left( - \left[ \varphi_a^a + \frac{3}{2} m^2 \tilde{V} - \frac{3}{2} \epsilon \right] \left( \delta C' - \kappa \epsilon' \right) + 2i \beta k^a \delta G'_i \right. \\
&\quad \left. \times \left[ \varphi_a^b \delta_b^i + \left( \frac{1}{2} m^2 \tilde{V} - \frac{1}{2} \epsilon \right) \delta_a^i \right] \right) + O(\delta^3, \kappa^2), \tag{A.104}
\end{aligned}$$

where  $\epsilon = \epsilon(\phi^{GI}, \pi^{GI})$  and  $p_b = p_b(\phi^{GI}, \pi^{GI})$ . We used here a notation where no argument means dependency on  $(\vec{k}, t)$  and a prime after a quantity means dependency on  $(-\vec{k}, t)$ . The last line in (A.101) yields  $O_{\mathcal{H}_{\text{geo}}, \{T\}}$ :

$$\begin{aligned}
O_{\mathcal{H}_{\text{geo}}, \{T\}} &= \frac{\kappa}{2\beta^2} \int d^3k \sum_{r \in \{\pm\}} \left[ \delta A^r \delta A'^r + \|\vec{k}\|^2 (\beta^2 + 1) \delta E^r \delta E'^r + 2r \|\vec{k}\| \delta E^r \delta A'^r \right] \\
&- \kappa \int \frac{d^3k}{\|\vec{k}\|^2} \left\{ - \left( \frac{1}{\kappa} \delta C_a - p_a \right) \left( \frac{1}{\kappa} \delta C'_b - p'_b \right) \left( \delta^{ab} - \frac{1}{4} \hat{k}^a \hat{k}^b \right) \right. \\
&\quad \left. + \frac{1}{4} \left( \frac{1}{\kappa} \delta C - \epsilon \right) \left( \frac{1}{\kappa} \delta C' - \epsilon' \right) \right\} \\
&- \kappa \int d^3k \delta G_i \left[ \frac{\beta^2 + 1}{\kappa^2} \hat{k}^i \hat{k}^j \delta G'_j + \frac{i\beta}{\kappa \|\vec{k}\|} \hat{k}^i \left( \frac{1}{\kappa} \delta C' - \epsilon' \right) \right. \\
&\quad \left. - \frac{2i}{\kappa \|\vec{k}\|} \epsilon_{abi} \hat{k}^a \left( \frac{1}{\kappa} \delta C'_b - p'_b \right) \right] \\
&- \kappa \int d^3k \left[ i \kappa \|\vec{k}\| \delta \tau \hat{k}^b \left( \frac{1}{\kappa} \delta C'_b - p'_b \right) \right] + O(\delta^3, \kappa^2). \tag{A.105}
\end{aligned}$$

As discussed in the main text in section 2.4 in (2.75) the physical Hamiltonian  $\delta \mathbf{H}$  consists in zeroth order of the Hamiltonian of the scalar field on Minkowski spacetime plus the Dirac observable associated with  $\delta^2 C^{\text{geo}}$  as well as the Dirac observable corresponding to the

interaction Hamiltonian  $\mathcal{H}_I$ . In the following equations these contributions have been marked in blue:

$$\begin{aligned}
\delta\mathbf{H}_{\text{can}} = & \int_{\mathbb{R}^3} d^3x \epsilon(\vec{x}, t) + \kappa \int_{\mathbb{R}^3} d^3k \frac{1}{2\beta^2} \sum_{r \in \{\pm\}} \left( \delta A^r(\vec{k}, t) \delta A^r(-\vec{k}, t) + 2r \|\vec{k}\| \delta E^r(\vec{k}, t) \delta A^r(-\vec{k}, t) \right. \\
& \left. + (\beta^2 + 1) \|\vec{k}\|^2 \delta E^r(\vec{k}, t) \delta E^r(-\vec{k}, t) \right) \\
& + \kappa \int_{\mathbb{R}^3} d^3k \delta_b^i \varphi_a^b(\vec{k}, t) \delta \mathcal{E}^a_i(-\vec{k}, t) \\
& + \kappa \int d^3k i k_b \kappa \delta \sigma'^a \left[ \varphi_a^b - \delta_a^b \left( \varphi_c^c + m^2 \tilde{V} - \epsilon \right) \right] + \kappa \int d^3k \left[ i \kappa \|\vec{k}\| \delta \tau \hat{k}^b p'_b \right] \\
& - \kappa \int_{\mathbb{R}^3} \frac{d^3k}{\|\vec{k}\|^2} \left( - p_c(\vec{k}, t) p_d(-\vec{k}, t) \left[ \delta^{cd} - \frac{1}{4} \hat{k}^c \hat{k}^d \right] + \epsilon(\vec{k}, t) \epsilon(-\vec{k}, t) \left[ \frac{1}{4} + \frac{3}{2} \right] \right. \\
& \quad \left. - \frac{3}{2} m^2 \tilde{V}(\vec{k}, t) \epsilon(-\vec{k}, t) - \varphi_a^a(\vec{k}, t) \epsilon(-\vec{k}, t) \right) \\
& + \int d^3k \delta G_i \left( \delta \Lambda'^i - 2i\beta k^a \frac{1}{\|\vec{k}\|^2} \left[ \varphi'^b_a \delta_b^i + \frac{1}{2} (m^2 \tilde{V}' - \epsilon') \delta_a^i \right] - \frac{\beta^2 + 1}{\kappa^2} \hat{k}^i \hat{k}^j \delta G'_j \right. \\
& \quad \left. - \frac{i\beta}{\kappa \|\vec{k}\|} \hat{k}^i \left( \frac{1}{\kappa} \delta C' - \epsilon' \right) + \frac{2i}{\|\vec{k}\|} \epsilon_{bia} \hat{k}^a \left( \frac{1}{\kappa} \delta C'_b - p'_b \right) \right) \\
& + \int d^3k \delta C_a \left( \delta N'^a + \frac{1}{\|\vec{k}\|^2} \left( \frac{1}{\kappa} \delta C'_b - 2p'_b \right) \left( \delta^{ab} - \frac{1}{4} \hat{k}^a \hat{k}^b \right) \right) \\
& + \int d^3k \delta C \left( \delta N' - \frac{1}{\|\vec{k}\|^2} \left[ \varphi'^a_a + \frac{3}{2} m^2 \tilde{V}' - \frac{3}{2} \epsilon' \right] - \frac{1}{4\|\vec{k}\|^2} \left( \frac{1}{\kappa} \delta C' - 2\epsilon' \right) \right) \\
& - \kappa \int d^3k \left[ i \|\vec{k}\| \delta \tau \hat{k}^b \delta C'_b \right] + O(\delta^3, \kappa^2) \tag{A.106}
\end{aligned}$$

with

$$\begin{aligned}
\delta\mathbf{H} = & \int_{\mathbb{R}^3} d^3x \epsilon(\vec{x}, t) + \kappa \int_{\mathbb{R}^3} d^3k \frac{1}{2\beta^2} \sum_{r \in \{\pm\}} \left( \delta A^r(\vec{k}, t) \delta A^r(-\vec{k}, t) + 2r \|\vec{k}\| \delta E^r(\vec{k}, t) \delta A^r(-\vec{k}, t) \right. \\
& \left. + (\beta^2 + 1) \|\vec{k}\|^2 \delta E^r(\vec{k}, t) \delta E^r(-\vec{k}, t) \right) \\
& + \kappa \int_{\mathbb{R}^3} d^3k \delta_b^i \varphi_a^b(\vec{k}, t) \delta \mathcal{E}^a_i(-\vec{k}, t) \\
& + \kappa \int d^3k i k_b \kappa \delta \sigma'^a \left[ \varphi_a^b - \delta_a^b \left( \varphi_c^c + m^2 \tilde{V} - \epsilon \right) \right] + \kappa \int d^3k \left[ i \kappa \|\vec{k}\| \delta \tau \hat{k}^b p'_b \right] \\
& - \kappa \int_{\mathbb{R}^3} \frac{d^3k}{\|\vec{k}\|^2} \left( - p_c(\vec{k}, t) p_d(-\vec{k}, t) \left[ \delta^{cd} - \frac{1}{4} \hat{k}^c \hat{k}^d \right] + \epsilon(\vec{k}, t) \epsilon(-\vec{k}, t) \left[ \frac{1}{4} + \frac{3}{2} \right] \right. \\
& \quad \left. - \frac{3}{2} m^2 \tilde{V}(\vec{k}, t) \epsilon(-\vec{k}, t) - \varphi_a^a(\vec{k}, t) \epsilon(-\vec{k}, t) \right) + O(\delta^3, \kappa^2). \tag{A.107}
\end{aligned}$$

As expected the physical Hamiltonian differs from  $\delta\mathbf{H}_{\text{can}}$  only by terms that involve the linearised constraints at least linearly and  $\delta\mathbf{H}$  agrees exactly with the expression for the physical

Hamiltonian shown in (2.75) in the main text. As discussed in the main text in section 2.4 below (2.67) as well as (2.76) the fall-off behaviour of the physical Hamiltonian is suitable for the choices of  $\delta\tau$  and  $\delta\sigma^c$  in (2.67) where we also discuss the fall-off behaviour of the gauge invariant variables for these choices.

## Appendix B. Two possible choices for a gauge fixing

### B.1. Gauge fixing following from the clocks introduced in section 2.4

In this appendix we present a possible gauge fixing that naturally arises from the choice of the clocks introduced in section 2.4 in the context of constructing Dirac observables. The induced gauge fixing is obtained by setting the gauge fixing conditions  $\mathcal{G}^I$  shown in (2.74) in section 2.4 equal to zero. Evaluating the constraints in momentum space, using the basis introduced in appendix A.2, yields the following seven relations among the elementary phase space variables at every point  $(\vec{k}, t)$ :

$$\|\vec{k}\|\delta\mathcal{E}^{\bar{1}} = \delta\mathcal{A}^{\bar{1}} - \delta\mathcal{A}^{\bar{2}} \quad i\|\vec{k}\|(\delta\mathcal{A}^4 + \delta\mathcal{A}^5) = -\beta\hat{k}^a p_a \quad (\text{B.1})$$

$$\|\vec{k}\|\delta\mathcal{E}^1 = \delta\mathcal{A}^2 - \delta\mathcal{A}^1 \quad i\|\vec{k}\|\delta\mathcal{A}^{\bar{2}} = \beta m^a(\vec{k}) p_a \quad (\text{B.2})$$

$$\|\vec{k}\|\delta\mathcal{E}^3 = \delta\mathcal{A}^5 - \delta\mathcal{A}^4 \quad i\|\vec{k}\|\delta\mathcal{A}^2 = \beta \bar{m}^a(\vec{k}) p_a \quad (\text{B.3})$$

$$\|\vec{k}\|(\delta\mathcal{A}^4 - \delta\mathcal{A}^5) = -\epsilon, \quad (\text{B.4})$$

where  $\epsilon$  and  $p_a$  denote the three-dimensional Fourier transforms of the energy and momentum density of the scalar field. The gauge fixing conditions, obtained by requiring  $\mathcal{G}^I$  to vanish and in addition choosing  $\tau(t, \vec{x}) = t, \sigma^a = x_\sigma^a = \text{const}$  and  $\xi^j = 0$ , hence  $\delta\tau = \frac{t}{\kappa}, \delta\sigma^a = \frac{x_\sigma^a}{\kappa} = \text{const}$ , give seven additional conditions in momentum space for every  $\vec{k}$  and  $t$ :

$$\delta\mathcal{A}^3 = 0 \quad \delta\mathcal{E}^2 = -\delta\mathcal{E}^1 \quad (\text{B.5})$$

$$\delta\mathcal{A}^1 = 0 \quad \delta\mathcal{E}^{\bar{2}} = -\delta\mathcal{E}^{\bar{1}} \quad (\text{B.6})$$

$$\delta\mathcal{A}^{\bar{1}} = 0 \quad 3\delta\mathcal{E}^3 - \delta\mathcal{E}^4 - \delta\mathcal{E}^5 = \frac{1}{\|\vec{k}\|^2} \epsilon \quad (\text{B.7})$$

$$\delta\mathcal{E}^5 - \delta\mathcal{E}^4 = \frac{2}{\|\vec{k}\|} (\delta\mathcal{A}^3 + \delta\mathcal{A}^4 + \delta\mathcal{A}^5). \quad (\text{B.8})$$

Note that  $\delta\tau$  or  $\delta\sigma$  were chosen to be constant in position and hence vanish in all appearing combinations in the gauge fixing conditions. Substituting these results into (A.40) and (A.41) yields an expression for  $\delta\mathcal{E}$  and  $\delta\mathcal{A}$  in terms of the physical degrees of freedom:

$$\begin{aligned} \delta\mathcal{E}^a_i &= \delta\mathcal{E}^+ m^a m_i + \delta\mathcal{E}^- \bar{m}^a \bar{m}_i + \frac{\epsilon}{\|\vec{k}\|^2} \delta_i^a + \frac{\beta i}{\|\vec{k}\|^2} p_c \\ &\quad \times \left[ m^c (\hat{k}^a \bar{m}_i - \bar{m}^a \hat{k}_i) - \bar{m}^c (\hat{k}^a m_i - m^a \hat{k}_i) - \hat{k}^c (m^a \bar{m}_i - \bar{m}^a m_i) \right] \\ &= \delta\mathcal{E}^+ m^a m_i + \delta\mathcal{E}^- \bar{m}^a \bar{m}_i + \frac{\epsilon}{\|\vec{k}\|^2} \delta_i^a - \frac{\beta}{\|\vec{k}\|^2} p_c \epsilon^{ca}_i \end{aligned} \quad (\text{B.9})$$

$$\begin{aligned}
\delta A_a^i &= \delta A^+ m_a m^i + \delta A^- \bar{m}_a \bar{m}^i - \frac{\beta i}{\|\vec{k}\|} p_c \left[ \bar{m}^c m_a \hat{k}^i + m^c \bar{m}_a \hat{k}^i - \frac{1}{2} \hat{k}^c (m_a \bar{m}^i + \bar{m}_a m^i) \right] \\
&\quad + \frac{\epsilon}{2\|\vec{k}\|} (\bar{m}_a m^i - m_a \bar{m}^i) \\
&= \delta A^+ m_a m^i + \delta A^- \bar{m}_a \bar{m}^i - \frac{\beta i}{\|\vec{k}\|} p_c \left[ \hat{k}^i \delta_a^c - \frac{1}{2} \hat{k}^c \delta_a^i - \frac{1}{2} \hat{k}^c \hat{k}_a \hat{k}^i \right] + \frac{i\epsilon}{2\|\vec{k}\|} \epsilon_{ba}^i \hat{k}^b. \quad (\text{B.10})
\end{aligned}$$

Note that one can also obtain setting the constraints equal to zero in the expressions shown in (A.102) and (A.103). The independent physical degrees of freedom are encoded in the matter fields  $(\phi, \pi)$  as well as the symmetric transverse-traceless parts of the connection and densitised triads that we denote as

$$\delta \mathcal{A}_a^i(\vec{k}, t) := \delta A^+(\vec{k}, t) m_a(\vec{k}) m^i(\vec{k}) + \delta A^-(\vec{k}, t) \bar{m}_a(\vec{k}) \bar{m}^i(\vec{k}) = \underline{P}_a^{i,j}(\vec{k}) \delta A_b^j(\vec{k}, t) \quad (\text{B.11})$$

$$\delta \mathcal{E}^a_i(\vec{k}, t) := \delta E^+(\vec{k}, t) m^a(\vec{k}) m_i(\vec{k}) + \delta E^-(\vec{k}, t) \bar{m}^a(\vec{k}) \bar{m}_i(\vec{k}) = \underline{P}^a_j{}^i(\vec{k}) \delta E^b_j(\vec{k}, t) \quad (\text{B.12})$$

with the projectors  $\underline{P}^a_j{}^i(\vec{k})$  and  $\underline{P}_a^{i,j}(\vec{k})$  onto the symmetric transverse-traceless parts that were defined in (A.42). As expected, with the components of  $\delta \mathcal{A}_a^i, \delta \mathcal{E}^a_i(\vec{k}, t)$  we end up with the six physical components given by

$$\delta E^+, \delta E^-, \delta A^+, \delta A^-, \phi, \pi. \quad (\text{B.13})$$

Without the matter fields, one gets the well-known four phase space field degrees of freedom of gravitational waves in vacuum, corresponding to two field degrees of freedom in the Lagrangian framework, and the real scalar matter field leads to two additional phase space field degrees of freedom. Reinserting (B.9) and (B.10) into the Hamiltonian (2.40) yields the total Hamiltonian of the linearised theory on the reduced phase space, (2.76). For this, we defined the transverse-traceless projectors

$$[P^\pm(\vec{k})]_a^b := \bar{m}_a(\pm\vec{k}) \bar{m}^b(\pm\vec{k}), \quad (\text{B.14})$$

such that  $[P^\pm(\vec{k})]_a^b \delta_b^i \delta E^a_i = \delta E^\pm$ ,  $[P^\pm(\vec{k})]_b^a \delta_a^i \delta A^i = \delta A^\pm$  and  $\underline{P}^a_j{}^i(\vec{k}) = \sum_{r \in \{\pm\}} [P^r(\vec{k})]_i^a [P^{-r}(\vec{k})]_b^j$ .

## B.2. A further gauge fixing often used for ADM variables introduced in [74]

At this point, we would like to make a short comparison to the clocks introduced in [69] based on seminal work in [74, 102]. Likewise to [69] we denote this set of clocks as ADM clocks. These clocks, similarly to the ones we use in this work, also do not contain the symmetric transverse-traceless gravitational degrees of freedom  $\delta A^\pm$  and  $\delta E^\pm$ . Hence, these four phase space degrees of freedom are again identified as the physical degrees of freedom in the gravitational sector, as in our case. However, the gauge fixing induced by setting the corresponding gauge fixing conditions  $\delta \mathcal{G}^I = \delta T^I - \tau^I$  used in [69] to zero is different from the one used by in section 2.4 in this work here, hence this set of clocks is not equivalent to the one discussed in section 2.4. This can be easily seen by considering the clock  ${}^G T^i$  associated with the Gauß constraint from [69] in momentum space. Requiring that the components  $m_i(\vec{k}) {}^G T^i(\vec{k})$  and  $\bar{m}_i(\vec{k}) {}^G T^i(\vec{k})$  vanish yields  $\delta E^I = \delta E^{\bar{I}} = 0$ .

In contrast the gauge fixing obtained by setting the constraints as well as gauge fixing conditions  $\delta\mathcal{G}^I$  used in section 2.4 equal to zero, we can read off from the transformation introduced in appendix A.4 that  $\delta\mathcal{E}^1 = -\frac{i\beta}{||\vec{k}||^2}\overline{m}^a p_a$  and  $\delta\mathcal{E}^{\overline{1}} = \frac{i\beta}{||\vec{k}||^2}m^a p_a$ . This gauge fixing condition would be equivalent to the Lorenz-like condition for the Gaußconstraint chosen in this work in section 2.4 in 2.49 only for the vacuum case, as can be seen in [61].<sup>9</sup> Setting the gauge fixing conditions  $\mathcal{G}_{ADM}^I$  induced by the ADM clocks used in [69] equal to zero imposes the following set of conditions for the phase space variables:

$$\delta\mathcal{E}^1 = \delta\mathcal{E}^{\overline{1}} = \delta\mathcal{E}^2 = \delta\mathcal{E}^{\overline{2}} = 2\delta\mathcal{A}^3 + \delta\mathcal{A}^4 + \delta\mathcal{A}^5 = \delta\mathcal{E}^5 - \delta\mathcal{E}^4 = \delta\mathcal{E}^3 - \delta\mathcal{E}^4 - \delta\mathcal{E}^5 = 0, \quad (\text{B.15})$$

which implies on the constraint hypersurface that

$$\delta\mathcal{E}^a{}_i = \delta\mathcal{E}^a{}_i + \frac{\epsilon}{2||\vec{k}||^2}(\delta_i^a + \hat{k}^a \hat{k}_i) \quad (\text{B.16})$$

$$\delta\mathcal{A}_a{}^i = \delta\mathcal{A}_a{}^i - \frac{i\beta}{||\vec{k}||} p_b \left( \hat{k}^j \delta_a^b + \hat{k}_a \delta^{ib} - \frac{1}{2} \hat{k}^b \delta_a^i - \hat{k}_a \hat{k}^i \hat{k}^b \right) + \frac{i\epsilon}{2||\vec{k}||} \hat{k}^b \epsilon_{ba}{}^i. \quad (\text{B.17})$$

In this particular gauge, the physical Hamiltonian becomes

$$\begin{aligned} \delta\mathbf{H} = & \int_{\mathbb{R}^3} d^3x \epsilon(\vec{x}, t) + \kappa \int_{\mathbb{R}^3} d^3k \frac{1}{2\beta^2} \sum_{r \in \{\pm\}} \left( \delta\mathcal{A}^r(\vec{k}, t) \delta\mathcal{A}^r(-\vec{k}, t) + 2r||\vec{k}|| \delta\mathcal{E}^r(\vec{k}, t) \delta\mathcal{A}^r(-\vec{k}, t) \right. \\ & \left. + (\beta^2 + 1)||\vec{k}||^2 \delta\mathcal{E}^r(\vec{k}, t) \delta\mathcal{E}^r(-\vec{k}, t) \right) \\ & + \kappa \int_{\mathbb{R}^3} d^3k \delta_b^i \varphi_a^b(\vec{k}, t) \delta\mathcal{E}^a{}_i(-\vec{k}, t) \\ & - \kappa \int_{\mathbb{R}^3} \frac{d^3k}{||\vec{k}||^2} \left( -p_c(\vec{k}, t) p_d(-\vec{k}, t) \left[ \delta^{cd} - \frac{3}{4} \hat{k}^c \hat{k}^d \right] + \epsilon(\vec{k}, t) \epsilon(-\vec{k}, t) \left[ \frac{1}{4} + 1 \right] \right. \\ & \left. - m^2 \tilde{V}(\vec{k}, t) \epsilon(-\vec{k}, t) - \varphi_a^b(\vec{k}, t) \epsilon(-\vec{k}, t) \left[ \frac{1}{2} \delta_b^a + \frac{1}{2} \hat{k}^a \hat{k}_b \right] \right) + O(\delta^3, \kappa^2). \end{aligned} \quad (\text{B.18})$$

As can be readily seen, it differs from the Hamiltonian in the gauge used in this work in (A.107) only in the term encoding self-interaction of the scalar field. The reason is that by choosing the clocks, different choices for physical temporal and spatial coordinates were established and these are determined by the choice of different sets of clocks that on the gauge fixed surface are equal to  $\delta\tau$  and  $\delta\sigma^a$  respectively. As discussed in section 2.4 by an abuse of notation we understand the  $\vec{x}$  and  $t$  arguments of the fields to be the associated values of the diffeomorphism and Hamiltonian clock,  $\delta\tau$  and  $\delta\sigma^a$  respectively. By inspection of the clocks, it turns out that the ADM clocks used in this section, corresponding to the parameters  $x_{ADM}^a$  and  $t_{ADM}$  are related to the parameters used in this works,  $t$  and  $x^a$ , by

<sup>9</sup> Here only the vacuum case of linearised gravity is considered and the gauge fixing employed is not equivalent to the one used in this work and discussed in appendix B.1. In [61] one condition is that  $\delta E^a{}_i$  is traceless, that is  $\delta E^a{}_i \delta_a^i = 0$ , which is not satisfied for the gauge fixing discussed in appendix B.1 where we have  $\delta\mathcal{E}^a{}_i \delta_a^i = \frac{3\epsilon}{||\vec{k}||^2}$ .

$$t_{ADM} = \frac{1}{2}t \tag{B.19}$$

$$x_{ADM}^a = x^a + \frac{1}{2}\partial^a(\epsilon * G^{\Delta\Delta}), \tag{B.20}$$

where we neglected factors of  $\kappa$  and  $\beta$ , as the notation regarding these factors in [69] partly differs from the one used in this work.

We realise that in the case of vacuum gravity the choices for physical temporal and spatial coordinates agree but differ for non-vanishing momentum and energy density of the scalar field

Compared to the work in [35], where for the ADM constraints the same gauge fixing as in this section was used, their gauge fixed Hamiltonian appears with different prefactors in the self-interaction part compared to the one in (B.18). A reason might be, that we were not able to reproduce the expression for  $V$  in their equation (6) in [35] which consists of the second order terms of the expansion of  $\sqrt{q_{ab}} {}^{(3)}R$ , where  $q_{ab}$  denotes the spatial ADM metric and  ${}^{(3)}R$  the three-dimensional Ricci scalar. However, in contrast the result for this expansion given by the same authors in [38] could be reproduced by our computations.

### Appendix C. Local invertibility of $(\mathbb{1} - \Sigma(t, t_0))$ in the TCL master equation

An essential ingredient in the TCL master equation was the assumption that the superoperator  $(\mathbb{1} - \Sigma(t, t_0))$  was invertible in a suitable neighbourhood of  $t_0$  and we used that the inverse can be written in terms of a (perturbative) geometric series akin to

$$(\mathbb{1} - \Sigma(t, t_0))^{-1} = \sum_{n=0}^{\infty} (\Sigma(t, t_0))^n. \tag{C.1}$$

In order to represent the inverse in this way we need to ensure that the series actually converges which we will prove in this part of the appendix. The first thing to notice is the fact that  $\Sigma(t_0, t_0) = 0$ , which directly follows from the very definition (4.14). Secondly, all constituents of  $\Sigma(t, t_0)$  are continuous in  $t$ , which renders  $\Sigma(t, t_0)$  itself continuous. Consequently, there exists an interval  $[t_0, t]$  with  $t > t_0$  such that  $(\mathbb{1} - \Sigma(t, t_0))$  is invertible on the entire interval. We shall denote with  $\Sigma_R(t, t_0) \in \mathcal{B}(\mathcal{B}(\mathcal{H}))$  the restriction to precisely that neighbourhood, while the latter indicates that it is a bounded superoperator on the base Hilbert space. Hence, we can immediately define a norm of  $\Sigma_R(t, t_0)$  which is essentially inherited from the Hilbert space:

$$\|\Sigma_R(t, t_0)\|_{\mathcal{B}(\mathcal{B}(\mathcal{H}))} := \sup_{A \in \mathcal{B}(\mathcal{H})} \|\Sigma_R(t, t_0)A\|_{\mathcal{B}(\mathcal{H})}, \quad \|A\|_{\mathcal{B}(\mathcal{H})} := \sup_{\Psi \in \mathcal{H}} \|A\Psi\|_{\mathcal{H}} = 1, \quad \|\Psi\|_{\mathcal{H}} = 1.$$

By virtue of the definition of the restriction and the fact that  $(\mathbb{1} - \Sigma(t, t_0))^{-1}$  is invertible for all times in  $[t_0, t]$  we have  $\|\Sigma_R(t, t_0)\| := \lambda < 1$ , where we omit the explicit  $\mathcal{B}(\mathcal{B}(\mathcal{H}))$ -label since it follows from context. The time interval for which  $(\mathbb{1} - \Sigma(t, t_0))$  is invertible depends on the coupling strength of the interaction, which is immediately evident from the definition (4.14), which linearly depends on  $\alpha$  in the lowest orders of the expansion of the involved propagators. Heuristically speaking this means that the time interval  $[t_0, t]$  is longer the smaller the coupling constant is. In order to prove this, we need to establish the fact that the inversion can be indeed written in the form of a geometric series. Firstly, the series is absolutely convergent:

$$\left\| \sum_{n=0}^{\infty} (\Sigma_R(t, t_0))^n \right\| \leq \sum_{n=0}^{\infty} \|(\Sigma_R(t, t_0))^n\| \leq \sum_{n=0}^{\infty} \|\Sigma_R(t, t_0)\|^n = \sum_{n=0}^{\infty} \lambda^n < \infty.$$

Since  $\mathcal{B}(\mathcal{H})$  is a Banach space by construction and hence complete with respect to the induced norm, the series  $\sum_n (\Sigma_R(t, t_0))^n$  converges to an operator in  $\mathcal{B}(\mathcal{H})$ . In full analogy to the geometric series from standard calculus, let us introduce the following expression:

$$\mathfrak{R}(N) := (\Sigma_R(t, t_0))^N (\mathbb{1} - \Sigma_R(t, t_0))^{-1}. \tag{C.2}$$

With the help of this auxiliary quantity, we then establish a telescoping series:

$$(\Sigma_R(t, t_0))^M = (\Sigma_R(t, t_0))^M (\mathbb{1} - \Sigma_R(t, t_0)) (\mathbb{1} - \Sigma_R(t, t_0))^{-1} = \mathfrak{R}(M) - \mathfrak{R}(M + 1).$$

Consequently, the finite sum can be written in terms of two contributions, that is

$$\sum_{n=0}^N (\Sigma_R(t, t_0))^n = \mathfrak{R}(0) - \mathfrak{R}(N + 1),$$

where now it is possible to take the limit of  $N \rightarrow \infty$  and complete the proof:

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{n=0}^N (\Sigma_R(t, t_0))^n &= \sum_{n=0}^{\infty} (\Sigma_R(t, t_0))^n = \mathfrak{R}(0) - \lim_{N \rightarrow \infty} \mathfrak{R}(N + 1) = \mathfrak{R}(0) \\ &= (\mathbb{1} - \Sigma_R(t, t_0))^{-1}, \end{aligned}$$

where we used the definition of  $\mathfrak{R}(N)$  in (C.2) in the last step and moreover made use of the previous estimate:

$$\begin{aligned} \lim_{N \rightarrow \infty} \|\mathfrak{R}(N + 1)\| &\leq \lim_{N \rightarrow \infty} \|(\Sigma_R(t, t_0))^N\| \|(\mathbb{1} - \Sigma_R(t, t_0))^{-1}\| \\ &= \lim_{N \rightarrow \infty} \lambda^{N+1} \|(\mathbb{1} - \Sigma_R(t, t_0))^{-1}\| = 0, \end{aligned}$$

since  $\lambda < 1$  and  $\|(\mathbb{1} - \Sigma_R(t, t_0))^{-1}\|$  is finite on the interval  $[t_0, t]$  as we have shown already above.

### Appendix D. Evaluation of the correlation functions

In this appendix we will derive the correlation functions and their spectral representation, working with a regularisation established by formulating the theory in a box of finite volume introduced in equation (4.48). Given that the two number operators  $n_k^+ := (b_k^+)^\dagger b_k^+$  and  $n_k^- := (b_k^-)^\dagger b_k^-$  mutually commute, as they live on different Hilbert spaces, the density matrix of the environment can be written in terms of the following tensor product:

$$\rho_{\mathcal{E}} = \rho_{\mathcal{E}^+} \otimes \rho_{\mathcal{E}^-}, \tag{D.1}$$

where

$$\rho_{\mathcal{E}^r} = \frac{1}{Z_{\mathcal{E}^r}} \exp \left\{ -\beta \sum_{\vec{k} \in \mathbb{K}} \Omega_k n_k^r \right\}, \tag{D.2}$$

with  $r \in \{+, -\}$ ,  $Z_{\mathcal{E}^r} := \text{tr}_{\mathcal{E}^r} \{\rho_{\mathcal{E}^r}\}$  and  $\text{tr}_{\mathcal{E}^r}$  denoting the partial trace over the  $r$ -part of the environmental Hilbert space  $\mathcal{H}_{\mathcal{E}} = \mathcal{H}_{\mathcal{E}^+} \otimes \mathcal{H}_{\mathcal{E}^-}$ .

At this point we introduce an alternative notation for the summation over the discrete  $\vec{k}$ -vectors. Instead of summing over all elements of the set  $\mathbb{K}$  that contains the discrete, permitted  $\vec{k}$ -vectors, we want to sum over an index running over the natural numbers. Such a bijection exists, because each possible  $\vec{k}$  consists of three components  $k_x, k_y, k_z \in \frac{\pi\mathbb{Z}}{L}$  with  $L$  denoting the length of the box. Thus, we can identify an element  $\vec{k} \in \mathbb{K}$  uniquely by providing three numbers. A bijection between  $\mathbb{Z}^3$  and  $\mathbb{N}$  can be constructed therefore we can sum over  $j \in \mathbb{N}$  instead of  $\vec{k} \in \mathbb{K}$ .

We use this notation and rewrite the number operator in terms of the occupation number basis of the corresponding part of the Hilbert space:

$$n_i^r = \sum_{n_1^r=0}^{\infty} \sum_{n_2^r=0}^{\infty} \dots \sum_{n_i^r=1}^{\infty} \dots n_i^r |n_1^r, n_2^r, \dots, n_i^r, \dots\rangle \langle n_1^r, n_2^r, \dots, n_i^r, \dots|, \tag{D.3}$$

where the individual  $n_i^r \in \mathbb{N}_0$  on the right hand side are the eigenvalues of the occupation number operator and  $|n_1^r, n_2^r, \dots, n_i^r, \dots\rangle$  the corresponding eigenstates. Due to the orthonormality of different Fock states  $|n_1^r, n_2^r, \dots\rangle$  we obtain

$$\exp \left\{ -\beta \sum_{j \in \mathbb{N}} \Omega_j n_j^r \right\} = \sum_{n_1^r=0}^{\infty} \sum_{n_2^r=0}^{\infty} \dots \prod_{i \in \mathbb{N}} e^{-\beta n_i^r \Omega_i} |n_1^r, n_2^r, \dots\rangle \langle n_1^r, n_2^r, \dots|. \tag{D.4}$$

Due to the regularisation established by considering the system to be in a finite box, we can express the partial trace in terms of Fock states and find:

$$tr_{\mathcal{E}_r} \{ \cdot \} = \sum_{n_1^r=0}^{\infty} \sum_{n_2^r=0}^{\infty} \dots \langle n_1^r, n_2^r, \dots | \cdot | n_1^r, n_2^r, \dots \rangle. \tag{D.5}$$

We can use this expression to evaluate the partition sum of the Gibbs state:

$$\begin{aligned} Z_{\mathcal{E}_r} &= tr_{\mathcal{E}_r} \left( \exp \left\{ -\beta \sum_{j \in \mathbb{N}} \Omega_j n_j^r \right\} \right) = \sum_{n_1^r=0}^{\infty} \sum_{n_2^r=0}^{\infty} \dots \prod_{j \in \mathbb{N}} e^{-\beta n_j^r \Omega_j} = \sum_{n_1^r=0}^{\infty} \sum_{n_2^r=0}^{\infty} \dots \prod_{j \in \mathbb{N}} [e^{-\beta \Omega_j}]^{n_j^r} \\ &= \left( \sum_{n_1^r=0}^{\infty} [e^{-\beta \Omega_1}]^{n_1^r} \right) \cdot \left( \sum_{n_2^r=0}^{\infty} [e^{-\beta \Omega_2}]^{n_2^r} \right) \cdot \dots = \prod_{i \in \mathbb{N}} \left( \sum_{n_i^r=0}^{\infty} [e^{-\beta \Omega_i}]^{n_i^r} \right) = \prod_{i \in \mathbb{N}} \frac{1}{1 - e^{-\beta \Omega_i}}. \end{aligned} \tag{D.6}$$

In the last step we assumed that  $\vec{k} = 0$ , which corresponds to  $\Omega_k = 0$ , is not contained in the set  $\mathbb{K}$ , which is the usual infrared divergence present in quantum field theory. Then  $e^{-\beta \Omega_i} < 1$  always holds and the expression is a geometric series.

In actual applications it is important that the partition sum is finite. This is indeed the case, a proof can be found for instance in [120]. The final partition sum for part  $r$  of the environment in (D.6) is independent of this label  $r$ , so we get the same result for  $r = +$  and  $r = -$  respectively. From (D.1) one can then see that the partition sum of the total environmental Gibbs state is thus just given by the square of (D.6):

$$Z_{\mathcal{E}} = \left[ \prod_{j \in \mathbb{N}} \frac{1}{1 - e^{-\beta \Omega_j}} \right]^2. \tag{D.7}$$

Now with this result the thermal Wightman functions can be computed explicitly:

$$\begin{aligned}
 G^{>a,b}_i(\vec{x}-\vec{y},t-s) &= \langle \delta\mathcal{E}^a_i(\vec{x},t)\delta\mathcal{E}^b_j(\vec{y},s) \rangle = \text{tr}_{\mathcal{E}} \left\{ \delta\mathcal{E}^a_i(\vec{x},t)\delta\mathcal{E}^b_j(\vec{y},s)\rho_{\mathcal{E}} \right\} \\
 &= \int \frac{d^3k d^3p}{2(2\pi)^3 \sqrt{\Omega_k \Omega_p}} \sum_{r,u \in \{\pm\}} [P^{-r}(\vec{k})]_i^a [P^{-u}(\vec{p})]_j^b \\
 &\quad \cdot \text{tr}_{\mathcal{E}} \left\{ \left[ b_k^r e^{-i\Omega_k t + i\vec{k}\vec{x}} + (b_{-k}^r)^\dagger e^{i\Omega_k t + i\vec{k}\vec{x}} \right] \left[ b_p^u e^{-i\Omega_p s + i\vec{p}\vec{y}} + (b_{-p}^u)^\dagger e^{i\Omega_p s + i\vec{p}\vec{y}} \right] \rho_{\mathcal{E}} \right\},
 \end{aligned} \tag{D.8}$$

where  $[P^{-r}(\vec{k})]_i^a := [P^{-r}(\vec{k})]_b^a \delta_i^b$ . Using the explicit expressions given above for the Gibbs state and the trace in the occupation number basis, we obtain the following results:

$$\text{tr}_{\mathcal{E}} \{ b_k^r b_p^u \rho_{\mathcal{E}} \} = 0 \tag{D.9}$$

$$\text{tr}_{\mathcal{E}} \{ b_k^r (b_{-p}^u)^\dagger \rho_{\mathcal{E}} \} = \delta^{ru} \delta(\vec{k} + \vec{p}) \text{tr}_{\mathcal{E}} \{ b_k^r (b_k^r)^\dagger \rho_{\mathcal{E}} \} = \delta^{ru} \delta(\vec{k} + \vec{p}) \text{tr}_{\mathcal{E}} \{ [n_k^r + 1] \rho_{\mathcal{E}} \} \tag{D.10}$$

$$\text{tr}_{\mathcal{E}} \{ (b_{-k}^r)^\dagger b_p^u \rho_{\mathcal{E}} \} = \delta^{ru} \delta(\vec{k} + \vec{p}) \text{tr}_{\mathcal{E}} \{ (b_k^r)^\dagger b_k^r \rho_{\mathcal{E}} \} = \delta^{ru} \delta(\vec{k} + \vec{p}) \text{tr}_{\mathcal{E}} \{ n_k^r \rho_{\mathcal{E}} \} \tag{D.11}$$

$$\text{tr}_{\mathcal{E}} \{ (b_{-k}^r)^\dagger (b_{-p}^u)^\dagger \rho_{\mathcal{E}} \} = 0. \tag{D.12}$$

Reinserting this back into the Wightman functions in (D.8) they simplify to

$$\begin{aligned}
 G^{>a,b}_i(\vec{x}-\vec{y},t-s) &= \int \frac{d^3k}{2(2\pi)^3 \Omega_k} \sum_{r \in \{\pm\}} [P^{-r}(\vec{k})]_i^a [P^{-r}(-\vec{k})]_j^b \\
 &\quad \times \left[ (\text{tr}_{\mathcal{E}} \{ n_k^r \rho_{\mathcal{E}} \} + 1) e^{-i\Omega_k(t-s) + i\vec{k}(\vec{x}-\vec{y})} + \text{tr}_{\mathcal{E}} \{ n_k^r \rho_{\mathcal{E}} \} e^{-i\Omega_k(s-t) + i\vec{k}(\vec{y}-\vec{x})} \right].
 \end{aligned} \tag{D.13}$$

Now the remaining trace can be computed with respect to the occupation number basis and yields the expected Bose–Einstein distribution:

$$\begin{aligned}
 N(\Omega_k) := \text{tr}_{\mathcal{E}} \{ n_k^r \rho_{\mathcal{E}} \} &= \frac{1}{Z_{\mathcal{E}r}} \sum_{\substack{n_1^r=0 \\ \dots \\ n_K^r=0}}^{\infty} \dots \sum_{\substack{\tilde{n}_1^r=0 \\ \dots \\ \tilde{n}_K^r=0}}^{\infty} \left\langle n_1^r, \dots, n_K^r \left| n_k^r e^{-\beta \sum_{j=1}^K \tilde{n}_j^r \Omega_j} \right| \tilde{n}_1^r, \dots, \tilde{n}_K^r \right\rangle \langle \tilde{n}_1^r, \dots, \tilde{n}_K^r | n_1^r, \dots, n_K^r \rangle \\
 &= \frac{1}{Z_{\mathcal{E}r}} \sum_{n_1^r=0}^{\infty} \dots \sum_{n_K^r=0}^{\infty} \left\langle n_1^r, \dots, n_K^r \left| n_k^r e^{-\beta \sum_{j=1}^K n_j^r \Omega_j} \right| n_1^r, \dots, n_K^r \right\rangle \\
 &= \frac{1}{-\beta} \frac{1}{Z_{\mathcal{E}r}} \frac{\partial}{\partial \Omega_k} \sum_{n_1^r=0}^{\infty} \dots \sum_{n_K^r=0}^{\infty} \left\langle n_1^r, \dots, n_K^r \left| e^{-\beta \sum_{j=1}^K n_j^r \Omega_j} \right| n_1^r, \dots, n_K^r \right\rangle \\
 &= \frac{1}{-\beta} \frac{\partial Z_{\mathcal{E}r}}{\partial \Omega_k} = \frac{1}{-\beta} \frac{\partial}{\partial \Omega_k} \ln(Z_{\mathcal{E}r}) = \frac{1}{-\beta} \frac{\partial}{\partial \Omega_k} \left[ -\sum_{j=1}^K \ln(1 - e^{-\beta \Omega_j}) \right] \\
 &= \frac{e^{-\beta \Omega_k}}{1 - e^{-\beta \Omega_k}} = \frac{1}{e^{\beta \Omega_k} - 1}.
 \end{aligned} \tag{D.14}$$

This result is independent of the polarisation, hence we can evaluate the sum over polarisations separately and obtain the transverse-traceless projector

$$\sum_{r \in \{\pm\}} [P^{-r}(\vec{k})]_i^a [P^{-r}(-\vec{k})]_j^b = \underline{P}^{ab}_{ij}(\vec{k}). \tag{D.15}$$

Hence we can express the Wightman function in the following manner:

$$G^{>a\ b}_i(\vec{x}-\vec{y}, t-s) =: P^{a\ b}_i G^{>}(\vec{x}-\vec{y}, t-s) \quad (\text{D.16})$$

with

$$\begin{aligned} G^{>}(\vec{x}-\vec{y}, t-s) &= \int \frac{d^3k}{2(2\pi)^{\frac{3}{2}}\Omega_k} \left[ (N(\Omega_k) + 1)e^{-i\Omega_k(t-s)+i\vec{k}(\vec{x}-\vec{y})} + N(\Omega_k)e^{-i\Omega_k(s-t)+i\vec{k}(\vec{y}-\vec{x})} \right] \\ &= \int \frac{d^4k}{2(2\pi)^{\frac{3}{2}}\Omega_k} \left[ (N(\Omega_k) + 1)\delta(k^0 - \Omega_k) + N(\Omega_k)\delta(k^0 + \Omega_k) \right] e^{-ik^0(t-s)+i\vec{k}(\vec{x}-\vec{y})} \\ &= \int \frac{d^4k}{2(2\pi)^{\frac{3}{2}}\Omega_k} [N(k^0) + 1] \left[ \delta(k^0 - \Omega_k) + \delta(k^0 + \Omega_k) \right] e^{-ik^0(t-s)+i\vec{k}(\vec{x}-\vec{y})}, \quad (\text{D.17}) \end{aligned}$$

where we used that  $N(-\Omega_k) = -[N(\Omega_k) + 1]$ . Defining

$$\rho^{>}(k^0, \|\vec{k}\|) = [1 + N(k^0)]\rho(k^0, \|\vec{k}\|) \quad \text{and} \quad \rho(k^0, \|\vec{k}\|) = \sqrt{\frac{\pi}{2}} \frac{1}{\Omega_k} [\delta(k^0 - \Omega_k) - \delta(k^0 + \Omega_k)],$$

we can write down the Wightman function in its spectral decomposition:

$$G^{>}(\vec{x}-\vec{y}, t-s) = \int \frac{d^4k}{(2\pi)^{\frac{4}{2}}} \rho^{>}(k^0, \|\vec{k}\|) e^{-ik^0(t-s)+i\vec{k}(\vec{x}-\vec{y})}. \quad (\text{D.18})$$

For  $G^{<}(\vec{x}-\vec{y}, t-s)$ , defined as

$$G^{<a\ b}_i(\vec{x}-\vec{y}, t-s) =: P^{a\ b}_i G^{<}(\vec{x}-\vec{y}, t-s) \quad (\text{D.19})$$

a similar derivation yields

$$G^{<}(\vec{x}-\vec{y}, t-s) = \int \frac{d^4k}{(2\pi)^{\frac{4}{2}}} \rho^{<}(k^0, \|\vec{k}\|) e^{-ik^0(t-s)+i\vec{k}(\vec{x}-\vec{y})} \quad (\text{D.20})$$

with

$$\rho^{<}(k^0, \|\vec{k}\|) = N(k^0)\rho(k^0, \|\vec{k}\|).$$

From the equalities  $G^{<}(\vec{x}-\vec{y}, t-s) = G^{>}(\vec{y}-\vec{x}, s-t)$  and  $P^{b\ a}_j = P^{a\ b}_i$ , one can directly read off  $G^{<}(\vec{x}-\vec{y}, t-s)$  and construct the thermal Green's function

$$\begin{aligned} G^{(F)a\ b}_i(\vec{x}-\vec{y}, t-s) &:= P^{a\ b}_i G^{>}(\vec{x}-\vec{y}, t-s)\theta(t-s) + P^{a\ b}_i G^{<}(\vec{x}-\vec{y}, t-s)\theta(t-s) \\ &= P^{a\ b}_i \left[ \theta(t-s) \int \frac{d^3k}{2(2\pi)^{\frac{3}{2}}\Omega_k} e^{-i\Omega_k(t-s)+i\vec{k}(\vec{x}-\vec{y})} + \theta(s-t) \right. \\ &\quad \left. \times \int \frac{d^3k}{2(2\pi)^{\frac{3}{2}}\Omega_k} e^{-i\Omega_k(s-t)+i\vec{k}(\vec{y}-\vec{x})} \right] \\ &\quad + P^{a\ b}_i \int \frac{d^3k}{(2\pi)^{\frac{3}{2}}\Omega_k} N(\Omega_k) \cos[(\Omega_k(t-s) - \vec{k}(\vec{x}-\vec{y}))] \\ &= G^{(F)}_{\Theta=0}{}^{a\ b}_i(\vec{x}-\vec{y}, t-s) + P^{a\ b}_i \int \frac{d^3k}{(2\pi)^{\frac{3}{2}}\Omega_k} N(\Omega_k) \\ &\quad \times \cos[(\Omega_k(t-s) - \vec{k}(\vec{x}-\vec{y}))] \quad (\text{D.21}) \end{aligned}$$

with the Heaviside function  $\theta(\tau)$ . The first part is the usual Green's function one obtains for a temperature parameter  $\Theta$  that vanishes and a second thermal contribution present in the finite

temperature case. In case of vanishing temperature parameter  $\Theta = 0$ , we obtain  $N(\Omega_k) = 0$  and find the Green's function of the zero temperature case that takes the following form:

$$\begin{aligned}
 G_{\Theta=0}^{(F) a b; i j}(\vec{x} - \vec{y}, t - s) &= P^a b; i j \left[ \theta(t - s) \int \frac{d^3 k}{2(2\pi)^{\frac{3}{2}} \Omega_k} e^{-i\Omega_k(t-s) + i\vec{k}(\vec{x}-\vec{y})} + \theta(s - t) \int \frac{d^3 k}{2(2\pi)^{\frac{3}{2}} \Omega_k} e^{-i\Omega_k(s-t) + i\vec{k}(\vec{y}-\vec{x})} \right] \\
 &= P^a b; i j \int_{\mathbb{R}^4} \frac{d^4 k}{(2\pi)^{\frac{4}{2}}} \frac{-i}{k^2 - i\epsilon} e^{-ik^0(t-s) + i\vec{k}(\vec{x}-\vec{y})} \\
 &= P^a c; i l P^b e; j m \delta^{dl} \delta^{fm} [\delta_{ce} \delta_{df} + \delta_{cf} \delta_{de} - \delta_{cd} \delta_{ef}] \frac{1}{2} \int_{\mathbb{R}^4} \frac{d^4 k}{(2\pi)^{\frac{4}{2}}} \frac{-i}{k^2 - i\epsilon} e^{-ik^0(t-s) + i\vec{k}(\vec{x}-\vec{y})}. \quad (D.22)
 \end{aligned}$$

Thus, the results is just given by the transverse-traceless projection of the spatial part of the graviton propagator in harmonic gauge, see for instance [121] for the explicit form of the graviton propagator. Note that we chose the integration contour in the second step such that we obtain the Feynman propagator. After a four dimensional Fourier transformation of the entire thermal Green's function and again choosing the Feynman prescription we obtain:

$$G^{(F) a b; i j}(p) = P^a b; i j(\vec{p}) \left\{ \frac{-i}{p^\mu p_\mu - i\epsilon} + 2\pi N(\Omega_p) \delta(p^\mu p_\mu) \right\}, \quad (D.23)$$

which indeed has the expected form, see for instance [122].

### ORCID iD

Kristina Giesel  <https://orcid.org/0000-0002-0617-906X>

### References

- [1] Cohen Tannoudji C, Grynberg G and Dupont-Roe J 1992 *Atom-Photon Interactions* (New York: Wiley)
- [2] Carmichael H J 1999 *Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck Equations* vol 1 (Berlin: Springer)
- [3] Dattagupta S and Puri S 2004 *Dissipative Phenomena in Condensed Matter: Some Applications* vol 71 (Heidelberg: Springer)
- [4] Heiss D 2008 *Fundamentals of Quantum Information: Quantum Computation, Communication, Decoherence and all That* vol 587 (Berlin: Springer)
- [5] Orszag M 2016 *Quantum Optics: Including Noise Reduction, Trapped Ions, Quantum Trajectories and Decoherence* (Berlin: Springer)
- [6] Jonathan J H 1989 Decoherence in quantum cosmology *Phys. Rev. D* **39** 2912–23
- [7] Polarski D and Alexei A S 1996 Semiclassicality and decoherence of cosmological perturbations *Class. Quantum Grav.* **13** 377–92
- [8] LaFlamme R and Louko J 1991 Reduced density matrices and decoherence in quantum cosmology *Phys. Rev. D* **43** 3317–31
- [9] Boyanovsky D 2015 Effective field theory during inflation: reduced density matrix and its quantum master equation *Phys. Rev. D* **92** 023527
- [10] Hollowood T J and McDonald J I 2017 Decoherence, discord and the quantum master equation for cosmological perturbations *Phys. Rev. D* **95** 103521
- [11] Bahns D, Pohl A and Witt I (eds) 2019 *Open Quantum Systems: A Mathematical Perspective (Tutorials, Schools, and Workshops in the Mathematical Sciences)* (Cham: Birkhäuser) (<https://doi.org/10.1007/978-3-030-13046-6>)

- [12] Breuer H-P et al 2007 Heinz-peter breuer and francesco petruccione *The Theory of Open Quantum Systems* (Oxford: Oxford University Press)
- [13] Breuer H-P and Petruccione F 2003 Concepts and methods in the theory of open quantum systems *Irreversible Quantum Dynamics* (Berlin: Springer) pp 65–79
- [14] Rivas Angel and Huelga S F 2012 *Open Quantum Systems* (Berlin: Springer)
- [15] Hubert Zurek W 2003 Decoherence, einselection and the quantum origins of the classical *Rev. Mod. Phys.* **75** 715–75
- [16] Giulini D, Kiefer C, Joos E, Kupsch J, Stamatescu I O, and Zeh H D 1996 Decoherence and the appearance of a classical world in quantum theory
- [17] Schlosshauer M 2014 The quantum-to-classical transition and decoherence (arXiv:1404.2635)
- [18] Caldeira A O and Leggett A J 1981 Influence of dissipation on quantum tunneling in macroscopic systems *Phys. Rev. Lett.* **46** 211–4
- [19] Qidong X and Blencowe M P 2020 Toy models for gravitational and scalar qed decoherence (arXiv:2005.02554)
- [20] Klauder J R and Skagerstam B-S 1985 *Coherent States - Applications in Physics and Mathematical Physics* 1st edn (World Scientific Publishing Company)
- [21] Perelomov A 1986 *Generalized Coherent States and Their Applications* (Berlin: Springer)
- [22] Zhang W-M, Feng D H and Gilmore R 1990 Coherent states: theory and some applications *Rev. Mod. Phys.* **62** 867–927
- [23] Ali S T, Antoine J-P and Gazeau J-P 2014 *Coherent States, Wavelets and Their Generalizations (Theoretical and Mathematical Physics)* 2nd edn (Berlin: Springer)
- [24] Ellis J, Hagelin J S, Nanopoulos D V and Srednicki M 1984 Search for violations of quantum mechanics *Nucl. Phys. B* **241** 381–405
- [25] Ellis J, Lopez J L, Mavromatos N E and Nanopoulos D V 1996 Precision tests of CPT symmetry and quantum mechanics in the neutral kaon system *Phys. Rev. D* **53** 3846
- [26] Benatti F and Floreanini R 1999 Complete positivity and neutron interferometry *Phys. Lett. B* **451** 422–9
- [27] Benatti F and Floreanini R 2000 Open system approach to neutrino oscillations *J. High Energy Phys.* **JHEP02(2000)032**
- [28] Lisi E, Marrone A and Montanino D 2000 Probing possible decoherence effects in atmospheric neutrino oscillations *Phys. Rev. Lett.* **85** 1166
- [29] Morgan D, Winstanley E, Brunner J and Thompson L F 2006 Probing quantum decoherence in atmospheric neutrino oscillations with a neutrino telescope *Astropart. Phys.* **25** 311–27
- [30] Guzzo M M, de Holanda P C and Oliveira R L N 2016 Quantum dissipation in a neutrino system propagating in vacuum and in matter *Nucl. Phys. B* **908** 408–22
- [31] Coelho Jão A B and Anthony Mann W 2017 Decoherence, matter effect and neutrino hierarchy signature in long baseline experiments *Phys. Rev. D* **96** 093009
- [32] Carpio J A, Massoni E and Gago A M 2018 Revisiting quantum decoherence for neutrino oscillations in matter with constant density *Phys. Rev. D* **97** 115017
- [33] Balieiro Gomes G, Forero D V, Guzzo M M, De Holanda P C and Oliveira R L N 2019 Quantum decoherence effects in neutrino oscillations at dune *Phys. Rev. D* **100** 055023
- [34] Hellmann D, Päs H and Rani E 2022 Searching new particles at neutrino telescopes with quantum-gravitational decoherence *Phys. Rev. D* **105** 055007
- [35] Anastopoulos C and Hu B L 2013 A master equation for gravitational decoherence: probing the textures of spacetime *Class. Quantum Grav.* **30** 165007
- [36] Blencowe M P 2013 Effective field theory approach to gravitationally induced decoherence *Phys. Rev. Lett.* **111** 021302
- [37] Oniga T and Charles H T W 2016 Quantum gravitational decoherence of light and matter *Phys. Rev. D* **93** 044027
- [38] Lagouvardos M and Anastopoulos C 2021 Gravitational decoherence of photons *Class. Quantum Grav.* **38** 115012
- [39] Bassi A, Großardt A and Ulbricht H 2017 Gravitational decoherence *Class. Quantum Grav.* **34** 193002
- [40] Anastopoulos C and Bei-Lok H 2022 Gravitational decoherence: a thematic overview *AVS Quantum Sci.* **4** 015602
- [41] Donadi S and Bassi A 2022 Seven non-standard models coupling quantum matter and gravity (arXiv:2202.13542)
- [42] Kok P and Yurtsever U 2003 Gravitational decoherence *Phys. Rev. D* **68** 085006

- [43] Breuer H P, Goklu E and Lammerzähl C 2009 Metric fluctuations and decoherence *Class. Quantum Grav.* **26** 105012
- [44] Asprea L 2021 Gravitational decoherence *PhD Thesis* Trieste U
- [45] Gorini V, Kossakowski A and Chandy George Sudarshan E 1976 Completely positive dynamical semigroups of N-level systems *J. Math. Phys.* **17** 821–5
- [46] Lindblad G 1976 On the generators of quantum dynamical semigroups *Commun. Math. Phys.* **48** 119–30
- [47] Rovelli C 2004 *Quantum Gravity (Cambridge Monographs on Mathematical Physics)* (Cambridge: Cambridge University Press)
- [48] Thiemann T 2007 *Modern Canonical Quantum General Relativity (Cambridge Monographs on Mathematical Physics)* (Cambridge: Cambridge University Press)
- [49] Rovelli C and Vidotto F 2014 *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory* (Cambridge: Cambridge University Press)
- [50] Thiemann T 2006 Complexifier coherent states for quantum general relativity *Class. Quantum Grav.* **23** 2063–118
- [51] Freidel L and Etera R L 2011 U(N) Coherent States for loop quantum gravity *J. Math. Phys.* **52** 052502
- [52] Stottmeister A and Thiemann T 2016 Coherent states, quantum gravity and the Born-Oppenheimer approximation, I: general considerations *J. Math. Phys.* **57** 063509
- [53] Calcinari A, Freidel L, Livine E and Speziale S 2020 Twisted geometries coherent states for loop quantum gravity *Class. Quantum Grav.* **38** 025004
- [54] Dittrich B 2017 The continuum limit of loop quantum gravity—a framework for solving the theory pp 153–79
- [55] Steinhaus S 2020 Coarse graining spin foam quantum gravity—a review *Front. Phys.* **8** 295
- [56] Feller A and Etera R L 2017 Surface state decoherence in loop quantum gravity, a first toy model *Class. Quantum Grav.* **34** 045004
- [57] Ansel Q 2021 A model of spinfoam coupled with an environment *Gen. Relativ. Gravit.* **53** 39
- [58] Azouit R, Sarlette A and Rouchon P 2016 Adiabatic elimination for open quantum systems with effective Lindblad master equations (arXiv:1603.04630)
- [59] Ashtekar A 1986 New variables for classical and quantum gravity *Phys. Rev. Lett.* **57** 2244–7
- [60] Ashtekar A, Romano J D and Tate R S 1989 New variables for gravity: inclusion of matter *Phys. Rev. D* **40** 2572
- [61] Ashtekar A, Rovelli C and Smolin L 1991 Gravitons and loops *Phys. Rev. D* **44** 1740–55
- [62] Varadarajan M 2002 Gravitons from a loop representation of linearized gravity *Phys. Rev. D* **66** 024017
- [63] Rovelli C 1991 What is observable in classical and quantum gravity? *Class. Quantum Grav.* **8** 297–316
- [64] Rovelli C 1991 Quantum reference systems *Class. Quantum Grav.* **8** 317–32
- [65] Rovelli C 2002 Partial observables *Phys. Rev. D* **65** 124013
- [66] Dittrich B 2007 Partial and complete observables for Hamiltonian constrained systems *Gen. Relativ. Gravit.* **39** 1891–927
- [67] Dittrich B 2006 Partial and complete observables for canonical general relativity *Class. Quantum Grav.* **23** 6155–84
- [68] Vytheeswaran A S 1994 Gauge unfixing in second class constrained systems *Ann. Phys.* **236** 297–324
- [69] Dittrich B and Tambornino J 2007 A Perturbative approach to Dirac observables and their space-time algebra *Class. Quantum Grav.* **24** 757–84
- [70] Dittrich B and Tambornino J 2007 Gauge invariant perturbations around symmetry reduced sectors of general relativity: applications to cosmology *Class. Quantum Grav.* **24** 4543–86
- [71] Giesel K, Herzog A and Singh P 2018 Gauge invariant variables for cosmological perturbation theory using geometrical clocks *Class. Quantum Grav.* **35** 155012
- [72] Markus B Fob and William C C L 2022 Cosmological perturbations and invariant observables in geodesic lightcone coordinates *J. Cosmol. Astropart. Phys.* **JCAP01(2022)034**
- [73] Anastopoulos C and Hu B L 2007 Decoherence in quantum gravity: issues and critiques *J. Phys.: Conf. Ser.* **67** 012012
- [74] Arnowitt R, Deser S and Misner C W 2008 The dynamics of general relativity *Gen. Relativ. Gravit.* **40** 1997–2027

- [75] Gambini R, Pedro Garcia-Pintos L and Pullin J 2011 An axiomatic formulation of the Montevideo interpretation of quantum mechanics *Stud. Hist. Phil. Sci. B* **42** 256–63
- [76] Gambini R and Pullin J 2018 The montevideo interpretation of quantum mechanics: a short review *Entropy* **20** 413
- [77] Gambini R and Pullin J 2020 The montevideo interpretation: how the inclusion of a quantum gravitational notion of time solves the measurement problem *Universe* **6** 236
- [78] Hoehn P A, Smith A R H and Lock M P E 2021 Trinity of relational quantum dynamics *Phys. Rev. D* **104** 066001
- [79] Feynman R P and Vernon F L 1963 The theory of a general quantum system interacting with a linear dissipative system *Ann. Phys., NY* **24** 118–73
- [80] Nakajima S 1958 On quantum theory of transport phenomena steady diffusion *Prog. Theor. Phys.* **20** 948–59
- [81] Zwanzig R 1960 Ensemble method in the theory of irreversibility *J. Chem. Phys.* **33** 1338–41
- [82] Shibata F, Takahashi Y and Hashitume N 1977 A generalized stochastic liouville equation. non-markovian versus memoryless master equations *J. Stat. Phys.* **17** 171–87
- [83] Chaturvedi S and Shibata F 1979 Time-convolutionless projection operator formalism for elimination of fast variables. Applications to Brownian motion *Z. Phys. B* **35** 297–308
- [84] Matsubara T 1955 A new approach to quantum statistical mechanics *Prog. Theor. Phys.* **14** 351–78
- [85] Kubo R 1957 Statistical mechanical theory of irreversible processes I. General theory and simple applications in magnetic and conduction problems *J. Phys. Soc. Japan* **12** 570–86
- [86] Martin P C and Schwinger J 1959 Theory of many-particle systems. I *Phys. Rev.* **115** 1342–73
- [87] Kuchař K 1972 A bubble-time canonical formalism for geometrodynamics *J. Math. Phys.* **13** 768–81
- [88] Kuchař K 1999 Covariant gauge fixing and kuchař decomposition *Phys. Rev. D* **61** 024037
- [89] Boldrin A and Małkiewicz P 2023 Gauge-fixing and spacetime reconstruction in the Hamiltonian theory of cosmological perturbations *Class. Quantum Grav.* **40** 015003
- [90] Kaplanek G and Tjoa E 2023 Effective master equations for two accelerated qubits *Phys. Rev. A* **107** 012208
- [91] Stanley Deser R A and Misner C W 1960 Consistency of canonical reduction of general relativity *J. Math. Phys.* **1** 434
- [92] Fernando J and Barbero G 1995 Real Ashtekar variables for Lorentzian signature space times *Phys. Rev. D* **51** 5507–10
- [93] Immirzi G 1997 Real and complex connections for canonical gravity *Class. Quantum Grav.* **14** L177–81
- [94] Thiemann T 1995 Generalized boundary conditions for general relativity for the asymptotically flat case in terms of Ashtekar's variables *Class. Quantum Grav.* **12** 181–98
- [95] Corichi A, Rubalcava I and Vukasinac T 2014 Hamiltonian and Noether charges in first order gravity *Gen. Relativ. Gravit.* **46** 1813
- [96] Campiglia M 2015 Note on the phase space of asymptotically flat gravity in Ashtekar–Barbero variables *Class. Quantum Grav.* **32** 145011
- [97] Regge T and Teitelboim C 1974 Role of surface integrals in the Hamiltonian formulation of general relativity *Ann. Phys.* **88** 286
- [98] Dittrich B and Thiemann T 2006 Testing the master constraint programme for loop quantum gravity. IV. Free field theories *Class. Quantum Grav.* **23** 1121–42
- [99] Poisson E and Will C M 2014 *Gravity: Newtonian, Post-Newtonian, Relativistic* (Cambridge: Cambridge University Press)
- [100] Giddings S and Weinberg S 2020 Gauge-invariant observables in gravity and electromagnetism: black hole backgrounds and null dressings *Phys. Rev. D* **102** 026010
- [101] Thiemann T 2006 Reduced phase space quantization and Dirac observables *Class. Quantum Grav.* **23** 1163–80
- [102] Kuchař K 1970 Ground state functional of the linearized gravitational field *J. Math. Phys.* **11** 3322–34
- [103] Ashtekar A and Bonga Beatrice 2017 On the ambiguity in the notion of transverse traceless modes of gravitational waves *Gen. Relativ. Gravit.* **49** 122
- [104] Giesel K, Hofmann S, Thiemann T and Winkler O 2010 Manifestly gauge-invariant general relativistic perturbation theory. I. Foundations *Class. Quantum Grav.* **27** 055005
- [105] Van Kampen N G 1974 A cumulant expansion for stochastic linear differential equations. I *Physica* **74** 215–38

- [106] Van Kampen N G 1974 A cumulant expansion for stochastic linear differential equations. II *Physica* **74** 239–47
- [107] Breuer H-P, Andrea M and Petruccione F 2002 Time-local master equations: influence functional and cumulant expansion
- [108] Homa G, Zsolt Bernád J and Lisztes L 2019 Positivity violations of the density operator in the Caldeira-Leggett master equation *Eur. Phys. J. D* **73** 1–13
- [109] Chris Fleming N I C, Anastopoulos C and Bei-Lok H 2010 The rotating-wave approximation: consistency and applicability from an open quantum system analysis *J. Phys. A: Math. Theor.* **43** 405304
- [110] Boyanovsky D 2015 Effective field theory out of equilibrium: Brownian quantum fields *New J. Phys.* **17** 063017
- [111] Hornberger K 2009 *Introduction to Decoherence Theory* (Berlin: Springer) pp 221–76
- [112] Evans T S and Steer D A 1996 Wick’s theorem at finite temperature *Nucl. Phys. B* **474** 481–96
- [113] Burrows B L and Colwell D J 1990 The fourier transform of the unit step function *Int. J. Math. Educ. Sci. Technol.* **21** 629–35
- [114] Fahn M J, Giesel K and Kobler M 2022 A gravitationally induced decoherence model using Ashtekar variables: one-particle sector (in preparation)
- [115] Wilson-Gerow J 2017 A study of the quantum-to-classical transition in gravity, and a study of the consequences of constraints in gauge theory path-integrals *Master’s Thesis* Philip Stamp, University of British Columbia
- [116] Wilson-Gerow J and Stamp P C E 2020 Gauge invariant propagators and states in quantum electrodynamics (arXiv:2011.05305)
- [117] Anastopoulos C 1996 Quantum theory of nonrelativistic particles interacting with gravity *Phys. Rev. D* **54** 1600–5
- [118] Asprea L, Gasbarri G and Bassi A 2021 Gravitational decoherence: a general nonrelativistic model *Phys. Rev. D* **103** 104041
- [119] Asprea L and Gasbarri G 2021 Gravitational decoherence: a nonrelativistic spin 1/2 fermionic model *Phys. Rev. D* **104** 024043
- [120] Joseph Fahn M 2020 Gravitationally induced decoherence in open quantum systems using linearised gravity formulated in Ashtekar variables *Master’s Thesis* Kristina Giesel, Friedrich-Alexander-Universität Erlangen-Nürnberg
- [121] Donoghue J F, Ivanov M M, and Shkerin A 2017 Epfl lectures on general relativity as a quantum field theory (arXiv:1702.00319)
- [122] Koksma J F, Prokopec T and Schmidt M G 2011 Decoherence in an interacting quantum field theory: thermal case *Phys. Rev. D* **83** 085011