September 2, 2025

Lectures on low-energy quantum gravity and its experimental tests

Daniel Carney

Physics Division, Lawrence Berkeley National Laboratory
Berkeley, CA

Email: carney@lbl.gov

Abstract

These are notes for a series of graduate summer school lectures on quantum gravity at scales well below the Planck energy density. The focus is on how gravity can be perturbatively quantized in this regime, and in what sense we can test whether gravity is actually quantized this way in nature. I begin with a detailed treatment of the quantization of perturbative gravity as a standard effective quantum field theory and a brief survey of some possible alternative phenomenologies. I then give an overview of experiments which are aiming to test aspects of quantum gravity, and what exactly such tests can really determine about the nature of gravity in the real world.

Contents

Meta

My goal with these notes has been, roughly, to provide an updated introductory review of the field of testing quantum gravity. Our original review, the "User's Manual" [1], holds up well despite nearly a decade of progress on these questions, and I would still recommend it as a first resource for someone starting to think about these things.

However, there were a number of items I wanted to improve. Topics which have substantially developed since then and are now covered here include a variety of new models of gravity (in particular, the classical-quantum models of Oppenheim et al. in Lec. 2.2 and the entropic gravity construction in Lec. 2.3), and a detailed treatment of graviton and graviton-based noise detection and its implications. The treatment of perturbative quantum gravity is also much more thorough, and the discussion of "classical" gravity models is substantially improved compared to [1].

Some acknowledgements: I would like to thank the organizers and students at the 2025 Saalburg summer school for the chance to put these materials together and discuss them, and to the Heraeus Foundation that funds this lovely event. Many thanks are also due to a long list of collaborators and colleagues over the years that have led to the development of the ideas presented here, particularly Markus Aspelmeyer, Valerie Domcke, Manthos Karydas, Akira Matsumura, Jess Riedel, Nick Rodd, Roshni Singh, Philip Stamp, and Jake Taylor. I would also like to heartily thank my student Kai-Isaak Ellers and postdoc Giuseppe Fabiano, who helped prepare the materials in Lecture 1, which will be published in a separate tutorial paper. Finally, I am hugely grateful to the people and organizations who fund this work: the U.S. Department of Energy (Office of High Energy Physics, under Contract No. DEAC02-05CH11231) and the Heising-Simons Foundation (under Award No. 2023-4467).

Throughout this document, I use the following conventions: the metric signature is $(-+++)$, $\hbar = c = 1$ but G_N and k_B are kept explicit, the reduced Planck mass $M_{\rm pl}^2 = 1/8\pi G_N$, and plane wave states are non-relativistically normalized $\langle \mathbf{k}|\mathbf{k}'\rangle = \delta^3(\mathbf{k}-\mathbf{k}')$. The point with this last choice is for ease of transition between relativistic and non-relativistic calculations, and to keep compatibility with Weinberg's textbooks (especially his wildly underrated Lectures on Quantum Mechanics, the latter chapters of which largely inspired the material in Lecture 1). Gravitational perturbations are normalized to be dimensionless $[h_{\mu\nu}] = 1$, i.e., in units of strain as usually reported in gravitational wave detection. Numerical values are quoted in a random mix of SI and natural units based on their typical use in various subfields, without apology.

0. Why study quantum gravity way below the Planck scale?

Most of the historical and deep questions about quantum gravity are ultraviolet in nature: what happens at singularities, what happens to spacetime at very fine distances, how is matter unified with gravity at very high energies, what happens at the endpoint of black hole evaporation, etc.

In these lectures, we will instead study quantum gravity at low energies. By low energy, I mean situations where curvature is small in Planck units. We are going to be focused on an elementary, observationally-minded question: is gravity actually quantized in nature? And if so, is it quantized as a theory of gravitons by analogy with the other gauge forces, or in some more sophisticated way?

Since the other fundamental forces are known to be quantized, it may seem almost certain that gravity should be similarly quantized, at least perturbatively. And, as I will emphasize in Lecture 1, quantizing gravity this way produces a perfectly self-consistent, predictive model. This is in contrast to statements one often hears that "we don't know how to quantize gravity". So why study this? Let me borrow some philosophy from Steve Weinberg [2]:

> It would be difficult to pretend that the gravitational infrared divergence problem is very urgent. My reasons for now attacking this question are: (1) Because I can. There still does not exist any satisfactory quantum theory of gravitation, and in lieu of such a theory it would seem well to gain what experience we can by solving any problems that can be solved with the limited formal apparatus already at our disposal... (2) Because something might go wrong, and that would be interesting.

In particular, we are entering an era where the question of whether gravity is quantized or not is no longer an issue of theoretical consistency but can be addressed by real experiments [1]. And so we can really check if something "goes wrong".

The experiments that have been conceived so far generally fall into two categories. The first are tests of "non-classical" properties of the gravitational interaction, such as tests of whether gravity can entangle a pair of massive systems. The other are tests of the reversibility or unitarity of time evolution under gravity. These are depicted schematically in Fig. 1. We will study a number of concrete realizations in detail in Lectures 3 and 4.

In my opinion, the right way to address the question of whether gravity is quantized and the experiments addressing it is by analogy with tests of quantum me-

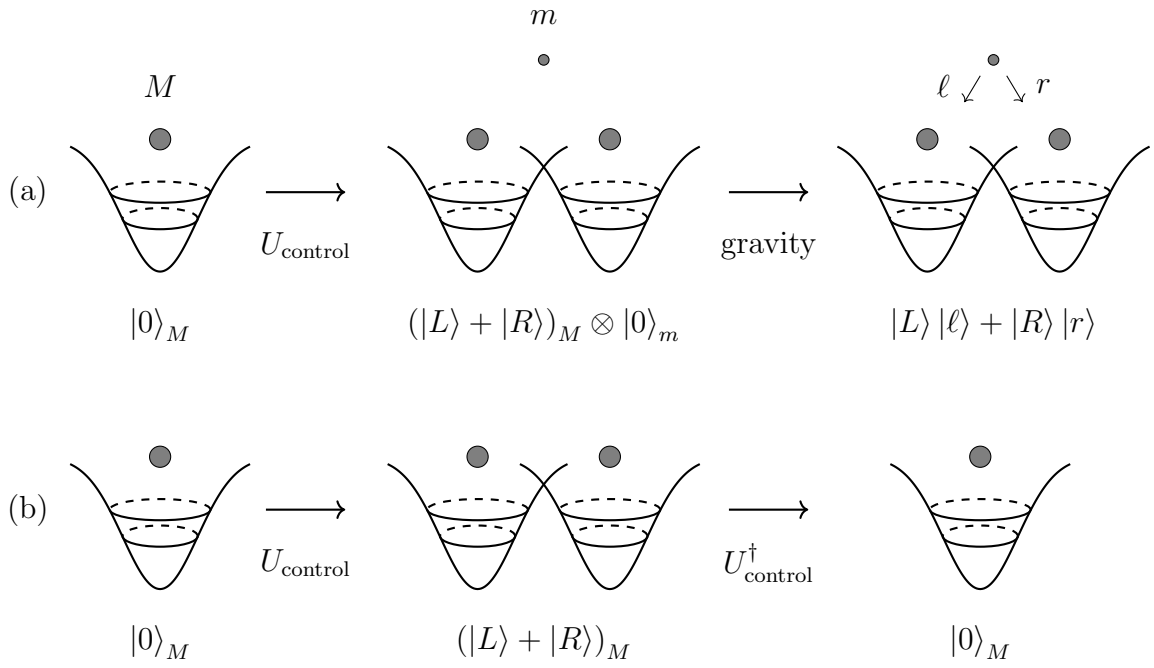

Figure 1: (a) Gravitational entanglement test. Mass M is superposed by some experimental control protocol, and a test mass m interacts. The question is whether the gravitational time evolution generates the entangled state shown in the third panel. (b) Gravitational reversibility test. The question is now whether gravity caused some non-unitary evolution during the second panel, which would ruin our ability to coherently recombine the two branches with $U^\dagger_{\text{control}}$ in the third panel.

chanics like Bell inequality measurements. No experiment can "prove that gravity is quantized", much like Bell tests do not "prove that the world is quantum mechanical". What experiments do is distinguish between null and alternative hypotheses. What Bell did was construct an alternative hypothesis to quantum mechanics, which he called local hidden variables. He showed that if the world was described by local hidden variables, certain correlation functions had to satisfy a particular upper bound [3]. The experiments of Clauser [4] and Aspect [5] then showed that this alternative hypothesis was inconsistent with observation, by obtaining a larger value of the correlation function, and thus local hidden variables was ruled out.

Very similarly, here we will compare a baseline null hypothesis to a suite of possible alternatives, and analyze our ability to experimentally distinguish these. The null hypothesis is ordinary perturbative quantum gravity, viewed as an effective field

theory with action

$$S = \frac{M_{\rm pl}^2}{2}\int d^4x\sqrt{-g}\left(R + \frac{c_1}{M_{\rm pl}^2}R^2 + \frac{c_2}{M_{\rm pl}^2}R_{\mu\nu}R^{\mu\nu} + \cdots\right) + S_{\rm matter}. \tag{0.1}$$

To really define the model, we assume the metric is perturbative around some background, say flat spacetime,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad h_{\mu\nu} \to \hat{h}_{\mu\nu} \tag{0.2}$$

where the hat means we promote the perturbations to a field. As we will discuss extensively, this model is internally self-consistent and makes detailed predictions as long as the curvature scale $R/M_{\rm pl}^2 \ll 1$, so that the higher order terms can be neglected [6, 7]. For example, in a gravitational wave detectable at LIGO, we have $R/M_{\rm pl}^2 \approx 10^{-100}$. At the surface of the binary merger creating this wave, $R/M_{\rm pl}^2 \approx 10^{-75}$. The largest plausible value I know of this ratio is in inflation assuming a very high scale, when $R/M_{\rm pl}^2$ could be as large as about 10^{-10}.

What about the alternative hypotheses? Here the landscape is much more wild, and we will have to analyze a variety of phenomenologies. It is tempting to say something like, "well, what if gravity is just classical"? But that is ill-defined, because it does not specify how matter would couple to gravity. Nevertheless the basic idea for alternatives is simple: you have to violate one of the basic rules of effective quantum field theory where $\hat{h}_{\mu\nu}$ is the fundamental field. For example, in Lecture 2, we will analyze models where time evolution is fundamentally non-unitary [8–12], or in which $h_{\mu\nu}$ arises as a kind of thermodynamic order parameter rather than a fundamental field [13]. These models are distinguished by their predictions in the kinds of experiments shown in Fig. 1, where they give different phenomenologies than perturbative quantum gravity. They are either non-entangling, non-reversible, or both.

At present, we have never done an experiment that can definitively distinguish between these hypotheses. There are plenty of experiments where an external, classical gravitational field acts on a quantum system: among others, matter-wave interferometry [14] and gravitational wave detection [15] fall into this category. And of course, these experiments are also consistent with perturbative quantum gravity. But what we really need to do is some experiment that would require the gravitational field to be quantized to explain the results (or a measurement inconsistent with that hypothesis!). Any such experiment will be exceedingly difficult, but as we will see in Lectures 3 and 4, we are already able to make limited experimental statements. And in my opinion, definitive statements are not far beyond the horizon.

1. Fundamentals: how to quantize gravity at low energy

To begin these lectures, we are going to discuss the most conservative approach to quantum gravity at low energies. The idea is natural and straightforward: the other gauge forces in nature are known to be quantized and we will simply quantize gravity the same way. More specifically, we can quantize small gravitational perturbations around a fixed classical background metric $\overline{g}_{\mu\nu}$. Schematically, putting hats on operators, we will study a model based on the expansion

$$\hat{g}_{\mu\nu} = \overline{g}_{\mu\nu} + \hat{h}_{\mu\nu}, \tag{1.1}$$

where h is going to be small in the sense that it has curvature $R/M_{\rm pl}^2 \ll 1$ as discussed in the introduction. The quantized perturbations are called gravitons.

While actually setting up this model in detail is somewhat involved, the basic punchline is that we can consistently view this framework, which we will refer to as perturbative quantum gravity, as a self-consistent effective quantum field theory. By self-consistent I mean there is a clear set of rules that one can use to make predictions for experiments, and these predictions are finite and so forth, just like predictions from quantum electrodynamics or the Standard Model. This is usually viewed as totally obvious by people trained in high energy theory, but often mystical or even wrong by people from other communities, e.g., quantum foundations.

In this lecture, I am going to present perturbative quantum gravity in the most concrete, simple setting possible: a Hamiltonian framework for gravitons coupled to non-relativistic point particles. The reason for this is that the whole model is *exactly* identical to the theory of non-relativistic point charges coupled to photons, and at this point in history there can be no doubt that this latter model is well-defined and useful for experimental predictions![1] Moreover, all the experiments we will discuss later are in the non-relativistic limit (at least for the detectors) and so this turns out to be the most practically useful setting for real calculations.

[1]The one important difference is background dependence, as discussed in the introduction. In QED, you can expand $\hat{A}_\mu = \overline{A}_\mu + \delta\hat{A}_\mu$ in analogy with Eq. (1.1). The difference in QED is that we know exactly how $\overline{A}_\mu$ arises as a coherent state solution of QED itself, whereas in gravity our understanding of what quantum model $\overline{g}_{\mu\nu}$ solves is much less clear. Thus Eq. (1.1) should be viewed as a fundamental assumption in perturbative quantum gravity.

1.1. Derivation of Hamiltonian

We will study non-relativistic point particles coupled to perturbations around flat spacetime $\overline{g}_{\mu\nu} = \eta_{\mu\nu}$. Our goal is to derive a Hamiltonian for the joint system of these particles and the gravitational perturbations. In this section everything will be classical, and then in Sec. 1.3 we will see how to quantize the model.

There are many ways to derive the Hamiltonian we want. In a modern effective field theory approach, the way to proceed would be to identify all the symmetries of the problem and then write all the possible terms allowed by these symmetries. A wonderful treatment of this approach for non-relativistic QED can be found in [**?**]. Here the symmetries would be Galilean invariance and the non-relativistic limit of diffeomorphism invariance. It turns out to be pretty subtle to carefully codify all of that. Thus, here we will follow a more traditional approach and just do the canonical analysis of the Einstein-Hilbert action expanded around flat spacetime. Since this is a gauge theory we will have to deal with the usual difficulties of taking the Legendre transform from the Lagrangian to Hamiltonian in a constrained system [**?**].

The action is

$$S = \int d^4x\,\sqrt{-g}\left(M_{\rm pl}^2 R + \mathcal{L}_{\rm counterterms}\right) + S_{\rm mat}, \quad S_{\rm mat} = -\sum_n m_n \int dt\,\sqrt{-g_{\mu\nu}\dot{x}_n^\mu \dot{x}_n^\nu}. \tag{1.2}$$

We will discuss the counter-terms below. The $x_n^\mu = x_n^\mu(t)$ in the matter action are worldlines, here parametrized by an arbitrary (not necessarily proper) time t.[2] Trying to follow standard canonical methods, one immediately encounters the difficulty of deciding what is meant by the canonical momenta $\pi \sim \delta\mathcal{L}/\delta\dot{g}$ for the metric components: what time coordinate does the dot represent?

The usual solution is to work with the ADM framework and we will follow that here. I will try to speedrun through this formalism a bit, but interested readers are referred to the excellent textbook by Poisson [] for a detailed treatment. Poisson treats the problem non-perturbatively and includes boundary terms. Here I will drop the boundary terms, and work perturbatively. The perturbative treatment given here will appear in [**?**]. To my great surprise, I could not find a perturbative treatment like this anywhere in the literature, which is why we worked it out ourselves. However, the analogous treatment for small perturbations in inflation can be found in [], and the treatment for the particles is inspired by [**?**].

[2]The overall sign is negative to give a positive rest mass in the energy, as we will see shortly.

We write the metric in the form

$$ds^2 = -N^2 dt^2 + \gamma_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \tag{1.3}$$

where N, N_i, and γ_{ij} are functions of all four coordinates, known as the lapse, shift, and induced metric respectively. This decomposition is completely general as long as spacetime is globally hyperbolic $\mathcal{M} = \mathbf{R} \times \Sigma$. The advantage of this decomposition is that we now have a preferred time coordinate t and so canonical variables are defined in straightforward fashion. Of course this choice of t was totally arbitrary, reflecting coordinate invariance, and we will see this gauge symmetry show up as a constraint shortly.

First we want to take the metric (1.3) and rewrite the action (1.2) in terms of quantities on spatial slices. This part can be done non-perturbatively. The 3d geometry is described by the induced covariant derivative and connection coefficients

$$D_i A_j = \partial_i A_j - \Gamma^k_{ij} A_k, \quad \Gamma^k_{ij} = \frac{1}{2}\gamma^{k\ell}\left(\partial_i \gamma_{j\ell} + \partial_j \gamma_{\ell i} - \partial_\ell \gamma_{ij}\right), \tag{1.4}$$

where A_i is any spatial vector field, and the generalization to the derivative acting on any tensor is the same as in ordinary 4d geometry. The indices i, j run only over spatial coordinates. In terms of this derivative, the extrinsic curvature of the spatial slices is given by

$$K_{ij} = \frac{1}{2N}\left(\dot{\gamma}_{ij} - D_i N_j - D_j N_i\right). \tag{1.5}$$

A somewhat involved exercise then shows that the 4d Ricci scalar can be written as

$$R = {}^{(3)}R + K^2 - K^{ij}K_{ij} + \text{total derivative}, \tag{1.6}$$

with $K = \gamma^{ij}K_{ij}$, and ${}^{(3)}R$ the spatial Ricci scalar. This is defined as usual but with D_i instead of ∇_μ, i.e., the 3d Riemann tensor is

$${}^{(3)}R^i_{jk\ell} = \partial_k \Gamma^i_{j\ell} - \partial_\ell \Gamma^i_{jk} + \Gamma^i_{km}\Gamma^m_{j\ell} - \Gamma^i_{\ell m}\Gamma^m_{jk}, \tag{1.7}$$

the 3d Ricci tensor is ${}^{(3)}R_{j\ell} = {}^{(3)}R^k_{jk\ell}$, and so ${}^{(3)}R = \gamma^{j\ell}\,{}^{(3)}R_{j\ell}$. Finally, the metric determinant is $\sqrt{-g} = N\sqrt{\gamma}$. Again, see Poisson for derivations of all these geometric identities. We also want to write the matter action in these variables. The proper length appearing in the action can be written

$$s^2 := -g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu = \left(N^2 - \gamma_{ij}N^i N^j\right) - 2\gamma_{ij}N^i \dot{x}^j - \gamma_{ij}\dot{x}^i \dot{x}^j \tag{1.8}$$

where we are normalizing $\dot{x}^0 = 1$.[3] So in total, the action is now written

$$S = M_{\rm pl}^2 \int d^4x\, N\sqrt{\gamma}\left({}^{(3)}\!R + K^2 - K^{ij}K_{ij}\right) - \sum_n m_n \int dt\, s_n + S_{\rm counterterms} \tag{1.9}$$

with s_n given by Eq. (1.8), one for each particle n.

At this stage it is helpful to move to perturbation theory to make progress. This means we want to expand the metric around flat space plus small corrections, as in Eq. (1.1), with $\overline{g}_{\mu\nu} = \eta_{\mu\nu}$. In the ADM parametrization (1.3), this means that we perturb

$$N = 1 + n, \quad N_i = n_i, \quad \gamma_{ij} = \delta_{ij} + h_{ij} \tag{1.10}$$

where n, n_i, and h_{ij} are all small. We will think of all of them at the same order, so I will schematically denote expansions in these quantities as $\mathcal{O}(h, h^2, \ldots)$ in the sense that a quantity like n^2 is $\mathcal{O}(h^2)$. To get linearized gravity, we need the action evaluated up to $\mathcal{O}(h^2)$. It is straightforward to use Eqs. (1.4), (1.5), and (1.10) to obtain

$$K_{ij} = \frac{1}{2}\left(\dot{h}_{ij} - \partial_i n_j - \partial_j n_i\right) + \mathcal{O}(h^2), \quad K = \frac{1}{2}\left(\dot{h} - 2\partial_i n^i\right) + \mathcal{O}(h^2), \tag{1.11}$$

which is all we need since these appear squared in the action. It takes a little more work to get ${}^{(3)}\!R$ to $\mathcal{O}(h^2)$. I record the detailed calculation in Appendix A. The main point is that ${}^{(3)}\!R$ itself has terms of both $\mathcal{O}(h)$ and $\mathcal{O}(h^2)$, and then when we multiply the $\mathcal{O}(h)$ by the determinant

$$\sqrt{-g} = N\sqrt{\gamma} = 1 + n + \frac{1}{2}h + \mathcal{O}(h^2) \tag{1.12}$$

some stay $\mathcal{O}(h)$ and others become $\mathcal{O}(h^2)$. In any case, the result is

$$\sqrt{-g}\,{}^{(3)}\!R = n\left(\partial^i\partial^j h_{ij} - \nabla^2 h\right) + F(h) + \mathcal{O}(h^3), \tag{1.13}$$

with

$$F(h) = \frac{1}{4}(\nabla h)^2 - \frac{1}{4}(\nabla h_{ij})^2 - \frac{1}{2}\partial_i h \partial_j h^{ij} + \frac{1}{2}(\partial_i h^{ij})^2. \tag{1.14}$$

Here and afterward $\nabla_i = \partial_i$ just means the spatial derivative. In writing this, we have integrated by parts in various places and dropped all total derivatives, including the $\mathcal{O}(h)$ terms, so everything is now $\mathcal{O}(h^2)$. Finally, we will want the coupling of

[3] **(Probably more rigorous to also let $\dot{x}^0$ be a variable and then talk about reparametrization invariance –dc)**

matter to gravity. For our purposes, the linear h coupling is sufficient. The worldline Lagrangian in Eqs. (1.8) and (1.9) expands out to

$$s = 1 - \frac{1}{2}\dot{\mathbf{x}}^2 + n - n_i\dot{x}^i - \frac{n}{2}\dot{\mathbf{x}}^2 - \frac{1}{2}h_{ij}\dot{x}^i\dot{x}^j + \mathcal{O}(h^2). \tag{1.15}$$

Note that I expanded the pure kinetic term here also, i.e., took the non-relativistic limit $|\dot{\mathbf{x}}|^2 \ll 1$.

Finally, we can compute the canonical momenta needed to go to the Hamiltonian framework. Putting the above results together we obtain the action to desired order in the gravitational perturbations:

$$\begin{aligned} S = M_{\rm pl}^2 \int d^4x \left[\frac{1}{4}\left(\dot{h} - 2\partial_i n^i\right)^2 + \frac{1}{4}\left(\dot{h}_{ij} - \partial_i n_j - \partial_j n_i\right)^2 + n\left(\partial^i\partial^j h_{ij} - \partial_i\partial^i h\right) + F(h)\right. \\ \left. + \sum_n m_n \int dt \left[1 - \frac{1}{2}\dot{\mathbf{x}}^2 + n - n_i\dot{x}^i - \frac{n}{2}\dot{\mathbf{x}}^2 - \frac{1}{2}h_{ij}\dot{x}^i\dot{x}^j\right]. \end{aligned} \tag{1.16}$$

Note that the lapse n and shift n_i appear without time derivatives, so they are non-dynamical in the sense that

$$\pi_n = \frac{\partial\mathcal{L}}{\partial\dot{n}} = 0, \quad \pi_{n_i} = \frac{\partial\mathcal{L}}{\partial\dot{n}^i} = 0. \tag{1.17}$$

The momenta for the spatial metric are

$$\pi_{ij} = \frac{\partial\mathcal{L}}{\partial\dot{h}_{ij}} = \frac{1}{2}(\dot{h}_{ij} - \delta_{ij}\dot{h} - \partial_i n_j - \partial_j n_i + 2\delta_{ij}\partial_k n^k). \tag{1.18}$$

This equation and its trace can be easily inverted to write $\dot{h}_{ij}$ as a function of the momenta,

$$\dot{h}_{ij} = 2\pi_{ij} - \pi\delta_{ij} + \partial_i n_j + \partial_j n_i. \tag{1.19}$$

The momenta for the matter particles, from Eq. (1.16), is given by

$$p_n^i = \frac{\partial\mathcal{L}}{\partial\dot{x}_{n,i}} = m\left[\dot{x}_n^i - n\dot{x}_n^i + n^i + h^{ij}\dot{x}_{n,j}\right]. \tag{1.20}$$

which can similarly be inverted to obtain $\dot{x}$ as a function of p:

$$\dot{x}_n^i = \frac{p_n^i}{m} + n\frac{p_n^i}{m} - n^i - \frac{h^{ij}p_{n,j}}{m} + \mathcal{O}(h^2). \tag{1.21}$$

It it worth checking this formula by plugging it back into Eq. (1.20).

Using these formulas for the momenta, we can now compute the Hamiltonian. The Legendre transformation is as usual

$$H = \int d^3\mathbf{x}\, \pi_{ij}\dot{h}^{ij} + \sum_n p_{n,i}\dot{x}^i_n - L. \tag{1.22}$$

Let's do this piece by piece. Using Eq. (1.19) and the first line of Eq. (1.16), a short calculation gives the pure metric part of the Hamiltonian:

$$H_{\rm grav} = M_{\rm pl}^2 \int d^3\mathbf{x}\, \left[\pi_{ij}\pi^{ij} - \frac{1}{2}\pi^2 - F(h) + 2\pi_{ij}\partial^i n^j \pi_{ij}\right]. \tag{1.23}$$

Using Eq. (1.21) and the second line of Eq. (1.16), the matter part is, again to $\mathcal{O}(h)$ and in the small-velocity limit,

$$H_{\rm mat} = \sum_a m_a + \frac{\mathbf{p}_a^2}{2m_a} - \frac{h_{ij}(\mathbf{x}_a)p_a^i p_a^j}{2m_a} + n\left[m_a + \frac{\mathbf{p}_a^2}{2m_a}\right] - n^i p_{a,i}. \tag{1.24}$$

We see the usual rest mass and kinetic energy of the particles, their linearized coupling to the metric perturbations, plus the energy and momentum multiplying the lapse and shift respectively. The third line of Eq. (1.16) transforms trivially since it does not involve momenta. Putting all these results together, and collecting the terms involving the lapse and shift, we have the full Hamiltonian:

$$\begin{aligned} H = {} & M_{\rm pl}^2 \int d^3\mathbf{x}\, \left[\pi_{ij}\pi^{ij} - \frac{1}{2}\pi^2 - F(h)\right] + \sum_a \frac{\mathbf{p}_a^2}{2m_a} - \frac{h_{ij}(\mathbf{x}_a)p_a^i p_a^j}{2m_a} \\ & + \int d^3\mathbf{x}\, n \left[M_{\rm pl}^2\left(\partial^i\partial^j h_{ij} - \partial_i\partial^i h\right) + \sum_a \left(m_a + \frac{\mathbf{p}_a^2}{2m_a}\right)\delta^3(\mathbf{x} - \mathbf{x}_a(t))\right] \\ & + \int d^3\mathbf{x}\, n^i \left[-2M_{\rm pl}^2\partial^j\pi_{ij} + \sum_a p_{a,i}\delta^3(\mathbf{x} - \mathbf{x}_a(t))\right]. \end{aligned} \tag{1.25}$$

To get the shift vector appearing as an overall prefactor, we integrated by parts on $\pi_{ij}\partial^i n^j$. We also dropped the rest mass term in the matter Hamiltonian since it is just an overall constant.

In our expression for the Hamiltonian (1.25), the first line is what you should expect by analogy with, say, QED. It has the kinetic terms for both the gravitational perturbations and matter, and their linearized coupling. The other two lines are pure constraints: since $\pi_n = \pi_{n_i} = 0$, the terms in brackets next to these have to vanish. What we need to do to get a final Hamiltonian is to find some parametrization of the space of solutions to the constraints. Plugging these solutions back in then gives a Hamiltonian that acts only on the "physical" degrees of freedom, i.e., those that are

free to vary. This is a notoriously painful process in general and this setting will not be an exception.

Here is what we will do. First of all, so far we have not fixed a gauge. Let's use this one:

$$\partial_i h^{ij} = 0, \quad \pi = 0. \tag{1.26}$$

Notice that, in particular, since we have matter, we cannot use the commonly seen transverse-traceless (TT) gauge, where we would also set $h = 0$. This would be inconsistent with the n constraint, which is basically Gauss' law. However, with $\partial_i h^{ij} = 0$, we can solve this constraint explicitly

$$h(\mathbf{x}) = \frac{1}{M_{\rm pl}^2} \sum_a \frac{m_a}{|\mathbf{x} - \mathbf{x}_a|} + \mathcal{O}\left(\frac{p^2}{m^2}\right). \tag{1.27}$$

Including the momentum piece of this Gauss law, as well as solving the n^i constraint, is possible but painful. I relegate the details to an appendix. However, one can see what will basically happen from the structure of Eq. (1.27). Plugging this back into the Hamiltonian will give the Newtonian interactions. For example in the $F(h)$ term we get a contribution

$$M_{\rm pl}^2 \int d^3\mathbf{x}\, h\nabla^2 h = \frac{1}{M_{\rm pl}^2} \sum_{a\neq b} \frac{m_a m_b}{|\mathbf{x}_a - \mathbf{x}_b|} + \mathcal{O}\left(\frac{p^2}{m^2}\right). \tag{1.28}$$

and so forth. Here we dropped the diagonal terms again because they are just an overall constant in the energy. In any case, following this through, we end up with the following Hamiltonian:

$$\begin{aligned} H = {} & \frac{M_{\rm pl}^2}{2} \int d^3\mathbf{x} \left[\overline{\pi}_{ij}\overline{\pi}^{ij} + \overline{h}_{ij}\nabla^2 \overline{h}^{ij}\right] + \sum_a \frac{\mathbf{p}_a^2}{2m_a} \\ & - \sum_a \frac{\overline{h}_{ij}(\mathbf{x}_a) p_a^i p_a^j}{2m_a} - \frac{1}{M_{\rm pl}^2} \sum_{a\neq b} \frac{m_a m_b}{|\mathbf{x}_a - \mathbf{x}_b|} + H_{\rm counterterms} \\ & + \frac{1}{M_{\rm pl}^2} \sum_{a\neq b} \frac{m_a m_b}{|\mathbf{x}_a - \mathbf{x}_b|} \left(\frac{3\mathbf{p}_a^2}{2m_a^2} - \frac{3\mathbf{p}_b^2}{2m_b^2} + \frac{7\mathbf{p}_a\cdot\mathbf{p}_b}{2m_a m_b} + \cdots\right). \end{aligned} \tag{1.29}$$

The overlines denote the transverse traceless components of the fields.[4] In what follows we will usually drop the overline.

In the rest of these lectures, we are going to use the top two lines of Eq. (1.29) as our basic theory of quantum gravity. Well, we have to quantize it, which we'll do

[4]Formally, this is defined via the general decomposition $h_{ij} = \overline{h}_{ij} + \partial_i V_j + \partial_j V_i + \frac{1}{3}\delta_{ij}\left(h - 2\partial_k V^k\right)$, where V_i is a vector and h is the trace. We will say a bit more in the next section.

next; this is purely classical so far. But the form should look extremely familiar and comfortable. The first line is just kinetic energies for the field and particles, and then the second line gives the graviton-matter and Newtonian matter-matter interactions.

The dots in the last line of Eq. (1.29) represent a further set of terms of $\mathcal{O}(p^2/m^2)$ which are correcting the Newtonian interaction. We will not use these explicitly anywhere in what follows; they are just the leading relativistic corrections to the $1/r$ interaction. I just highlight them to compare to the kinds of expressions you will find in other reviews on perturbative quantum gravity as an effective QFT [6, 7]. Notice that all of these terms here are still purely classical! There are also quantum corrections. We will look at how terms of that type are generated later in this lecture.

1.2. Expansion into gravitational waves

We can write the field in terms of a mode expansion to connect to usual gravitational wave/graviton physics. This will be useful when we do perturbation theory by treating the first line of Eq. (1.29) as the free Hamiltonian. The equations of motion for h_{ij} and π_{ij} under just the free Hamiltonian are of course[5]

$$\ddot{\overline{h}}_{ij} = \nabla^2 \overline{h}_{ij} \tag{1.30}$$

which can be solved by plane waves

$$\overline{h}_{ij}(\mathbf{x},t) = \int d^3\mathbf{k}\, \alpha_{ij}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}-\omega_k t} + \text{c.c.}, \quad \omega_k = |\mathbf{k}|, \quad k_i \alpha^{ij}(\mathbf{k}) = 0, \quad \alpha^i_i(\mathbf{k}) = 0. \tag{1.31}$$

The last two equalities are the tracelessness of $\overline{h}_{ij}$. As is well known, for a given spatial vector $\mathbf{k}$, we can find orthogonal basis traceless tensors $\epsilon_{ij,s}(\mathbf{k})$ with $s = 1, 2$. These satisfy the completeness relation

$$\epsilon_{ij,s}(\mathbf{k})\epsilon_{k\ell,s}(\mathbf{k})^* = T_{ijk\ell}(\mathbf{k}) = \frac{1}{2}\left[P_{ik}P_{j\ell} + P_{i\ell}P_{jk} - P_{ij}P_{k\ell}\right], \quad P_{ij} = \delta_{ij} - \frac{k_i k_j}{k^2}. \tag{1.32}$$

In terms of these, we can write the mode expansion as

$$h_{ij}(\mathbf{x},t) = \sum_{s=1,2} \int \frac{d^3\mathbf{k}}{\sqrt{2\omega_k} M_{\rm pl}} \epsilon_{ij,s}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}-\omega_k t} b_{\mathbf{k},s} + \text{c.c.}, \tag{1.33}$$

where finally we have dropped the overline in the notation as promised. At this stage, $b_{\mathbf{k},s}$ is just a classical Fourier coefficient, normalized here to have dimensions of $1/M^{-3/2}$.

[5] **(Should justify this by working through the Poisson brackets –dc)**

1.3. Quantization

Finally, we will quantize this model. The reason for waiting so long is because we are going to want to compare "classical" gravity models to the quantum one, and all of the steps we have just gone through work just the same if we treat gravity classically or quantum mechanically. The heart of quantum mechanics is to promote the dynamical degrees of freedom into operators, and to assign them canonical commutation relations:

$$[\hat{b}_{\mathbf{k},s}, \hat{b}^\dagger_{\mathbf{k}',s'}] = \delta_{ss'}\delta^3(\mathbf{k}-\mathbf{k}'), \quad [\hat{\mathbf{x}}_a, \hat{\mathbf{p}}_b] = i\delta_{ab} \tag{1.34}$$

and all other commutators vanishing.

1.4. Renormalization

Armed with the quantum theory of gravitons coupled to point particles, let's calculate something non-trivial. In particular, to emphasize the way this model gives finite predictions out of naively infinite calculations, let's study an example where we have to analyze renormalization: the gravitational Lamb shift. We are going to follow the treatment by Bethe [16], which is slightly heuristic but gets the essential physics (and the numerical value in QED!) basically correct. A much more thorough calculation in QED can be found in Weinberg's QFT Volume 1 [17].

Consider a single particle of mass m gravitationally bound to a large system of mass M. In the standard Lamb shift, this would be an electron and proton, respectively. In gravity, a cute example would be an axion bound to a black hole, a subject of much recent study (see e.g. [18, 19]). The bound state wavefunctions $|n\rangle$ of the relative position coordinate (i.e., the position of the light particle in the frame of the heavy one) have a discrete spectrum, of course, given by

$$E_n^{(0)} = -\frac{m\alpha^2}{2n^2}, \tag{1.35}$$

plus perturbative corrections. For ease of comparison to the standard hydrogen calculations: in units where $V_{\text{Coulomb}} = -e^2/4\pi r$, the levels are $E_n^{(0)} = m\alpha_{\text{EM}}^2/2n^2$ with $\alpha_{\text{EM}}^2 = e^2/4\pi$. So in our conventions for gravity $V_{\text{Newton}} = -G_N Mm/r$, we have $\alpha_{\text{grav}}^2 = G_N Mm$. The superscript 0 means these are the eigenvalues of the "unperturbed" Hamiltonian

$$H_0 = \frac{\hat{\mathbf{p}}^2}{2m} - \frac{\alpha^2}{|\hat{\mathbf{x}}|}. \tag{1.36}$$

The Lamb shift is the perturbation to the energy levels ΔE_n due to virtual fluctuations of the electromagnetic or gravitational field. More specifically, in hydrogen-like atoms,

as we will see below, the 2S state ($|n\ell m\rangle = |200\rangle$) gets a shift to its energy level, whereas the 2P state ($|21 \pm 1\rangle$) does not. Thus the frequency of light that is resonant on the 2S $\to$ 2P transition is slightly different than one would expect from Eq. (1.35), by a factor of about 0.1% [16] and this can be measured in experiments by trying to drive the transition with a microwave field of controllable frequency.

Let us study the calculation of the gravitational version of this phenomenon. In both QED and gravity, the energy shift starts at second order in the coupling α^2 (i.e. at $\mathcal{O}(\alpha^4)$), and is ultraviolet divergent. We are going to consider the effects of two terms in the Hamiltonian:

$$\delta H = V + c_1(\mathbf{p}^2)^2, \quad V = \frac{1}{m}\hat{h}_{ij}(\hat{\mathbf{x}})\hat{p}^i\hat{p}^j \tag{1.37}$$

where $\hat{h}_{ij}(\hat{\mathbf{x}})$ has the mode expansion given in Eq. (1.33). In Eq. (1.37), the first term is the ordinary linearized gravity-matter interaction. The second term is a counterterm which will be needed for renormalization. The coefficient c_1 at this stage is a free parameter. It will actually be useful to begin with free particle states (rather than the bound ones), which we can use to fix the divergent part of the counterterm coefficient c_1.

With the perturbation (1.37), the first-order energy shift in the $|\mathbf{p}\rangle$ state is

$$\Delta E^{(1)}_{\mathbf{p}}\delta^3(0) = \frac{1}{m}\langle\mathbf{p}|\hat{h}_{ij}(\hat{\mathbf{x}})\hat{p}^i\hat{p}^j|\mathbf{p}\rangle + c_1\langle\mathbf{p}|(\hat{\mathbf{p}}^2)^2|\mathbf{p}\rangle = c_1(\mathbf{p}^2)^2. \tag{1.38}$$

The $\delta^3(0)$ is needed because we are using continuum-normalized states $\langle\mathbf{p}|\mathbf{p}'\rangle = \delta^3(\mathbf{p}-\mathbf{p})$. In this formula, the state $|\mathbf{p}\rangle$ really means $|\mathbf{p}\rangle_{\text{matter}}\otimes|0\rangle_{\text{gravitons}}$ but we will suppress the graviton state when it is just the vacuum. Thus the first term in Eq. (1.38) vanishes, because the graviton part $\sim \langle 0|\hat{b}_{\mathrm{k,s}}|0\rangle = 0$. The same statement holds in QED, of course. The second term, however, gives the final non-zero contribution $c_1(\mathbf{p}^2)^2$. This will play an essential role in obtaining a finite answer.

Now we calculate the second order term. In a bound state, this will give the Lamb shift; here we begin with a free particle to show how the renormalization works. The second order perturbation to the energy of a free particle of momentum $\mathbf{p}$ is

$$\Delta E^{(2)}_{\mathbf{p}}\delta^3(0) = \sum_{\beta}\frac{\langle\mathbf{p}|V|\beta\rangle\langle\beta|V|\mathbf{p}\rangle}{E^{(0)}_{\mathbf{p}} - E^{(0)}_{\beta}}, \tag{1.39}$$

where $|\beta\rangle$ is, in principle, a sum over all intermediate states. We only need the ones connected to $|\mathbf{p}\rangle_{\text{matter}}\otimes|0\rangle_{\text{gravitons}}$ through a single insertion of V. That means $|\beta\rangle = |\mathbf{q}\rangle_{\text{matter}}\otimes|\mathbf{k},s\rangle_{\text{gravitons}}$, i.e.,

$$\Delta E^{(2)}_{\mathbf{p}}\delta^3(0) = \sum_{s}\int d^3\mathbf{k}d^3\mathbf{q}\,\frac{\langle\mathbf{p}|V|\mathbf{q},\mathbf{k}s\rangle\langle\mathbf{q},\mathbf{k}s|V|\mathbf{p}\rangle}{E^{(0)}_{\mathbf{p}} - E^{(0)}_{\mathbf{q}} - \omega_k}, \tag{1.40}$$

with $\omega_k = |\mathbf{k}|$ the energy of the graviton. We will evaluate the matrix element in the approximation that $e^{i\mathbf{k}\cdot\mathbf{x}} \approx 1$ in the $h_{ij}(\mathbf{x})$. In QED this is called the dipole approximation and for gravity it would analogously be called the quadrupole approximation, a piece of terminology we will justify retroactively. In this approximation, the matrix element factorizes and is simple to evaluate:

$$\begin{aligned}\langle \mathbf{q}, \mathbf{k}s|V|\mathbf{p}\rangle &\approx \frac{1}{m}\langle \mathbf{q}|\hat{p}^i\hat{p}^j|\mathbf{p}\rangle \times \sum_{s'}\int \frac{d^3\mathbf{k}'}{\sqrt{2\omega_k}M_{\rm pl}}\langle \mathbf{k}s|\epsilon^s_{ij}(\mathbf{k})\hat{b}^\dagger_{\mathbf{k}',s'}|0\rangle \\ &= \frac{p^ip^j\epsilon^s_{ij}(\mathbf{k})}{mM_{\rm pl}\sqrt{2\omega_k}}\delta^3(\mathbf{q}-\mathbf{p}).\end{aligned} \tag{1.41}$$

Putting this back into Eq. (1.40), we obtain

$$\begin{aligned}\Delta E^{(2)}_{\mathbf{p}} &= \frac{1}{m^2M_{\rm pl}^2}\sum_s\int\frac{d^3\mathbf{k}}{2\omega_k^2}p^ip^jp^kp^\ell\epsilon^s_{ij}(\mathbf{k})\epsilon^{s*}_{k\ell}(\mathbf{k}) \\ &= \frac{1}{m^2M_{\rm pl}^2}\int\frac{d^3\mathbf{k}}{4\omega_k^2}\left[\mathbf{p}^2-\left(\frac{\mathbf{p}\cdot\mathbf{k}}{k^2}\right)^2\right] \\ &= \frac{1}{m^2M_{\rm pl}^2}\frac{16\pi}{15}(\mathbf{p}^2)^2\int_0^\Lambda dk.\end{aligned} \tag{1.42}$$

To get the second line, we used the polarization sum in Eq. (1.32), and the third line is a simple integration over the angular variables. We see that there is a linearly UV divergent contribution to this energy shift, who overall coefficient goes like $(\mathbf{p}^2)^2$. This is where the counterterm contribution in the *first*-order calculation, Eq. (1.38), comes in: we choose $c_1 = c_1(\Lambda)$ so that the total shift is finite. Specifically, if we choose

$$c_1(\Lambda) = -\frac{1}{m^2M_{\rm pl}^2}\frac{16\pi}{15}\Lambda \tag{1.43}$$

then the total shift to the energy is

$$\Delta E_{\mathbf{p}} = \Delta E^{(1)}_{\mathbf{p}} + \Delta E^{(2)}_{\mathbf{p}} + \cdots = 0 + \cdots, \tag{1.44}$$

which is not only finite but actually vanishes. The dots represent higher order terms in perturbation theory which we are of course ignoring. Setting c_1 this way is the same as the usual procedure used to define masses of particles with an on-shell renormalization condition in quantum field theory.

With the counterterm $c_1(\Lambda)$ chosen, we can compute the Lamb shift in the bound state energies E_n. The first order correction again is due entirely to the counterterm

$$\Delta E^{(1)}_n = c_1\langle n|(\hat{\mathbf{p}}^2)^2|n\rangle. \tag{1.45}$$

The second order term is more complicated. We have

$$\begin{aligned}\Delta E_n^{(2)} &= \sum_{n'}\sum_s \int d^3\mathbf{k}\, \frac{\langle n|\hat{V}|n',\mathbf{k}s\rangle\,\langle n',\mathbf{k}s|\hat{V}|n\rangle}{E_n^{(0)}-E_{n'}^{(0)}-\omega_k} \\ &= \frac{1}{m^2M_{\rm pl}^2}\sum_{n'}\int \frac{d^3\mathbf{k}}{2\omega_k}\frac{\langle n|\hat{p}^i\hat{p}^j|n'\rangle\,\langle n'|\hat{p}^k\hat{p}^\ell|n\rangle}{E_n^{(0)}-E_{n'}^{(0)}-\omega_k}T_{ijk\ell}(\mathbf{k}),\end{aligned} \tag{1.46}$$

where again the polarization sum $T_{ijk\ell}$ is given in Eq. (1.32). There are two things to notice here compared to the free-particle calculation. One is that we used the bound states $|n'\rangle$ as the intermediate states on the matter. The free particle plane waves $|\mathbf{p}\rangle$ do not contribute because they are orthogonal to the $|n\rangle$ and the perturbation is diagonal in $|\mathbf{p}\rangle$. The other is that now we have a much less trivial tensor structure in the integrand, because the perturbation is not diagonal in the $|n\rangle$ basis.

To get further, we begin with a nice trick that separates the log and linear UV divergences at the $|\mathbf{k}|\to\infty$ part of the integral. Note that

$$\frac{1}{E-\omega} = \frac{E}{\omega(E-\omega)} - \frac{1}{\omega}, \tag{1.47}$$

and using this we can break up the integral (1.46) into two pieces. The part with the $1/\omega$ factor is

$$\Delta E_n^{(2)} \supset -\frac{1}{m^2M_{\rm pl}^2}\sum_{n'}\int \frac{d^3\mathbf{k}}{2\omega_k^2}\,\langle n|\hat{p}^i\hat{p}^j|n'\rangle\,\langle n'|\hat{p}^k\hat{p}^\ell|n\rangle\,T_{ijk\ell}(\mathbf{k}). \tag{1.48}$$

The sum over intermediate states $|n'\rangle$ now collapses to an identity operator, and the integrand is rotationally symmetric in $\mathbf{k}$ other than $T_{ijk\ell}$. The angular integral over that is given by the simple formula

$$I_{ijk\ell} = \int d\Omega\, T_{ijk\ell}(\mathbf{k}) = \frac{4\pi}{15}\left[3\delta_{ik}\delta_{j\ell}+3\delta_{i\ell}\delta_{jk}-2\delta_{ij}\delta_{k\ell}\right], \tag{1.49}$$

so in total we get the simple linear divergence

$$\Delta E_n^{(2)} \supset -\frac{1}{m^2M_{\rm pl}^2}\frac{8\pi}{15}\,\langle n|(\mathbf{p}^2)^2|n\rangle\,\Lambda. \tag{1.50}$$

But this is exactly cancelled by the counterterm in the first-order calculation, Eq. (1.45), using *the same value* of the Wilson coefficient $c_1(\Lambda)$ that we needed to cancel the free-particle shift in Eqs. (1.43) and (1.44). **(I seem to actually be off by a factor of 1/2, need to track that down –dc)** This is the beauty and, perhaps miracle, of renormalization: one counterterm suffices to cancel the UV divergences appearing in many different observables.

Finally, we want to get the log divergent part of Eq. (1.46), which is what we will use to estimate the Lamb shift. This comes from the first term in Eq. (1.47), and gives the shift

$$\Delta E_n^{(2)} = \frac{1}{m^2 M_{\rm pl}^2} \sum_{n'} \int \frac{d^3\mathbf{k}}{2\omega_k^2} \frac{E_n^{(0)} - E_{n'}^{(0)}}{E_n^{(0)} - E_{n'}^{(0)} - \omega_k} \langle n|\hat{p}^i \hat{p}^j|n'\rangle \langle n'|\hat{p}^k \hat{p}^\ell|n\rangle \, T_{ijk\ell}(\mathbf{k}). \quad (1.51)$$

Once again, the integrand is rotationally symmetric in $\mathbf{k}$ other than the polarization sum $T_{ijk\ell}$. We can use Eq. (1.49) to do the angular integral. The remaining dk integral is easy, and we obtain

$$\Delta E_n^{(2)} = \frac{1}{m^2 M_{\rm pl}^2} \frac{2\pi}{15} \sum_{n'} (E_n^{(0)} - E_{n'}^{(0)}) \ln\left(\frac{\Lambda}{E_n^{(0)} - E_{n'}^{(0)}}\right) \langle n|\hat{p}^i \hat{p}^j|n'\rangle \langle n'|\hat{p}^k \hat{p}^\ell|n\rangle \, I_{ijk\ell}. \quad (1.52)$$

We can further simplify this expression in a few ways. First, the n' dependence in the log is very slow, so we can again follow Bethe and estimate this by just putting an expectation value around the denominator [16]. Next, we can put the linear E terms into the matrix elements by writing them as H_0 appropriately, e.g.,

$$(E_n^{(0)} - E_{n'}^{(0)}) \langle n|\hat{p}^i \hat{p}^j|n'\rangle \langle n'|\hat{p}^k \hat{p}^\ell|n\rangle = \langle n|H_0 \hat{p}^i \hat{p}^j|n'\rangle \langle n'|\hat{p}^k \hat{p}^\ell|n\rangle - \langle n|\hat{p}^i \hat{p}^j H_0|n'\rangle \langle n'|\hat{p}^k \hat{p}^\ell|n\rangle \,. \quad (1.53)$$

With this, the sum over n' again collapses to an identity operator. A little algebra and the explicit form of $I_{ijk\ell}$ lets us write the result as

$$\Delta E_n^{(2)} \approx \frac{1}{m^2 M_{\rm pl}^2} \frac{2\pi}{15} \ln\left(\frac{\Lambda}{\langle E_n^{(0)} - E_{n'}^{(0)}\rangle}\right) \left(3 \langle n|[[H_0, \hat{p}^i \hat{p}^j], \hat{p}_i \hat{p}_j]|n\rangle - 2 \langle n|[[H_0, \hat{\mathbf{p}}^2], \hat{\mathbf{p}}^2|n\rangle\right). \quad (1.54)$$

In the QED case, you can further simplify the analogous matrix element down to just $|\psi_n(0)|^2$, the value of the wavefunction at the origin $r = 0$ squared [16]. It seems like there must be some similarly clever expression one could get here—maybe $|\nabla\psi_n(0)|^2$—but I have been unable to find it.

As promised, our answer for the gravitational Lamb shift in Eq. (1.54) is log divergent. In proper modern treatment we should further renormalize this to get a finite final value. Here however, we will again follow Bethe and try to get an estimate of this effect by simply setting $\Lambda = m$, i.e., cutting off the virtual fluctuations at the mass of the light particle. Since this is in a log, the precise value is not so important.

A very crude estimate would be to take the momenta to be over order the inverse "Bohr radius" for the problem

$$p \sim \frac{1}{a_0} = \frac{1}{\alpha m} \tag{1.55}$$

and the H_0 factors to be of order the unperturbed energies $H_0 \sim \alpha\alpha^2 m$. Doing this, one finds that the fractional change to the energy levels is of order

$$\frac{\Delta E_n^{(2)}}{E_n^{(0)}} \sim G_N M m \frac{m}{M} \ln\left(\frac{\Lambda}{\langle E_n^{(0)} - E_{n'}^{(0)} \rangle}\right). \tag{1.56}$$

For our black hole-axion example, this is an insanely small number, far too small to be observable. But the key point is that it is "UV finite". Naive perturbation theory would have produced a linearly UV-divergent answer, but standard regularization and renormalization procedures bring this into a finite (or at worst, log divergent) quantity.

2. What if gravity isn't quantized this way?

In the first lecture, we studied how gravity can be self-consistently, perturbatively quantized as a theory of gravitons. In this lecture, we are going to ask a seemingly innocuous question: what if gravity *isn't* quantized that way?

The central motivation for this, at least to me, is that we can try to do experimental tests of aspects of these ideas. Going back to the analogy with Bell violation tests, the graviton picture is the "null hypothesis", the analogue of quantum mechanics. In this lecture, we are going to study a variety of "alternative hypotheses", the analogues of the local hidden variables model. So what we are trying to do here is give a flavor, or ideally even a parametrization, of the kinds of possible behaviors of quantum gravity that we can distinguish in experiments.

Any such alternative model has to be somewhat strange, in a precise sense. As Weinberg showed in the 60's, any model obeying the following basic rules:

- Unitary time evolution
- Lorentz invariance
- Locality (specifically, cluster decomposition)
- Massless spin-2 states in the scattering spectrum

must necessarily have Einstein's equations as Heisenberg equations of motion [20, 21]. Since we have already observed classical spin-2 gravitational waves, this means that if gravity is not perturbatively quantized as an EFT, it must somehow violate one of the first three rules. This would certainly be exotic!

In particular, what we will study are a set of ideas about what could happen if gravity is somehow fundamentally irreversible, i.e., does not generate unitary time evolution. Very generally, non-unitary time evolution on a quantum system is described by a channel

$$\rho \to \rho' = \sum_\alpha K_\alpha \rho K_\alpha^\dagger, \quad \sum_\alpha K_\alpha^\dagger K_\alpha = 1, \tag{2.1}$$

which basically follows from the requirement that the evolution takes the density matrix ρ into a new density matrix [22]. The unitary case is when the sum α only has one term. If there are more, then a theorem of Stinespring shows that you can always "dilate" the channel, meaning you can find another system S, state $|0\rangle_S$ of

that system, and unitary evolution U on everything such that Eq. (2.1) is just the partial trace of the total unitary evolution:

$$\rho \otimes |0\rangle \langle 0|_S \to U(\rho \otimes |0\rangle \langle 0|_S)U^\dagger. \tag{2.2}$$

(Concretely: if $|\alpha\rangle$ labels a complete set of basis states of S, then $K_\alpha = \langle \alpha|U|0\rangle_S$.) This is a formal way of saying that if gravity generates non-unitary evolution, then we can always view this as arising from unitary evolution on some larger system, part of which we cannot observe, i.e. some kind of "bath". See Fig. **??**. If this kind of open evolution is happening continuously, then under a Markovian assumption, it means that the system obeys a Lindblad equation [22]

$$\dot{\rho} = i[\hat{H}, \rho] + \sum_\alpha L_\alpha^\dagger \rho L_\alpha - \frac{1}{2}\left\{L_\alpha^\dagger L_\alpha, \rho\right\}. \tag{2.3}$$

The first term gives unitary evolution, while the remaining terms represent noise.

We will look at two concrete mechanisms where gravity arises as a kind of noisy evolution represented by equations like these. One is a set of models in which the gravitational field is a classical, albeit stochastic and random object. The other involves gravity arising as a thermal, entropic effect. While the detailed models are very different, they share the same basic feature discussed above, namely the generation of non-reversible time evolution on matter. This is a very definitely observable prediction, and we can look for it experimentally. Another experimental signature present in these models is that they produce less (sometimes zero) entanglement between massive systems than perturbative quantum gravity. We will study these kinds of predictions in detail in lectures 3 and 4.

The models we are going to study here are best viewed as phenomenological, at least at the time of this writing. What I mean is that their purpose is to give some sense for what a non-graviton model might look like, both in its underlying mechanics and its experimental predictions. I would not recommend viewing them as real candidates for fundamental theories. Indeed, with one notable exception, they are not even relativistic. For an ambitious, open-minded theorist, this is an opportunity: it is a very interesting problem to try to make better versions of these ideas.

Finally, before moving on, I want to offer a somewhat heterodox, personal, second motivation for thinking about these kinds of models. Of course the experiments are the truest motivation, but here is another: background independence. Our fundamental starting point for perturbative quantum gravity was to think of the metric as an operator expanded around a background $\hat{g}_{\mu\nu} = \overline{g}_{\mu\nu} + \hat{h}_{\mu\nu}$. This works in other

gauge theories: for example, in QED, we can expand $\hat{A}_\mu = \langle \hat{A}_\mu \rangle + \delta \hat{A}_\mu$ in fluctuations around a constant field, a laser, etc. But there, we know that these effectively classical background fields can be viewed consistently as a limit of an underlying quantum model. In gravity, this is somewhat more murky: are we really sure that the UV theory admits classical background-like solutions $\overline{g}_{\mu\nu} = \langle \hat{g}_{\mu\nu} \rangle$ for all of the relevant spacetimes we might want to expand around? For example, what about fluctuations in de Sitter?

2.1. "Classical" gravity

The simplest thing people often ask is: what if gravity is just classical? But this isn't even defined: probably it means that the metric is a classical variable, but how are you supposed to couple that to matter?

Suppose we try something like this:

$$G_{\mu\nu} = 8\pi G_N \langle \hat{T}_{\mu\nu} \rangle \,. \tag{2.4}$$

Here, the left hand side is supposed to involve only classical variables, in particular, a classical metric $g_{\mu\nu}$. On the right hand side, we have the stress tensor operator acting on quantized matter variables. This equation tells us how the metric responds to some matter distribution, but it doesn't tell us how the matter state evolves, for which we need a Schrödinger equation or something equivalent. It turns out to be tricky to make this work. One simple thing you could try would be to take the non-relativistic limit of Eq. (2.4). Taking the usual weak field metric ansatz $ds^2 = -(1+2\Phi)dt^2 + \cdots$, the 00 component of Eq. (2.4) gives

$$\nabla^2 \Phi(\mathbf{x}) = 4\pi G_N \langle \hat{T}_{00}(\mathbf{x}) \rangle \,, \quad \hat{T}_{00}(\mathbf{x}) = \sum_a m_a \delta^3(\mathbf{x} - \hat{\mathbf{x}}_a), \tag{2.5}$$

where as usual we're taking matter to consist of a bunch of point particles with position operators $\hat{\mathbf{x}}_i$. We can then try to insert this into a normal Schrödinger equation for the matter,

$$i\partial_t |\psi\rangle = \hat{H}_0 |\psi\rangle + \int d^3\mathbf{x}\, \hat{\rho}(\mathbf{x}) \Phi(\mathbf{x}) |\psi\rangle \,, \tag{2.6}$$

where Φ solves Eq. (2.5). Notice that $\Phi = \Phi(\langle \hat{T}_{00} \rangle)$ depends on the state $|\psi\rangle$ through the expectation value, so this is now a "Schrödinger" equation that is non-linear in the quantum state. The non-linear equation Eq. (2.6) is usually referred to as the Schrödinger-Newton equation [23–28].

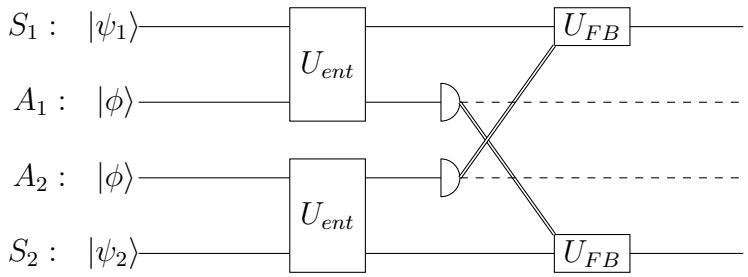

Figure 2: Measurement-and-feedback representation of "classical" gravity models coupled to quantum matter, from [1,10]. Here we show a single time step in the evolution process for two massive bodies S_1 and S_2. In the first part of the time step, the positions $\mathbf{x}_{1,2}$ of $S_{1,2}$ are measured by coupling to local ancillas $A_{1,2}$, which are then projectively measured (caps). The feedback control then uses this classical measurement data to generate local feedback unitaries $U_{\rm FB}$ [Eq. (**??**)], which generate an effective Newtonian gravitational interaction. Notice that no quantum information can flow between S_1 to S_2, which is why this model cannot entangle $S_{1,2}$ with each other.

Evolution equations that are non-linear in the quantum state, taken as fundamental, have serious pathologies. In particular, they generically lead to superluminal signaling [29, 30]. However, it is also possible to embed a model like this into a perfectly normal, unitary, non-superluminal quantum system, which can alleviate the pathologies. To see the basic idea, consider a quantum computer, which consists of some qubits coupled to some effectively classical readout and control systems, like lasers and classical computers. To do many quantum computational tasks, for example error correction, the qubits need to be measured by these control systems. This extracts classical information about them, say $\langle\sigma_z\rangle$, which can then be used to generate feedback unitaries on the qubits $U = U(\langle\sigma_z\rangle)$. The resulting evolution on the qubits is thus non-linear in their state! But nothing weird happened: the point is just that we are looking at reduced evolution on a subsystem (the qubits). The measurement and feedback system acts like a "bath".

(Need to fill this out still, sorry. But the canonical references for how to apply this to gravity are [10, 31–33] –dc)

2.2. "Classical-quantum" gravity

The models of "classical gravity" introduced in the previous section are only defined in the non-relativistic limit. Could a similar idea work in the fully relativistic setting?

I would say that at present, the answer is still unknown. However, over the past few years there has been an interesting attempt due to Oppenheim, Weller-Davies, and collaborators [12, 34–36]. They call the model "classical-quantum" (CQ) gravity, and I will follow that terminology here.

To set expectations: the simplest versions of CQ gravity are in conflict with observations in nature [37], as we will discuss a bit later. However, the idea is worth studying, since it is far from obvious that it would even be possible to formulate a self-consistent, relativistic model of classical gravity coupled to quantum matter. While I don't immediately see how, I think it's also possible that the basic model could be tweaked in some way to be consistent with current observations.

In the non-relativistic models of the previous section, the gravitational interaction is purely Newtonian, so in principle we do not need to discuss the dynamics of the gravitational field itself. In CQ gravity, the gravitational field $g_{\mu\nu}$ is a bona fide degree of freedom. It is classical in the sense that it has a trivial commutator algebra, but it is still drawn from a classical probability distribution $\mathcal{P}(g_{\mu\nu})$, and this distribution evolves stochastically. The gravitational field couples to quantized matter in a covariant way I will make precise below. This is very much aligned with the measurement-and-feedback models discussed in the previous section, and my personal opinion is that CQ gravity is a relativistic generalization of the measurement-feedback gravity paradigm.

Before we show how the dynamics works in a CQ model, we have to specify the space of states. We can consider this quite generally for a classical system with canonical variables (Q, P) coupled to a quantum system with canonical variables (q, p). We want an object that reduces to the density matrix of the quantum system when the classical system is marginalized over, and that reduces to the classical probability distribution for the classical system when the quantum system is marginalized over. An appropriate mathematical object, the state ϱ in CQ theory, is a density matrix-valued function of the classical variables:

$$\varrho(Q, P) := \int dq d\underline{q}\, \varrho(q, \underline{q}; Q, P)\, |q\rangle \langle \underline{q}| \,. \tag{2.7}$$

For a fixed classical configuration (Q, P), this object is the density matrix of the quantum system conditioned on that classical configuration. Conversely, for a fixed quantum state $|q\rangle$, $\varrho(q, q; Q, P) = \mathcal{P}(Q, P|q)$ is the classical probability distribution on (Q, P) conditioned on that quantum state $|q\rangle$. If we are not measuring the classical system or quantum system at all, the reduced states for the other systems are the

averages over these conditional states:

$$\varrho \to \begin{cases} \rho = \int dQdP\, \varrho(Q,P), & \text{classical system unobserved} \\ \mathcal{P}(Q,P) = \int dq\, \varrho(q,q;Q,P), & \text{quantum system unobserved.} \end{cases} \tag{2.8}$$

In the CQ gravity models, Q, P represent the canonical gravitational degrees of freedom (e.g., in the ADM framework), and q represents the quantum matter.

Evolution of the complete CQ state is specified by a path integral:

$$\varrho(q_f, \underline{q}_f; Q_f, P_f; t_f) = \int_{q_i,\underline{q}_i,Q_i,P_i}^{q_f,\underline{q}_f,Q_f,P_f} DQDPDq\, \delta[P - P(Q,\dot{Q})] e^{I_{CQ}[q,\underline{q},Q,P]} \varrho(q_i, \underline{q}_i; Q_i, P_i; t_i). \tag{2.9}$$

Note that this is a Lorentzian path integral. The delta functional enforces the definition of the canonical momentum. The CQ functional I_{CQ} generally takes the form

$$I_{CQ}[q,\underline{q},Q,P] = \underbrace{i\left\{S_0[q] - S_0[\underline{q}]\right\}}_{\text{Schrodinger evolution on } \rho} + \underbrace{I_{FP}[Q,P]}_{\text{Fokker-Planck evolution on } \mathcal{P}} + \underbrace{iS_{\text{int}}[q,\underline{q},Q,P]}_{\text{CQ interaction}}. \tag{2.10}$$

Let's break down what each term does, and then work out some examples and their physical consequences.

The first term in Eq. (2.10) generates ordinary, unitary time evolution in the quantum system $\rho \to U\rho U^\dagger$. The first iS acts on the ket and the second acts on the bra in the density matrix elements. Using this kind of path integral to evolve a density matrix is sometimes referred to as the Schwinger-Keldysh formalism [38, 39], although the idea goes back to Feynman and Vernon [40].

The second term in Eq. (2.10) generates stochastic evolution on the classical variables. This is probably less familiar to most people, but within statistical mechanics it is a reasonably standard setup. As a concrete example, consider a Brownian particle undergoing diffusion. Its probability distribution $\mathcal{P} = \mathcal{P}(Q,P)$ satisfies the Fokker-Planck equation

$$\dot{\mathcal{P}} = -\frac{P}{m}\frac{\partial}{\partial Q}\mathcal{P} + \frac{D_2}{2}\frac{\partial^2}{\partial P^2}\mathcal{P}. \tag{2.11}$$

The coefficient D_2 sets the rate of diffusion. This equation admits a path integral solution [41],

$$\mathcal{P}(Q_f, P_f, t_f) = N \int DQDP\, \delta\left[\dot{Q} - \frac{P}{m}\right] e^{-\frac{1}{D_2}\int dt\, [m\ddot{Q}]^2} \mathcal{P}(Q_0, P_0, t_0), \tag{2.12}$$

where N is a normalization factor. To get some intuition for what is going on here, let me relate a lesson I got from Shivaji Sondhi. A Fokker-Planck equation represents

the stochastic time evolution of a probability distribution, so it is like a classical statistical mechanical version of the Schrodinger picture. Equivalently, we can work with the analogue of the Heisenberg evolution, where we stochastically time evolve the observables. The equation of motion in that picture is called a Langevin equation, and in this specific example, the equivalent Langevin equations are

$$\dot{Q} = \frac{P}{m}, \quad \dot{P} = \sqrt{D_2}\xi \implies \ddot{Q} = \frac{\sqrt{D_2}}{m}\xi, \tag{2.13}$$

where $\xi = \xi(t)$ is a Gaussian-distributed random white noise variable

$$\mathcal{P}[\xi(t)] = Ne^{-\int dt\, \xi^2(t)}. \tag{2.14}$$

Thus, the probability distribution of a given particle configuration Q_f at t_f can be obtained by just solving $\xi = m\ddot{Q}/\sqrt{D_2}$ and inserting it into this distribution, yielding

$$\mathcal{P}(Q_f, t_f) = N \int DQ e^{-\frac{1}{D_2}\int dt\, \left[m\ddot{Q}\right]^2} \mathcal{P}(Q_0, t_0), \tag{2.15}$$

which is just Eq. (2.12) with the DP integral evaluated. Again, notice that this is a real (Lorentzian) time path integral. The weight in the exponential is, however, real, unlike quantum mechanics. What it does is to weight the paths with zero noise $\xi = 0$ the most strongly and then suppress those with $\xi \neq 0$ by the Gaussian weighting of Eq. (2.14). Note that this path integral is quartic in derivatives, but this does not mean anything acausal is happening, unlike a quantum-mechanical path integral. In the CQ gravity model, what this term is going to do is produce stochastic evolution of the gravitational field, where the dominant behavior obeys the Einstein equations and then there are noise-induced fluctuations away from this. Notice that these stochastic kicks ξ are here inserted as a fundamental feature of the model, but can also be interpreted as coming from a bath, reinforcing the general story of this lecture.

Finally, the last term in Eq. (2.10) tells us how to couple the classical and quantum degrees of freedom. There is no particularly simple interpretation of this, but one important thing to note is that it includes a "coupling" of the bra and ket evolution since it depends on both q and $\underline{q}$. This is a hallmark of noise acting on a quantum system, and in that sense, this kind of term is very standard [40]. Since Q, P are random classical variables, the effect on the quantum system here is just that of a random classical external field. What is less common is the interpretation of its effect on the *classical* system, but what we will see below is that we can roughly interpret this as giving a kind of semiclassical backreaction similar to that in Eq. (2.4).

With this basic structure in hand, we can work out some basic physics of CQ gravity. In fact, we are going to work out something much simpler: a classical Yukawa field ϕ interacting with a pair of quantized scalar fields $\hat{\chi}_{1,2}$. Akira Matsumura and I studied this model in detail in [42]. The pair of fields is just for convenience, so that we can treat two particle scattering with distinguishable particles. The interaction is as usual

$$S_{\rm int} = \lambda \int d^4x\, \phi \hat{\boldsymbol{\chi}}^2, \tag{2.16}$$

where λ has dimensions of mass and the bold $\boldsymbol{\chi} = (\chi_1, \chi_2)$. This is much simpler than doing full blown gravity because it doesn't involve any gauge issues. However, we can recover the Newtonian limit of gravity by taking

$$\lambda \to \sqrt{G_N} m^2, \quad m_\phi \to 0 \implies V_{\rm Yukawa} = -\frac{G_N m^2}{r}, \tag{2.17}$$

so in particular we can analyze how this model compares to the non-relativistic classical gravity models in the previous section. Using the interaction (2.16), we take the CQ path integral action (2.10) for this model to be[6]

$$\begin{aligned} I_{CQ}[\boldsymbol{\chi}, \underline{\boldsymbol{\chi}}, \phi, \pi] = {} & i(S_0[\boldsymbol{\chi}] + S_{\rm int}[\boldsymbol{\chi}, \phi] - S_0[\underline{\boldsymbol{\chi}}] - S_{\rm int}[\underline{\boldsymbol{\chi}}, \phi]) + \ln(\delta[\dot{\phi} - \pi]) \\ & - \frac{1}{2D_2} \int_{t_0}^{t_f} d^4x \left(\dot{\pi} - (\nabla^2 - m_\phi^2)\phi - \frac{1}{2}\frac{\delta S_{\rm int}}{\delta \phi}[\boldsymbol{\chi}, \phi] - \frac{1}{2}\frac{\delta S_{\rm int}}{\delta \phi}[\underline{\boldsymbol{\chi}}, \phi] \right)^2 \\ & - \frac{D_0}{2} \int_{t_0}^{t_f} d^4x \left[\frac{\delta S_{\rm int}}{\delta \phi}[\boldsymbol{\chi}, \phi] - \frac{\delta S_{\rm int}}{\delta \phi}[\underline{\boldsymbol{\chi}}, \phi] \right]^2 . \end{aligned} \tag{2.18}$$

Here $\boldsymbol{\chi} = (\chi_1, \chi_2)$ is a multiplet of the two scalars, and π is the canonical momentum for the classical ϕ field. The log term is just another way to write the delta functional that enforces the canonical variable definition. The S_0 are just the ordinary free actions for the fields $\mathcal{L} \sim (\partial\phi)^2 + m^2\phi^2$ and similarly for $\chi_{1,2}$.

We could say many things about the model defined by Eq. (2.21). The basic interpretation follows from the general remarks above. In particular, the terms proportional to $1/D_2$ in the second line generate diffusive stochastic evolution of ϕ, just like the Fokker-Planck model shown above. Note that D_2 has units of M^2. The Fokker-Planck equation for ϕ is sourced by specific configurations of the quantum

[6] This specific form ensures that the resulting dynamics takes density matrices to density matrices. A proof of this is well beyond the scope of these notes but can be found in [12, 37]. In particular, requiring this self-consistent evolution also requires $D_0 D_2 \geq 1$, which these authors refer to as a decoherence-diffusion tradeoff for reasons that will become clear soon.

fields $\chi_{1,2}$ (i.e., the classical field is back-reacted by the quantum fields, in this specific sense). Finally, the third line causes decoherence of the matter fields into the $|\chi\rangle$ field configuration basis, by suppressing any density matrix element $|\boldsymbol{\chi}\rangle\langle\underline{\boldsymbol{\chi}}|$ with $\boldsymbol{\chi} \neq \underline{\boldsymbol{\chi}}$. This decoherence occurs at a rate proportional to D_0, which here has units of $1/M^2$.

I would like to make two remarks about the phenomenology of this model, and then finish by showing how it reduces to the same kind of Schrödinger-Newton type of evolution from the previous section. First, a positive remark: the path integral in Eq. (2.21) really does appear to produce a Lorentz-covariant time evolution for a classical system (here, ϕ) coupled to quantum matter (here, $\hat{\chi}_{1,2}$). In particular, it is possible to define scattering theory, at least for $\chi\chi \to \chi\chi$ at tree level, and the resulting scattering probabilities are Lorentz-invariant [42]. The calculation and setup is involved, but the result is easy enough to state:

$$\begin{aligned}
&\mathcal{P}(\mathbf{p}_1\mathbf{p}_2 \to \mathbf{p}_1'\mathbf{p}_2') = \\
&= \frac{VT}{(2\pi)^4}\frac{\lambda^4\delta^4(p_1'+p_2'-p_1-p_2)}{(2\pi)^4\prod_{i=1,2}\omega_{\mathbf{p}_i}\omega_{\mathbf{p}_i'}}\left[\left(\frac{1}{(p_1'-p_1)^2+m_\phi^2}\right)^2+\left(\frac{D_2}{[(p_1'-p_1)^2+m_\phi^2]^2}+D_0\right)^2\right] \\
&+ \frac{VT}{(2\pi)^4}\frac{\lambda^4\delta^4(p_1'-p_1-(p_2'-p_2))}{(2\pi)^4\prod_{i=1,2}\omega_{\mathbf{p}_i}\omega_{\mathbf{p}_i'}}\left(\frac{D_2}{[(p_1'-p_1)^2+m_\phi^2]^2}+D_0\right)^2 .
\end{aligned} \tag{2.19}$$

Here, we are quoting the probability — there is no amplitude, since this is fundamentally an open system. The overall factor $VT/(2\pi)^4 = \delta^4(0)$ is just a normalization that drops out when you convert this into a cross-section as usual. This probability transforms properly under the Lorentz group, but notice that it allows for processes that violate four-momentum conservation! The reason that this is possible is because there is no Noether's theorem for open systems, since the bath can exchange energy and momentum with the system [43, 44]. In particular, the second line comes from events where both particles are kicked with the same momentum.

The more negative comment that has to be made is that this model is inconsistent with observation. In particular, the diffusive evolution of ϕ means that the amplitude is random and is growing in variance, with $\Delta\phi^2 \sim D_2$. Applied to the gravitational field, this means that the universe would be filled with a background of gravitational waves. Simultaneously, the D_0 term means that matter in the universe is decohering at all times. Since we can do experiments where matter is not decohering for ~ 60 seconds (see lecture 4), this puts a lower bound on D_0, and then the consistency

condition $D_0 D_2 \geq 1$ means there is a lower bound on D_2 as well. Putting in numbers, one finds that the universe would have a stochastic background of gravitational waves that is inconsistent with current observations [45].[7] However, this conclusion is only strictly nkown in the simple version shown here, where $D_{0,2}$ are constants. There are two possible outs: $D_{0,2}$ can be non-trivial functions of spacetime [12], and it may also be possible to add a damping term (which has a coefficient called D_1 in the Fokker-Planck framework [41]). Whether either of these really works is at present an open question.

With those observational comments in mind, let's finish by demonstrating that this model is, in fact, a relativistic version of the measurement-feedback "Schrödinger-Newton" models from the previous section.[8] To do this we will work out the evolution law for the matter density matrix in the non-relativistic limit. In the non-relativistic limit, the Yukawa interaction between the ϕ field and χ particles reduces to

$$V_{\rm int} = \frac{\lambda}{m} \sum_{a=1,2} \phi(\hat{\mathbf{x}}_a), \tag{2.20}$$

where $\mathbf{x}_a$ is the position of particle $a = 1, 2$. You can derive this by considering one-particle matrix elements and matching. Using this, we can reduce the CQ path integral action (2.10) to its non-relativistic form

$$\begin{aligned} I_{CQ}[\mathbf{x}_a, \underline{\mathbf{x}}_a, \phi, \pi] = i(S_0[\mathbf{x}_a] + S_{\rm int}[\mathbf{x}_a, \phi] - S_0[\underline{\mathbf{x}}_a] - S_{\rm int}[\underline{\mathbf{x}}_a, \phi]) + \ln(\delta[\dot{\phi} - \pi]) \\ - \frac{1}{2D_2} \int_{t_0}^{t_f} d^4x \left(\dot{\pi} - (\nabla^2 - m_\phi^2)\phi - \frac{1}{2}\frac{\delta S_{\rm int}}{\delta \phi}[\mathbf{x}_a, \phi] - \frac{1}{2}\frac{\delta S_{\rm int}}{\delta \phi}[\underline{\mathbf{x}}_a, \phi] \right)^2 \\ - \frac{D_0}{2} \int_{t_0}^{t_f} d^4x \left[\frac{\delta S_{\rm int}}{\delta \phi}[\mathbf{x}_a, \phi] - \frac{\delta S_{\rm int}}{\delta \phi}[\underline{\mathbf{x}}_a, \phi] \right]^2 . \end{aligned} \tag{2.21}$$

Here $\mathbf{x}_a$ schematically denotes both $\mathbf{x}_1$ and $\mathbf{x}_2$, and the underline as usual denotes the bra part of the state for these. In this limit the free and interaction terms can be written

$$S_0[\mathbf{x}_a] = \int_{t_0}^{t_f} dt \sum_{a=1,2} \frac{1}{2} m \dot{\mathbf{x}}_a^2, \quad S_{\rm int}[\mathbf{x}_a, \phi] = \lambda \int_{t_0}^{t_f} d^4x \, J(x)\phi(x), \quad J(x) = \sum_{a=1,2} \delta^3(\mathbf{x} - \mathbf{x}_a(t)). \tag{2.22}$$

[7] I first learned of this estimate from my former colleague Geoff Penington on Twitter [46], but a more formalized version can also be found in Zach Weller-Davies' PhD thesis [37].

[8] The argument given here is due to Akira Matsumura, and will appear in a future publication. This argument is slightly heuristic, but a more rigorous discussion can be found in [35].

From these expressions, we see that the path integral is a Gaussian function of ϕ and π. The $D\phi D\pi$ integral can therefore be evaluated exactly, leading to a path integral that only acts on the matter $\mathbf{x}_a$. The detailed calculation of the resulting matter path integral can be found in [42]. The result is

$$\rho(\mathbf{x}_{a,f}, \underline{\mathbf{x}}_{a,f}, t_f) = \int D\mathbf{x}_a D\underline{\mathbf{x}}_a e^{iS_{IF}[\mathbf{x}_a, \underline{\mathbf{x}}_a]} e^{iS_0[\mathbf{x}_a] - iS_0[\underline{\mathbf{x}}_a]} \rho(\mathbf{x}_{a,0}, \underline{\mathbf{x}}_{a,0}, t_0) \tag{2.23}$$

where the Feynman-Vernon influence functional S_{IF} is

$$\begin{aligned} iS_{IF} = -\frac{\lambda^2}{2} \int_{t_0}^{t_f} dt \int d^3\mathbf{x} d^3\mathbf{y} \Big[& J(x) F_1(\mathbf{x}-\mathbf{y}) J(y) + \underline{J}(x) F_1^*(\mathbf{x}-\mathbf{y}) \underline{J}(y) \\ & - 2J(x) F_2(\mathbf{x}-\mathbf{y}) \underline{J}(y) \Big]. \end{aligned} \tag{2.24}$$

Here the underline on J means with the $\underline{\mathbf{x}}_a$ variables, and the $F_{1,2}$ functions are linear combinations of Green's functions:

$$F_1(\mathbf{x}) = -\frac{i}{2} \left[G_R(\mathbf{x}) + G_R(-\mathbf{x}) \right] + F_2(\mathbf{x}), \quad F_2(\mathbf{x}) = D_2 G_C(\mathbf{x}) + D_0 \delta^3(\mathbf{x}) \tag{2.25}$$

where G_R and G_C are a retarded "quantum" and a "classical" Fokker-Planck Green's function respectively:

$$G_R(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{\mathbf{k}^2 + m_\phi^2}, \quad G_C(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{[\mathbf{k}^2 + m_\phi^2]^2}. \tag{2.26}$$

In particular, the $1/k^4$ behavior of G_C and its overall D_2 coefficient reflect the quartic derivative structure of the Fokker-Planck evolution for ϕ.

Eq. (2.23) gives us the time-integrated evolution of the density matrix for just the matter variables, i.e., the reduced evolution with the gravitational field averaged out. We can use the structure of this to read off an equivalent differential equation for $\dot{\rho}$. First, note that anything involving $\mathbf{x}_a$ acts on the ket part of the density matrix, i.e., from the left, while anything with $\underline{\mathbf{x}}_a$ acts on the bra, i.e., from the right. In particular, the terms in the exponential with an overall factor of i

$$iS_{\text{imag}} = iS_0[\mathbf{x}_a] - iS_0[\underline{\mathbf{x}}_a] + \frac{i\lambda^2}{4} \int dt d^3\mathbf{x} d^3\mathbf{y}\, J(x) G_R(\mathbf{x}-\mathbf{y}) J(y) - \underline{J}(x) G_R(\mathbf{x}-\mathbf{y}) \underline{J}(y) \tag{2.27}$$

generate unitary evolution, i.e., act as a Hamiltonian $\dot{\rho} \supset i \left[\hat{H}, \rho \right]$, with

$$\hat{H} = \hat{H}_0 + \lambda^2 \int d^3\mathbf{x} d^3\mathbf{y} \hat{J}(\mathbf{x}) G_R(\mathbf{x}-\mathbf{y}) \hat{J}(\mathbf{y}) = \hat{H}_0 - \sum_{a \neq b} \frac{G_N m^2}{|\hat{\mathbf{x}}_a - \hat{\mathbf{x}}_b|}. \tag{2.28}$$

To get the last equality, we used Eqs. (2.26), took the limit $m_\phi \to 0$, and identified $\lambda = \sqrt{G_N} m$. In other words, this reproduces ordinary Newtonian gravity, including an entangling potential operator.[9] The remaining terms in the exponential are proportional to F_2, which is real. Since they are real, they give noise terms in the Lindblad language. Putting everything together, the answer takes the form

$$\begin{aligned}\dot{\rho} = i\left[\hat{H}, \rho\right] &+ G_N D_2 \int d^3\mathbf{x} d^3\mathbf{y}\, G_C(\mathbf{x}-\mathbf{y}) \left(\hat{\mu}(\mathbf{x})\rho\hat{\mu}(\mathbf{y}) - \frac{1}{2}\left\{\hat{\mu}(\mathbf{x})\hat{\mu}(\mathbf{y}), \rho\right\}\right) \\ &+ G_N D_0 \int d^3\mathbf{x} \left(\hat{\mu}(\mathbf{x})\rho\hat{\mu}(\mathbf{x}) - \frac{1}{2}\left\{\hat{\mu}(\mathbf{x})\hat{\mu}(\mathbf{x}), \rho\right\}\right)\end{aligned} \tag{2.29}$$

where the mass density operator $\hat{\mu}$ is a sum of delta functions of the positions

$$\hat{\mu}(\mathbf{x}) = \sum_a m_a \delta^3(\mathbf{x} - \hat{\mathbf{x}}_a), \tag{2.30}$$

and G_C is given in Eq. (2.26).

Eq. (2.29) is a Lindblad evolution law for the massive particles. The Hamiltonian term has an ordinary, entangling, two-body Newtonian potential operator. This term is precisely what one would get in perturbative quantum gravity, but there it would be the end of the story. Here instead we have the remaining Lindbladian terms. These act diagonally in the position basis $|\mathbf{x}_a\rangle$. This means that particles will be decohered into position eigenstates. Notice that this happens at a rate $\sim G_N m^2 D_2$, as well as $\sim G_N m^2 D_0$. If we saturate the "diffusion-decoherence tradeoff" inequality $D_0 D_2 \geq 1$, this means that there is an irreducible level of noise in the system.

2.3. Newton's law as an entropic force

The models in the previous sections have all violated the EFT rules by generating non-unitary time evolution. In the Schrödinger-Newton and measurement-feedback models, this non-unitarity can be viewed as coming from measurements and control done by some system external to massive objects undergoing gravity. In the CQ model, there is an additional, irreducible noise driving the classical gravitational field.

To round out this discussion, then, let us turn to a third proposal, in which gravity arises as some kind of emergent thermal or entropic effect. The core idea here is that massive bodies undergoing gravity still obey a non-unitary evolution, but here

[9] Note that there are also diagonal $a = b$ terms that we dropped. This is because those just contribute an overall (infinite) constant in the Hamiltonian.

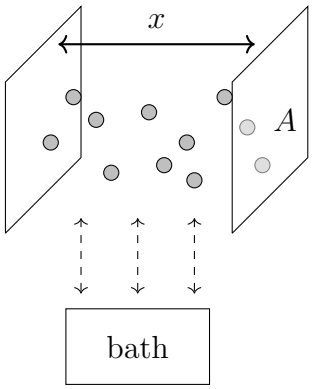

Figure 3: A simple example of an entropic force: the pressure exerted on piston walls by an ideal gas. From [13].

the reason is that gravity itself arises from some complex microscopic many-body thermalizing system. Thus, the coarse-grained evolution appears non-reversible.

The notion that gravity could be fundamentally some kind of emergent thermal system goes back to Jacobson [47], and more recently, Verlinde has given the specific proposal that gravity is an "entropic force" [48]. To understand the basic idea of an entropic force, consider an ideal gas between two movable pistons. See Fig. 3. For simplicity, suppose we hold the gas at fixed temperature T by coupling it to an external reservoir. If x is the separation between the walls, the Helmholtz free energy $\mathcal{A}$ of the gas is a function of x:

$$\mathcal{A} = \mathcal{A}(x) = U(x) - TS(x). \tag{2.31}$$

For a non-interacting ideal gas, the internal energy U is actually independent of x. The entropy, meanwhile, obeys the Sackur-Tetrode law

$$S \approx Nk_{\mathrm{B}} \ln \left[\frac{V}{N} (mT)^{3/2} \right] + \text{const.} \tag{2.32}$$

As per the usual laws of statistical mechanics, interactions with the heat bath will drive the system towards the extremum of the free energy. This depends on the length x since the volume $V = Ax$, so the minimum is when

$$0 = \frac{d\mathcal{A}}{dx} = -T\frac{dS}{dx} = -\frac{Nk_{\mathrm{B}}T}{x}. \tag{2.33}$$

What we have discovered is that there is an entropic force, which in this case is simply the usual pressure $P = Nk_{\mathrm{B}}T/V$ given by the ideal gas law:

$$F_{\mathrm{ent}} := -T\frac{dS}{dx} = PA. \tag{2.34}$$

The point of this is that the force between the walls is not generated by exchange of virtual quanta or anything like that. It arises instead from the entropic dynamics of an intervening many-body system.

Could Einstein gravity arise in a similar fashion? Once again, I would say that the current state of the art is that we simply do not know. Neither Jacobson nor Verlinde provided a complete microscopic model that gives rise to gravity in this way. Rather, they provide very general requirements on a set of microscopic degrees of freedom (analogous to the gas particles above) under which one should be able to reproduce at least semiclassical general relativity in some thermodynamic limit. However, so far, no one has provided such a set of degrees of freedom.[10]

What we can do, at least, is reproduce Newtonian gravity through a mechanism like this [13]. In particular, it is possible to construct a totally explicit N-body system that evolves unitarily, which in the thermodynamic limit gives rise to the Newtonian force between masses.

Consider once again a collection of point masses m_a with coordinates $\mathbf{x}_a$. We write the following Hamiltonian:

$$H = H_S + H_M + H_B + V_{SM} + V_{MB}. \tag{2.35}$$

The subscripts stand for the system S of the particles, a mediator M which is the analogue of the gas particles, a heat bath B, and then couplings between the particles and mediator V_{SM} and between the mediator and heat bath V_{MB}. Very concretely, we will take the mediator to consist of a collection of qubits, and the bath to consist of a collection of harmonic oscillators. In the simplest implementation, these terms are given by

$$\begin{aligned} H_S &= \sum_{i=a} \frac{\mathbf{p}_a^2}{2m_a} \\ H_M &= 0 \\ H_B &= \sum_{p=1}^{\infty} \omega_p b_p^\dagger b_p \\ V_{SM} &= \sum_{\alpha=1} \omega_\alpha(\mathbf{x}_1, \mathbf{x}_2, \ldots) N_\alpha, \quad N_\alpha = \frac{Z_\alpha - 1}{2} \\ V_{MB} &= \sum_{\alpha,p} g_{\alpha,p} \left[\sigma_\alpha^- b_p^\dagger + \sigma_\alpha^+ b_p\right]. \end{aligned} \tag{2.36}$$

[10]The possible exception to this statement is that it may be possible to view AdS/CFT as arising in this fashion, but it is fair to say that this is currently not a consensus view.

I am showing all of these terms explicitly to make the point that the model is completely specified at the microscopic level, and the fundamental time evolution is unitary. Here $\alpha = 1, 2, \ldots, N$ labels the qubits (we will take $N \to \infty$ eventually), and Z_α is the Pauli-Z operator on the αth qubit. The $\sigma^\pm = X \pm iY$ raise and lower the qubit value between $0, 1$, so what V_{MB} does is exchange energy between the qubit and the heat bath. The key physics is in V_{SM}, which says that the qubit frequencies ω_α depend on the positions of the masses $\mathbf{x}_i$. We will take the qubits to have a harmonic spectrum in the sense that

$$\omega_\alpha(\mathbf{x}_1, \mathbf{x}_2, \ldots) = \alpha f(\mathbf{x}_1, \mathbf{x}_2, \ldots). \tag{2.37}$$

I do not know if this is essential, but at present it is the only way I know to make the construction work, and it makes the calculations very straightforward. We will specify f shortly.

Consider the Heisenberg equations of motion for the ath mass. In particular, the force on this mass is given by

$$\dot{\mathbf{p}}_a = i[H, \mathbf{p}_a] = \sum_\alpha \nabla_{\mathbf{x}_a} \omega_\alpha(\mathbf{x}_1, \mathbf{x}_2, \ldots) N_\alpha. \tag{2.38}$$

Let's initially look at this in the limit that the masses m_i are large and the positions are well-locallized so that we can treat the $\mathbf{x}_a$ as c-numbers. Suppose that the qubits are prepared in the thermal state ρ_T defined by a fixed set of positions $\mathbf{x}_a$. Then the thermally-averaged force on the mass is given by $\mathbf{F}_a = \mathrm{tr}_M[\dot{\mathbf{p}}_a \rho_T]$, which can be evaluated directly:

$$\begin{aligned} \mathbf{F}_a &= \sum_\alpha \nabla_{\mathbf{x}_a} \omega_\alpha(\mathbf{x}_1, \mathbf{x}_2, \ldots) \frac{e^{-\omega_\alpha/T}}{1 - e^{-\omega_\alpha/T}} \\ &\approx \frac{\nabla_{\mathbf{x}_a} f}{f^2} \int_0^\infty d\omega\, \omega \frac{e^{-\omega/T}}{1 - e^{-\omega/T}} \\ &= \frac{\pi^2}{12} T^2 \nabla_{\mathbf{x}_i} \frac{1}{f}. \end{aligned} \tag{2.39}$$

From this, we can see how to choose f in order to reproduce Newton's law. To get a $1/r^2$ force, we need $1/f \sim 1/r$. The general answer is to take

$$\frac{1}{f(\mathbf{x}_1, \mathbf{x}_2, \ldots)} = \lambda + \frac{1}{2} \sum_{a \neq b} \frac{\ell_a \ell_b}{|\mathbf{x}_{ab}|}, \quad \mathbf{x}_{ab} = \mathbf{x}_a - \mathbf{x}_b, \tag{2.40}$$

with λ and the ℓ_i all free parameters with units of length. This yields

$$\mathbf{F}_a = \frac{\pi^2 T^2}{12} \sum_{b \neq a} \ell_a \ell_b \frac{\hat{\mathbf{x}}_{ab}}{|\mathbf{x}_{ab}|^2}. \tag{2.41}$$

We see that we have recovered a $1/r^2$ force law. The last step is to identify the emergent G_N as follows:

$$\frac{\pi^2 T^2}{12} \equiv \frac{G_N}{L^4}, \quad \ell_a \equiv m_a L^2, \tag{2.42}$$

which gives the Newtonian gravitational force:

$$\mathbf{F}_a = \frac{1}{2} \sum_{a \neq b} G_N m_a m_b \frac{\hat{\mathbf{x}}_{ab}}{|\mathbf{x}_{ab}|^2}. \tag{2.43}$$

We are left with three overall free parameters: the temperature T and two length scales λ and L. The two parameters T and L are constrained in terms of G_N and the masses m_i, but λ is completely free at this stage. Physically, T and L are fixed by the requirement that the Newtonian gravitational force arises in expectation value, whereas we will see shortly that λ enters only in the *fluctuations* of the force.

What is the interpretation of Eq. (2.43)? It shows that it is possible for a microscopic, thermal system of qubits to act together to produce Newton's law of gravity. There are two basic simplifications we used to get to this conclusion that need to be addressed. The first is that the derivation above assumed that the masses undergoing gravity are well-localized and heavy enough that we could approximate their positions as c-numbers. What happens if we need to understand the quantum behavior of the masses? The second is that this argument only works at some instant of time: we simply posited that the qubits are in a thermal state and then computed the average thermal force. But once the masses start moving, why would the qubits stay thermalized?

Both of these questions can be answered in detail by solving the complete model, including the bath-mediator interaction, given by Eq. (2.35). The detailed derivation is somewhat involved, but the basic limit in which it works is easy to state. We need to assume a kind of adiabaticity: thermalization of the bath and mediator (to fixed temperature T) is much faster than the motion of the masses. This is analogous to the ideal gas case, where we assumed isothermal evolution as the pistons change the volume of the gas.

In this adiabatic limit, the evolution of the quantum state of the masses can be calculated explicitly. The derivation is somewhat difficult and beyond the scope of these notes, but can be found in [13]. One finds a Lindblad equation for the density matrix,

$$\dot{\rho} = i[H_S + V_N, \rho] + \sum_{\alpha,\pm} L^\dagger_{\alpha,\pm} \rho L_{\alpha,\pm} + \frac{1}{2} \left\{ L^\dagger_{\alpha,\pm} L_{\alpha,\pm}, \rho \right\}. \tag{2.44}$$

Here V_N is the ordinary Newtonian potential operator, so the commutator term generates the exact same, quantum-coherent (!) evolution as perturbative quantum gravity. The difference is contained in the other terms, which generate noise. Specifically, the Lindblad noise operators are given by

$$\begin{aligned} L_{\alpha,+} &= L_{\alpha,+}(\mathbf{x}) = \sqrt{2\zeta T}\int_0^{\omega_\alpha(\mathbf{x})/T} d\nu\,\sqrt{g_+(\nu)} \\ L_{\alpha,-} &= L_{\alpha,-}(\mathbf{x}) = \sqrt{\frac{2T}{\zeta}}\int_0^{\omega_\alpha(\mathbf{x})/T} d\nu\,\sqrt{g_-(\nu)}. \end{aligned} \tag{2.45}$$

These represent events where the αth mediator qubit gains or loses a quanta of energy via the bath, respectively. The notation $\mathbf{x} = \mathbf{x}_1, \mathbf{x}_2, \ldots$ means a dependence on all of the masses, via Eq. (2.37). The dependence on the dimensionless functions $g_\pm$ is not particularly important for our purposes. The Boltzmann distribution $n_B(\omega) = (e^{\omega/T} - 1)^{-1}$ has appeared because of the bath of oscillators. Finally, ζ is a dimensionless parameter that controls the rate at which the mediator qubits exchange energy with the bath. Thus in total we have four free parameters: T, L, λ, and ζ, constrained by one equation which defines the emergent G_N.

The noise operators $L = L(\mathbf{x})$ depend on the mass positions. What this means is that the noise acts to decohere the masses into the position basis. It will also mean that the noise acts to generate fluctuations in the mass momenta. The intuition for this is simple, and follows the same basic idea as the measurement-feedback models given above. The mediator qubits are measuring the mass positions and using this information to adjust their frequencies. These measurements, plus thermal fluctuations in the bath, add up to a seemingly irreducible source of noise. However, there is a key subtlety: notice that $L \sim \sqrt{\omega} \sim 1/\lambda$, using Eqs. (2.37) and (2.40). Thus it appears that we can simply dial the free parameter λ as large as we want to eliminate this noise! And in the limit $\lambda \to \infty$, Eq. (2.44) simply reduces to ordinary perturbative quantum gravity, at least at the level of the Newton interaction. We will study the experimental implications of this range of phenomenologies in the next sections.

Problems

1. **Interferometry, decoherence, and the Lindblad equation.**

   (a) Calculate the probabilities of outcomes $P(0,1)$ in interferometer, with initial state $|0\rangle$. An interferometer involves sending the system through a beamsplitter, letting the state propagate freely for time Δt, inverting the beamsplitter, and then measuring in the $|0,1\rangle$ basis. A simple 50/50 beamsplitter is

   $$U_{\text{control}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \tag{2.46}$$

   (b) Now add an external potential $V = \alpha Z$ with α some constant, and calculate $P(0,1)$. Give an interpretation of this. Why is this called interferometry?

   (c) Now we will consider the effects of decoherence from continuous monitoring by a bath. A simple model of this is to have the state evolve under the Lindblad evolution

   $$\dot{\rho} = i[V,\rho] + \sum_{i=1,2} L_i^\dagger \rho L_i - \frac{1}{2}\left\{L_i^\dagger L_i, \rho\right\} \tag{2.47}$$

   with $L_1 = \sqrt{\Gamma}\,|0\rangle\langle 0|$, $L_2 = \sqrt{\Gamma}\,|1\rangle\langle 1|$. First, take a general density matrix

   $$\rho(0) = \begin{pmatrix} a & b \\ b^* & c \end{pmatrix} \tag{2.48}$$

   and find $\rho(t)$. Hint: calculate it component by component. Now use this result to compute $P(0,1)$ in the interferometer. Probably easiest to first do it with $V = 0$. Give an interpretation of the result.

2. **Application to gravity tests.** Now we apply the previous problem to the material in the lectures.

   (a) Consider implementing an interferometer by taking a beam of spin-polarized neutrons and sending them horizontally over the Earth, then using magnetic field gradients to create Stern-Gerlach type gratings that move the neutrons vertically. Draw this and map these words onto the mathematical objects in 1a above (i.e., identify U_{control} and the states $|0,1\rangle$.)

(b) Now identify the gravitational V operator, including the coefficient α, and compute $P(0,1)$. Treat V as just an external classical field. You can ignore the motion of the neutrons due to gravity i.e. just treat them as moving at fixed z.

(c) Now instead consider the toy entropic gravity model from the lecture. Answer qualitatively: Does the potential change? What is the effect of the noise terms?[11]

(d) Suppose you do this experiment (by the way, it was done long ago by Colella, Overhauser, and Werner, PRL 1975). Suppose you see perfect recombination of the neutrons. Is this consistent with perturbative quantum gravity? Is it consistent with entropic gravity?

(e) Just kidding: no experiment will ever see perfect recombination, because there's always noise causing decoherence from totally banal things – for example, stray random magnetic fields acting on the neutrons. So what you actually get is some slightly degraded interference fringes. (For an example, see the COW paper cited just above). Does this change your answer to the previous question?

[11] If you want be ambitious, you can try this quantitatively, e.g., estimate $P(0,1)$ in the entropic gravity model. You can approximate $g_\pm$ appearing in the Lindblad operators as just equal to 1.